Essays in Financial Economics

by

Winston Wei Dou

Submitted to the Department of Finance
on September 16, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three essays that theoretically and empirically investigate the asset pricing and macroeconomic implications of uncertainty shocks, propose new measures for model robustness, explain the joint dynamics on equity excess returns and real exchange rates.

In the first chapter, I show that the effect of uncertainty shocks on asset prices and macroeconomic dynamics depends on the degree of risk sharing in the economy and the origin of uncertainty. I develop a general equilibrium model with imperfect risk sharing and two sources of uncertainty shocks: (i) cash-flow uncertainty shocks, which affect the idiosyncratic volatility of firms' productivity, and (ii) growth uncertainty shocks, which affect the idiosyncratic variability of firms' investment opportunities. My model deviates from the neoclassical setting in one respect: firms' investment policies are set by the experts who are subject to a moral hazard problem and thus must maintain an non-diversified ownership stake in the firm. As a result, risk sharing between experts and other investors is imperfect. Limited risk sharing distorts equilibrium investment choices, firm valuation, and prices of risk in equilibrium relative to the frictionless benchmark. In the calibrated model, the risk premium on growth uncertainty shocks is negative under poor risk sharing conditions and positive otherwise. Moreover, the cross-sectional spread in valuations between value and growth stocks loads positively on the growth uncertainty shocks under poor risk sharing conditions and negatively otherwise. Empirical tests support these predictions of the model.

The second chapter is based on the joint work Chen, Dou, and Kogan (2015), in which we propose a new quantitative measure of model fragility, based on the tendency of a model to over-fit the data in sample with poor out-of-sample performance. We formally show that structural economic models are fragile when the cross-equation restrictions they impose on the baseline statistical model appear excessively informative about combinations of model parameters that are otherwise difficult to estimate. We develop an analytically tractable asymptotic approximation to our fragility measure which we use to identify the problematic parameter combinations. Using these asymptotic results, we diagnose fragility in asset pricing models with rare disasters

and long-run consumption risk.

The third chapter is based on the joint work Dou and Verdelhan (2015), which presents a two-good, two-country real model that replicates the basic stylized facts on equity excess returns and real interest rates. In the model, markets are incomplete. In each country, workers cannot participate in financial markets whereas investors trade domestic and foreign stocks, as well as an international bond. The investors' asset positions are subject to a borrowing constraint, along with a short-selling constraint on equity. Foreign and domestic agents differ in their elasticity of inter temporal substitution and in their risk-aversion. A time-varying probability of a global disaster implies time-varying risk premia in asset markets, and therefore large and time-varying expected valuation effects on international asset positions. The model highlights the role of market incompleteness and heterogeneity across countries in accounting for the volatility of equity and debt international capital flows.

Thesis Supervisor: Leonid Kogan
Title: Professor of Finance

Acknowledgments

I wish to thank my advisers, Hui Chen, Leonid Kogan, Andrew W. Lo, and Adrien Verdelhan, for their generous gifts of time and advice throughout the course of this research. This work is an outgrowth of the journey that I have been on over the past six years in the Department of Finance at MIT Sloan School of Management.

I can never be more grateful for the efforts and help offered by all my advisers in the whole process of writing this dissertation. It was their indefatigable support that raised me up to this point of achievement. Without their careful reading of the preliminary drafts and valuable feedbacks, this version of dissertation would never have been achieved. However, that's not all they have given me; their influence has been great, not only on the style and substance of this work, but also on my attitude toward research and passion on intellectual inquiry.

There are countless souls who have touched me along my way of studying and pursuing academic excellence at MIT. My advisers, as well as faculty members including Ricardo Caballero, Stewart Myers, Andrey Malenko, Jun Pan, Jonathan Parker, Antoinette Schoar, and Steve Ross, among others, are at the top of the list. They brought me into the world of serious research on financial economics. Their support, guidance, and encouragement kept me on track over each and every fork in the road. To me, they are not only responsible advisers and professors, but also best friends in life and obliging mentors in my career.

I also wish to thank the entire finance department – faculty, staff and students – for making my life at MIT an enriched and enjoyable one. Particularly, I wish to say "thanks" to Hillary Ross who has been helping and protecting doctoral students.

Finally, for my parents, I guess I will never know how they put up with my long absence; for my wife, Joyce, I will never know how she endured my irregular schedule and nocturnal work habit. But I do know their constant prayers and encouragement have given me the strength to persist and have this project finished. To them, "thanks" is far from enough, but it will have to suffice for now, and it truly comes from the bottom of my heart.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis consists of three main parts. The first one is a theoretical and empirical study of impact of uncertainty, especially when there are multiple sources of uncertainty and when it is interacted with limited risk sharing. The second part is a methodological study in which a measure of model fragility is proposed and justified. The third part is a quantitative study of international asset pricing, capital flows, and asset holdings. In this chapter, I introduce the motivation, results, and key mechanism of each part.

1.1 Embrace or Fear Uncertainty

This chapter is based on Dou (2016). The volatility of idiosyncratic shocks can affect agents' economic behaviors when markets are incomplete or technologies contain optionality features. The literature refers to the aggregate shock to the common component of idiosyncratic volatilities as an uncertainty shock, since it alters agents' information sets about future economic outcomes altogether.[1] Uncertainty shocks have proven

[1]There has been a fast growing literature studying the aggregate effects of such uncertainty shocks (e.g., Pástor and Veronesi, 2006, 2009; Bloom, 2009; Arellano, Bai, and Kehoe, 2011; Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry, 2013; Bachmann and Bayer, 2014; Christiano, Motto, and Rostagno, 2010, 2014; Bundick and Basu, 2014; Gilchrist, Sim, and Zakrajsek, 2014; Herskovic, Kelly, Lustig, and Nieuwerburgh, 2014). Here, the use of the term *uncertainty* is different from *Knightian uncertainty*, which emphasizes the situations where agents cannot know all the information they need to set accurate odds in the first place (e.g., Knight, 1921; Hansen and Sargent, 2008a). Also, uncertainty here is different

useful in explaining macroeconomic fluctuations and have been adopted as a standard feature of dynamic stochastic general equilibrium (DSGE) models for policy analysis.[2]

Despite substantial advances in understanding the economic impact of uncertainty shocks, two central and fundamental questions remain unsolved: first, how to reconcile the mixed empirical evidence about the effect of uncertainty shocks on asset prices and investment in one coherent framework; second, what are the factors underpinning the impact of uncertainty shocks. In particular, whether a positive uncertainty shock benefits or harms growth firms relative to value firms remains debatable, as does whether a rise in uncertainty boosts or curtails aggregate investment. The stylized facts are summarized in Figure 1-1.[3]

To address these questions, I develop a tractable investment-based general equilibrium model of asset prices with heterogeneous firms and agents in incomplete markets. Using the model as a guide, I revisit the link between asset prices and uncertainty with an explicit emphasis on the interaction with risk sharing conditions in the economy. The model not only provides a theoretical framework that quantitatively makes sense of these seemingly contradictory empirical findings; its main contribution is to do so by providing a fundamental economic mechanism through the explicit modeling of

from aggregate volatility, which has also been extensively studied in the literature (Bansal and Yaron, 2004a; Drechsler and Yaron, 2011; Shaliastovich, 2015; Campbell, Giglio, and Polk, 2013; Campbell, Giglio, Polk, and Turley, 2015; Fernandez-Villaverde, Guerron-Quintana, Rubio-Ramirez, and Uribe, 2011; Nakamura, Sergeyev, and Steinsson, 2014; Segal, Shaliasovich, and Yaron, 2015; Gourio, Siemer, and Verdelhan, 2015; Ai and Kiku, 2015).

[2]Policy authorities, including the Federal Reserve Board and the European Central Bank have claimed that uncertainty has an adverse effect on economy, and they have built uncertainty shocks into their core DSGE models as a main driver of the aggregate fluctuations (Christiano, Motto, and Rostagno, 2010, 2014). For example, at the 2013 Causes and Macroeconomic Consequences of Uncertainty conference, Federal Reserve Bank of Dallas President Richard Fisher gave a formal speech titled "Uncertainty matters. A lot." It emphasized that uncertainty could worsen the Great Recession and the ongoing recovery.

[3]I use the average idiosyncratic volatility across U.S. public firms' stock returns as a proxy for the total uncertainty. In Panels A and B of Figure 1-1, the high uncertainty in the late 1980s occurs with positive value spreads (i.e., cross-sectional spreads between value and growth stock returns) and high investment, and the high uncertainty in the late 1990s occurs with negative value spreads and high investment. However, the high uncertainty in the early 1990s and the late 2000s accompanies negative value spreads and low investment. Panel C shows that aggregate market volatility is almost perfectly correlated with total uncertainty over the period 1980 - 2014. The mixed empirical evidence on total uncertainty's effects illustrated in Panels A and B are thus linked to the ambiguous impacts of market volatility on asset prices and macroeconomic dynamics documented by Bansal, Kiku, Shaliastovich, and Yaron (2014) and Campbell, Giglio, Polk, and Turley (2015), among others.

endogenous imperfect risk sharing. The model recognizes two key elements shaping the impact of uncertainty shocks: (i) the risk sharing condition of the economy and (ii) the sources of uncertainty shocks.

Let me describe the main features of my model, starting with the characteristics of firms' technologies then turning to the characteristics of the agents. In the model, firms produce consumption goods using production units, which are building blocks of assets in place. The existing assets in place depreciate over time. Firms invest to create new assets in place using growth options. Growth options are intangible assets associated with innovations such as blueprints and research and development (R&D) projects. The investment decision is an option that the firm exercises optimally only when it receives an investment opportunity. Their investment opportunities arrive randomly over time and are subject to firm-specific shocks.

Firms' technologies feature two sources of uncertainty shocks: the cash-flow uncertainty shock and the growth uncertainty shock. Cash-flow uncertainty captures the variation in idiosyncratic volatility of assets-in-place productivity; growth uncertainty captures the variation in idiosyncratic volatility of investment-opportunity quality. Growth uncertainty can have a very different effect than cash-flow uncertainty due to the optionality embedded in growth options. This optionality arises from the flexibility in the innovation process. Simply put, if the quality of the investment opportunity turns out to be exceedingly good, the firm has the flexibility to dial up investment to exploit the beneficial realization of the investment shock; alternatively, the firm can tune down investment to insure against the adverse realization of the investment shock. The optionality makes the benefit of growth options a convex function of the underlying shock. As a result, growth uncertainty increases the value of growth options and the aggregate investment. Effectively, the growth uncertainty shock affects the economy in the same way as a simple aggregate investment-specific technological (IST) shock, which directly alters the economy's real investment environment.[4]

[4]The aggregate investment-specific technological shock has become a standard feature of real business cycle models (Greenwood, Hercowitz, and Krusell, 1997, 2000; Fisher, 2006; Justiniano, Primiceri, and Tambalotti, 2011). Moreover, recent papers show that the aggregate investment-specific technologi-

By their nature, the two uncertainty shocks cause different impacts in complete markets; moreover, their effects on the economy can be altered by the interactions between imperfect risk sharing and uncertainty shocks. To understand the interactions, I introduce financial frictions that endogenously arise from a standard moral hazard problem. Specifically, agents in my economy are either experts or households. Each expert is a representative agent for a team of managers and active insiders, who are usually financial intermediaries. Within each team, the managers and active insiders perfectly insure each other's consumption risks.[5] Each expert uses her unique skills to manage a particular firm's assets. In other words, the expert is the key talent without whose efforts the particular firm would cease to perform. To invest, each expert raises funds from the capital markets by issuing equity. However, experts face a moral hazard problem that imposes a co-investment or skin-in-the-game constraint: each expert must retain an undiversified ownership stake in the firm as a commitment not to make managerial decisions that maximize private benefits at the cost of reduced firm value. This incentive constraint limits an expert's capacity to insure against idiosyncratic cash flow and investment risks. On the other hand, households cannot run firms or trade assets, but they can invest in financial securities and therefore partially share risks with experts.

My model therefore deviates from the neoclassical setting in one key respect: firms' investment policies are set by experts who are subject to background risks imposed by incentive constraints.[6] This friction distorts experts' investment decisions and portfolio allocation from the first-best benchmark; households provide risk sharing to ex-

cal shock can help explain asset pricing puzzles (Christiano and Fisher, 2003; Papanikolaou, 2011; Kogan and Papanikolaou, 2013, 2014; Garlappi and Song, 2014; Kogan, Papanikolaou, and Stoffman, 2015).

[5]This is a simplification assumption widely adopted in the macroeconomic models with financial sectors (e.g., Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011; Brunnermeier and Sannikov, 2014). In other words, like mine here, these models focus on the financial frictions between households versus insiders of the corporate and financial sectors.

[6]The implications of background risks for asset prices and firms' financing and investment behavior have been investigated by Heaton and Lucas (1997, 2000a,b); Miao and Wang (2007); Chen, Miao, and Wang (2010), among others. In my model, the background risks are endogenously derived from a moral hazard problem, in an explicit and coherent way, within a general equilibrium macroeconomic framework. I focus on investigating the general equilibrium implications of the endogenous background risks.

perts through financial markets trying to mitigate the distortion and smooth their own consumption. When experts' balance sheets are well capitalized, they bear a small amount of implied idiosyncratic wealth risks and thus have a higher capacity to share risks with households; otherwise, they bear a large amount of implied idiosyncratic wealth risks and thus have a lower capacity to share risks with households, because they require more insurance from households to keep up real investment. The theoretical concept of the risk sharing condition can be interpreted as the condition of the financial sector in the data; in reality, the financial sector plays the largest role in determining the degree of risk sharing in the economy.

Due to experts' endogenous background risks, the impact of uncertainty shocks on asset prices and investment depends on the degree of risk sharing. When risk sharing is limited, positive uncertainty shocks dramatically increase the severity of background risks to experts, who become implicitly more risk averse. More precisely, in response to a positive cash-flow uncertainty shock, experts require higher risk premia on assets in place; for a rise in growth uncertainty, experts require higher risk premia on growth options. Yet whether the implied higher risk premia eventually increase or curtail experts' willingness to invest depends on the specification of preferences. More precisely, it depends on whether the intertemporal substitution effect dominates the wealth effect. In general, the intertemporal substitution effect dominates when the elasticity of intertemporal substitution (EIS) is larger than one.[7] In such cases, a rise in uncertainty induces experts to invest less in firms' assets today and more in the future, since their desire for a better investment environment dominates that for consumption smoothing. The interest rate tends to decline due to the fly-to-quality effect, and yet it remains stable because of the high EIS coefficient. As a result, assets' prices have to drop to provide higher risk premia. Specifically, a rise in cash-flow uncertainty always decreases the prices of assets in place. However, a rise in growth uncertainty can have an ambiguous impact on growth option's prices. The net effect of higher growth uncertainty depends on the competition between the positive force of the optionality and the neg-

[7] The discussion on the relationship between the EIS coefficient and the dominance of intertemporal substitution effect can be found in Weil (1990) and Bhamra and Uppal (2006).

ative force of the precautionary saving motive. In times when risk sharing is limited, the precautionary saving motive becomes strong enough to dominate the option effect.

Compared to cash-flow uncertainty shocks, higher growth uncertainty causes an additional risk to experts. It is the risk of increasing inequality in the distribution of innovation benefits from growth options. The skewness in the distribution of innovation benefits matters when the risk sharing of idiosyncratic investment shocks is limited. Such a high-moment risk thus becomes particularly devastating when the risk sharing condition is poor. The intuition is further elaborated as follows. Most of the benefits from innovation accrue to a small fraction of experts, while the majority of experts bear the cost of creative destruction since they need to pay for the new assets in place to keep up their production levels. Wealth is reallocated from the experts who do not invest to those who receive high-quality investment opportunities. This reallocation becomes more skewed when growth uncertainty becomes higher, since growth-option benefits are asymmetric. Technically speaking, each expert faces a more skewed idiosyncratic investment risk.[8] For a risk averse expert, the higher skewness in the idiosyncratic risk leads to lower certainty equivalent wealth. Therefore, the growth uncertainty shock contributes to an adverse redistribution risk: the displacement risk.[9]

The risk sharing condition, moreover, is endogenous and affected by uncertainty shocks. When the intertemporal substitution effect dominates, experts charge higher risk premia and want to sell assets to reduce their exposure to idiosyncratic risks, in response to a rise in uncertainty. This leads to a plunge in asset prices. Experts are atomistic, so they do not take into account the general equilibrium effect of their own asset sales on asset prices, even though they are aware of the adverse effect of plunging asset prices on their risk sharing conditions. This pecuniary externality arises from financial constraints, together with competitive asset markets.[10] Due to such a pecu-

[8] In contrast, the idiosyncratic cash flow risk is always symmetric.

[9] There is more discussion and a literature review on the asset pricing implications of displacement risks in Section 1.1.1.

[10] This particular pecuniary externality has been explicitly investigated and highlighted by Lorenzoni (2008). The models studying financial stability and its macroeconomic implications are mainly built on this basic mechanism (e.g., Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997; Bernanke, Gertler, and Gilchrist, 1999; He and Krishnamurthy, 2011, 2013; Brunnermeier and Sannikov, 2014; Di Tella,

niary externality, the adverse feedback loop between plunging asset prices (soaring risk premia) and deteriorating risk sharing conditions characterizes the equilibrium.

This fundamental economic mechanism has important asset pricing implications. To understand them, it is necessary to establish how these shocks affect the marginal investor's utility and determine how uncertainty shocks affect the cross section of firms.

The cash-flow uncertainty shock always carries a negative market price of risk,[11] because the cash-flow uncertainty decreases experts' current and future consumption. However, when growth uncertainty rises, experts face three endogenous risks: the investment risk, the endogenous financial risk, and the displacement risk. When the EIS coefficient is sufficiently large, the latter two contribute to a negative market price of risk for growth uncertainty shocks, whereas the investment risk contributes to a positive market price of risk. The risk sharing condition determines the net effect between the two countervailing forces. The positive force of investment risk dominates when risk sharing is efficient; the negative force of the endogenous financial risk and the displacement risk dominates otherwise.

Uncertainty shocks do not affect all firms equally in the cross section. The heterogeneous impacts are time varying; they depend on risk sharing conditions. A positive cash-flow uncertainty increases the value of growth options relative to assets in place. This is because a higher cash-flow uncertainty immediately increases the riskiness of assets in place. As a result, experts gravitate to safer assets, including growth options. This portfolio rebalancing tendency increases the price of growth options. Meanwhile, the price of investment goods decreases, which provides a hedge for growth options against the drop in the value of assets in place. A rise in growth uncertainty can increase or decrease the value of growth options relative to assets in place; the sign depends on the risk sharing condition. When risk sharing is efficient, the growth uncertainty increases the investment risk attached to the growth options, which increases the value of growth options relative to assets in place; otherwise, the growth uncertainty

2014).

[11] Recall that the formal (technical) definition of market price of risk for a shock is the negative contemporaneous response of a marginal investor's marginal utility to a unit increase in such a shock.

shock increases the endogenous financial risk and the displacement risk attached to growth options, which pushes experts to safer assets, including assets in place. The portfolio rebalancing tendency has an increasing effect on the value of assets in place. In summary, uncertainty shocks affect firms differently depending on whether they derive most of their value from growth options or assets in place, and depending on whether the risk sharing condition is good or poor.

Using state-of-the-art techniques, I solve the model globally to capture the nonlinearity of economic dynamics and the endogenous fluctuations of risk sharing conditions. I calibrate the model to match the moments of macroeconomic variables and check whether the calibrated model can provide reasonable asset pricing moments and cross-sectional dynamics of firms. In my calibration, the model has a reasonable quantitative performance, which is summarized as follows. First, the model reproduces a sizable equity premium, mainly attributed to the market incompleteness; it also reproduces a large value premium, mainly attributed to the heterogeneous effects of cash-flow uncertainty shocks. Second, in the model as in the data, the sales dispersion is countercyclical, while the investment dispersion is pro-cyclical. This empirical pattern is highlighted in Bachmann and Bayer (2014) as an important cross-equation restriction for the macroeconomic models with uncertainty shocks. My model provides a novel reconciliation for the two dispersion processes within a unified framework. In this framework, the sales dispersion is driven by the cash-flow uncertainty shock, but not by the growth uncertainty shock; it is countercyclical because the cash-flow uncertainty leads to economic downturns. On the other hand, the investment dispersion is driven by the growth uncertainty shock, but not by the cash-flow uncertainty shock; it is pro-cyclical because the impact of growth uncertainty shocks on the investment dispersion is asymmetric: the effect is larger when the risk sharing condition is good. These connections between uncertainty and dispersion are verified in the data using estimated uncertainty shocks.

I empirically test the model's main predictions. I first set up a regime-switching model in which the exposure of value spreads to growth uncertainty shocks is time-varying and characterized by a latent Markovian state variable. My theory implies that the latent state in which the exposure is higher should correspond to the state in which risk sharing is limited. I use the credit spread (e.g., Gilchrist and Zakrajsek, 2012) and the chronologies of financial crisis constructed by Reinhart and Rogoff (2009) as proxies for the risk sharing condition in the data. The empirical evidence is consistent with the model: the estimated likelihood of being in the latent state of higher growth uncertainty shock exposure is significantly, positively associated with the proxies of risk sharing conditions. I also provide additional empirical tests verifying this particular prediction; the results of statistical tests are significant. Then, I verify the predictions of the market price of risk for uncertainty shocks in the data.

In summary, this paper casts light on the recent debate on the role of uncertainty shocks in explaining asset pricing phenomena and macroeconomic dynamics, and on how the cross section of asset returns can identify uncertainty shocks from different sources. Moreover, the time-varying cross-sectional moments of asset prices, depending on the degree of risk sharing, impose additional cross-equation restrictions on the properties of uncertainty shocks used in macroeconomic models and thus can provide extra insights on the origins of aggregate fluctuations. Further, as both the model and empirical evidence highlight the importance of sources and risk sharing conditions for determining how the economy reacts to uncertainty shocks and the endogeneity of aggregate volatilities driven by different underlying uncertainty shocks, this paper provides a cautionary note to empirical studies using one aggregate volatility index to draw conclusions on the economic impact of uncertainty.

1.1.1 Related Literature

The idea that uncertainty shocks affect investment and asset prices dates back at least to the literature exploring the (implicit) optionality associated with production and investment technologies (e.g., Oi, 1961; Hartman, 1972; Abel, 1983; Caballero, 1991; Dixit and Pindyck, 1994; Bar-Ilan and Strange, 1996; Abel, Dixit, Eberly, and Pindyck, 1996). Since then, many different dynamic structural models have been developed based on these ideas trying to quantify the relevance of uncertainty shocks in the data.

The technical challenge of analyzing stochastic dynamic general equilibrium models with structural links between uncertainty shocks and the data is well known. The existing literature tries to make progress by focusing on a single isolated channel in each model. One strand of literature investigates the wait-and-see effect by introducing decreasing-scale-to-return production, sizable adjustment costs, and irreversibility into the dynamic setting (e.g., Bloom, 2009; Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry, 2013; Bachmann and Bayer, 2014). The asymmetric effect of uncertainty on benefits and costs of waiting captures the essence of the waiting option effect. This is referred to as the *bad news principle* by Bernanke (1983). However, the waiting option effect can be mitigated or even turned over when some environmental variables shift. For example, this idea has been demonstrated in Miao and Wang (2007) and Bolton, Wang, and Yang (2013) under partial equilibrium frameworks. For investors bearing uninsurable idiosyncratic risks and firms being financially constrained, the uncertainty shock can have both a positive and a negative effect on investment and financing decisions. My model deliberately brings the idea of financial friction and imperfect risk sharing into a general equilibrium framework in which the opposite impacts of uncertainty shocks emerge endogenously.

Another strand of literature explores the credit risk premium channel (e.g., Christiano, Motto, and Rostagno, 2010, 2014; Arellano, Bai, and Kehoe, 2011; Gilchrist, Sim, and Zakrajsek, 2014). The key idea is that in an economy with corporate debt and costly default, higher uncertainty lifts the default probability for firms that are already near default boundaries, and hence the cost of debt financing increases. This in turn reduces the investment and increases the default probabilities for firms that are originally not so close to the default boundaries. As a result of the ripple effect, aggregate hiring decreases, which leads to lower household consumption and thus feeds back to a higher credit risk premium. This adverse feedback loop reinforces the ripple effect, dragging the whole economy into recessions and creating high credit spreads. It is clear that if the financial sector is strong and very few firms are close to financially binding constraints, the adverse risk premium effect will be largely dampened.

A third strand of literature investigates the interaction between learning and uncertainty shocks. One interaction is the learning-by-doing mechanism, which assumes that investors have imperfect information about the underlying state and that the only way to achieve extra signals about the true state is through a sequence of real investments. Naturally, in a high uncertainty environment, investors conduct earlier and more intensive investment to learn the underlying state (e.g., Roberts and Weitzman, 1981; Pindyck, 1993; Pavlova, 2002). Moreover, in Pástor and Veronesi (2006, 2009), the authors show that the uncertainty shock increases the value of growth options relative to assets in place, and this effect is particularly large when uncertainty shocks are convolved with Bayesian learning. On the other hand, uncertainty shocks, interacting with learning, can also depress asset prices and investment. In Van Nieuwerburgh and Veldkamp (2006), if acquiring information becomes slower and belief uncertainty becomes higher during economic downturns, the learning mechanism generates slow recoveries and countercyclical asset pricing dynamics. Moreover, Fajgelbaum, Schaal, and Taschereau-Dumouchel (2013) show that low activity and slower learning can form an unpleasant feedback loop. The fixed point for this feedback loop is the equilibrium that displays uncertainty traps: self-reinforcing episodes of high uncertainty and low activity. The uncertainty trap can substantially worsen recessions and increase their duration.

A main contribution of this paper is to introduce two sources of uncertainty shocks into one unified theoretical framework in which the impact of uncertainty shocks varies endogenously, governed by a macroeconomic condition: the degree of risk sharing in the economy. Importantly, the theoretical framework is tractable, which allows for accurate global solutions. This model is motivated by several strands of literature. Basically, I incorporate the models of heterogeneous agents bearing undiversified idiosyncratic risks and the macroeconomic models of financial stability into an investment-based general equilibrium model for asset prices. Therefore, my paper is also deeply connected to the following three strands of literature.

The asset pricing literature on heterogeneous agents with undiversified idiosyncratic risks explores the possibility of solving the equity premium puzzle based on market incompleteness. This literature goes back to Mankiw (1986) and Constantinides and Duffie (1996). The key idea is that the time-varying cross-sectional dispersion of consumption can increase the volatility of the stochastic discount factor (e.g., Constantinides and Duffie, 1996; Storesletten, Telmer, and Yaron, 2007; Herskovic, Kelly, Lustig, and Nieuwerburgh, 2014; Ghosh and Constantinides, 2015), and the undiversified idiosyncratic investment risks increase the correlation between the individual consumption growth and the asset return (e.g., Heaton and Lucas, 1997, 2000a,b). In my model, both effects arise endogenously from a moral hazard problem. The resulting effects are further amplified by endogenous financial frictions. Most importantly, a key difference of my model is that the marginal investors of the aggregate equity have fully diversified portfolios. Here, the undiversified idiosyncratic shocks affect the economy initially through the real investment channel; then, the distorted real investment deteriorates agents' risk sharing on aggregate shocks due to the limited market participation. More broadly, my model is connected to the papers trying to rationalize the volatile stochastic discount factors through market incompleteness, such as Alvarez and Jermann (2000, 2001), Chien and Lustig (2010), Chien, Cole, and Lustig (2012), and Dou and Verdelhan (2015).

The idea of undiversified idiosyncratic risks has also been adopted in dynamic structural corporate models (or partial equilibrium dynamic macroeconomic models) to study firm's investment and financing behavior (e.g., Miao and Wang, 2007; Chen, Miao, and Wang, 2010; Panousi and Papanikolaou, 2012; Glover and Levine, 2015). My model incorporates these partial equilibrium mechanisms, together with asset pricing channels, into a general equilibrium model to study their aggregate implications.

The macroeconomic literature on financial stability builds financial frictions into otherwise standard neoclassical models. This literature started from Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Bernanke, Gertler, and Gilchrist (1999). Recent advances explore the concentration of aggregate risk and its role in creating systemic risks and nonlinear risk premia dynamics through the balance sheet channel (e.g., Adrian and Boyarchenko, 2012; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014; Di Tella, 2014; Haddad, 2014; Drechsler, Savov, and Schnabl, 2014). One contribution of my paper to this literature is to quantitatively examine the asset pricing implications of financial frictions in the cross section of various types of assets.

My model fits within the literature studying asset prices in investment-based general equilibrium models. It is most closely related to the papers explicitly modeling assets in place and growth options and focusing on the cross section of asset prices. Gomes, Kogan, and Zhang (2003) study a model in which book-to-market ratios are positively associated with average returns. But, growth options are riskier than assets in place in the model. Papanikolaou (2011) presents a model with aggregate investment shocks, which by nature affect assets in place and growth options differently. In this calibrated model, the aggregate investment shock benefits growth options relative to assets in place, and carries a negative market price of risk if the late resolution of uncertainty is preferred by investors. Pástor and Veronesi (2009) and Gârleanu, Panageas, and Yu (2012) study the asset pricing dynamics in models with episodes of endogenous technology adoption. Ai and Kiku (2013) study a model in which the cost of option exercise is pro-cyclical and thus the assets in place are riskier. Ai, Croce, and Li (2012) study a model in which the younger vintages of assets in place have lower exposure to aggregate productivity shocks and thus growth options are less risky. These papers all assume perfect risk sharing. However, Gârleanu, Kogan, and Panageas (2012) and Kogan, Papanikolaou, and Stoffman (2015) rationalize the negative price of risk for the aggregate investment shock by introducing displacement risks that arise from market incompleteness. Moreover, Opp (2014) explicitly incorporates the venture capital intermediation into an otherwise standard dynamic general equilibrium macroeconomic model of asset prices and focuses on the asset pricing phenomenon of venture capital cycles. Despite perfect risk sharing, the informational friction causes costly external financing for new ventures and hence distorts investment; the venture capital firms alleviate such information frictions in the economy. Also, the displacement risk of technological innovations plays an important role in determining the risk premia in the model. My model studies the unequal effects of uncertainty shocks on asset returns in the cross section and the time variation of these effects driven by endogenous imperfect risk sharing. In my model, growth uncertainty shocks endogenously cause displacement risks, especially when the risk sharing condition is poor.

1.2 Measuring the "Dark Matter" in Asset Pricing Models

This chapter is based on the joint work Chen, Dou, and Kogan (2015). When building and evaluating a quantitative economic model, we care about how the model performs out of sample in addition to how well it fits the past data. This dual concern gives rise to the classic tradeoff between the accuracy of in-sample fit and the tendency of over-fitting. Too much emphasis on in-sample fit favors complex models, which are prone to over-fit the data in sample and likely to have poor out-of-sample performance. Precisely, those complex models over-utilize degrees of freedom of some parameters to accommodate certain functional-form assumptions to obtain accurate in-sample fit. We refer to such functional forms of models as fragile and such parameters as "dark matter". A model, containing such fragile functional-form assumption or "dark matter", is also referred to as fragile. Model fragility is a property of a model which captures its tendency to over-fit the past data, or in other words, captures the unreliability to conclude its out-of-sample performance based on the accuracy in-sample fit. Thus, models with higher fragility should be less favored among a set of candidate models that fit the past data well.

The above tradeoff is intuitive but not easy to implement in practice. As we build increasingly sophisticated quantitative models, the need for a systematic way to quantify model fragility also grows. Traditional over-fitting tendency measures, including the Akaike Information Criterion (AIC) and its various extensions, focus on the number of free parameters in a model used to accommodate its functional forms. Such measures potentially miss the implicit complexity: the effective number of degrees of freedom in a model depends not only on the number of free parameters, but also on the sensitivity of the key implications of the model with respect to "reasonable" changes in the parameter values. If its implications are highly sensitive, a particular functional-form assumption of the model can always fit the data by choosing specific parameter values within a "reasonable" range, and thus tends to over-fit the data in sample. In such case, the accuracy of in-sample fit becomes unreliable for assessing the particular functional-form assumption and, of course, the full model.

In this paper, we propose a new quantitative measure of model fragility. Our measure is constructed based on Fisher information matrices, so we refer to it as Fisher fragility measure. Consider a typical structural model as a combination of functional-form specifications and parameters implied by economic theories and statistical distributions. The model describes a joint distribution of variables $\mathbf{x}_t$ and $\mathbf{y}_t$. The baseline model describes the distribution of sample $\mathbf{x}^\mathbf{n} \equiv (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ using the parameter vector $\theta \in \Theta$. The functional form assumed on top of the baseline model is chosen to be evaluated for its fragility. We are therefore measuring the "dark matter" of parameters θ of the baseline model. The functional-form assumption, on top of the baseline model, introduces additional ingredients that establish a joint distribution of $(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$.[12] In this setting, we think of the additional functional forms implied by economic theories as adding cross-equation restrictions to the system of moments based solely on the baseline model. By definition, our Fisher fragility measure effectively compares the inverse Fisher information matrices for the baseline model and the full structural model along the directions associated with these linear subspaces and aggregates the differences.

Our Fisher fragility measure provides a simple decomposition that attributes the sources of model fragility (i.e. "dark matter") to a set of 1-dimensional linear subspaces of the parameter space. This decomposition offers an intuitive sample size interpretation. Each 1-dimensional linear subspace, indexed by j, corresponds to a particular

[12] As an example, consider a Lucas economy. There is a representative agent with certain preferences, and the growth rate of endowment is IID normal. A structural model in this case can be the model for the joint dynamics of the exogenous endowment ($\mathbf{x}_t$) and endogenous return on the endowment claim ($\mathbf{y}_t$), and θ includes the mean and volatility of endowment growth.

linear combination of model parameters, $v_j\theta$. We assume that model parameters are estimated by GMM, with the fitted parameter vector $\widehat{\theta}$, and use the GMM J-distance, $J(\widehat{\theta}; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$, as the quantitative measure of model's in-sample fit given observations $(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$. Asymptotically, our measure corresponds to the amount of extra data needed to lower the asymptotic variance of $v_j\widehat{\theta}$ under the baseline model to the level of its variance under the full structural model (with the original data).

It is worth highlighting that when measuring model fragility, the goal is not to prove a model wrong. It is true that if a model is misspecified, further testable restrictions may reveal that. However, as Box (1976) and Hansen (2014) stress, all models are simplifications of reality that can eventually be rejected with sufficient data. Hansen (2014) states that "the important criticisms are whether our models are wrong in having missed something essential to the questions under consideration." This is why we formulate our measure using the GMM framework. Through the selection of moment conditions, the econometrician has the ability to determine what the essential predictions of the model are.

It is also worth emphasizing that when measuring model fragility, the goal is not to estimating parameters. This paper is not really proposing new estimation procedures or drawing statistical inferences of any point estimators. Rather, the main purpose is to provide a new model fragility measure facilitating structural model selection when there are multiple candidates that fit a common set of fixed observations well in sample. Our model fragility measure is in the same spirit of those penalization procedures based on statistical fragility measures and adopted in statistical model selections such as AIC, BIC, and LASSO procedures. However, differently, our measure is specifically constructed for structural economic models. The over-fitting (model fragility) evaluation is opposite to the goodness-of-fit consideration; the latter takes a parametric model as fixed and focuses on the distribution of possible sample generated from it, whereas the former takes a sample as given and focuses on the sensitivity of various parametric models that fit in sample.

How to justify that our measure is indeed a measure of model fragility? To answer the important question, we extend a popular measure of statistical over-fitting tendency to our structural setting,[13] and we establish an asymptotic equivalence result showing that the over-fitting tendency measure is actually equivalent to our Fisher fragility measure. Let θ_0 denote the parameter value for the true model. The corresponding value of the J-distance, $J(\theta_0; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$ is generally higher than the fitted value $J(\widehat{\theta}(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}); \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$, because the latter is chosen to minimize the J-statistic in sample. Then, the gap between the J-distance for the true model and the fitted structural model, $d\{\theta_0; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\} = J(\theta_0; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}) - J(\widehat{\theta}(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}); \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$, measures the degree of over-fitting by the estimated model. Not knowing what the true model is, we follow the common approach and average the degree of over-fitting over a set of possible true models, $\int_{\theta \in \Theta} \xi(\theta) d\{\theta; \mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}\}\, d\theta$, where $\xi(\theta)$ assigns relative weights to alternative models. There is no broadly accepted choice of how to weigh the alternative models, and the exact specification depends on the context. We first specify the baseline model for $\mathbf{x}_t$. We then use the posterior distribution for θ implied by the baseline model, $\pi(\theta|\mathbf{x}^\mathbf{n})$, as the distribution over the alternative models $\xi(\theta)$. With this definition, we are assuming that inference based on the baseline model is reliable. We are therefore measuring the fragility of the full structural model relative to the baseline model.

Our measure of the average degree of over-fitting is fundamentally connected to model complexity. In addition, as we show, it is closely linked to model sensitivity to parameter perturbations.

Sensitivity analysis is a common technique for assessing model robustness. Intuitively, a model is considered robust if its key implications are not excessively sensitive to small perturbations of model parameters. In practical applications, one must specify the relevant perturbations and quantify "excessive sensitivity." As a result, it is diffi-

[13] Spiegelhalter, Best, Carlin, and van der Linde (2002) measures over-fitting in a similar way, but using the log-likelihood instead of the J-distance. Also, they use arbitrary prior for generating the "reasonable" alternative models to assess a statistical model's over-fitting tendency; however, we argue it's crucial to choose a baseline model and a self-coherent posterior for generating the "reasonable" alternative functional-form specifications in economic modeling evaluation. This procedure allows for more economic-meaningful model assessment, beyond pure statistical considerations. Using this procedure, economists can focus on the fragility of certain functional forms implied by economic theories and the "dark matter" of certain parameter space, not necessarily the whole model and all its parameters.

cult to generalize the traditional sensitivity analysis to multivariate settings. Model fragility may not be fully revealed by perturbing individual parameters – one must contemplate all possible multivariate perturbations, making the common approach impractical for high-dimensional problems.

Our methodology is not subject to such limitations. We use the posterior associated with the baseline model to weigh the relevant perturbations, and use the variance of the moments in the structural model to judge the degree of sensitivity of the moments. This eliminates the need for ad hoc choices associated with traditional sensitivity analysis. In addition, the asymptotic measure helps diagnose the sources of model fragility. Knowing the relative importance of each subspace for the overall fragility of the model effectively reduces the dimensionality of the multivariate sensitivity analysis. For example, if a single 1-dimensional subspace is dominant in terms of its contribution to overall model fragility, one only needs to examine the sensitivity of various moments to the perturbation of parameters in this particular subspace to quantify the main aspects of model fragility.

We also provide an information-theoretic interpretation for our model fragility measure. It answers the other fundamental question: what information is really captured by our Fisher fragility measure, and in what economic sense? We argue that our fragility measure is connected to the informativeness of the economic restrictions on the model parameters. To introduce the concept of informativeness, consider an example of a model that links the observations of the stock price P to the parameter θ describing the distribution of cash flows through a restriction: $\mathbb{E}[P] = \bar{P}(\theta)$. An econometrician starts with a baseline statistical model for cash flows and forms an (unconstrained) posterior belief about θ based on the observed cash-flow data and the baseline model, which is depicted in the left panel of Figure 1-2. The flatness of this posterior distribution indicates that there is nontrivial uncertainty about the true value of θ according to the baseline model.

The middle panel plots the model-implied price function $\bar{P}(\theta)$. Due to the high sensitivity of $\bar{P}$ to θ (the derivative $\partial\bar{P}/\partial\theta$ is large), there is only a narrow set of values of

θ (highlighted by the shaded region) for which the observed price data are statistically close to the model-implied prices. This has two implications. First, by imposing the economic model, the econometrician obtains a posterior for θ (see the right panel) that is much more concentrated than the posterior distribution under the baseline model. In this case, we say that the economic restriction $\mathbb{E}[P] = \bar{P}(\theta)$ is highly informative about θ.

Second, for values of θ away from the shaded region in the middle panel but still in the range of values considered highly likely under the unconstrained posterior, the fit between the model and the observed price data deteriorates drastically, which is a sign of model fragility. Thus, high informativeness of the economic restrictions is closely linked to model fragility. Economic parameter restrictions are highly informative when they can significantly influence inference about certain combinations of model parameters that are relatively difficult to estimate statistically without such restrictions. Such parameter combinations are where the "dark matter" concentrates in the parameter space.

We formalize the above notion of informativeness of economic restrictions in an information-theoretic framework with an intuitive effective sample size interpretation. The informativeness of cross-equation restrictions relative to the baseline model is also reflected in the effect of the former on the posterior distribution of model parameters. We quantify the discrepancy between the posteriors of model parameters under the baseline model and under the model with further economic restrictions using relative entropy. We then define an effective sample size measure of informativeness of cross-equation restrictions as the average amount of extra data that, under the baseline model, generates the same magnitude of the shift in the posterior distribution. In other words, we equate the information content of the economic restrictions with the information content of additional data under the baseline statistical model. We show that the resulting measure of informativeness of cross-equation restrictions is related asymptotically to our measure of model fragility.

An important class of applications of our measure is to structural models that involve agents' beliefs. One prevalent approach to discipline beliefs is by imposing the rational expectations (RE) assumption. The RE assumption ties down the beliefs of economic agents by endowing them with precise knowledge of the probability law implied by an economic model. A common example of a RE model is a setting in which the agents know the true parameter value θ_0. The assumption of such precise knowledge is usually justified as a limiting approximation to the beliefs formed by learning from a sufficiently long history of data (see Hansen, 2007). If the posterior distribution of θ given $\mathbf{x}^\mathbf{n}$ under the baseline model serves to describe the outcome of such learning, then high fragility of the economic model means that the model moments are highly sensitive to the exact choice among the likely values of the model parameter vector. In that case, assuming that the agents know the true parameter vector may be a poor approximation to a broader class of models in which agents maintain nontrivial uncertainty about the probability law of the model.

We apply the fragility measure to two examples from the asset pricing literature. The first example is a rare-disaster model. In this model, parameters describing the likelihood and the magnitude of economic disasters are relatively difficult to estimate from the data unless one uses information in asset prices.[14] We describe the fragility measure in this example analytically. We also illustrate how to incorporate uncertainty about the structural parameters (preference parameters in this context) when computing model fragility. The second example is a long-run risk model with a six-dimensional parameter space. We use this example to illustrate how to systematically diagnose the sources of fragility in a complex model.

1.2.1 Related Literature

The idea that model fragility is connected to complexity dates back at least to Fisher (1922). Model complexity is traditionally measured by the number of parameters in

[14] A few papers have pointed out the challenges in testing disaster risk models. Zin (2002) shows that certain specifications of higher-order moments in the endowment growth distribution can help the model fit the empirical evidence while being difficult to reject in the data. In his 2008 Princeton Finance Lectures, John Campbell suggests that variable risk of rare disasters might be the "dark matter for economists."

the model, because of the coincidence of the two quantities in Gaussian-linear models (see, e.g. Ye, 1998; Efron, 2004). Numerous statistical model selection procedures are based on this idea.[15]

The limitations of using the number of parameters to measure model complexity are well known. Extant literature covers several alternative approaches to measuring the "implicit model complexity." Ye (1998), Shen and Ye (2002), and Efron (2004) propose to measure complexity (or "generalized degrees of freedom" in their terminology) for Gaussian-linear models using the sensitivity of fitted values with respect to the observed data. Gentzkow and Shapiro (2013) apply a similar idea to examine identification issues in complex structural models. Spiegelhalter, Best, Carlin, and van der Linde (2002), Ando (2007) and Gelman, Hwang, and Vehtari (2013), among others, propose a Bayesian complexity measure they call "the effective number of parameters," which is based on out-of-sample model performance. These methods measure the sensitivity of model implications to parameter perturbations. The important common feature of these proposals is that they rely on the model being evaluated to determine the magnitude of necessary parameter perturbations. This is potentially problematic when evaluating economic models that are fragile according to our definition. For such models, the posterior distribution over the parameters is highly concentrated as a result of excessive model sensitivity to its parameters. Relying on this posterior to generate parameter perturbations can under-represent the true extent of model fragility. In contrast, we propose to use the baseline model to determine the distribution $\xi(\theta)$ over the potential alternative models.

Hansen (2007) discusses extensively concerns about the informational burden that rational expectations models place on the agents, which is one of the key motivations for research in Bayesian learning, model ambiguity, and robustness.[16] In particular, the literature on robustness in macroeconomic models (see Hansen and Sargent, 2008b; Ep-

[15]Examples include the Akaike Information Criterion (AIC) (Akaike, 1973), the Bayesian Information Criterion (BIC) (Schwarz, 1978), the Risk Inflation Criterion (RIC) (Foster and George, 1994), and the Covariance Inflation Criterion (CIC) (Tibshirani and Knight, 1999).

[16]See Gilboa and Schmeidler (1989), Epstein and Schneider (2003), Hansen and Sargent (2001, 2008b), and Klibanoff, Marinacci, and Mukerji (2005), among others.

stein and Schneider, 2010, for recent surveys) recognizes that the traditional assumption of agents' precise knowledge of the relevant probability distributions is not reasonable in certain contexts. This literature explicitly incorporates robustness considerations into agents' decision problems. Our approach is complementary to this line of research in that we propose a general methodology for measuring and detecting fragility of economic models, thus identifying situations in which parameter uncertainty and robustness could be particularly important.

Our work is connected to the literature in rational expectations econometrics, where economic assumptions (the cross-equation restrictions) have been used extensively to gain efficiency in estimating the structural parameters.[17] When imposing such assumptions results in a fragile model, standard inference may result in excessively small confidence regions for the parameters, with low coverage probability under reasonable parameter perturbations. Related, fragile models tend to generate excessively high quality of in-sample fit, which biases model selection in their favor. The combination of these two effects makes common practice of post-selection inference misleading in the presence of "dark matter".

1.3 The Volatility of International Capital Flows and Foreign Assets

This chapter is based on Dou and Verdelhan (2015). After decades of financial liberalization, foreign assets represent now a large fraction of aggregate wealth. For the U.S., the gross foreign equity and bond holdings amount to 83% of GDP in 2010 (Lane and Milesi-Ferretti, 2007, updated). Foreign holdings are volatile because their unit value changes, through valuation effects, and their quantities changes, through international capital flows. During the recent Great Recession for example, the value of the net U.S. foreign equity and bond holdings decreased by 51%, while at the same time,

[17]For classic examples, see Saracoglu and Sargent (1978), Hansen and Sargent (1980), Campbell and Shiller (1988a), among others, and textbook treatments by Lucas and Sargent (1981), Hansen and Sargent (1991).

international capital flows dried up. From the perspective of the benchmark models in international economics, such large valuation changes and such volatile capital flows are puzzling. In this paper, we propose a two-good, two-country model that is consistent with the basic stylized facts in equity and interest rate markets. With a model consistent with asset prices in hand, we turn to the macroeconomic quantities: we use the model to assess the volatility of international capital flows and foreign assets.

Our model has four main characteristics: a rich endowment process, general recursive preferences with heterogenous agents, limited market participation, and short-selling and borrowing constraints.

The total endowment process has a global and a country-specific component. Both components are described by Markov processes. The growth rate of the global component is subject to disaster risk: with a small, time-varying probability, the world growth rate may fall. The country-specific endowment is persistent, but only subject to Gaussian risk. The total endowment is levered and divided into a labor income stream and a dividend stream. The leverage is also time-varying: as in the data, in bad times, leverage is large (Longstaff and Piazzesi, 2004a). With these features and risk-averse agents, the model delivers large and time-varying risk premia in line with the empirical evidence on equity and bond markets.

The agents are characterized by Epstein and Zin (1989b) preferences, which disentangle risk-aversion from the inter-temporal elasticity of substitution. The domestic (i.e. U.S.) agent is less risk-averse than her foreign (i.e. rest-of-the-world, denoted ROW) counterpart, but has a higher inter-temporal elasticity of substitution. The differences across agents lead to large gross foreign asset positions. As in the data, the U.S. tends to borrow from the ROW and invests in the foreign stock market, therefore providing insurance to the ROW. International trade is frictionless and each agent consumes both domestic and foreign goods.

In each country, some agents participate in international financial markets, while others do not. The workers, who do not participate, consume all of their labor income each period. The investors, who do participate, choose optimally the quantity of domestic and foreign stocks as well as their net borrowing or lending positions. Their investment decisions are subject to two constraints: they cannot short stocks and their borrowing is limited by the amount they can reimburse the next period in the worst state of the world. These constraints rule out defaults and ensure that the equilibrium solution of the model is stationary even if agents with Epstein and Zin (1989b) preferences differ in their risk-aversion and inter-temporal elasticity of substitution. These constraints would not be necessary if agents would share the same preference parameters or if agents were characterized by constant relative risk-aversion preferences, but they are necessary in our model to obtain a stationary equilibrium.

In the model, markets are incomplete, even for the agents who participate in financial markets. There are five different endowment shocks (the global Gaussian growth rate, the global disaster state, the disaster probability shock, and two country-specific endowment shocks), but there are only three assets traded (two stocks and one bond). Moreover, borrowing and short-selling constraints sometimes, but not always, bind. Market incompleteness is a key feature of our model. While investors can choose optimally their portfolio positions to mitigate the impact of market incompleteness, workers can not work around their participation constraint.

Such a rich model has never been simulated before. Building on the results of Kubler and Schmedders (2003) and Duffie, Geanakoplos, Mas-Colell, and McLennan (1994), we show that the model has a wealth-recursive Markov solution. The proof extends previous results on heterogenous agent models to the case of Epstein and Zin (1989b) preferences and stochastic growth. Knowing that a wealth-recursive Markov solution exists, the model is simulated at the quarterly frequency. Our solution method relies on three ingredients: a time-shift, as proposed by Dumas and Lyasoff (2012), a wealth-recursive equilibrium, and a finite-period approximation of the infinite-horizon problem. The simulated moments are then compared to their empirical counterparts. The data sample focuses on the U.S. and an aggregate of the other G10 countries to build the ROW. The sample period is 1973.IV–2010.IV.

In the simulation, the model matches the characteristics of the U.S. and ROW GDP and aggregate consumption, as well the equity and risk-free bond returns. The endowment process matches the mean, standard deviation, and autocorrelation of the growth rates and H.P-filtered series of U.S. GDP, as well as its cross-country correlation with the ROW GDP. The model produces equity excess returns that are large and volatile in both countries. Equity excess returns are also predictable, using the price-dividend ratio and the wealth-consumption ratio, as in the data. The mean and the volatility of the risk-free rates are also in line with their empirical counterpart. The exchange rate is slightly less volatile than in the data, but the average return on the currency carry trade is in line with the data. The exchange rate change exhibits a low, negative correlation with relative consumption growth. The next exports, as a fraction of GDP, however, is less volatile in the model than in the data.

The model is used to assess the magnitude and volatility of international capital flows and foreign holdings. The model features not only unexpected valuation changes but also expected returns on foreign investments; the model can thus shed light on the current debate on the size of expected valuation effects and their importance in assessing the sustainability of the U.S. current account.

In the simulation, the U.S. invests in ROW equity and the ROW invests in U.S. equity. But the magnitudes of these gross positions differ: the U.S. holds more foreign equity assets than foreign equity liabilities. The reverse is true for bonds, and the U.S. in a net borrower. Overall the U.S. borrows from the ROW and invests in the ROW equity. The U.S. gross equity positions are even more volatile in the model than in the data, reflecting both the expected and unexpected valuation shocks. The bonds positions, to the contrary, are not volatile, in line with their empirical counterpart. The changes in expected excess returns lead to changes in optimal portfolio holdings and thus international capital flows. In our calibration, the gross equity flows are more volatile than in the data, although the volatility of the net equity positions is close to the one in the data. In comparison, the net debt flows are as smooth in the model as in the data.

The model thus highlights the key role of expected returns, i.e. expected valuation changes, in the volatility of international capital flows. The volatility of expected and unexpected equity returns seems to account, to a first order, for the volatility of international capital positions and flows in the data.

A study of the volatility of equity and bond assets and flows requires four features: (i) the markets must be incomplete such that equity and bond gross asset positions and flows can be defined separately in a meaningful way; (ii) portfolio holdings must be time-varying such that capital flows exist; (iii) expected returns must be large and time-varying for the model to be consistent with the prices of the underlying assets; and (iv) the model must be solved globally. A very large literature studies international holdings and capital flows, but few papers satisfy the four conditions above. Let us rapidly review the most relevant strands of the literature.

A large literature studies the equity home bias — a statement about the puzzlingly low amount of international diversification in the data compared to the one implied by standard neoclassical models. Important contributions include Baxter and Jermann (1997), Lewis (1999), Coeurdacier (2009), Nieuwerburgh and Veldkamp (2009), Coeurdacier and Gourinchas (2011) and Heathcote and Perri (2013). This literature is too large to be summarized here — the database Scopus returns more than 230 published articles over the last 25 years with the expressions "home bias" and "international" in the title or abstract; we refer the reader to the recent and excellent survey proposed by Coeurdacier and Rey (2013). Few papers in this literature feature large and time-varying risk premia: exceptions are Stathopoulos (2012), who considers habit preferences, and Benigno and Nisticò (2012), who introduce model uncertainty and long run consumption risk. Colacito, Croce, Ho, and Howard (2014) study international capital flows in a production economy in the spirit of Backus, Kehoe, and Kydland (1992).

Another large literature studies the sustainability of the current account imbalances and the size of potential valuation effects on foreign holdings. In a seminal paper, Gourinchas and Rey (2007) find a higher return on US external assets than on its external liabilities. Curcuru, Dvorak, and Warnock (2010) offer alternative estimates. Ahmed, Curcuru, Warnock, and Zlate (2015) describe the different components of international portfolio flows. Important contributions on the current account imbalances include Kraay and Ventura (2000), Ventura (2001), Caballero, Farhi, and Gourinchas (2008), and Devereux and Sutherland (2010).

Finally, a recent literature studies the impact of market incompleteness on the capital flows and exchange rate puzzles, notably the Backus and Smith (1993) puzzle (see Backus and Smith, 1993) and the forward premium puzzle. The Backus and Smith (1993) puzzle refers to the perfect correlation between exchange rate changes and relative consumption in a complete market model with CRRA preferences. In the data, the correlation is small and negative. The forward premium puzzle refers to the deviations from the uncovered interest rate parity and the large currency carry trade excess returns (see Fama, 1984; Tryon, 1979) Notable contributions in this literature include the work by Alvarez, Atkeson, and Kehoe (2002), Chari, Kehoe, and McGrattan (2002), Bacchetta and Wincoop (2006), Corsetti, Dedola, and Leduc (2008), Alvarez, Atkeson, and Kehoe (2009), Pavlova and Rigobon (2010), Pavlova and Rigobon (2012), Bruno and Shin (2014), Maggiori (2015), and Favilukis, Garlappi, and Neamati (2015). Solving optimal portfolio problems in incomplete markets is challenging. Earlier solutions in the context of closed economies with specific preferences (e.g., log utility) or endowment processes include Dumas (1989), Wang (1996), Cochrane, Longstaff, and Santa-Clara (2008), Longstaff and Wang (2012), and Martin (2013). Our model, existence theorem, and solution method can be used in the context of closed economies with heterogenous agents.

Recent attempts have been made to improve the solution method. Devereux and Sutherland (2011) and Tille and van Wincoop (2010) propose a second-order approximation method, subsequently used in several papers. In a key contribution, Rabitsch, Stepanchuk, and Tsyrennikov (2015) however, show that this solution method is inaccurate in the presence of heteroscedasticity and nonlinearities, which are key features of our model. Our solution method therefore is global and does not require any second-order approximation. Evans and Hnatkovska (2005) suggest a different approximation based on a constant wealth ratio, which is not applicable in our case.

The papers closer to ours are Gourinchas, Rey, and Govillot (2010), Stepanchuk and Tsyrennikov (2015), Dumas, Lewis, and Osambela (2014), Maggiori (2015), and Chien, Lustig, and Naknoi (2015): the first two consider differences in risk-aversion across countries when markets are, respectively, complete or incomplete; the third one studies differences of opinion in complete markets; the last two papers feature incomplete markets to study respectively the impact of differences in financial development or the Backus and Smith puzzle (1993) puzzle. These authors only consider constant risk premia. Our work builds on these papers to deliver an incomplete market model with time-varying risk premia. The time-variation in expected return is key, as changes in expected returns translate into changes in optimal portfolio holdings and therefore capital flows.

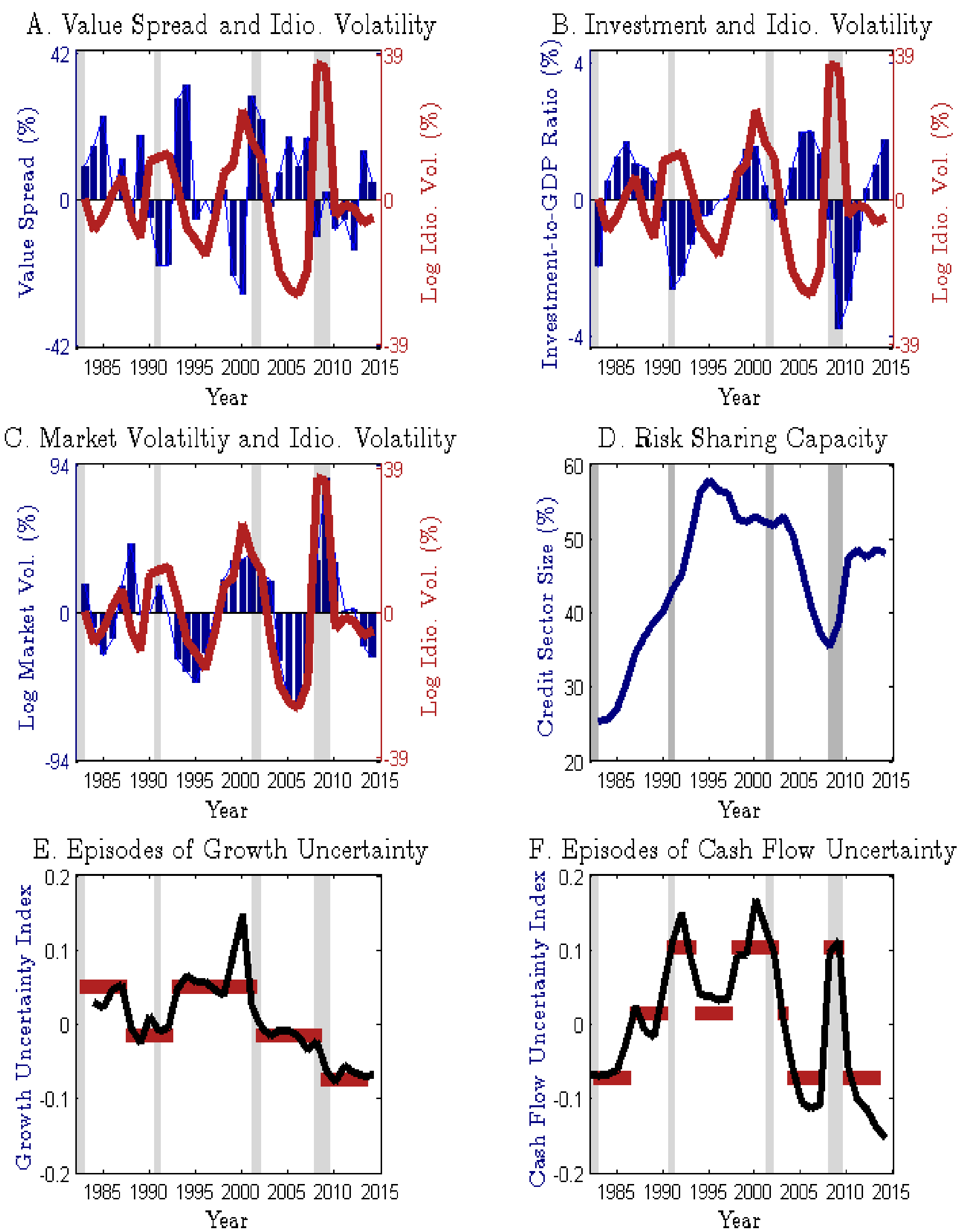

Figure 1-1: Impact of Idiosyncratic Volatility

This figure illustrates the dynamics of idiosyncratic volatility of stock returns. It highlights the comovement pattern of the average idiosyncratic volatility with the cross-sectional spread between value and growth stock returns (i.e., value spreads) and the aggregate investment.

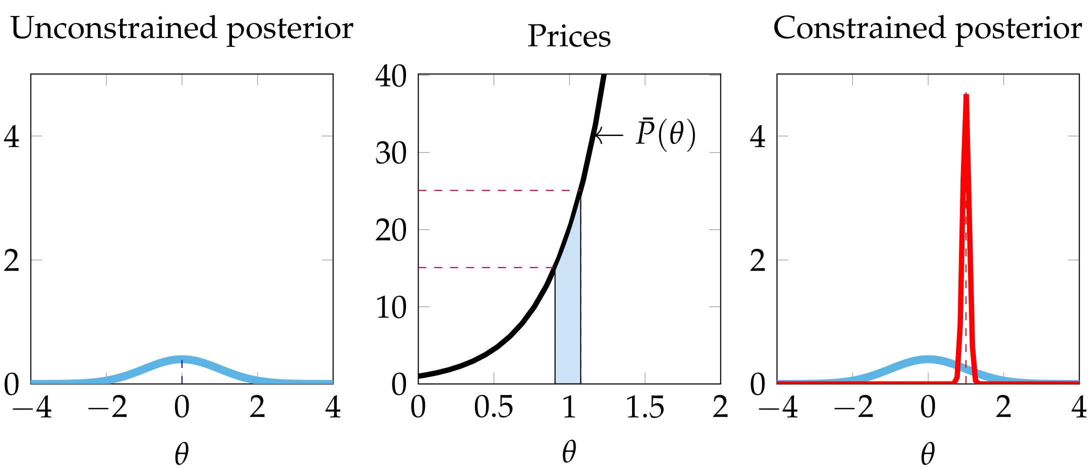

Figure 1-2: An example of an "informative" economic restriction on the parameters.

The left panel plots the unconstrained posterior about θ based on cash-flow data. The middle panel plots the price function $\bar{P}(\theta)$. The dashed lines represent the confidence band for the mean of price observations. The right panel plots the constrained posterior about θ based on both cash-flow and price data.

Chapter 2

Embrace or Fear Uncertainty: Growth Options, Limited Risk Sharing, and Asset Prices

2.1 Model

In this section, I develop a continuous-time general equilibrium model with two sectors: the consumption goods sector and the investment goods sector. I summarize the model's features as follows. First, there are two types of agents in the economy: experts and households. Experts with population $\varkappa$ indexed by $f \in \mathbb{F} \equiv [-\varkappa, 0]$ are the only ones who can manage and trade firm's assets; households with population 1 indexed by $h \in \mathbb{H} \equiv [0, 1]$ provide labor. Second, firms hold two classes of assets: assets in place generate consumption goods; growth options create new assets in place. Although assets are irreversible at the aggregate level, they can be continuously traded among firms. Third, outputs are affected by firm-specific idiosyncratic shocks, which are unobservable to agents except the expert who literally manages the assets. The information asymmetry makes it possible for the expert to take hidden actions such as shirking efforts or stealing for the private benefit at the expense of diffused shareholders. To deal with the agency problem, the expert is restricted to become a blockholder who owns a significant fraction of the firm's equity. Fourth, the volatilities of idiosyncratic shocks are time varying and driven by aggregate shocks. These are the uncertainty shocks. Experts respond optimally to the uncertainty shocks in making decisions on investment and hiring for the firm. Fifth, all agents can trade financial contracts in capital markets where a full set of Arrow-Debreu securities are available. Sixth, I deliberately cast the model in continuous time, because the continuous-time formation allows me to characterize the key equilibrium relationships by cleaner expressions and conveniently summarize the equilibrium conditions by a set of coupled ordinary differential equations.

2.1.1 Firms and Technologies

There is a continuum of infinitely-lived firms in the consumption goods sector. Each firm is managed by an expert and indexed by $f \in \mathbb{F}$. Existing assets in place depreciate with a constant rate δ, and new assets in place are built based on a combination of existing growth options and investment goods newly produced in the investment goods sector. Growth options can be used to create new assets in place when investment opportunities arrive.

Consumption goods firms. Each firm's assets consist of assets in place and growth options. The equity of the firm f is freely traded, and it is the claim on the dividends generated by the assets in place and the value added by the creation of new assets in place from growth options.

Assets in place. Denote by k_t the aggregate amount of assets in place in the economy and by $k_{f,t}$ the amount of assets in place held by the individual firm f, where $t \in [0, \infty)$ is the time index. Assets in place $k_{f,t}$ held by the firm f generates output at rate $y_{f,t}$, over the period $[t, t + \mathrm{d}t]$,

$$y_{f,t} = k_{f,t}^{\varphi} \ell_{f,c,t}^{1-\varphi}, \tag{2.1}$$

where $\varphi \in (0, 1)$ captures the capital share in production and $\ell_{f,c,t}$ represents the labor input for production. When held by the expert f, the existing assets in place evolves according to

$$\frac{\mathrm{d}k_{f,t}}{k_{f,t}} = -\delta \mathrm{d}t + \sigma \mathrm{d}Z_t + \mathrm{d}A_{f,t},$$

where δ is the constant depreciation rate, Z_t is a Brownian motion describing an aggregate shock in the economy, and $A_{f,t}$ is a cumulative firm-specific process describing the idiosyncratic cash flows. The shocks $\mathrm{d}Z_t$ and $\mathrm{d}A_{f,t}$ can be interpreted as the aggregate and the idiosyncratic (short-term) cash flow shocks, respectively.[1] The cash-flow uncertainty is defined as the volatility of the idiosyncratic shock $\mathrm{d}A_{f,t}$:

$$\nu_{c,t} \equiv \mathrm{vol}\left(\mathrm{d}A_{f,t}\right).$$

The exposure to the aggregate shock is constant σ; however, the exposure to the idiosyncratic shock, denoted by $\nu_{c,t}$, is stochastic. The idiosyncratic volatility $\nu_{c,t}$ represents an aggregate economic condition because the prospects of short-term cash flows become blurred when $\nu_{c,t}$ increases.

Growth options, investment opportunities, and new assets in place. Growth options allow the firm to create new assets in place when investment opportunities arrive. Specifically, the growth options are intangible assets associated with ideas of technological innovations such as R&D projects, blueprints, and patents; however, these innovative ideas are necessary but not alone sufficient to realize the final commercial benefits. The investment opportunities are business opportunities or ideas to commercialize the technological innovations and turn them into commercial benefits through making real investment. The arrivals of investment opportunities are firm-specific, so the model has the feature that investment is lumpy at the firm level but smooth at the aggregate level, which is consistent with the data. This modeling feature is cru-

[1] This way of modeling capital accumulation and production is actually equivalent to the conventional TFP shock method, where the adjustment cost function is not only homogeneous with respect to capital stock k_t but also the TFP shock a_t, i.e., the adjustment cost, is $\iota(g_t)a_t k_t$, if there is an adjustment cost function.

cial since it allows me to study the time-series properties of cross-sectional investment dispersions.[2]

I denote it by $s_{f,t}$ the amount of growth options held by the firm f and denote it by p_t the unit price of growth options. Although the aggregate amount of growth options is assume to be constant $s_t = \bar{s}$,[3] the firms can freely trade growth options with each other at the price p_t. The existing stock of growth options stay constant over period $[t, t + \mathrm{d}t]$; that is, $\mathrm{d}s_{f,t} = 0$.

Let $M_{f,t}$ be the firm-specific point process that describes the number of investment opportunities obtained by firm f up to time t. Upon the reception of a new investment opportunity at time t (i.e., $\mathrm{d}M_{f,t} = 1$), the firm f decides whether to invest or not. This is similar to Khan and Thomas (2008), which explicitly accounts for the micro-level investment spikes and the fluctuation of the extensive margin of investments. The firm can undertake an investment only upon payment of its fixed adjustment cost ϖ, specifically by forfeiting $\varpi p_t s_{f,t}$ of current consumption goods. The fixed adjustment cost denominated in the units of growth options captures the essence of the real-option model of investment in Jovanovic (2009) and Ai and Kiku (2013), among others.[4] Denote by $u_{f,t}$ the variable characterizing whether the firm f undertakes an investment or not. If it is undertaken, $u_{f,t} = 1$; otherwise, $u_{f,t} = 0$. Upon $u_{f,t} = 1$, the firm creates new assets in place $k_{f,t}^{new}$ using the technology:

$$k_{f,t}^{new} = \underbrace{\varepsilon_{f,t}}_{\text{idio. IST shock}} \times \underbrace{m\left(s_{f,t}, g_{f,t}\right) k_t}_{\text{inputs and scaling}}, \tag{2.2}$$

[2]This follows the standard modeling in the literature on the implications of lumpy investment for aggregate macroeconomic dynamics, such as Khan and Thomas (2008).

[3]The exogenous stationarity in the relative-growth option scale is standard in the asset pricing literature and growth literature (e.g., Gomes, Kogan, and Zhang, 2003; Ai and Kiku, 2013). In order to avoid tracking an extra endogenous state variable, models either assume constant growth options or assume that the growth options grow proportionally to the total assets in place. In my model, the value of growth options is linear in k_t. Then, the relative growth option scale is $p_t s_t / (q_t k_t) = \bar{s}\hat{p}_t / \hat{q}_t$, where p_t and q_t are prices of growth options and assets in place, respectively. In equilibrium, $\bar{s}\hat{p}_t / \hat{q}_t$ is a stationary process.

[4]In Jovanovic (2009) and Ai and Kiku (2013), the growth options fully depreciate after being used for investment. In macroeconomic models studying the role of micro-level nonconvex costs of investment adjustment in generating nonlinear aggregate investment dynamics, the fixed adjustment costs are usually denominated by profits (e.g., Bloom, 2009) or denominated by labor (e.g., Khan and Thomas, 2008).

where $\varepsilon_{f,t}$ is the idiosyncratic investment-specific (IST) shock to capture the idiosyncratic shock on the quality of investment opportunities, $s_{f,t}$ is the amount of existing growth options, and $g_{f,t}$ is the input of investment goods. To create new assets in place with the amount of $\varepsilon_{f,t}m(s_{f,t}, g_{f,t})k_t$, the capital stock of growth options $s_{f,t}$ is prefixed (i.e., not adjustable at time t after the realization of $\varepsilon_{f,t}$); however, the firm can choose the investment goods input $g_{f,t}$ optimally conditional on the realization of $\varepsilon_{f,t}$. The cost of purchasing investment goods is $\tau_t g_{f,t}$, where τ_t is the equilibrium market price of investment goods.[5] The production function $m(s, g)$ is a constant-elasticity-of-substitute (CES) function. In particular, I assume that $m(s, g)$ has the Cobb-Douglas functional form with the share of capital to be α; that is, $m(s, g) \equiv s^{1-\alpha}g^{\alpha}$.[6]

Once the firm f receives an investment opportunity at time t (i.e. $\mathrm{d}M_{f,t} = 1$) and implements it (i.e. $u_{f,t} = 1$), the new assets in place $k^{new}_{f,t}$ are created from the growth options with the rate:

$$i_{f,t} \equiv k^{new}_{f,t}/k_t = \varepsilon_{f,t} m\left(s_{f,t}, g_{f,t}\right).$$

Growth uncertainty. The idiosyncratic IST shock $\varepsilon_{f,t}$ in (2.2) is assumed to be independently distributed over time and across firms, to avoid having to keep track of the distribution of $\varepsilon_{f,t}$ as an infinitely-dimensional state variable. The assumption of idiosyncratic investment risks have been adopted by the macroeconomics literature (e.g., Khan and Thomas, 2008; Bachmann and Bayer, 2014), and by the asset pricing

[5]It should be noted that, similar to Gomes, Kogan, and Zhang (2003), although the tangible assets is complementary to the intangible assets investment at the aggregate level, each individual expert cannot really internalize the aggregate impact of their tangible asset holdings, and hence in the decentralized economy the assets in place investment has zero complementarity for the R&D investments. Therefore, there is an externality in the economy, which makes the allocations in a competitive equilibrium not necessarily identical to those solved by the social planner's problem.

[6]Similar to Gomes, Kogan, and Zhang (2003), I assume that the scale of new assets in place created from growth options is linear in the aggregate assets in place. This guarantees that the ratios of the aggregate new to the aggregate existing assets in place and of the aggregate value of growth options to the aggregate value of assets in place are both stationary over time. Other examples include Ai, Croce, and Li (2012) where the aggregate investment is assumed to be a deterministic function of the aggregate investment goods by restricting the cross-sectional distribution of idiosyncratic investment shocks, and Ai and Kiku (2013) where the aggregate assets in place and the aggregate growth options are assumed to follow an exogenous common stochastic trend which is the arrival intensity of new growth options. Other general equilibrium asset pricing models with growth options, such as Papanikolaou (2011) and Kogan, Papanikolaou, and Stoffman (2015), have aggregate assets in place to follow a mean-reverting stationary process in equilibrium, so the stationary ratios are guaranteed endogenously.

literature (e.g., Gomes, Kogan, and Zhang, 2003; Ai, Croce, and Li, 2012). However, a key difference in this model is that the variance of the distribution of idiosyncratic growth opportunity quality shocks is time varying. More precisely, I assume that $\varepsilon_{f,t}$ has a symmetric distribution $\varepsilon_{f,t} \sim \mathbf{N}\left(0, \nu_{g,t}^2\right)$. The growth uncertainty is the standard deviation of the IST shock

$$\nu_{g,t} \equiv \text{std}\left(\varepsilon_{f,t}\right)$$

where the growth uncertainty $\nu_{g,t}$ evolves randomly over time. The idiosyncratic volatility $\nu_{g,t}$ represents an aggregate economic condition because the prospects of investment opportunities become blurred when $\nu_{g,t}$ increases.

Optimal investment. Here, I describe the investment decision of an expert when an investment opportunity arrives. Because experts can choose the variable utilization rate of growth options $u_{f,t}$, the optimal investment decision-making can be decomposed into two steps. First, conditioning on the full utilization (i.e., $u_{f,t} = 1$), the expert maximizes the net present value $\Pi_{f,t}$ by choosing investment goods input $g_{f,t}$. Given the price of assets in place, denoted by q_t, and the price of investment goods τ_t, the optimization problem and the net present value $\Pi_{f,t}$ can be expressed as

$$\max_{g_{f,t}} \Pi_{f,t} \equiv q_t k_{f,t}^{new} - \tau_t g_{f,t}, \quad \text{with } k_{f,t}^{new} \equiv i_{f,t} k_t \text{ and } i_{f,t} \equiv \varepsilon_{f,t} s_{f,t}^{1-\alpha} g_{f,t}^{\alpha}. \tag{2.3}$$

In other words, the net present value $\Pi_{f,t}$ is the market value of the new assets in place k_t^{new} minus its investment cost $\tau_t g_{f,t}$. The optimal input of investment goods is strictly convex in the idiosyncratic investment shock $\varepsilon_{f,t}$ and is linear in the stock of existing growth options $s_{f,t}$:

$$g_{f,t} = o_g s_{f,t} \varepsilon_{f,t}^{\frac{1}{1-\alpha}} \left(\frac{q_t k_t}{\tau_t}\right)^{\frac{1}{1-\alpha}}, \quad \text{with constant } o_g \equiv \alpha^{\frac{1}{1-\alpha}}. \tag{2.4}$$

This is the result of a simple intratemporal optimization based on (2.3). The optimal investment condition (2.4) is similar to the standard q-theory of investment developed by Hayashi (1982) where the optimal investment is directly linked to the marginal q of assets in place ($\frac{q_t k_t}{\tau_t}$), denominated by the investment goods. Yet there is one key difference. Because $0 < \alpha < 1$, the optimal investment goods demand $g_{f,t}$ is a convex function of marginal q instead of a concave function, which is a direct result of the Oi-Hartman-Abel-Caballero channel. Given the price of growth options, denoted by p_t, the optimal present value of newly created assets in place can be expressed as $\Pi_{f,t} \equiv \overline{\pi}_{f,t} s_{f,t} p_t$ where the optimal net present value rate $\overline{\pi}_{f,t}$ has the analytical expression:

$$\overline{\pi}_{f,t} = o_\pi \varepsilon_{f,t}^{\frac{1}{1-\alpha}} \left(\frac{q_t k_t}{p_t^{1-\alpha} \tau_t^{\alpha}} \right)^{\frac{1}{1-\alpha}}, \text{ where } o_\pi \equiv (1-\alpha) o_g \text{ is a constant.} \tag{2.5}$$

And the optimal investment rate is $i_{f,t} = o_\iota \varepsilon_{f,t}^{\frac{1}{1-\alpha}} \left(\frac{q_t k_t}{\tau_t} \right)^{\frac{\alpha}{1-\alpha}} s_{f,t}$, where $o_\iota = o_g^{\alpha}$ is a constant. In the second step, the expert chooses the utilization rate $u_{f,t} \in \{0, 1\}$ to maximize the profits from creating new assets in place. It is clear that a firm will absorb its fixed cost $\varpi p_t s_{f,t}$ to undertake the investment opportunity if the investment profit rate $\overline{\pi}_{f,t}$ is at least ϖ. It follows immediately that a firm will undertake the investment opportunity if its idiosyncratic IST shock $\varepsilon_{f,t}$ lies at or above some threshold values. Because all agents face the same option-exercising problem, the threshold value only depends on the aggregate state variables. I denote the exercising boundary by ξ_t, which is characterized as follows:

$$\varepsilon_{f,t} \geq \xi_t \quad \text{if and only if} \quad \overline{\pi}_{f,t} = o_\pi \varepsilon_{f,t}^{\frac{1}{1-\alpha}} \left(\frac{q_t k_t}{p_t^{1-\alpha} \tau_t^{\alpha}} \right)^{\frac{1}{1-\alpha}} \geq \varpi.$$

From (2.5), it leads the analytical expression for the exercising threshold ξ_t:

$$\xi_t = o_\xi \varpi^{1-\alpha} \left(\frac{q_t k_t}{p_t^{1-\alpha} \tau_t^{\alpha}} \right)^{-1}, \quad \text{where} \quad o_\xi \equiv o_\pi^{\alpha-1} \quad \text{is a constant.} \tag{2.6}$$

Thus, the profit rate of growth options for the firm f is

$$\pi_{f,t} = (\overline{\pi}_{f,t} - \varpi)\mathbf{1}_{\{\varepsilon_{f,t} \geq \xi_t\}}.$$

Investment goods firms. There is a representative firm in the investment goods sector. It uses the labor of households to produce the investment goods needed to create new assets in place in the consumption goods sector. More precisely, the production function for the investment goods output rate over the infinitesimal interval $[t, t + \mathrm{d}t)$ is

$$g_t = z_\iota \ell_{\iota,t}, \tag{2.7}$$

where z_ι is the average total productivity factor in the investment goods sector and $\ell_{\iota,t}$ is the total labor demand to produce investment goods g_t. I assume constant return to scale for labor input for simplification.[7]

Spot markets. The outputs (consumption goods and investment goods) and the firm's assets (assets in place and growth options) are traded in perfectly competitive spot markets. There is one spot price in each market, and this spot price is only determined by the aggregate state of the economy, even though the participants are heterogeneous. The spot prices are market-clearing prices for which each single participant is a price taker.

2.1.2 Uncertainty Shocks

The cash-flow uncertainty $\nu_{c,t}$ and the growth uncertainty $\nu_{g,t}$ move stochastically. The uncertainty shocks are large shocks driving the state variable ν_t, which has a one-to-one correspondence to the 2-tuple $(\nu_{g,t}, \nu_{c,t})$. I assume that the growth uncertainty $\nu_{g,t}$

[7] Similar to Papanikolaou (2011) and Kogan, Papanikolaou, and Stoffman (2015), the production function of the investment goods only works with fixed amount of capital input. But, to guarantee profits on the capital input and thereby generate meaningful share prices of investment goods firms, Papanikolaou (2011) assumes decreasing returns to scale for the labor input. Like Kogan, Papanikolaou, and Stoffman (2015), my focus is not to link investment-minus-consumption (IMC) portfolio returns to aggregate shocks in the economy. So, I also assume constant return to scale for the labor input.

follows a 2-state homogeneous continuous-time Markov chain taking values in the set $\mathcal{V}_g \equiv \left\{\nu_g^L, \nu_g^H\right\}$, where $\nu_g^L < \nu_g^H$. Similarly, I assume that the cash-flow uncertainty $\nu_{c,t}$ follows a 2-state homogeneous continuous-time Markov chain taking values in the set $\mathcal{V}_c \equiv \{\nu_c^L, \nu_c^H\}$ where $\nu_c^L < \nu_c^H$. For simplicity, the growth uncertainty process and the cash-flow uncertainty process are assumed to move independently with the transition rate matrices $\mathcal{Q}_g$ and $\mathcal{Q}_c$, respectively,

$$\mathcal{Q}_g \equiv \begin{bmatrix} \lambda^{(\nu_g^L,\nu_g^H)} & -\lambda^{(\nu_g^L,\nu_g^H)} \\ -\lambda^{(\nu_g^H,\nu_g^L)} & \lambda^{(\nu_g^H,\nu_g^L)} \end{bmatrix} \quad \text{and} \quad \mathcal{Q}_c \equiv \begin{bmatrix} \lambda^{(\nu_c^L,\nu_c^H)} & -\lambda^{(\nu_c^L,\nu_c^H)} \\ -\lambda^{(\nu_c^H,\nu_c^L)} & \lambda^{(\nu_c^H,\nu_c^L)} \end{bmatrix}.$$

The transition intensity for ν_t is denoted as $\lambda^{(\nu_t,\nu')}$ which only depends on $\mathcal{Q}_g$ and $\mathcal{Q}_c$.

2.1.3 Preferences

Both experts and households have stochastic differential utility of Duffie and Epstein (1992a,b). This preference is a continuous-time version of the recursive preferences proposed by Kreps and Porteus (1978b), Epstein and Zin (1989c), and Weil (1990). The Epstein-Zin-Weil recursive preference has become a standard preference in asset pricing and macro literature to capture the reasonable joint behavior of asset prices and macroeconomic quantities. More precisely, the utility is defined recursively as follows:

$$U_0 = \mathbb{E}_0\left[\int_0^\infty \mathbf{f}(c_t, U_t)\mathrm{d}t\right],$$

where

$$\mathbf{f}(c_t, U_t) \equiv \rho\left[\frac{u(c_t)}{((1-\gamma)U_t)^{\theta^{-1}-1}} - \theta U_t\right], \text{ with } \theta \equiv \frac{1-\gamma}{1-\psi^{-1}}$$

and the felicity function $\mathbf{f}(c_t, U_t)$ is an aggregator over current consumption rate c_t and future utility level U_t. The coefficient ρ is the rate parameter of time preference, γ is the risk-aversion parameter for one-period consumption, and ψ is the parameter of elasticity of intertemporal substitution (EIS) for deterministic consumption paths. The period utility function has the form:

$$u(c_t) = \frac{c_t^{1-\psi^{-1}}}{1-\psi^{-1}}.$$

The preference between consumption and leisure can be viewed as a special case of the KPR preference (King, Plosser, and Rebelo, 1988, 2002) and the GHH preference (Greenwood, Hercowitz, and Huffman, 1988), where leisure is not appreciated or work is not undervalued. Thus, the labor supply is inelastic.[8]

To ensure stationarity between experts and households, I assume that agents die independently of each other according to a Poisson process with constant intensity μ. New agents are born at the same rate μ with a fraction $\frac{\varkappa}{1+\varkappa}$ as experts and $\frac{1}{1+\varkappa}$ as households, so the measure of households and the measure of experts both remain constant. The wealth of agents who die is bestowed on the newly born on a per-capita basis. The subjective discount factor ρ captures the effective time preference because I make it include the adjustment for the likelihood of death for each agent (see Gârleanu and Panageas, 2015).

2.1.4 Labor Markets

The aggregate labor supply is one since each household inelastically supplies their labor-hours endowment. On the demand side, the labor choices are endogenous in both the consumption goods sector and the investment goods sector. Driven by the ag-

[8] The inelastic labor supply is adopted for several reasons: (1) this is a useful benchmark that allows a direct comparison to the existing literature on production-based asset pricing and investment in incomplete markets where inelastic labor supply is the most common assumption (e.g., Danthine and Donaldson, 2002; Angeletos, 2007; Guvenen, 2009; Kogan, Papanikolaou, and Stoffman, 2015); (2) this allows us to focus on illustrating our key mechanism that results from the financial friction; (3) in the literature, it is shown that this assumption together with limited risk sharing can provide reasonable asset pricing implications; and (4) this is actually a not-far-off approximation to the reality. Wages have risen in the U.S. over long periods of time, but the proportion of time spent working has not changed very much. This old stylized fact has recently been reconfirmed by Ramey and Francis (2009). I can investigate the extent to which labor supply choice can be endogenized without compromising the performance on the asset pricing side, and study how frictions in labor markets can help improve the quantitative performance of our model. These are definitely important questions to understand but out of the scope of this paper. I leave them as the future research agenda.

gregate shocks in the economy, the share of aggregate labor supply allocated between the two sectors is time-varying. The labor demand in the investment goods sector is straightforward. It is determined by the aggregate investment goods demand. According to (2.7), it holds that the aggregate labor demand in the investment goods sector is

$$\ell_{\iota,t} = z_{\iota}^{-1} g_t.$$

The optimal labor demand of each firm is a static (i.e., state-by-state) optimization problem. This is because the firm's employment $\ell_{f,c,t}$ affects only the profits rate $y_{f,t} - w_t \ell_{f,c,t}$ at time t. As a result, the optimal labor demand maximizes profits state-by-state at time t. As a result, given wages, the optimal labor demand can be solved only based on the intratemporal Euler equation, which is independent of the intertemporal optimizations. They are summarized by Proposition 1. All the detailed proofs of propositions and corollaries can be found in the online appendix.

Proposition 1 (Optimal Labor Demand and Output)**.** *Given w_t and $k_{f,t}$, labor demand and output are linear in $k_{f,t}$ and decreasing in w_t: $\ell_{c,f,t} = \ell(w_t) k_{f,t}$ and $y_{f,t} = y(w_t) k_{f,t}$, where $\ell(w_t) \equiv \left[\frac{(1-\varphi)}{w_t}\right]^{1/\varphi}$ and $y(w_t) \equiv \left[\frac{(1-\varphi)}{w_t}\right]^{\frac{1-\varphi}{\varphi}}$.*

From Proposition 1, the aggregate labor demand and the aggregate output, by the Law of Large Numbers, are

$$\ell_{c,t} \equiv \int_{f \in \mathbb{F}} \ell_{c,f,t} \mathrm{d}f = \ell(w_t) k_t \quad \text{and} \quad y_t \equiv \int_{f \in \mathbb{F}} y_{f,t} \mathrm{d}f = y(w_t) k_t, \quad \text{respectively.}$$

The endogenous labor reallocation between the consumption goods sector ($\ell_{c,t}$) and the investment goods sector ($\ell_{\iota,t}$) play a crucial role in understanding the cross-section of stock returns.

2.1.5 Firms' Payouts and Assets' Holding Returns

Because firms face no financing frictions, the irrelevance theorems (see Modigliani and Miller, 1958; Miller and Modigliani, 1961) hold for firms' capital structures and payout policies. Similar to most macroeconomic and asset pricing models, I assume that the firms are all-equity firms and pay out earnings. So, for each firm f, its payout is equal to the profits from assets in place $\varphi y(w_t)k_{f,t}\mathrm{d}t$ plus value added by growth options $\pi_{f,t}s_{f,t}p_t\mathrm{d}M_{f,t}$ minus expenditures for assets in place $q_t i_t k_{f,t}\mathrm{d}t$:

$$\mathrm{d}D_{f,t} = \underbrace{\varphi y(w_t)k_{f,t}\mathrm{d}t + \pi_{f,t}s_{f,t}p_t\mathrm{d}M_{f,t}}_{\text{total profits}} \quad - \quad \underbrace{q_t i_t k_{f,t}\mathrm{d}t.}_{\text{expenditure for new assets in place}}$$

Here, $D_{f,t}$ is the cumulative payout of firm f and the incremental payout $\mathrm{d}D_{f,t}$ can theoretically be negative in the model. Because I do not particularly specify the external financing frictions of the firm, the negative payout can be interpreted as issuing new equity by the firm. In fact, numerically, under my baseline calibration, the payout turns out to be negative only in the extreme ranges of the state space which are visited by the economy very rarely in simulations.

The total payout of a firm can be decomposed into two components. One is due to the capital stock of assets in place and the other is due to the capital stock of growth options. They are relevant for the valuation of assets in place and growth options, respectively. More precisely, the decomposition based on the accounting for assets in place and growth options is as follows:

$$\mathrm{d}D_{f,t} = \underbrace{\left[\varphi y(w_t)k_{f,t} - q_t i_t k_{f,t}\right]\mathrm{d}t}_{\text{payout due to assets in place}} + \underbrace{\pi_{f,t}p_t s_{f,t}\mathrm{d}M_{f,t}.}_{\text{payout due to growth options}}$$

Moreover, the instantaneous holding returns of assets in place and growth options for experts are, respectively,

$$\mathrm{d}R^{\mathbf{k}}_{f,t} = \mu^{\mathbf{k}}_{f,t}\mathrm{d}t + (\sigma^q_t + \varphi\sigma)\mathrm{d}Z_t + \sum_{\nu\neq\nu_t} \varsigma_t^{q,(\nu_t,\nu)}\mathrm{d}N_t^{(\nu_t,\nu)} + \mathrm{d}A_{f,t}, \tag{2.8}$$

and

$$\mathrm{d}R^{\mathbf{s}}_{f,t} = \mu^{\mathbf{s}}_{f,t}\mathrm{d}t + (\sigma^p_t + \varphi\sigma)\mathrm{d}Z_t + \sum_{\nu\neq\nu_t} \varsigma_t^{p,(\nu_t,\nu)}\mathrm{d}N_t^{(\nu_t,\nu)} + \pi_{f,t}\mathrm{d}M_{f,t}, \tag{2.9}$$

where the drift terms $\mu^{\mathbf{k}}_{f,t}$ and $\mu^{\mathbf{s}}_{f,t}$ can be found in the online appendix and the diffusion terms σ^p_t and σ^q_t and the jump size terms $\varsigma^{q,(\nu_t,\nu)}_t$ and $\varsigma^{p,(\nu_t,\nu)}_t$ are defined in the beginning of Section 4.3.

2.1.6 Financial Markets

There is a full set of short-term financial contracts available to all agents. Each financial contract has zero net supply. And their prices are always normalized at one. All agents trade those short-term contracts in a perfectly competitive capital market. The contracts are traded continuously at time t with the payoffs realized at the end of the infinitesimal interval $[t, t + \mathrm{d}t]$. Among the financial contracts, one is the short-term risk-free bond with payoff $1 + r_t \mathrm{d}t$, one is traded on the aggregate Z_t shock with a contingent payoff $1 + r^Z_t \mathrm{d}t + \mathrm{d}Z_t$, one is traded on the growth uncertainty shock with a contingent payoff $1 + r^{\nu_g}_t \mathrm{d}t + \left[\mathrm{d}N^{(\nu_{g,t},\nu'_g)}_t - \lambda^{(\nu_{g,t},\nu'_g)}\mathrm{d}t\right]$, one is traded on the cash-flow uncertainty shock with a contingent payoff $1 + r^{\nu_c}_t \mathrm{d}t + \left[\mathrm{d}N^{(\nu_{c,t},\nu'_c)}_t - \lambda^{(\nu_{c,t},\nu'_c)}\mathrm{d}t\right]$, a continuum of short-term contracts are traded on idiosyncratic cash flow shocks $W_{f,t}$ with payoffs $1 + r^W_{f,t}\mathrm{d}t + \mathrm{d}W_{f,t}$ for all $f \in \mathbb{F}$, and a continuum of short-term contracts are traded on idiosyncratic investment shocks $\varepsilon_{f,t}\mathrm{d}N_{f,t}$ with payoffs $1 + r^N_{f,t}\mathrm{d}t + \left[\varepsilon_{f,t}\mathrm{d}N_{f,t} - \mathbb{E}(\varepsilon_{f,t})\lambda \mathrm{d}t\right]$ for all $f \in \mathbb{F}$. In sum, the financial market is complete.

The expected payoffs $r_t, r^Z_t, r^{\nu_g}_t, r^{\nu_c}_t, r^W_{f,t}$, and $r^N_{f,t}$ are endogenously determined by the market clearing conditions. Importantly, later I shall show that the expected rate of returns are time varying, driven by the cash-flow uncertainty shocks and the growth uncertainty. Moreover, each firm's equity can be freely traded. However, because a full set of contingent claims are already available to all agents, the equities of firms become redundant in terms of spanning the contingent space. Without loss of generality, I assume that a firm's equity on its assets in place and equity on its growth options can be traded separately.

Although a full set of contingent claims are available, the market can be endogenously incomplete due to lack of commitments. Later I show that due to zero commitment in long-term contracts and a moral hazard problem, experts face portfolio constraints including limited access to short-term financial contracts on particular idiosyncratic risks.

2.1.7 Moral Hazard

I now introduce an agency conflict induced by the separation of ownership and control. The diffused investors fund the firm controlled by the expert. In contrast to the neoclassical model in which the firm-specific cash flow process $A_{f,t}$ and the investment opportunity process $M_{f,t}$ are exogenously specified, those processes in my model are affected by expert's unobservable actions. Specifically, the expert is able to secretly divert cash flows and investment opportunities from the firm under her control, which I describe explicitly as follows.

Hidden actions in cash flows. The expert f's hidden action $a^A_{f,t} \in \left[0, \bar{a}^A\right]$ determines the expected rate of idiosyncratic cash flow shock $\mathrm{d}A_{f,t}$, so that

$$\mathrm{d}A_{f,t} = -a^A_{f,t}\mathrm{d}t + \nu_{c,t}\mathrm{d}W_{f,t},$$

where $W_{f,t}$ is a Brownian motion capturing the firm f's underlying (short-term) idiosyncratic cash flows. The expert controls the drift, but not the idiosyncratic volatility of the process $A_{f,t}$.[9] When the expert takes the action $a^A_{f,t}$, she enjoys a flow of pecuniary private benefits with intensity $a^A_{f,t}\phi q_t k_{f,t}$ over $[t, t+\mathrm{d}t]$. Here, $0 \leq \phi < 1$, which means that the stealing is inefficient. More precisely, the variable $a^A_{f,t}$ can be interpreted

[9] A common setting is that there is a menu of projects whose risk characteristics are common knowledge and yet experts can choose which to be undertaken (e.g., Cadenillas, Cvitanic, and Zapatero, 2007). My model can be extended to allow the expert to choose among multiple projects and the main mechanism is not altered. Moreover, the expert can also affect the volatility by secretly injecting funds from her own hidden saving accounts. This is not the focus on this paper. To rule out the possibility of altering the idiosyncratic volatility secretly through injecting cash flows from the hidden saving account, I assume that the expert cannot affect the idiosyncratic volatility of (short-term) cash flows and that her net worth is observable, which is without loss of generality due to the Revelation-Principle type of results (e.g., DeMarzo and Fishman, 2007). The similar assumptions are also adopted in DeMarzo, Fishman, He, and Wang (2012), among others. In particular, DeMarzo and Sannikov (2006) restrict the stealing process to be Lipschitz continuous. And, it is well known that all sample paths of a standard Brownian motion have infinite total variation. Thus, idiosyncratic volatility cannot be secretly altered in their model.

as the fraction of cash flows that the expert diverts for her pecuniary private benefits and the parameter ϕ captures the expert's net pecuniary benefits per dollar diverted. Given the linearity, this framework of stealing is effectively equivalent to the binary setup in which the expert can steal (i.e., $a^A_{f,t} = \overline{a}^A$) or not steal (i.e., $a^A_{f,t} = 0$).

Hidden actions in growth options. Similarly, I assume that the investment opportunity $M_{f,t}$ is affected by the expert's unobserved action in the following way,

$$\mathrm{d}M_{f,t} = (1 - a^M_{f,t})\mathrm{d}N_{f,t}, \tag{2.10}$$

where $N_{f,t}$ is a Poisson count process that describes the number of investment opportunities of firm f that arrive up to time t. The intensity of the underlying Poisson process $N_{f,t}$ is λ. The action $a^M_{f,t}$ is binary.[10] In particular, the expert does not steal when $a^M_{f,t} = 0$ and steals when $a^M_{f,t} = 1$. When the expert takes the action $a^M_{f,t} = 0$, she obtains zero pecuniary private benefit. By contrast, when the action of $a^M_{f,t} = 1$ is taken by the expert, she steals the investment opportunity from the firm to launch new ventures in her own private account.[11] The lumpy pecuniary benefit is $\phi\pi_{f,t}p_t s_{f,t}$, where the coefficient ϕ equals to the expert's net pecuniary benefits per dollar diverted.

Severity of agency problem. Here, $1 - \phi$ can be interpreted as the deadweight loss rate of stealing incurred by the expert. Thus, ϕ represents the severity of the agency problem and, as I show later, captures the minimum levels of incentives required to prevent the expert from stealing.

Formulating the optimal contracting problem. The history paths in

$$\mathcal{H}_t \equiv \sigma\left(\{Z_{t'}, \nu_{g,t'}, \nu_{c,t'}, A_{f,t'}, M_{f,t'} : 0 \le t' \le t, f \in \mathbb{F}\}, \{\varepsilon_{f,t'} : 0 \le t' < t, f \in \mathbb{F}\}\right)$$

[10]As in the free cash flow case, the binary-action setting is equivalent to the continuous-action setting when pecuniary private benefit is linear in actions. However, the binary-action setting has a more natural interpretation for the diversion of investment opportunities.

[11]The investment opportunity is non-replicable; otherwise, the value of growth options is infinity, which is pathological.

are observable and contractable. Denote H_t to be a particular history path in $\mathcal{H}_t$. Similar to He and Krishnamurthy (2011), Brunnermeier and Sannikov (2014) and Di Tella (2014), I take the approach of short-term contracts: the relation only lasts from t to $t + \mathrm{d}t$; at time $t + \mathrm{d}t$, the contract (relation) ends. In fact, the optimal contract can be implemented by a sequence of short-term contracts even when long-term contracts are available in my setting, if experts are assumed to have zero commitment to long-term contracts and to be able to modify the older contracts and offer new contracts at any time in a costless manner. The intuition is that the participation constraint for the diffused investors is always binding in each short period $[t, t + \mathrm{d}t]$, which simply is the capital market non-arbitrage condition, and the incentive compatibility of the contracts in each short period $[t, t + \mathrm{d}t]$ is not affected by the history; hence the current contract is always subject to being replaced by new contracts and hence recontracting continuously is optimal.[12]

Right after the realization of the history H_t, the expert and her diffused investors meet up and enter contracts for $[t, t + \mathrm{d}t]$. The expert f offers contracts to her diffused investors (the principals in the contracting relation), which specifies the upfront lumpy payment $P_{f,t}$ collected from the diffused investors and the cash payment $P_{f,t} + \mathrm{d}F_{f,t}$ paid from the expert to her diffused investors over $[t, t + \mathrm{d}t]$. Here, $\mathrm{d}F_{f,t}$ is the net cash payment by the expert over $[t, t + \mathrm{d}t]$. The cumulative net payment process $F_{f,t}$ and the upfront payment $P_{f,t}$ are required to be adapted to the filtration $\mathcal{H}_t$. Thus, a short-term contract consists of a pair of functions $(P_{f,t}, \mathrm{d}F_{f,t})$ specifying the investors' upfront payment to the expert at t and the net cash payment of the expert to the investors over $[t, t + \mathrm{d}t]$. Let $\mathcal{C}_{f,t} \equiv \mathcal{C}_f(\mathcal{H}^t) \equiv (P_{f,t}, \mathrm{d}F_{f,t})$ represent the contract offered by the expert.

[12] There three important points here. First, it is worth pointing out that this result is very different from the equivalence results of long- and short-term contracts such as Fudenberg, Holmstrom, and Milgrom (1990). Those papers investigate sufficient conditions under which a sequence of short-term contracts can achieve the same efficiency level for long-term contracts where commitment is nonzero. Second, if I assume the expert is committed to long-term contracts, like in DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin, and Rochet (2007), DeMarzo and Fishman (2007), and DeMarzo, Fishman, He, and Wang (2012), the tractability will be worsened with the main mechanism remaining unchanged. Third, the short-term contracting problem I focus on in this paper is analogous to the contracting problem in a one-period principal-agent problem (e.g., Holmstrom and Tirole, 1997).

The participation constraint for the diffused investors is

$$0 = \mathbb{E}_t^a \left[\mathrm{d}F_{f,t} + \left(P_{f,t} + \mathrm{d}F_{f,t}\right) \frac{\mathrm{d}\Lambda_t}{\Lambda_t} \right], \tag{2.11}$$

where Λ_t is the stochastic discount factor of households and is determined in the Walrasian equilibrium with details illustrated in Section 2.3.3 and $\mathbb{E}^a$ is the expectation operator under the probability measure that is induced by the hidden action processes. The participation constraint for the expert is endogenously mingled with her occupational choice: she endogenously decides whether to become a household by selling off all productive assets (assets in place and growth options). That is, by choosing $k_{f,t} = s_{f,t} = 0$, the expert f endogenously becomes a household. However, in the equilibrium, the expert never converts herself to a household; the expert is always offered a high enough risk premia for holding the productive assets. This is a result of the limited market participation assumption that households cannot choose to become experts due to the lack of the specialized knowledge or skills.[13]

Given any sequence of contracts characterized by $\mathcal{C}_f \equiv \{\mathcal{C}_{f,t} : t \geq 0\}$, the expert will choose an optimal sequence of strategies $\mathcal{S}_f \equiv \{\mathcal{S}_{f,t} : t \geq 0\}$ that specifies the hidden actions, the consumption, and investment choices

$$\mathcal{S}_{f,t} \equiv \left(a_{f,t}^A, a_{f,t}^M, c_{f,t}^{\mathrm{e}}, k_{f,t}, s_{f,t}, g_{f,t}\right).$$

More precisely, for a sequence of contracts $\mathcal{C}_f$, the expert f's net worth follows the law

[13] This is different from the limited market participation of certain financial markets for risky financial securities (e.g., Mankiw and Zeldes, 1991; Allen and Gale, 1994; Basak and Cuoco, 1998; Vissing-Jorgensen, 2002a; Guvenen, 2009) in two folds: first, households cannot invest or manage firms' assets and thus the economy stops functioning without experts; second, households can freely trade all financial securities in capital markets.

of motion,

$$\mathrm{d}n^{\mathrm{e}}_{f,t} = -\ c^{\mathrm{e}}_{f,t}\mathrm{d}t\ +\ \underbrace{q_t k_{f,t}\mathrm{d}R^{\mathbf{k}}_{f,t} + p_t k_t s_{f,t}\mathrm{d}R^{\mathbf{s}}_{f,t}}_{\text{gains from assets holdings}}\ -\ \underbrace{\mathrm{d}F_{f,t}}_{\text{contract pay}} \tag{2.12}$$
$$+\ \underbrace{a^{A}_{f,t}\phi q_t k_{f,t}\mathrm{d}t + a^{M}_{f,t}\phi\pi_{f,t}p_t s_{f,t}\mathrm{d}N_{f,t}}_{\text{privite benefits from shirking}},$$

where the instantaneous returns from holding the assets can be found in Equation (2.8). Further, given prices and wages, the expert f chooses the strategies $\mathcal{S}_f$ to solve

$$U\left(H_0, n^{\mathrm{e}}_{f,0}; \mathcal{C}_f\right) = \max_{\mathcal{S}_f} \mathbb{E}^{a}_{0}\left[\int_0^\infty \mathbf{f}\left[c^{\mathrm{e}}_{f,t'}, U(H_{t'}, n^{\mathrm{e}}_{f,t'}; \mathcal{C}_f)\right]\mathrm{d}t'\right],$$

where the net worth process $\left\{n^{\mathrm{e}}_{f,t'} : t' \geq 0\right\}$ includes the potential private benefits from taking a sequence of actions $\left\{a^{A}_{f,t}, a^{M}_{f,t} : t' \geq 0\right\}$ and the gain from the holding of firms' assets by taking choosing $\left\{k_{f,t'}, s_{f,t'}, g_{f,t'} : t' \geq 0\right\}$.

The contract-strategy pair $(\mathcal{C}_f, \mathcal{S}_f)$ is *feasible* if it satisfies the solvency constraint $n^{\mathrm{e}}_{f,t} \geq 0$ for all history paths $H_t \in \mathcal{H}_t$. A feasible contract-strategy pair $(\mathcal{C}_f, \mathcal{S}_f)$ is *optimal* if there is no other pair that provides the same payoff to the diffused investors and a higher expected utility to the expert. And, a feasible pair $(\mathcal{C}_f, \mathcal{S}_f)$ is *incentive compatible* if the optimal strategy $\mathcal{S}_f$ implements the efficient actions $a^{A}_{f,t} = a^{M}_{f,t} = 0$ all the time given the contracts $\mathcal{C}_f$. To characterize an optimal contract-strategy pair, I start with a Revelation-Principle type result as in the context of mechanism design: given any contract-strategy pair $(\mathcal{C}_f, \mathcal{S}_f)$ for the expert, there exits an incentive-compatible contract-strategy pair $(\mathcal{C}^*_f, \mathcal{S}^*_f)$ with the same payoff to diffused investors and a weakly higher expected utility for the expert. It allows me to focus on the incentive-compatible contract-strategy pairs for finding optimal contracts. The intuition is straightforward (e.g., DeMarzo and Fishman, 2007) and the rigorous proof is in the online appendix. I denote $\mathbb{E}$ to be the expectation operator under the probability measure induced by the efficient actions.

More precisely, an incentive-compatible contract-strategy pair $(\mathcal{C}_f, \mathcal{S}_f)$ is optimal if it maximizes the value function of the expert f, given prices and wages,

$$U\left(H_t, n^{\mathrm{e}}_{f,t}\right) = \max_{\mathcal{C}_f} U\left(H_t, n^{\mathrm{e}}_{f,t}; \mathcal{C}_f\right) \tag{2.13}$$

subject to the participation constraint of diffused investors in (2.11), where the value function $U\left(H_t, n^{\mathrm{e}}_{f,t}; \mathcal{C}_f\right)$ is the optimal utility achieved by the optimal strategy $\mathcal{S}_f$ given the contracts $\mathcal{C}_f$ with

$$U\left(H_t, n^{\mathrm{e}}_{f,t}; \mathcal{C}_f\right) = \mathbb{E}_t\left[\int_t^{\infty} \mathbf{f}\left[c^{\mathrm{e}}_{f,t'}, U^{\mathrm{e}}(H_{t'}, n^{\mathrm{e}}_{f,t'}; \mathcal{C}_f)\right] \mathrm{d}t'\right].$$

In summary, I incorporate the optimal contracting problem into a dynamic general equilibrium framework, and thus the optimal contracts are part of the fixed point solution for a (Walrasian) general equilibrium. More precisely, at the decentralized level, optimal contracts are derived as if agents take the aggregate price and wage dynamics as given; in turn, the aggregate level, the demand and supply formed from the aggregation of decentralized optimal contracts need to match so that the markets are cleared. To finally solve the optimal contracts and the general equilibrium, it is useful to first provide a characterization (i.e., a necessary condition for the optimal contracts) and an implementation mechanism for the optimal contracts. After incorporating the characterization and the implementation, the general equilibrium framework with optimal contracting becomes a rather standard model for asset pricing and risk sharing in incomplete markets.

2.1.8 Concentrated Risk: the Optimal Contracts and Implementations

Characterization of optimal contracts. Because the cumulative payment process $F_{f,t}$ is adapted to $\mathcal{H}_t$, the net cash payment specified by the contract can be formulated as follows[14]

$$\mathrm{d}F_{f,t} = \mu^F_{f,t}\mathrm{d}t \;+\; \underbrace{(1-\beta^A_{f,t})q_t k_{f,t}\mathrm{d}A_{f,t} + \left[\pi_{f,t} - \beta^M_{f,t}(\varepsilon_{f,t})\right] p_t s_{f,t}\mathrm{d}M_{f,t}}_{\text{contingent payments on idiosyncratic shocks}}$$
$$+ \;\underbrace{\beta^Z_{f,t}\sigma\mathrm{d}Z_t + \sum_{\nu\neq\nu_t}\beta^{V,(\nu_t,\nu)}_{f,t}\mathrm{d}N^{(\nu_t,\nu)}_t,}_{\text{contingent payments on aggregate shocks}}$$

where the (functional) processes $\mu^F_{f,t}$, $\beta^A_{f,t}$, $\beta^M_{f,t}(\cdot)$, $\beta^Z_{f,t}$, and $\beta^{V,(\nu_t,\nu)}_{f,t}$ are adaptive to the filtration $\mathcal{H}_t$. Particularly, the function $\beta^M_{f,t}(\cdot)$ can be nonlinear. Plugging the expression of $\mathrm{d}F_{f,t}$ above into (2.12), the dynamics of the net worth of expert f can be rewritten as follows

$$\mathrm{d}n^{\mathrm{e}}_{f,t} = \underbrace{(\phi - \beta^A_{f,t})q_t k_{f,t} a^A_{f,t}\mathrm{d}t + \left[\phi\pi_{f,t} - \beta^M_{f,t}(\varepsilon_{f,t})\right] p_t s_{f,t} a^M_{f,t}\mathrm{d}N_{f,t}}_{\text{terms altered by } a^A_{f,t} \text{ and } a^M_{f,t}}$$
$$+ \left[\text{terms independent of } a^A_{f,t} \text{ or } a^M_{f,t}\right].$$

Thus, for any incentive-compatible contracts (i.e., satisfying $a^A_{f,t} = a^M_{f,t} = 0$), it must satisfy the following two conditions:

$$\beta^A_{f,t} \geq \phi \quad \text{and} \quad \beta^M_{f,t}(\varepsilon) \geq \phi\pi_{f,t}(\varepsilon) \text{ for all } \varepsilon.$$

It is straightforward that the optimal contracts must satisfy that $\beta^A_{f,t} \equiv \phi$ and $\beta^M_{f,t} \equiv \phi\pi_{f,t}$ for all f and t. This is because the expert is risk averse and hence wishes to dump all the idiosyncratic risks $\mathrm{d}W_{f,t}$ and $\mathrm{d}N_{f,t} - \lambda\mathrm{d}t$, while at the same time households

[14] The net payment $\mathrm{d}F_{f,t}$ does not depend on the idiosyncratic shocks not associated with the firm f, because all agents are risk averse and avoid unnecessary idiosyncratic risk exposures. Another important feature is that jumps with random sizes affect the payoff process (e.g., Sung, 1997; Biais, Mariotti, Rochet, and Villeneuve, 2010; Hoffmann and Pfeil, 2010). In general, it leads to nonlinear optimal contracts. However, the linearity of optimal contracts in this setting is due to two main reasons: first, it follows the timing convention of taking hidden actions after the realization of shocks (e.g., DeMarzo and Fishman, 2007; Edmans and Gabaix, 2011; Edmans, Gabaix, Sadzik, and Sannikov, 2012); second, the private pecuniary benefit is contingent and proportional to the payoff.

can buy it for free due to their capacity to fully diversify any idiosyncratic risks.

Implementation of optimal contracts. I now characterize the optimal contracts in terms of an optimal mechanism. In particular, I consider the implementation of optimal contracts based on simple financial contracts, including firms' stock shares, options, risk free bond, and indices tracking aggregate states. Specifically, the expert f achieves her optimal incentive-compatible contracting results in the following ways: (1) she buys and manages assets in place $k_{f,t}$ and growth options $s_{f,t}$; (2) she sells $1-\phi$ fraction of the firm's equity to her diffused shareholders; and (3) she trades indices in perfect financial markets. In summary, this implementation features blockholding and active trading on indices.[15] Rather than attempting to describe all possible implementations, I shall focus on this simple yet empirically relevant mechanism.

The following proposition describes the detailed specifications of the implementations and establishes their optimality.

Proposition 2 (Blockholding and Indexation)**.** *For each $f \in \mathbb{F}$, suppose the expert f has initial net worth $n_{f,0}$. She gets infinite penalty unless the solvency condition $n_{j,t} \geq 0$ holds. The expert f is required to hold ϕ share of the firm f's equity. The expert f is not allowed to diversify or hedge away the idiosyncratic risks of firm f as a blockholder. She can trade a risk-free bond and financial indices tracking aggregate shocks. Under the capital market configuration, it is optimal for each expert f to choose actions $a^A_{f,t} = a^M_{f,t} = 0$.*

[15] An alternative theory that generates the same results is that the experts bargain with diffused shareholders' for the rents, subject to some capital market constraints. Rents can be efficient. For example, Myers (2000) and Lambrecht and Myers (2007, 2008, 2012) show how rents can align managers' and shareholders' interests if the managers maximize the present value of rents subject to a capital market constraint. Also, Eisfeldt and Papanikolaou (2013) develop a model in which the outside option of the key talent determines the share of firm cash flows that accrue to shareholders. This outside option varies systematically and renders firms depending more on the key talents riskier from shareholders' perspective.

2.1.9 Aggregation: Investments and Productions

In this section, I discuss the aggregation results on the production and investment side of the economy.[16] An important feature of our model is that, the evolution of the aggregate assets in place follows the standard process as in the neoclassical growth model, though heterogeneous firms make decentralized investment decisions in my economy.

More precisely, under incentive-compatible optimal contracts, the law of motion for the aggregate capital stock of assets in place $k_t = \int_{f\in\mathbb{F}} k_{f,t}\mathrm{d}f$ is not affected by any particular idiosyncratic shocks; it can be characterized as follows:

$$\mathrm{d}k_t = (i_t - \delta)k_t\mathrm{d}t + \sigma k_t \mathrm{d}Z_t.$$

Here $i_t \equiv \int_{f\in\mathbb{F}} i_{f,t}\mathbf{1}_{\{\varepsilon_{f,t}>\xi_t\}}\mathrm{d}N_{f,t}$ is the aggregate investment rate with the analytical formula:

$$i_t = \underbrace{\overbrace{\mathcal{G}_\alpha\left(\nu_{g,t};\xi_t\right)}^{\uparrow \text{ in } \nu_{g,t}}}_{\text{marginal efficiency of investment}} \times \underbrace{\overbrace{o_\iota\left(\frac{q_t k_t}{\tau_t}\right)^{\frac{\alpha}{1-\alpha}}}^{\uparrow \text{ in } q_t/\tau_t}}_{\text{conventional q theory}}. \tag{2.14}$$

The term $\mathcal{G}_\alpha(\nu_{g,t};\xi_t)$ acts as the endogenous marginal efficiency of investment and the shocks that drive its fluctuations are endogenous aggregate investment shocks. As shown in Proposition 3, the endogenous investment shock $\mathcal{G}_\alpha(\nu_{g,t};\xi_t)$ is increasing in $\nu_{g,t}$ and decreasing in ξ_t. In fact, it has the following analytical expression

$$\mathcal{G}_\alpha(\nu_{g,t};\xi_t) \equiv \lambda \times \nu_{g,t}^{\frac{1}{1-\alpha}} \times \bar{\Gamma}_\alpha\left(\xi_t/\nu_{g,t}\right), \tag{2.15}$$

and the function $\bar{\Gamma}_\alpha(\cdot)$ is defined as

$$\bar{\Gamma}_\alpha(\xi_t/\nu_{g,t}) \equiv o_\alpha \times \bar{\Gamma}\left(\frac{1}{2}\left(\xi_t/\nu_{g,t}\right)^2, \frac{2-\alpha}{2-2\alpha}\right) \tag{2.16}$$

where $\bar{\Gamma}(\cdot,\cdot)$ is the standard upper incomplete gamma function and o_α is a universal

[16] The detailed derivations in this section can be found in the online appendix.

constant.[17]

The function $\mathcal{G}_\alpha(\nu_{g,t};\xi_t)$ is the key to understand how growth uncertainty can increase aggregate investment. More precisely, I decompose the function into a multiplication of two terms that capture the complementary effect and the option effect of growth uncertainty on the aggregate investment

$$\underbrace{\mathcal{G}_\alpha(\nu_{g,t};\xi_t)}_{\text{marginal efficiency of investment}} = \underbrace{\nu_{g,t}^{\frac{\alpha}{1-\alpha}}}_{\text{complementary effect}} \times \underbrace{\left[\overbrace{\nu_{g,t}}^{\text{intensive margin}} \times \overbrace{\bar{\Gamma}_\alpha(\xi_t/\nu_{g,t}) \times \lambda}^{\text{extensive margin}}\right]}_{\text{option effect}}. \tag{2.17}$$

The first term $\nu_{g,t}^{\frac{\alpha}{1-\alpha}}$ captures the exogenous positive effect of growth uncertainty on aggregate investment. It is similar to the Oi-Hartman-Abel-Caballero effect: the flexible inputs, which can be adjusted after productivity shocks are realized and are complementary to the productivity of the capital, create optionality in the capital. In my case, when $\alpha = 0$, investment goods are not needed in creating new assets in place (i.e. zero complementarity). As a result, the Oi-Hartman-Abel-Caballero effect disappears. The second term $\nu_{g,t} \times \bar{\Gamma}_\alpha(\xi_t/\nu_{g,t}) \times \lambda$ captures the option effect of exercising investment opportunities. The variable $\nu_{g,t}$ captures the intensive margin effect caused by growth uncertainty shocks: the high-quality investment opportunities are likely to be more profitable when growth uncertainty increases. Moreover, the function $\bar{\Gamma}_\alpha(\xi_t/\nu_{g,t})$ captures the extensive margin effect caused by growth uncertainty shocks: more experts endogenously choose to make investment for fixed exercising boundary ξ_t. However, the exercising boundary is endogenously adjusted in the economy, which can partly offset the exogenous effect of increasing growth uncertainty; this is called wait-and-see effect (e.g., Miao and Wang, 2007; Bloom, 2009).

[17] The upper incomplete gamma function is defined as $\bar{\Gamma}(x_1, a_1) = \int_{x_1}^{\infty} x^{a_1 - 1} e^{-x} \mathrm{d}x$ and $o_\alpha \equiv 2^{(2\alpha-1)/(2-2\alpha)} \pi^{-1/2}$ where π is the mathematical constant but not the profit rate of growth options π.

Another important feature of our model is that, the aggregate output is Cobb-Douglas with diminishing return to scale in the aggregate assets in place as in a standard neoclassical growth model, though each firm's optimal output is linear in terms of its own assets in place. More precisely, the aggregate output of the consumption goods sector is

$$y_t = k_t^{\varphi} \ell_{c,t}^{1-\varphi},$$

and under incentive-compatible optimal contracts, the aggregate output of the investment goods sector is

$$g_t = \mathcal{G}_\alpha(\nu_{g,t}; \xi_t) \times o_g \left(\frac{q_t k_t}{\tau_t} \right)^{\frac{1}{1-\alpha}}.$$

Intuitively, the aggregate investment goods demand g_t is affected by the growth uncertainty $\nu_{g,t}$ similarly to the aggregate investment rate i_t through the function $\mathcal{G}_\alpha(\nu_{g,t}; \xi_t)$.

The aggregate payout from assets in place and the aggregate profit from growth options are summarized as follows. Particularly, the analytical formula of aggregate profit from growth options provides intuitions that help understand how growth uncertainty shocks affect the value of growth options. More precisely, under incentive-compatible optimal contracts, the aggregate net payout due to assets in place is

$$d_t = \varphi y_t - q_t i_t k_t$$

and the aggregate profit of growth options is $\Pi_t \equiv \pi_t p_t \bar{s}$ where

$$\frac{\pi_t}{\lambda} = \underbrace{\varpi \left(\frac{\nu_{g,t}}{\xi_t} \right)^{\frac{1}{1-\alpha}}}_{\text{effective payoff}} \times \underbrace{\bar{\Gamma}_\alpha(\xi_t / \nu_{g,t})}_{\text{adj. prob. of exercising}} - \underbrace{\varpi}_{\text{strike price}} \times \underbrace{\bar{\Phi}(\xi_t / \nu_{g,t}),}_{\text{prob. of exercising}} \tag{2.18}$$

where the function $\bar{\Gamma}_\alpha(\cdot)$ is defined in (2.16), and the function $\bar{\Phi}(\cdot)$ is the complementary cumulative distribution function (CCDF) of a standard normal variable. The net profit rate of growth options derived in (2.18) resembles the well-known Black-Scholes-Merton option pricing formula (Black and Scholes, 1973; Merton, 1973b). In the following decomposition, the term $\varpi\left(\frac{\nu_{g,t}}{\xi_t}\right)^{\frac{1}{1-\alpha}}$ can be viewed as the effective payoff when the option is exercised, the term $\bar{\Gamma}_\alpha(\xi_t/\nu_{g,t})$ can be interpreted as the adjusted likelihood of exercising the option ($\varepsilon_{f,t} > \xi_t$), the term ϖ is strike price, and the term $\bar{\Phi}(\xi_t/\nu_{g,t})$ is the actual probability of exercising the growth option ($\varepsilon_{f,t} > \xi_t$).

Thus, keeping the exercising boundary ξ_t fixed, the profit rate of growth options is monotonically increasing in growth uncertainty. This is summarized in the following proposition.

Proposition 3 (Optionality)**.** *Under incentive-compatible optimal contracts, the aggregate profit rate of growth options (π_t) is strictly increasing in growth uncertainty ($\nu_{g,t}$) and strictly decreasing in the exercising boundary ξ_t fixed. At the same time, the endogenous investment efficiency That is, the partial derivatives always hold the following signs: $\partial\pi_t/\partial\nu_{g,t} > 0$, $\partial\pi_t/\partial\xi_t < 0$, $\partial\mathcal{G}_\alpha/\partial\nu_{g,t} > 0$, and $\partial\mathcal{G}_\alpha/\partial\xi_t < 0$.*

2.2 Equilibrium

I denote η_t to be the market price of risk for the aggregate shock z_t, and denote $\kappa_t^{(\nu_t,\nu)}$ to be the market price of risk for the uncertainty shock $N_t^{(\nu_t,\nu)}$. The market prices of the aggregate shocks depend only upon the aggregate state variables, though the economy is full of idiosyncratic shocks. I define the de-trended asset prices and human capital after taking out the economy's balanced growth path as follows: $\tilde{p}_t \equiv p_t/k_t^{\varphi}$, $\tilde{q}_t \equiv q_t/k_t^{\varphi-1}$, and $\tilde{\hbar}_t \equiv \hbar_t/k_t^{\varphi}$. I conjecture that the prices $\tilde{q}_t$, $\tilde{p}_t$, and the human capital $\tilde{\hbar}_t$ follow the Ito processes with jumps

$$\frac{\mathrm{d}\tilde{q}_t}{\tilde{q}_t} = \mu_t^q \mathrm{d}t + \sigma_t^q \mathrm{d}Z_t + \textstyle\sum_{\nu\neq\nu_t} \varsigma^{q,(\nu_t,\nu)} \mathrm{d}N_t^{(\nu_t,\nu)},$$

and

$$\frac{\mathrm{d}\tilde{p}_t}{\tilde{p}_t} = \mu_t^p \mathrm{d}t + \sigma_t^p \mathrm{d}Z_t + \textstyle\sum_{\nu\neq\nu_t} \varsigma^{p,(\nu_t,\nu)} \mathrm{d}N_t^{(\nu_t,\nu)},$$

and

$$\frac{\mathrm{d}\tilde{\hbar}_t}{\tilde{\hbar}_t} = \mu_t^{\hbar}\mathrm{d}t + \sigma_t^{\hbar}\mathrm{d}Z_t + \textstyle\sum_{\nu \neq \nu_t} \varsigma^{\hbar,(\nu_t,\nu)}\mathrm{d}N_t^{(\nu_t,\nu)}.$$

Here, the coefficient functions μ_t^p, μ_t^q, $\mu_t^{\hbar}$, σ_t^p, σ_t^q, $\sigma_t^{\hbar}$, $\varsigma^{p,(\nu_t,\nu)}$, $\varsigma^{q,(\nu_t,\nu)}$, and $\varsigma^{\hbar,(\nu_t,\nu)}$ are endogenously determined in equilibrium. In equilibrium, the prices and human capital are driven by the aggregate shocks Z_t and $N_t^{(\nu_t,\nu)}$, but not by the idiosyncratic shocks $\{W_{f,t}\}_{f\in\mathbb{F}}$, $\{N_{f,t}\}_{f\in\mathbb{F}}$, or $\{\varepsilon_{f,t}\}_{f\in\mathbb{F}}$. Later, I shall show that the productivity shock $\mathrm{d}Z_t$ does not affect the fluctuations of the de-trended prices (see Corollary 2). So, it holds that $\sigma_t^q \equiv \sigma_t^p \equiv \sigma_t^{\hbar} \equiv 0$.

2.2.1 Households' Optimization Problem

Given prices and wages, households face a standard portfolio problem with labor income. Although they cannot manage or trade firm assets, they can freely access to a complete financial market. Taking the processes of market price of risk η_t and $\left\{\kappa^{(\nu_t,\nu)} : \nu_t, \nu \in \mathcal{V}\right\}$ and the prices p_t, q_t, τ_t and the wages w_t as given, they solve the following utility maximization problem

$$U_{h,0}^{\mathrm{h}} = \max_{\left\{c_{h,t}^{\mathrm{h}}, \hat{\vartheta}_{h,t}^{\mathrm{h}}, \hat{\vartheta}_{h,t}^{\mathrm{h},(\nu_t,\nu)}\right\}_{t\geq 0}} \mathbb{E}_0\left[\int_0^{\infty} \mathbf{f}(c_{h,t}^{\mathrm{h}}, U_{h,t}^{\mathrm{h}})\mathrm{d}t\right] \tag{2.19}$$

subject to the solvency constraint $n_{h,t}^{\mathrm{h}} \geq 0$ the dynamic budget constraint

$$\frac{\mathrm{d}n_{h,t}^{\mathrm{h}}}{n_{h,t}^{\mathrm{h}}} = \left[\mu_{h,n,t}^{\mathrm{h}} - \hat{c}_{h,t}^{\mathrm{h}}\right]\mathrm{d}t + \underbrace{\sigma_{h,n,t}^{\mathrm{h}}\mathrm{d}Z_t + \sum_{\nu\neq\nu_t} \varsigma_{h,n,t}^{\mathrm{h},(\nu_t,\nu)}\left[\mathrm{d}N_t^{(\nu_t,\nu)} - \lambda^{(\nu_t,\nu)}\mathrm{d}t\right]}_{\text{only aggregate risk expsoures}}, \tag{2.20}$$

where the expected growth rate on net worth (pre consumption) is $\mu_{h,n,t}^{\mathrm{h}}$ and the aggregate risk exposures are

$$\sigma_{h,n,t}^{\mathrm{h}} = \underbrace{\hat{\vartheta}_{h,t}^{\mathrm{h}}}_{\text{indices}} + \underbrace{(1-\phi)\frac{q_t k_t}{n_t^{\mathrm{h}}}(\sigma_t^q + \varphi\sigma) + (1-\phi)\frac{p_t\bar{s}}{n_t^{\mathrm{h}}}(\sigma_t^p + \varphi\sigma)}_{\text{diversified equity holdings}} + \underbrace{\varrho\frac{\hbar_t}{n_t^{\mathrm{h}}}(\sigma_t^{\hbar} + \varphi\sigma)}_{\text{human capital}},$$

and

$$\varsigma^{\mathrm{h},(\nu_t,\nu)}_{h,n,t} = \underbrace{\hat{\vartheta}^{\mathrm{h},(\nu_t,\nu)}_{h,t}}_{\text{indices}} + \underbrace{(1-\phi)\frac{q_t k_t}{n^{\mathrm{h}}_t}\varsigma^{q,(\nu_t,\nu)}_t + (1-\phi)\frac{p_t \bar{s}}{n^{\mathrm{h}}_t}\varsigma^{p,(\nu_t,\nu)}_t}_{\text{diversified equity holdings}} + \underbrace{\varrho\frac{\hbar_t}{n^{\mathrm{h}}_t}\varsigma^{\hbar,(\nu_t,\nu)}_t}_{\text{pledgeable human capital}}.$$

Here, the shares $\hat{\vartheta}^{\mathrm{h}}_{h,t} \equiv \vartheta^{\mathrm{h}}_{h,t}/n^{\mathrm{h}}_{h,t}$ and $\hat{\vartheta}^{\mathrm{h},(\nu_t,\nu)}_{h,t} \equiv \vartheta^{\mathrm{h},(\nu_t,\nu)}_{h,t}/n^{\mathrm{h}}_{h,t}$ characterize the household's positions in risky assets. Here, the hatted consumption rate $\hat{c}^{\mathrm{h}}_{h,t}$ denotes the consumption rate normalized by the household h's net worth, i.e. $\hat{c}^{\mathrm{h}}_{h,t} \equiv c^{\mathrm{h}}_{h,t}/n^{\mathrm{h}}_{h,t}$. Because all households are homogenous up to their net worth levels, they choose homogeneous risk exposures in equity holdings and pledgeable human capital holdings. In other words, they hold the diversified equity portfolios and the pledgeable human capital proportional to their net worth. The total net worth of all households is $n^{\mathrm{h}}_t \equiv \int_{h\in\mathbb{H}} n^{\mathrm{h}}_{h,t}\mathrm{d}h$. Because the firm-level idiosyncratic risks $\{W_{f,t}, N_{f,t}\}_{t\geq 0}$ are priced at zero by households in equilibrium, the risk averse household will never have any exposure to them in equilibrium. The expected growth rate $\mu^{\mathrm{h}}_{h,n,t}$ includes three components: (i) the expected returns from the index holdings, (ii) the expected returns from the diversified equity holdings, and (iii) the identical labor income rate $\hat{w}_t \equiv w_t/n^{\mathrm{h}}_t$, which is guaranteed by perfect labor insurances among all households. More detailed explanations are in the online appendix.

2.2.2 Experts' Optimization Problem

Given prices and wages, experts face a joint problem of optimal portfolio allocation and optimal real investment, subject to portfolio constraints. The portfolio constraints arise endogenously as a result of incentive compatibility constraints in a moral hazard problem (see Section 2.1.7). Experts can continuously trade firm's assets in spot markets. Meanwhile, they can also access to the short-term financial contracts in the capital markets. Taking the processes of market price of risk η_t and $\left\{\kappa^{(\nu_t,\nu)} : \nu_t, \nu \in \mathcal{V}\right\}$ and the prices p_t, q_t, τ_t and the wages w_t as given, the expert f maximizes the utility

$$U^{\mathrm{e}}_{f,0} = \max_{\left\{\hat{c}^{\mathrm{e}}_{f,t}, g_{f,t}, k_{f,t}, s_{f,t}, \hat{\vartheta}^{\mathrm{e}}_{f,t}, \hat{\vartheta}^{\mathrm{e},(\nu_t,\nu)}_{f,t}\right\}_{t\geq 0}} \mathbb{E}_0\left[\int_0^\infty \mathbf{f}(c^{\mathrm{e}}_{f,t}, U^{\mathrm{e}}_{f,t})\mathrm{d}t\right] \tag{2.21}$$

subject to the solvency constraint $n^{\mathrm{e}}_{f,t} \geq 0$ and the dynamic budget constraint

$$\frac{\mathrm{d}n^{\mathrm{e}}_{f,t}}{n^{\mathrm{e}}_{f,t}} = \left(\mu^{\mathrm{e}}_{f,n,t} - \hat{c}^{\mathrm{e}}_{f,t}\right)\mathrm{d}t + \underbrace{\sigma^{\mathrm{e}}_{f,n,t}\mathrm{d}Z_t + \textstyle\sum_{\nu\neq\nu_t}\varsigma^{\mathrm{e},(\nu_t,\nu)}_{f,n}\left[\mathrm{d}N^{(\nu_t,\nu)}_t - \lambda^{(\nu_t,\nu)}\mathrm{d}t\right]}_{\text{aggregate risk exposures}} \tag{2.22}$$
$$+ \underbrace{\sigma^{\mathrm{e}}_{f,n,W,t}\mathrm{d}W_{f,t} + \left[\varsigma^{\mathrm{e}}_{f,n,N,t}\mathrm{d}N_{f,t} - \mathbb{E}^{\varepsilon}\left(\varsigma^{\mathrm{e}}_{f,n,N,t}\right)\lambda\mathrm{d}t\right]}_{\text{idiosyncratic risk exposures}}$$

where the consumption rate is $\hat{c}^{\mathrm{e}}_{f,t} \equiv c^{\mathrm{e}}_{f,t}/n^{\mathrm{e}}_{f,t}$ and the expected growth rate on net worth (pre consumption) is $\mu^{\mathrm{e}}_{f,n,t}$. Furthermore, the exposure to the aggregate shock $\mathrm{d}Z_t$ is

$$\sigma^{\mathrm{e}}_{f,n,t} = \underbrace{\hat{\vartheta}^{\mathrm{e}}_{f,t}}_{\text{indices}} + \underbrace{\phi\frac{q_t k_{f,t}}{n^{\mathrm{e}}_{f,t}}(\sigma^q_t + \varphi\sigma) + \phi\frac{p_t s_{f,t}}{n^{\mathrm{e}}_{f,t}}(\sigma^p_t + \varphi\sigma)}_{\text{concentrated equity holdings}}, \tag{2.23}$$

and the exposure to the aggregate uncertainty risk $\mathrm{d}N^{(\nu_t,\nu)}_t$ is

$$\varsigma^{\mathrm{e},(\nu_t,\nu)}_{f,n} = \underbrace{\hat{\vartheta}^{\mathrm{e},(\nu_t,\nu)}_{f,t}}_{\text{indices}} + \underbrace{\phi\frac{q_t k_{f,t}}{n^{\mathrm{e}}_{f,t}}\varsigma^{q,(\nu_t,\nu)}_t + \phi\frac{p_t s_{f,t}}{n^{\mathrm{e}}_{f,t}}\varsigma^{p,(\nu_t,\nu)}_t}_{\text{concentrated equity holdings}}, \tag{2.24}$$

The exposures to the idiosyncratic risks are

$$\underbrace{\sigma^{\mathrm{e}}_{f,n,W,t} = \phi\frac{q_t k_{f,t}}{n^{\mathrm{e}}_{f,t}}\nu_{c,t} \quad\text{and}\quad \varsigma^{\mathrm{e}}_{f,n,N,t} = \phi\frac{p_t s_{f,t}}{n^{\mathrm{e}}_{f,t}}\pi_{f,t}.}_{\text{concentrated equity holdings}}$$

Here, the shares $\hat{\vartheta}^{\mathrm{e}}_{f,t} \equiv \vartheta^{\mathrm{e}}_{f,t}/n^{\mathrm{e}}_{f,t}$ and $\hat{\vartheta}^{\mathrm{e},(\nu_t,\nu)}_{f,t} \equiv \vartheta^{\mathrm{e},(\nu_t,\nu)}_{f,t}/n^{\mathrm{e}}_{f,t}$ characterize the expert's positions in risky short-term financial contracts. The portfolio constraints forced experts to bear uninsured idiosyncratic risks $\sigma^{\mathrm{e}}_{f,n,W,t}$ and $\varsigma^{\mathrm{e}}_{f,n,N,t}$. The implementation described in Proposition 2 requires expert f to retain ϕ fraction of firm f's equity stake. The concentrated holdings of the aggregate risks in firm f's equity can be offset by the holdings of aggregate indices as in (2.23) and (2.24). Thus, the true effect of the financial restriction is to force each expert to bear the background risks, which are the uninsurable idiosyncratic investment risks. The expected growth rate $\mu^{e}_{f,n,t}$ includes three components: (i) the expected returns from financial index holdings, (ii) the expected returns from firm's assets holdings, and (iii) minus the expected returns of firm's equity paid out to diffused shareholders.[18]

2.2.3 Competitive Equilibrium

Now, I provide the formal definition of the competitive equilibrium with incomplete markets.

Definition 1. *Given the initial aggregate assets in place* $k_0 > 0$ *and growth options* $s_0 > 0$ *and the distributions among agents which satisfy* $\int_{f\in\mathbb{F}} k_{f,0} df + \int_{h\in\mathbb{H}} k_{h,0} dh = k_0$ *and* $\int_{f\in\mathbb{F}} s_{f,0} df + \int_{h\in\mathbb{H}} s_{h,0} dh = s_0$. *Each agent starts with strictly positive and identical net worth* $k_{j,0} > 0$ *and* $s_{j,0} > 0$ *for all* $j \in \mathbb{F} \cup \mathbb{H}$. *Households sell their capital to experts immediately at time* 0. *A* **competitive equilibrium** *is a set of aggregate and idiosyncratic stochastic processes adapted to the filtration generated by aggregate and idiosyncratic stochastic processes* $\mathcal{F}_t \equiv \sigma\{Z_{t'}, N^{(\nu_{t'},\nu)}_{t'}, W_{f,t'}, N_{f,t'}, \varepsilon_{f,t'} : 0 \le t' \le t, f \in \mathbb{F}, \nu_{t'}, \nu \in \mathcal{V}\}$. *The set of aggregate stochastic processes include the prices of productive capitals* $\{q_t, p_t\}$, *the market prices of aggregate risks* $\{\eta_t, \kappa^{(\nu_t,\nu)}_t : \nu_t, \nu \in \mathcal{V}\}$, *the aggregate productive capital stocks* $\{k_t, s_t\}$, *the wage process* $\{w_t\}$, *the price of investment goods* $\{\tau_t\}$, *and the human capital* $\{\hbar_t\}$. *The set of agent-level stochastic processes include the net worth processes* $\{n^{e}_{f,t}, n^{h}_{h,t}\}$, *the consumptions* $\{c^{e}_{f,t}, c^{h}_{h,t}\}$, *the holdings of firm assets* $\{k_{f,t}, s_{f,t}\}$, *the investment rates* $\{i_{f,t}\}$, *the demands for the investment goods* $\{g_{f,t}\}$, *the labor demands* $\{\ell_{c,f,t}, \ell_{\iota,t}\}$, *and risk exposures* $\{\sigma^{e}_{f,n,t}, \sigma^{h}_{h,n,t}, \varsigma^{e,(\nu_t,\nu)}_{f,n,t}, \varsigma^{h,(\nu_t,\nu)}_{h,n,t}\}$, *for all* $f \in \mathbb{F}$ *and* $h \in \mathbb{H}$, *such that*

[18] More details on the budget constraint of the expert and the household can be found in the online appendix.

(i) *Initial expert net worth satisfies* $n^e_{f,0} = q_0 k^e_{f,0} + p_0 s^e_{f,0}$ *and initial household net worth satisfies* $n^h_{h,0} = q_0 k^h_{h,0} + p_0 s^h_{h,0}$.

(ii) *Given the aggregate dynamics, each household solves her utility optimization problem* (2.19) *and each expert solves her utility optimization problem* (2.21).

(iii) *Market clearing conditions:*

(a) *Assets in place market and growth options market:*

$$\int_{f\in\mathbb{F}} k_{f,t} df = k_t \ \ and \ \ \int_{f\in\mathbb{F}} s_{f,t} df = \bar{s}.$$

(b) *Consumption goods market:*

$$\int_{f\in\mathbb{F}} c^e_{f,t} df + \int_{h\in\mathbb{H}} c^h_{h,t} dh = \int_{f\in\mathbb{F}} k^{\varphi}_{f,t} \ell^{1-\varphi}_{c,f,t} df - \varpi p_t \bar{s} \lambda \bar{\Phi}(\xi_t / \nu_{g,t}).$$

(c) *Investment goods market:*

$$\int_{f\in\mathbb{F}} g_{f,t} dN_{f,t} = z_\iota \ell_{\iota,t}.$$

(d) *Labor markets:*

$$\int_{f\in\mathbb{F}} \ell_{c,f,t} df + \ell_{\iota,t} = 1.$$

(e) *Financial market for insurance* Z_t *risk:*

$$\int_{f\in\mathbb{F}} \sigma^e_{f,n,t} n^e_{f,t} df + \int_{h\in\mathbb{H}} \sigma^h_{h,n,t} n^h_{h,t} dh \\ = q_t k_t (\sigma^q_t + \varphi\sigma) + p_t \bar{s} (\sigma^p_t + \varphi\sigma) + \varrho \hbar_t (\sigma^{\hbar}_t + \varphi\sigma).$$

(f) *Financial market for insurance* $N^{(\nu_t,\nu)}_t$ *risk:*

$$\int_{f\in\mathbb{F}} \varsigma^{e,(\nu_t,\nu)}_{f,n,t} n^e_{f,t} df + \int_{h\in\mathbb{H}} \varsigma^{h,(\nu_t,\nu)}_{h,n,t} n^h_{h,t} dh \\ = q_t k_t \varsigma^{q,(\nu_t,\nu)}_t + p_t \bar{s} \varsigma^{p,(\nu_t,\nu)}_t + \varrho \hbar_t \varsigma^{\hbar,(\nu_t,\nu)}_t.$$

(iv) Law of motion of aggregate capital

$$dk_t = \left(\int_{f \in \mathbb{F}} i_{f,t} dN_{f,t} - \delta \right) k_t dt + \sigma k_t dZ_t \quad \textit{and} \quad ds_t = 0.$$

By Walras' law, the market for risk-free debt clears automatically.

2.2.4 Solving for the Equilibrium Recursively

In order to solve the competitive equilibrium, I have to determine how the prices, investments, and consumptions of all agents depend on the historical paths of the aggregate shock Z_t, the aggregate uncertainty shocks $N_t^{(\nu_t, \nu)}$ and idiosyncratic shocks $W_{f,t}, N_{f,t}, \varepsilon_{f,t}$. In fact, I show that the equilibrium can be characterized, in a recursive formulation, by policy functions of three exogenous state variables $(z_t, \nu_{g,t}, \nu_{c,t})$ and two endogenous state variables. One endogenous state variable is the cross-sectional distribution of net worth among experts and households. Because Epstein-Zin-Weil preference is homothetic, the optimal control variables are all linear in the agent's net worth. The linear property allows me to simplify the endogenous state space, from an infinite-dimensional state space to a one-dimensional space. More precisely, I only need to track the evolution of experts' net worth relative to the total net worth held by all agents in equilibrium $x_t = \frac{n_t^{\mathrm{e}}}{Q_t}$, where $n_t^{\mathrm{e}} = \int_{f \in \mathbb{F}} n_{f,t}^{\mathrm{e}} \mathrm{d}f$ and $Q_t \equiv q_t k_t + \bar{s} p_t + \varrho \hbar_t$. The other endogenous state variable is the aggregate assets in place k_t, which captures the stochastic trend of the economy. Thus, the equilibrium can be characterized by state variables (z_t, ν_t, k_t, x_t) where $\nu_t \equiv (\nu_{g,t}, \nu_{c,t})$. Moreover, the Brownian motion Z_t only affects the economy through the i.i.d. shocks ($\mathrm{d}Z_t$) driving the stochastic trend of the economy and it is independent of the state variable ν_t. So, the variable Z_t does not really serve as a state variable characterizing the equilibrium.[19] As a result, the equilibrium is characterized by (ν_t, k_t, x_t).

[19]The same feature of i.i.d. cash flow shocks is also adopted in Bolton, Chen, and Wang (2011, 2013) and Dou and Ji (2015).

Dynamic evolution of the economy In equilibrium, all variables evolve around the stochastic trend k_t^{φ}. Moreover, the transitory fluctuations along the stochastic trend can be characterized by the state variables (v_t, x_t). The uncertainty state variable v_t is stationary by assumption. The endogenous state variable x_t is also mean-reverting in equilibrium. The dynamics of the variables in equilibrium can be summarized in Proposition 4.

Proposition 4 (Growth-trending Variables)**.** *The price variables, the firm-level output and payout variables, and the agent-level net worth variables in equilibrium have the following forms:*

$$\begin{aligned} p_t =& \tilde{p}_t k_t^{\varphi},\ q_t = \tilde{q}_t k_t^{\varphi-1},\ w_t = \tilde{w}_t k_t^{\varphi},\ \tau_t = \tilde{\tau}_t k_t^{\varphi},\ \hbar_t = \tilde{\hbar}_t k_t^{\varphi},\ \text{and} \\ y_{f,t} =& \tilde{y}_t k_t^{\varphi},\ d_{f,t} = \tilde{d}_t k_t^{\varphi},\ n^e_{f,t} = \tilde{n}^e_{f,t} k_t^{\varphi},\ n^h_{h,t} = \tilde{n}^h_{h,t} k_t^{\varphi},\quad \text{for all } f \in \mathbb{F} \text{ and } h \in \mathbb{H}, \end{aligned}$$

where $\tilde{p}_t$, $\tilde{q}_t$, $\tilde{w}_t$, $\tilde{\tau}_t$, $\tilde{\hbar}_t$, $\tilde{y}_t$, $\tilde{d}_t$, $\tilde{n}^e_{f,t}$, and $\tilde{n}^h_{h,t}$ are independent of the state variables z_t and k_t and are only driven by the state variables v_t and x_t.

Corollary 1 (Stationary Variables)**.** *The firm-level profit rate of growth options $\pi_{f,t}$, labor demand for production $\ell_{c,f,t}$, investment goods demand $g_{f,t}$, and investment rate $i_{f,t}$ do not depend on the growth-trend state variable k_t. They depend only on the stationary state variables v_t and x_t.*

I now consider the agent-level consumption, real investment, and portfolio holdings. In equilibrium, as in classic consumption-portfolio problems studied by Samuelson (1969) and Merton (1969), the individual consumption, real investment, and portfolio holdings are linear in terms of the individual net worth. This is because Epstein-Zin-Weil preferences are homothetic. Moreover, the linearity and symmetry of an individual's decision makes it unnecessary to track either the cross-sectional distribution of experts' net worth or the cross-sectional distribution of households' net worth to characterize the equilibrium. It facilitates the aggregation by making the two infinite-dimensional cross-sectional distributions irrelevant in equilibrium.

Proposition 5 (Linearity and Symmetry). *In equilibrium, the agent-level consumptions $c^e_{f,t}$ and $c^h_{h,t}$, the firm assets held by individual experts $k_{f,t}$ and $s_{f,t}$, and the positions of financial short-term contracts shosen by individual agents $\vartheta^e_{f,t}$, $\vartheta^{e,(\nu_t,\nu)}_{f,t}$, $\vartheta^h_{h,t}$, and $\vartheta^{h,(\nu_t,\nu)}_{h,t}$ for any $\nu_t, \nu \in \mathcal{V}$, $f \in \mathbb{E}$ and $h \in \mathbb{H}$, have the following forms:*

$$
\begin{aligned}
&c^e_{f,t} = \hat{c}^e_t n^e_{f,t},\ c^h_{h,t} = \hat{c}^h_t n^h_{h,t},\ k_{f,t}/k_t = \hat{k}_t \tilde{n}^e_{f,t},\ s_{f,t} = \hat{s}_t \tilde{n}^e_{f,t},\ \text{and} \\
&\vartheta^e_{f,t} = \hat{\vartheta}^e_t n^e_{f,t},\ \vartheta^{e,(\nu_t,\nu)}_{f,t} = \hat{\vartheta}^{e,(\nu_t,\nu)}_t n^e_{f,t},\ \vartheta^h_{f,t} = \hat{\vartheta}^h_t n^h_{f,t},\ \vartheta^{h,(\nu_t,\nu)}_{f,t} = \hat{\vartheta}^{h,(\nu_t,\nu)}_t n^h_{f,t},
\end{aligned}
$$

for all $\nu_t, \nu \in \mathcal{V}$, $f \in \mathbb{F}$, $h \in \mathbb{H}$. Importantly, the hatted variables $\hat{c}^e_t$, $\hat{c}^h_t$, $\hat{k}_t$, $\hat{s}_t$, $\hat{\vartheta}^e_t$, $\hat{\vartheta}^h_t$, $\hat{\vartheta}^{e,(\nu_t,\nu)}_t$, and $\hat{\vartheta}^{h,(\nu_t,\nu)}_t$ are only dependent on the aggregate stationary state variables ν_t and x_t. The detrended net worth $\tilde{n}^e_{f,t}$ and $\tilde{n}^h_{h,t}$ are defined in Proposition 4.

Value functions. Due to homotheticity of EZW preferences, I know that the value function for an expert with net worth n^j_t takes the following power form:

$$
U^j(\zeta^j_t, n^j_t) = \frac{\left(\zeta^j_t n^j_t\right)^{1-\gamma}}{1-\gamma},
$$

where ζ^j_t is the marginal value of net worth for the agent $j \in \{e, h\}$. The marginal value ζ^j_t captures the general equilibrium investment environment the agent faces. In particular, a higher marginal value of net worth ζ^j_t means a better investment environment for the agent. I conjecture that ζ^j_t follows the dynamic

$$
\frac{d\zeta^j_t}{\zeta^j_t} = \mu^j_{\zeta,t} dt + \sigma^j_{\zeta,t} dZ_t + \sum_{\nu \neq \nu_t} \varsigma^{j,(\nu_t,\nu)}_{\zeta} dN^{(\nu_t,\nu)}_t, \tag{2.25}
$$

where all the coefficients $\mu^j_{\zeta,t}$, $\sigma^j_{\zeta,t}$, and $\varsigma^{j,(\nu_t,\nu)}_{\zeta}$ for $j \in \{e, h\}$ are determined in equilibrium.[20]

[20] The HJB equations for experts and households can be found in the online appendix. The expressions of the Ito coefficients in (2.25), (2.26), and (2.27) are also in the online appendix.

Wealth distribution dynamics. Due to the homogeneity of experts and the homogeneity of households up to their own individual net worth levels, I only need to track the distribution between the aggregate experts' net worth n_t^{e} and the aggregate households' net worth n_t^{h}. I define $\tilde{Q}_t \equiv Q_t/k_t^{\varphi} = \tilde{n}_t^{\mathrm{e}} + \tilde{n}_t^{\mathrm{h}}$ and conjecture that

$$\mathrm{d}\tilde{Q}_t/\tilde{Q}_t = \mu_t^Q \mathrm{d}t + \sigma_t^Q \mathrm{d}Z_t + \sum_{\nu \neq \nu_t} \varsigma^{Q,(\nu_t,\nu)} \mathrm{d}N_t^{(\nu_t,\nu)}, \tag{2.26}$$

where the coefficients depends on those of the prices q_t, p_t and human capital $\hbar_t$. Thus, in equilibrium, the law of motion of x_t can be characterized as follows:

$$\frac{\mathrm{d}x_t}{x_t} = \mu_{x,t}\mathrm{d}t + \sigma_{x,t}\mathrm{d}Z_t + \sum_{\nu \neq \nu_t} \varsigma_{x,t}^{(\nu_t,\nu)} \mathrm{d}N_t^{(\nu_t,\nu)}, \tag{2.27}$$

where $\mu_{x,t}$ is the expected growth rate and the volatility of wealth share $\sigma_{x,t}$ and the jump size $\varsigma_{x,t}^{(\nu_t,\nu)}$ are,

$$\sigma_{x,t} = \sigma_{f,n,t}^{\mathrm{e}} - \sigma_t^Q - \varphi\sigma, \quad \text{and} \quad \varsigma_{x,t}^{(\nu_t,\nu)} = \frac{\varsigma_{f,n,t}^{\mathrm{e},(\nu_t,\nu)} + 1}{\varsigma_t^{Q,(\nu_t,\nu)} + 1} - 1, \text{ respectively.}$$

Because the aggregate Z_t process characterizes i.i.d. shocks in the economy which are independent with all other aggregate shocks and it is not a state variable, it only affects agents' myopic portfolio decisions and hence is perfectly shared by agents using contract term contracts on the shock. Thus, in equilibrium, the aggregate shock $\mathrm{d}Z_t$ should have zero impact on the endogenous state variable x_t. In fact, it is not hard to show the following results.

Proposition 6. *In the equilibrium, the aggregate risk Z_t is perfectly shared. Thus, each agent's exposure $\sigma_{f,n,t}^{e}$ to the productivity shock dZ_t is simply the constant myopic component:*

$$\sigma_{x,t} = \sigma_{f,n,t}^{e} - \sigma_t^Q - \varphi\sigma = 0.$$

Corollary 2. *In the equilibrium, the loadings of de-trended variables on the productivity shock*

dZ_t are all zero, since the risk Z_t is perfectly shared among heterogeneous agents. In particular,

$$\sigma^e_{\xi,t} \equiv \sigma^h_{\xi,t} \equiv \sigma^Q_t \equiv \sigma^p_t \equiv \sigma^q_t \equiv \sigma^{\hbar}_t \equiv 0.$$

Recursive Markov equilibria.

Definition 2. *A Recursive Markov Equilibrium characterized by state variables (x_t, ν_t) is a set of aggregate functions: marginal values of net worth in value functions ζ^e, ζ^h, price functions $p, q, w, \hbar, \eta, r$, and $\kappa^{(\nu,\nu')}$ and policy functions $\hat{c}^e, \hat{c}^h, g, \hat{k}, \hat{s}, \hat{\theta}^e, \hat{\theta}^h, \hat{\vartheta}^{e,(\nu,\nu')}$, and $\hat{\vartheta}^{h,(\nu,\nu')}$, and law of motions for the endogenous state variable x_t such that*

(i) the marginal value of net worths ζ^e and ζ^h solve the experts' and households' HJB equations, and $\hat{c}^e, \hat{c}^h, g, \hat{k}, \hat{s}, \hat{\theta}^e, \hat{\theta}^h, \hat{\vartheta}^{e,(\nu,\nu')}, \hat{\vartheta}^{h,(\nu,\nu')}$ are the optimal control variables, taking prices $\tilde{q}, \tilde{p}, \tilde{w}, \tilde{\hbar}, r, \eta$, and $\kappa^{(\nu,\nu')}$ and the law of motion of state variables x_t and ν_t as given;

(ii) the market clearing conditions are satisfied:

$$\hat{c}^e\tilde{Q}x + \hat{c}^h\tilde{Q}(1-x) = \tilde{y} - \varpi\bar{s}\tilde{p}\lambda\bar{\Phi}(\xi/\nu_g) \quad \textit{(Consumption Goods)}$$

$$g = z_\iota \ell_\iota \quad \textit{(Investment Goods)}$$

$$\tilde{Q}\hat{k}x = 1 \text{ and } \tilde{Q}x\hat{s} = \bar{s} \quad \textit{(Tangible and Intangible Capitals)}$$

$$\ell_c + \ell_\iota = 1 \quad \textit{(Labor Hours)}$$

$$\sigma^e_n x + \sigma^h_n(1-x) = \varphi\sigma \quad \textit{(Financial Securities for } Z_t \textit{ and } \sigma^Q_t \equiv 0\textit{)}$$

$$\varsigma^{e,(\nu,\nu')}_n x + \varsigma^{h,(\nu,\nu')}_n(1-x) = \varsigma^{Q,(\nu,\nu')} \quad \textit{(Financial Securities for } N^{(\nu,\nu')}_t\textit{)}$$

(iii) the law of motion of endogenous state variable x_t is characterized as in (2.27).

The fixed-point conditions that characterize the Recursive Markov equilibrium can be summarized by a set of coupled highly-nonlinear ordinary differential equations, whose details can be found in the online appendix.

2.3 Quantitative Results

In this section, I first explore whether a real business cycle model with two sources of uncertainty shocks and imperfect risk sharing can simultaneously match the key moments of macroeconomic variables and asset returns. This exercise reveals the quantitative importance of two uncertainty shocks, interacting with endogenous imperfect risk sharing, as drivers of macroeconomic fluctuations and determinants of risk premia. Then, the calibrated general equilibrium model provides a laboratory allowing me to examine the quantitative relevance of the key mechanism discovered in this paper. I show that the implications of the key mechanism are quantitatively significant and coherent within such an empirically-validated framework. Furthermore, in Section 2.4, I explore whether the implications of the key mechanism is observed in the data.

2.3.1 Calibration and Parameter Choices

Table 4.4 summarizes the parameter choices used in my calibration. The key parameter that characterizes the risk sharing imperfection is the severity of agency problem, denoted by ϕ. In the model, the experts effectively constitute blockholders.[21] The blockholders (including the inside blockholders) control the firm: they can either directly or indirectly intervene in firm's operations (e.g., Edmans and Manso, 2011; Holderness, 2003, 2009). Holderness (2009) reports 96% of randomly selected U.S. firms in 1995 have blockholders, and the average percentage of the voting rights in common stocks held by all blockholders is 43%. Khan, Dharwadkar, and Brandes (2005) show that from 1992 to 1999, the total institutional ownership increases from 52.6% to 58.8%, and the CEO ownership is ranged from 2.17% and 2.94%, based on a complete 8-year

[21] Each expert is a representative agent of the managers and insiders who actively intervene in the management and hold significant stake of a firm. The degree of blockholding tends to underestimate the concentrated ownership of the experts, because experts do not only hold stake in the firm through common stocks but also through compensations and rents. On the other hand, not all blockholders are forced to bear uninsurable idiosyncratic risks of the firm's equity. Hoping the two forces cancel out each other, I take the blockholding level as an approximation for the parameter ϕ.

sample with 224 U.S. public firms. More recent data show that the institutional blockholders in U.S. equity markets holding over 66% of the total equity.[22] I choose $\phi = 0.5$ to provide a reasonable blockholding in my model.

To illustrate the role of the two uncertainty shocks, I need to choose the evolution rules governing how the uncertainty fluctuates over time and to choose the uncertainty levels that characterizing the scale of fluctuations in uncertainty. The transition intensities $\lambda^{(\nu_g,\nu_g')}$ and $\lambda^{(\nu_c,\nu_c')}$ are estimated based on the regime-switching dynamics of estimated growth uncertainty and cash-flow uncertainty, respectively. Specifically, they are estimated based on Table 2.6 for which the details are in the online appendix. The uncertainty levels are calibrated such that the interdecile range (IDR) of sales growth rates and the cross-sectional standard deviation (CSD) of investment rates have means and standard deviations that reasonably match the moments in the data summarized in Table 2.4. More precisely, the means of the IDR of sales growth rates are 53.02% in the model and 49.02% in the data; moreover, its standard deviations are 16.03% in the model and 12.32% in the data. The means of the CSD of investment rates are 45.12% in the model versus 40.85% in the data; and, its standard deviations are 13.50% in the model versus 7.25% in the data.

To calibrate the specification of preferences, I choose a value for EIS $\psi = 2$ consistent with Bansal and Yaron (2004a); Bansal, Kiku, and Yaron (2012a), who emphasize that the preference of early resolution of uncertainty is important to understand uncertainty shock's impact on asset prices. Consistent with macroeconomic models of asset prices such as Guvenen (2009), I choose a value for risk aversion no bigger than 10. Here, I use $\gamma = 6$ to provide a comparable capital-to-output ratio to the data as summarized in Table 2.3 (196.20% in the model versus 169.24% in the data). The subjective discount factor is chosen to be $\delta = 0.0111$ to help the model match the average level of risk-free rates as in Table 2.5 (1.53% in the model versus 1.31% in the data).

The average lifespan parameter is chosen to be $\mu = 1/40$, which is a standard choice

[22]See, e.g., Carolyn K. Brancato and Stephan Rabimov, "The 2008 Institutonal Investment Report: Trends in Institutional Investor Assets and Equity Ownership of U.S. Corporations" (The Conference Board, 2008).

since the average number of working years in U.S. is about 40. The pledgeability of human capital chosen to be $\varrho = 5\%$ which is consistent with Lustig and Nieuwerburgh (2010). The population of experts is estimated based on the U.S. income distribution observations provided by U.S. census. A simple linear extrapolation estimates that, on average, about 2% of U.S. households earn annual salary $30,000$ dollars. Presumably, the experts make at least $30,000$ dollars a year, so 2% is a reasonable approximation for the population share of experts in the economy.

Lastly, as for the production and investment of consumption goods firms, the productivity volatility σ is calibrated in a standard way. I choose $\sigma = 10\%$ to match the standard deviation of output log growth as summarized in Table 2.2 (1.92% in the model versus 1.67% in the data). The shares of capital are chosen to be $\varphi = 0.3$ and $1 - \alpha = 0.1$, which help match the relative size of the consumption goods sector and the investment goods sector (approximately 23% for the investment goods sector in terms of sectoral outputs in the data), while generating a labor share of output of approximately 75% as in Table 2.3 (75.25% in the model versus 75.26% in the data). This is also in line with Papanikolaou (2011). The constant arrival rate of investment opportunities is chosen at $\lambda = 3.33$ and the fixed adjustment cost rate is chosen at $\varpi = 0.83\%$ to match the average annual positive investment rate (approximately 79% in the data) and the standard deviation of aggregate investment log growth (Table 2.2) simultaneously. The standard deviations of log growth rates of aggregate investment are 55.38% in the model and 36.00% in the data. The average productivity in the investment goods sector $z_\iota = 1.03$ and the depreciation rate $\delta = 15\%$ help match the average investment-to-output ratio (Table 2.3: 16.60% in the model versus 16.47% in the data) and the average payout-to-consumption ratio (Table 2.3: 6.30% in the model versus 5.46% in the data).

2.3.2 Model Implications

Macroeconomic moments. I report the model-implied moments of the growth rates of log consumption, log investment, and log output in Panel B of Table 2.2; for comparison, I also report their empirical counterparts in Panel A of the same table. Columns (1) – (3) of the table report means, standard deviations, and autocorrelations for each growth variable; Columns (4) – (6) report the correlations among the three growth variables. For the simulated data in panel B, the table shows the average values across independent simulations, along with the 5th and 95th percentiles reported in brackets. For the moments in the data, the table reports the point estimates and the corresponding confidence intervals in brackets estimated by stationary block bootstrap methods.

The moments in blue and bold are those used for calibration. For most of the moments of interest in Table 2.2, the data and the model are close statistically. However, the model fails to produce the right pattern of comovement between investment and consumption growth (-0.44 in the model versus 0.83 in the data). The main reason is that the growth uncertainty resembles the investment shock, especially when the risk sharing condition is good. This can be seen from (2.14) and (2.17). The implied investment shock generates opposite responses in investment and consumption. The negative correlation arises for two reasons. First, the high value of EIS implies that consumption does fall heavily in response to an implied positive investment shock. It generates exceedingly negative correlation between consumption and investment. Second, the aggregate productivity shock dZ_t moves investment and consumption in the same direction. However, when EIS is large, the effect of implied investment shocks dominates, generating a negative correlation between investment and consumption. This is a well-known issue for real business cycle models with investment shocks. In general, labor market frictions can help restore the positive correlation.

For a macroeconomic growth model, the valid quantitative analysis requires the key macroeconomic ratios characterizing the steady state along the balanced growth path to be replicated by the model with reasonably small errors. Basically, the calibration of the model should be able to generate the steady-state ratios consistent with the data. Table 2.3 compares the empirical moments of investment-to-output ratios, net-payout-to-consumption ratios, wage-income-to-output ratios, and capital-to-output ratios with their correspondences in the simulated data generated from the model. The moments in blue and bold are used for calibration. It shows that the data and the model are statistically close for most of the moments of interest. However, the model fails to generate a high enough standard deviation of wage-income-to-output ratio (2.04% in the model versus 4.02% in the data); moreover, their confidence intervals are not even overlapped. The main reason is that the asymmetric and volatile fluctuations in unemployment is hard to be captured by a model without frictions in labor markets (e.g., Petrosky-Nadeau and Zhang, 2013). In the model with frictionless labor markets, agents can efficiently smooth out the labor income shocks over time.

Table 2.4 compares the empirical moments of cross-sectional dispersions to the simulated ones from the model. Panel A shows that the sales growth dispersion is countercyclical (the correlation with log output growth is -17.32%), while the investment rate dispersion is procyclical in the data (the correlation with log output growth is 43.28%). It is consistent with the findings in Bachmann and Bayer (2014) who emphasize that the uncertainty-driving real business cycle models need to reconcile the two prominent patterns. Column (4) of Panel B shows that the model generates countercyclical sales growth dispersions (the correlation with log output growth is -27.66%) and procyclical investment rate dispersion (the correlation with log output growth is 23.82%) simultaneously. In the model, the sales growth dispersion is mainly driven by cash-flow uncertainty shocks (as shown in (2.42)), which decrease the output and investment due to the imperfect risk sharing; the investment rate dispersion is mainly driven by growth uncertainty shocks, which has asymmetrically stronger effect when risk sharing is less limited. The asymmetric effect of growth uncertainty shocks on the investment rate dispersion implies procyclical dispersion in equilibrium.

Table 2.5 compares the asset pricing moments in the data to the simulated moments from the model. In particular, the sizable equity premium (4.95% in the model versus 4.47% in the data) is main a result of the market incompleteness and the amplification effect of financial frictions on the uncertainty shocks. The model also reproduces the sizable value premium (7.57% in the model versus 5.05% in the data). The large average value spread is mainly due to the cash-flow uncertainty shock which carries a negative market price of risk and decreases the value of assets in place relative to growth options.

Figure 2-1 and Figure 2-2 show the key results of this paper. In Figure 2-1, while the market price of risk for the cash-flow uncertainty shock is always negative, the market price of risk for the growth uncertainty shock changes from negative to positive as the risk sharing condition gets better. In Figure 2-2, while the exposure of value spreads to cash-flow uncertainty shocks is always negative, their exposure to the growth uncertainty shock changes from positive to negative as the risk sharing condition improves.

2.3.3 Basic Mechanisms

Stochastic discount factors and idiosyncratic risk premia There is a full menu of short-term contingent claims on both aggregate shocks and idiosyncratic shocks available to the agents. However, the moral hazard makes the enforcement of some contingent claim contracts on idiosyncratic shocks imperfect.

I denote by $\mathcal{M}^{e}_{f,t}$ the utility gradients of the expert f at her optimal consumption policy. According to Duffie and Skiadas (1994, Theorem 2), the utility gradient of expert f has the following expression:

$$\mathcal{M}^{e}_{f,t} = \exp\left[\int_0^t \mathbf{f}_U(c^{e}_{f,t'}, U^{e}_{f,t'})\mathrm{d}t'\right] \mathbf{f}_c(c^{e}_{f,t}, U^{e}_{f,t}).$$

Thus, the instantaneous intertemporal marginal rate of substitution (IMRS) of expert f is

$$\frac{\mathrm{d}\mathcal{M}^{e}_{f,t}}{\mathcal{M}^{e}_{f,t}} = -\mu^{e}_t - \eta^{e}_t \mathrm{d}Z_t - \sum_{\nu'_g \neq \nu_g} \kappa^{e,(\nu_g,\nu'_g)} \left[\mathrm{d}N^{(\nu_g,\nu'_g)}_t - \lambda^{(\nu_g,\nu'_g)}\mathrm{d}t\right] \tag{2.28}$$
$$- \gamma\sigma^{e}_{f,n,W,t}\mathrm{d}W_{f,t} - \gamma\sigma^{e}_{f,n,N,t}\mathrm{d}N_{f,t},$$

where the drift μ^{e}_t and the coefficients of aggregate shocks η^{e}_t and $\kappa^{e,(\nu_g,\nu'_g)}_t$ only depend on aggregate state variables in equilibrium. The coefficients of idiosyncratic shocks

$\sigma^{\mathrm{e}}_{f,n,W,t}$ and $\sigma^{\mathrm{e}}_{f,n,N,t}$ also only depend on aggregate state with the following expressions:

$$\sigma^{\mathrm{e}}_{f,n,W,t} \equiv \nu_{c,t} \frac{\phi}{x_t} F_{\mathbf{k}}(x_t, \nu_t) \quad \text{and} \quad \sigma^{\mathrm{e}}_{f,n,N,t} = \frac{1}{\gamma} \left[1 - (1 + \varsigma^{\mathrm{e}}_{f,n,N,t})^{-\gamma} \right].$$

Effectively, the term $\gamma \sigma_{f,n,W,t}$ is the market price of the idiosyncratic cash flow risk $\mathrm{d}W_{f,t}$ required by the expert f, while the term $\gamma \sigma^{\mathrm{e}}_{f,n,N,t}$ is the market price of the idiosyncratic growth risk $\mathrm{d}N_{f,t}$ required by the expert f. The term $\sigma_{f,n,W,t}$ is simply the loading of idiosyncratic cash flow risks. The term $\sigma^{\mathrm{e}}_{f,n,N,t}$ is approximately equal to $\varsigma^{\mathrm{e}}_{f,n,N,t}$ when the latter is small according to the Taylor expansion.

In an economy with complete and frictionless financial market, there is a unique stochastic discount factor which is equal to every agent's utility gradient. In an incomplete market, for any particular set of assets, according to the intertemporal Euler equations, the non-arbitrage condition implies that the stochastic discount factor is equal to the highest utility gradient across all agents who have access to the particular set of assets. In fact, for the unconstrained agent in some state, her utility gradient must equal to the stochastic discount factor in that state. This is the similar idea in Chien and Lustig (2010) and Alvarez and Jermann (2001) for asset pricing in an incomplete market.

Because all experts can freely access all financial assets whose payoffs are contingent on the aggregate shocks, the cross-sectional average of these individual experts' IMRS is a valid SPD for those financial in all states. Thus, the following results can be derived readily. For those financial assets whose payoffs depend only on aggregate states, one SPD that prices their returns is provided by the average IMRS of experts. More precisely, it is the SPD Λ_t such that

$$\begin{aligned} \frac{\mathrm{d}\Lambda_t}{\Lambda_t} &\equiv \frac{1}{\varkappa} \int_{f \in \mathbb{F}} \left[\frac{\mathrm{d}\mathcal{M}^{\mathrm{e}}_{f,t}}{\mathcal{M}^{\mathrm{e}}_{f,t}} \right] \mathrm{d}f \\ &= -r_t \mathrm{d}t - \eta_t \mathrm{d}Z_t - \sum_{\nu \neq \nu_t} \kappa_t^{(\nu_t,\nu)} \left[\mathrm{d}N_t^{(\nu_t,\nu)} - \lambda^{(\nu_t,\nu)} \mathrm{d}t \right]. \end{aligned} \tag{2.29}$$

Here $r_t \equiv \mu_t^{\mathrm{e}}$ is the risk-free interest rate, $\eta_t \equiv \eta_t^{\mathrm{e}}$ is the market price of aggregate cash flow risk Z_t, and $\kappa_t^{(\nu_t,\nu)} \equiv \kappa_t^{\mathrm{e},(\nu_t,\nu)}$ is the market price of uncertainty risk $N_t^{(\nu_t,\nu)} - \lambda^{(\nu_t,\nu)} t$ for each $\nu \in \mathcal{V}$. All market prices η_t, $\kappa^{(\nu_t,\nu)}$ for all $\nu \in \mathcal{V}$ and interest rate r_t are determined endogenously in equilibrium. Households agree on the market prices of risk. It is straightforward to derive that the market price of risk η_t is constant $\eta \equiv \gamma\varphi\sigma$, because all agents (experts and households) perfectly share the aggregate risk of productivity shock $\mathrm{d}Z_t$ by holding constant risk exposure $\varphi\sigma$.

However, experts are the only agents who can trade firm's assets freely. For each firm-specific asset, the following Euler equations hold. More precisely, for each $f \in \mathbb{F}$, it holds that

$$\mathbb{E}_t\left[\mathrm{d}R_t^{\mathbf{k}}\right]/\mathrm{d}t - r_t = -\mathbb{E}_t\left[\frac{\mathrm{d}\mathcal{M}_{f,t}^{\mathrm{e}}}{\mathcal{M}_{f,t}^{\mathrm{e}}} \times \frac{d_{f,t}\mathrm{d}t - (1-\phi)q_t k_{f,t}\nu_{c,t}\mathrm{d}W_{f,t} + \mathrm{d}\left(q_t k_{f,t}\right)}{q_t k_{f,t}}\right]$$

for all $f \in \mathbb{F}$, and

$$\mathbb{E}_t\left[\mathrm{d}R_t^{\mathbf{s}}\right]/\mathrm{d}t - r_t = -\mathbb{E}_t\left[\frac{\mathrm{d}\mathcal{M}_{f,t}^{\mathrm{e}}}{\mathcal{M}_{f,t}^{\mathrm{e}}} \times \frac{\phi p_t s_{f,t}\mathrm{d}N_{f,t} + \mathrm{d}\left(p_t s_{f,t}\right)}{p_t s_{f,t}}\right] \quad \text{for all } f \in \mathbb{F}.$$

Here, $d_{f,t}\mathrm{d}t - (1-\phi)q_t k_{f,t}\nu_{c,t}\mathrm{d}W_{f,t}$ is the effective consumption goods net payout of assets in place to expert f since she can dump the amount $(1-\phi)q_t k_{f,t}\nu_{c,t}\mathrm{d}W_{f,t}$ of the idiosyncratic cash flow exposure for free. And, $\phi\pi_{f,t}p_t\bar{s}\mathrm{d}N_{f,t}$ is the effective pecuniary net payout of growth options to expert f since she can dump the amount $(1-\phi)\pi_{f,t}p_t\bar{s}\mathrm{d}N_{f,t}$ of idiosyncratic growth exposure for free.

The relations of (2.28) – (2.31) leads to the following beta pricing rules for assets in place and growth options. The expected return from holding assets in place (i.e. assets in place) in excess of the risk-free rate equals

$$
\begin{aligned}
\underbrace{\mathbb{E}_t\left[\mathrm{d}R_t^{\mathbf{k}}\right]/\mathrm{d}t - r_t}_{\text{risk premium of holding } k_t} &= \overbrace{\underbrace{\varphi\sigma\eta + \sum_{\nu\neq\nu_t} \varsigma_t^{q,(\nu_t,\nu)}\kappa_t^{(\nu_t,\nu)}\lambda_t^{(\nu_t,\nu)}}_{\text{aggregate premium for all agents}}}^{\mathbb{E}_t\left[\mathrm{d}\tilde{R}_t^{\mathbf{k}}\right]/\mathrm{d}t - r_t} \\
&\quad + \underbrace{\gamma(\phi\nu_{c,t})^2\frac{\mathcal{F}_{\mathbf{k}}(x_t,\nu_t)}{x_t}}_{\text{idiosyncratic premium for experts}},
\end{aligned}
\tag{2.30}
$$

where $\tilde{R}_t^{\mathbf{k}}$ is the equity return on assets in place and $\mathcal{F}_{\mathbf{k}}(x_t,\nu_t) \equiv q_t k_t/Q_t$ is the assets-in-place share in the total net worth Q_t.

And, the expected return from holding growth options (i.e. growth options) in excess of the risk-free rate equals

$$
\begin{aligned}
\underbrace{\mathbb{E}_t\left[\mathrm{d}R_t^{\mathbf{s}}\right]/\mathrm{d}t - r_t}_{\text{risk premium of holding } s_t} &= \overbrace{\underbrace{\varphi\sigma\eta + \sum_{\nu\neq\nu_t} \varsigma_t^{p,(\nu_t,\nu)}\kappa_t^{(\nu_t,\nu)}\lambda_t^{(\nu_t,\nu)}}_{\text{aggregate premium for all agents}}}^{\mathbb{E}_t\left[\mathrm{d}\tilde{R}_t^{\mathbf{s}}\right]/\mathrm{d}t - r_t} \\
&\quad \underbrace{+\,\lambda\phi\mathbb{E}_t^{\varepsilon}\left\{\left[1-\left(1+\varsigma_{f,n,N,t}^{\mathrm{e}}\right)^{-\gamma}\right]\pi_{f,t}\right\}}_{\text{idiosyncratic premium for experts}},
\end{aligned}
\tag{2.31}
$$

where $\tilde{R}_t^{\mathbf{s}}$ is the equity return on growth options and $\mathcal{F}_{\mathbf{s}}(x_t,\nu_t) \equiv p_t s_t/Q_t$ is the growth-options share in the total net worth Q_t.

Alternatively, the beta pricing rules (2.30) – (2.31) can also be derived using the first-order conditions of experts' Hamilton-Jacobi-Bellman (HJB) equations, together with their dynamic budget constraints.

Using the Taylor-expansion approximation, the idiosyncratic risk premium required for holding growth options can be approximated by

$$\lambda\phi\mathbb{E}_t^{\varepsilon}\left\{\left[1-\left(1+\varsigma^{\mathrm{e}}_{f,n,N,t}\right)^{-\gamma}\right]\pi_{f,t}\right\}\approx\lambda\phi\gamma\mathbb{E}_t^{\varepsilon}\left[\pi_{f,t}\varsigma^{\mathrm{e}}_{f,n,N,t}\right]$$
$$=\gamma\left[\mathcal{I}_{\alpha}(\nu_{g,t}/\xi_t)\right]^2\frac{F_{\mathbf{s}}(x_t,\nu_t)}{x_t}\times\frac{\phi^2}{\lambda}.$$

Figure 2-3 illustrates the idiosyncratic risk premia under the calibration summarized in Table 4.4. The uncertainty shocks increase the risk premia on the idiosyncratic risks. There are several additional observations that worth mentioning. First, the effect of uncertainty shocks on the idiosyncratic risk premia increases nonlinearly as the risk sharing becomes more limited (i.e., x_t decreases). The reason is that the expert's net worth has larger exposure to the idiosyncratic shocks when x_t is lower. Second, the risk premium on the idiosyncratic cash-flow shock is mainly affected by the cash-flow uncertainty shock, while the risk premium on the idiosyncratic investment shock is mainly affected by the growth uncertainty shock. These heterogeneous impacts are due to the distinct nature of the two uncertainty shocks. Third, while the cash-flow uncertainty always has significant positive impact on the idiosyncratic cash flow risk premium, the growth uncertainty has almost no effect on idiosyncratic risk premia when the risk sharing condition is good. The reason is that the investment shock effect dominates when the risk sharing condition is good.

Amplification: uncertainty shocks compromise risk sharing conditions. In the model, the risk sharing condition is endogenously affected by the two uncertainty shocks. To establish the link between uncertainty shocks and the risk sharing condition, I consider two different types of measures of how much risk sharing is limited. The first type of measure is based on the idea that the covariance between an agent's net worth and idiosyncratic risks is always zero when the market is complete (i.e., risk sharing is perfect). When experts have a larger exposure to idiosyncratic shocks in their net worth, there is a larger cross-sectional dispersion in growth rates of individual consumption shares. So, the cross-sectional dispersion in growth rates of individual consumption shares provides a reasonable measure for the risk sharing imperfectness. The second type of measure is based on the idea that the marginal value of net worth should be identical across all agents when the market is complete. Therefore, the discrepancy between agents' marginal value of net worth serves as another natural measure of risk sharing imperfectness.[23]

Consumption dispersion. In equilibrium, the household's net worth is independent of all idiosyncratic risks, while the incentive constraints force each expert f to expose his own net worth to the particular idiosyncratic risk $\mathrm{d}W_{f,t}$. Importantly, the expert's idiosyncratic risk exposure is endogenous and hence time varying.

But, it only depends on the aggregate states in the economy. For each expert f, the conditional instantaneous covariance of net worth growth with the idiosyncratic shock $\mathrm{d}W_{f,t}$ is

$$\mathrm{Cov}_t\left(\frac{\mathrm{d}n^{\mathrm{e}}_{f,t}}{n^{\mathrm{e}}_{f,t}},\ \mathrm{d}W_{f,t}\right)/\mathrm{d}t = \nu_{c,t} \times \phi \times \frac{1}{x_t} \times \digamma_{\mathbf{k}}(x_t, \nu_t),$$

where $\digamma_{\mathbf{k}}(x_t, \nu_t) \equiv q_t/Q_t$ is the value share of asset in place in total financial wealth. Moreover, the conditional instantaneous covariance of net worth growth with the idiosyncratic standardized shock $\mathrm{d}\tilde{N}_{f,t}$, which is normalized by aggregate profit rate of growth options π_t, is

$$\mathrm{Cov}_t\left(\frac{\mathrm{d}n^{\mathrm{e}}_{f,t}}{n^{\mathrm{e}}_{f,t}},\ \mathrm{d}\tilde{N}_{f,t}\right)/\mathrm{d}t = \mathcal{I}_\alpha(\nu_{g,t}/\xi_t) \times \phi \times \frac{1}{x_t} \times \digamma_{\mathbf{s}}(x_t, \nu_t),$$

where $\digamma_{\mathbf{s}}(x_t, \nu_t) \equiv p_t\bar{s}/Q_t$ is the value share of growth options in total financial wealth, and $\mathcal{I}_\alpha(\cdot)$ is a deterministic function which is strictly increasing.

It can be seen that the severity of agency problems characterized by ϕ, the expert's wealth share x_t, and the uncertainties $\nu_{c,t}$ and $\nu_{g,t}$ have direct monotonic impact on the risk sharing capacity measures, up to some general equilibrium valuation effects $\digamma_{\mathbf{k}}(x_t, \nu_t)$ and $\digamma_{\mathbf{s}}(x_t, \nu_t)$. The severity of agency problem, the wealth share and the uncertainties can also affect the risk sharing capacity indirectly, which can be summarized

[23] The detailed derivations in this section can be found in the online appendix.

by the general equilibrium effects $\digamma_{\mathbf{k}}(x_t, \nu_t)$ and $\digamma_{\mathbf{s}}(x_t, \nu_t)$.

For each expert f, the consumption share is defined as $S^{\mathrm{e}}_{f,t} \equiv c^{\mathrm{e}}_{f,t}/c^{\mathrm{e}}_{t}$. I use the cross-sectional standard deviation (CSD) of consumption share growth rates to capture the dispersion. The cross-sectional dispersion of consumption growth depends on the idiosyncratic risk exposures. The instantaneous cross-sectional variance of consumption share growth rates has the following analytical expression

$$\mathrm{var}_t\left(\frac{\mathrm{d}S^{\mathrm{e}}_{f,t}}{S^{\mathrm{e}}_{f,t}}\right)/\mathrm{d}t = \left[\nu_{c,t}\frac{\phi}{x_t}\digamma_{\mathbf{k}}(x_t,\nu_t)\right]^2 + \left[\mathcal{I}_\alpha(\nu_{g,t}/\xi_t)\frac{\phi}{x_t}\digamma_{\mathbf{s}}(x_t,\nu_t)\right]^2.$$

The basic idea of the proof is that each individual expert's consumption share is equal to his net worth share in the equilibrium. That is, $c^{\mathrm{e}}_{f,t}/c^{\mathrm{e}}_{t} = n^{\mathrm{e}}_{f,t}/n^{\mathrm{e}}_{t}$. Therefore, the cross-sectional instantaneous variance of the growth rates of consumption shares is equal to the instantaneous idiosyncratic variance of individual consumption growth.

Therefore, the instantaneous cross-sectional variance of the experts' consumption share growth rates is linked to their exposures of idiosyncratic risks in the following way:

$$\mathrm{var}_t\left(\frac{\mathrm{d}S^{\mathrm{e}}_{f,t}}{S^{\mathrm{e}}_{f,t}}\right)/\mathrm{d}t = \left[\mathrm{Cov}_t\left(\frac{\mathrm{d}n^{\mathrm{e}}_{f,t}}{n^{\mathrm{e}}_{f,t}}, \mathrm{d}W_{f,t}\right)/\mathrm{d}t\right]^2 + \left[\mathrm{Cov}_t\left(\frac{\mathrm{d}n^{\mathrm{e}}_{f,t}}{n^{\mathrm{e}}_{f,t}}, \mathrm{d}\tilde{N}_{f,t}\right)/\mathrm{d}t\right]^2.$$

The instantaneous cross-sectional standard deviation of $\frac{\mathrm{d}S^{\mathrm{e}}_{f,t}}{S^{\mathrm{e}}_{f,t}}$ across all experts is defined as a measure for the risk sharing imperfection (i.e. the inverse of risk sharing condition). More precisely, I define

$$\Xi_t \equiv \sqrt{\mathrm{var}_t\left(\frac{\mathrm{d}S^{\mathrm{e}}_{f,t}}{S^{\mathrm{e}}_{f,t}}\right)} = \sqrt{\left[\nu_{c,t}\frac{\phi}{x_t}\digamma_{\mathbf{k}}(x_t,\nu_t)\right]^2 + \left[\mathcal{I}_\alpha(\nu_{g,t}/\xi_t)\frac{\phi}{x_t}\digamma_{\mathbf{s}}(x_t,\nu_t)\right]^2}.$$

Marginal value gap. In complete market, the marginal value of wealth for agents should be identical. Thus, the gap between two marginal values of wealth can serve as a index for risk sharing imperfection (i.e. the inverse of risk sharing condition). I define

$$\Theta_t \equiv \log(\zeta_t^{\mathrm{e}}) - \log(\zeta_t^{\mathrm{h}}). \tag{2.32}$$

The quantity is called marginal value gap. It is obvious that Θ_t in always nonnegative in the equilibrium. This is because experts have the access to investing in assets in place and growth options in the spot capital market, whereas households are excluded from such investment opportunities. As a result, experts always get more utility per unit of net worth than households. Thus, in equilibrium, it holds that $\Theta_t \equiv \log(\zeta_t^{\mathrm{e}}) - \log(\zeta_t^{\mathrm{h}}) \geq 0$.

Figure 2-4 illustrates the consumption dispersion and marginal value gap under the calibration summarized in Table 4.4. The uncertainty shocks deteriorate the risk sharing condition by increasing the consumption dispersion and the marginal value gap. The effect of uncertainty shocks on the consumption dispersion increases nonlinearly as the risk sharing becomes more limited (i.e., x_t decreases). The reason is that the expert's net worth has larger exposure to the idiosyncratic shocks when x_t is lower. This is particularly true for growth uncertainty shocks' impact on consumption dispersions.

Imperfect risk sharing on uncertainty shocks: from an optimal portfolio perspective. When uncertainty rises, the idiosyncratic risk premia go up. However, experts are the only ones who can take advantage of the higher idiosyncratic risk premia by investing more in real assets. As a result, experts' investment environment deteriorates relative to households. Therefore, the risk sharing between experts and households is endogenously imperfect due to the incomplete market faced by experts. The imperfect risk sharing on uncertainty shocks can be seen from the optimal portfolio holdings of households, which deviate from the market portfolio by significant hedging components. More precisely, each household's portfolio holding can be characterized by $\left(\varphi\sigma, \varsigma_{n,t}^{\mathrm{h},(\nu_{g,t},\nu_g)}, \varsigma_{n,t}^{\mathrm{h},(\nu_{c,t},\nu_c)}\right)$. The risk sharing on Z_t is perfect and thus the optimal holding is the market portfolio or the myopic component. However, the optimal exposure to uncertainty shocks features significant hedging components with the following analytical expressions:

$$\varsigma_{n,t}^{\mathrm{h},(\nu_t,\nu)} = \underbrace{\varsigma_t^{Q,(\nu_t,\nu)}}_{\text{market portfolio}} + \underbrace{x_t \frac{\gamma-1}{\gamma} \varsigma_t^{\Theta,(\nu_t,\nu)}}_{\text{hedging for relative environment}},$$

where $\varsigma_t^{\Theta,(\nu_t,\nu)}$ is the effect of uncertainty shocks on the marginal value gap Θ_t. Under the benchmark calibration, it holds that $\varsigma_t^{\Theta,(\nu_g^L,\nu_g^H)} > 0$ and $\varsigma_t^{\Theta,(\nu_c^L,\nu_c^H)} > 0$. It means that higher growth uncertainty or cash-flow uncertainty compromises the relative investment environment of households. If $\gamma > 1$ (intra-temporal wealth effect dominates), households have hedging motives to rising uncertainty; that is, $x_t \frac{\gamma-1}{\gamma} \varsigma_t^{\Theta,(\nu_g^L,\nu_g^H)} > 0$ and $x_t \frac{\gamma-1}{\gamma} \varsigma_t^{\Theta,(\nu_c^L,\nu_c^H)} > 0$.

There are three points that are worth mentioning. First, when x_t is large and thus the risk sharing is high, the hedging components for the relative investment environment are almost gone because $\varsigma_t^{\Theta,(\nu_t,\nu)} \approx 0$ for all ν_t, ν. Thus, the optimal holdings go back to the market portfolio $\varsigma_t^{Q,(\nu_t,\nu)}$. In the extreme where the risk sharing is perfect, households only hold the market portfolio. Second, hedging motives mirrored in the framework of inter-temporal capital asset pricing (ICAPM) models (e.g. Merton, 1973a; Campbell, 1993), the equity risk premium and the value premium are theoretically and quantitatively accounted by the covariance between stock returns and relative investment environment of experts Θ_t which is negatively driven by uncertainty shocks. Third, the market portfolio $\varsigma_t^{Q,(\nu_t,\nu)}$ does not only contain the myopic component but also hedging component, since uncertainty shocks affect the overall investment environment.

Displacement risks: why growth uncertainty can be so fearful? Compared to cash-flow uncertainty shocks, higher growth uncertainty causes an additional risk to experts, the risk of increasing inequality in the distribution of innovation benefits from growth options. The skewness in the distribution of innovation benefits matters when the risk sharing on idiosyncratic investment shocks is limited. Thus, this risk becomes particularly devastating when risk sharing condition is poor. The intuition can be seen clearly from the impulse-response analysis illustrated in Figure 2-5. Panel A shows the growth uncertainty shock that hits the economy. It is a temporal shock with half life 1.6 years. Panels B, C, and D show the responses of experts' aggregate consumption, experts' consumption dispersion, and the median of the cross-section of experts' consumption shares, respectively. The blue solid curve is the response when the risk sharing condition is good, while the red dashed curve is the response when the risk sharing condition is poor. In Panel A, it is clear that the growth uncertainty shock works as an investment shock which transfer current consumption to future with a larger amount when the risk sharing is not limited; the growth uncertainty shock destroys consumption and it takes a long time to recover when the risk sharing is seriously limited. In Panel C, the variance of the cross-section of experts' consumption shares increases dramatically with the growth uncertainty shock when risk sharing is limited; they are almost not affected when risk sharing is not limited. Panel D shows that the cross-sectional distribution of experts becomes permanently more skewed, though the conditional cross-sectional variance of consumption share growth comes back to the original level quickly. The distribution is extremely skewed when the risk sharing is limited. However, the distribution does not matter for the equilibrium. It is a manifestation of the skewed wealth transfer among experts. The reason for the skewed wealth transfer is that when risk sharing is limited, experts cannot efficiently insure the idiosyncratic risks in investment opportunities. Thus, most of the benefits from innovation accrue to a small fraction of experts, while the majority of experts bear the cost of creative destruction since they need to pay for the new assets in place. Essentially, the wealth is reallocated from those who do not invest to those who receive high-quality investment opportunities. This reallocation becomes more skewed when growth uncertainty is high, because of asymmetric benefits of growth options. In other words, each expert faces a more skewed idiosyncratic investment risk. Because experts are risk-averse, the higher skewness of idiosyncratic risk decreases the expert's certainty-equivalent wealth. The displacement risk interacting with financial constraints is amplified.

2.4 Empirical Evidence

In this section, I analyze the model's key predictions on the asset pricing implications of uncertainty shocks in the data, using empirical measures of growth uncertainty shocks and cash-flow uncertainty shocks. In Section 2.4.1, I empirically construct both growth uncertainty and cash-flow uncertainty based on idiosyncratic stock return volatilities in a panel of U.S. public firms. In Section 2.4.2, I provide an alternative measurement of growth uncertainty and cash-flow uncertainty based on the time-varying cross-sectional dispersion of fundamental cash flows and investments, respectively. I show that the two alternative approaches produce coherent measurements of uncertainty, as predicted by the model. In Section 2.4.3, I examine the role of risk sharing condition for determining the securities' exposures to two kinds of uncertainty shocks. I explore whether time-varying cross-sectional heterogeneous risk exposures to growth uncertainty shocks and cash-flow uncertainty shocks can rationale the observed differences in average returns between value and growth firms.

2.4.1 Measuring Uncertainty Based on Idiosyncratic Stock Returns

The idiosyncratic return volatility on assets in place. I denote $\mathrm{d}\tilde{R}^{\mathbf{k}}_{f,t}$ as the instantaneous return on the equity of assets in place. The return is exposed to all three aggregate shocks in the economy with the risk loadings determined in equilibrium. Its conditional expected return $\mathbb{E}_t\left[\mathrm{d}\tilde{R}^{\mathbf{k}}_{f,t}\right]$ is determined by these risk loadings and market price of risk for the aggregate shocks according to the beta pricing rule in Equations (2.30) and (2.31). The equity return exposes to the idiosyncratic cash flow shock $\mathrm{d}W_{f,t}$. In the model, the idiosyncratic equity return is captured by the term $\nu_{c,t}\mathrm{d}W_{f,t}$. The instantaneous equity return on assets in place is characterized as follows:

$$\mathrm{d}\tilde{R}^{\mathbf{k}}_{f,t} = \underbrace{\mathbb{E}_t\left[\mathrm{d}\tilde{R}^{\mathbf{k}}_{f,t}\right]}_{\text{expected return}} + \underbrace{\varphi\sigma\mathrm{d}Z_t + \sum_{\nu\neq\nu_t}\varsigma^{q,(\nu_t,\nu)}\left(\mathrm{d}N^{(\nu_t,\nu)}_t - \lambda^{(\nu_t,\nu)}\mathrm{d}t\right)}_{\text{aggregate return risk}}$$
$$+ \underbrace{\nu_{c,t}\mathrm{d}W_{f,t}.}_{\text{idiosyncratic return risk}}$$

It is obvious that the idiosyncratic volatility of $\mathrm{d}\tilde{R}^{\mathbf{k}}_{f,t}$ is simply $\nu_{c,t}$. That is, if I denote, by $\sigma_{\mathbf{k},t}$, the idiosyncratic volatility of equity return on assets in place, the following relationship holds

$$\sigma_{\mathbf{k},t} \equiv \mathrm{ivol}_t\left(\mathrm{d}\tilde{R}^{\mathbf{k}}_{f,t}\right) = \nu_{c,t}. \tag{2.33}$$

The relation in (2.33) shows that the cash-flow uncertainty $\nu_{c,t}$ can be identified and measured by the idiosyncratic volatility of equity return on assets in place $\sigma_{\mathbf{k},t}$.

The idiosyncratic return volatility on growth options. I denote $\mathrm{d}\tilde{R}^{\mathbf{s}}_{f,t}$ as the instantaneous return on the equity of growth options. The return is exposed to all three aggregate shocks in the economy with the risk loadings determined in equilibrium. Its conditional expected return $\mathbb{E}_t\left[\mathrm{d}\tilde{R}^{\mathbf{s}}_{f,t}\right]$ is determined by these risk loadings and market price of risk for the aggregate shocks according to the beta pricing rule in Equations (2.30) and (2.31). The equity return exposes to the idiosyncratic investment shock $\pi_{f,t}\mathrm{d}N_t - \mathbb{E}^{\varepsilon}(\pi_{f,t})\lambda\mathrm{d}t$ which compounds the idiosyncratic investment opportunity shock $\mathrm{d}N_{f,t}$ with the idiosyncratic IST shock $\varepsilon_{f,t}$. In the model, the instantaneous equity return on growth options is

$$\mathrm{d}\tilde{R}^{\mathbf{s}}_{f,t} = \underbrace{\mathbb{E}_t\left[\mathrm{d}\tilde{R}^{\mathbf{s}}_{f,t}\right]}_{\text{expected return}} + \underbrace{\varphi\sigma\mathrm{d}Z_t + \sum_{\nu\neq\nu_t}\varsigma^{p,(\nu_t,\nu')}\left(\mathrm{d}N^{(\nu_t,\nu)}_t - \lambda^{(\nu,\nu)}\mathrm{d}t\right)}_{\text{aggregate return risk}}$$
$$+ \underbrace{\left(\pi_{f,t}\mathrm{d}N_{f,t} - \mathbb{E}^{\varepsilon}(\pi_{f,t})\lambda\mathrm{d}t\right).}_{\text{idiosyncratic return risk}}$$

Denote, by $\sigma_{\mathbf{s},t}$, the idiosyncratic equity return volatility of growth options. It is defined as follows:

$$\sigma^2_{\mathbf{s},t} \equiv \text{ivol}_t \left(\mathrm{d}\tilde{R}^{\mathbf{s}}_{f,t} \right)^2 = \text{var}_t \left[\pi_{f,t} \mathrm{d}N_{f,t} - \mathbb{E}^{\varepsilon}(\pi_{f,t}) \lambda \mathrm{d}t \right] / \mathrm{d}t, \tag{2.34}$$

where $\text{var}_t \left[\pi_{f,t} \mathrm{d}N_t - \mathbb{E}^{\varepsilon}(\pi_{f,t}) \lambda \mathrm{d}t \right]$ is the variance of $\pi_{f,t} \mathrm{d}N_t - \mathbb{E}^{\varepsilon}(\pi_{f,t}) \lambda \mathrm{d}t$. In fact, under the model's assumptions, the idiosyncratic volatility can be expressed in terms of economically meaningful variables. I summarize this result in Proposition 7.

Proposition 7 (Identification of Growth Uncertainty)**.** *Suppose $\alpha \in (0,1)$. Then, the idiosyncratic volatility of equity returns on growth options can be expressed in the following form:*

$$\sigma_{\mathbf{s},t} = \mathcal{I}_{\alpha}(\nu_{g,t}/\xi_t), \tag{2.35}$$

where $\mathcal{I}_{\alpha}(\cdot)$ is a deterministic function which is strictly increasing.

The shock in the growth uncertainty $\nu_{g,t}$ can be captured by

$$\Delta \log(\nu_{g,t}) = \Delta \log \left[\mathcal{I}^{-1}_{\alpha}(\sigma_{\mathbf{s},t}) \right] + \Delta \log(\xi_t). \tag{2.36}$$

According to (2.6), the second term on the right hand of (2.36) can be expressed as follows

$$\Delta \log(\xi_t) = (1-\alpha) \Delta \log(\hat{q}_t / \hat{p}_t) + \alpha \Delta \log(\hat{\tau}_t / \hat{q}_t). \tag{2.37}$$

The variables in (2.36) and (2.37) can all be approximated empirically. The volatility of $\Delta \log(\xi_t)$ is smaller than that of $\Delta \log \left[\mathcal{I}^{-1}_{\alpha}(\sigma_{\mathbf{s},t}) \right]$ by an order of magnitude in the data.[24] Thus, approximately, the growth uncertainty shocks can be constructed based

[24] I approximate $\sigma_{\mathbf{s},t}$ by using the lowest 10% U.S. public firms according to their book-to-market ratios. Their annual returns are available from Ken French's web site and the annual idiosyncratic return volatility can be constructed as in (2.39). The valuation ratio between assets in place and growth options $\hat{q}_t / \hat{p}_t$ can be approximated by using annual Compustat data. More precisely, I follow Fazzari, Hubbard, and Petersen (1988) and Kaplan and Zingales (1997). The relative price of investment goods is measured by the relative price of new equipments as in Greenwood, Hercowitz, and Krusell (1997, 2000), Cummins and Violante (2002), and Papanikolaou (2011). Specifically, I follow the method adopted by Israelsen (2010) to extend the quality-adjusted relative price of investment goods proposed by Gordon (1990) and Cummins and Violante (2002) to 2014.

only on the idiosyncratic equity return volatility $\sigma_{\mathbf{s},t}$:

$$\Delta \log(\nu_{g,t}) \approx \Delta \log \left[\mathcal{I}_{\alpha}^{-1}(\sigma_{\mathbf{s},t}) \right]. \tag{2.38}$$

Methodology. Guided by the implications of the model above, I construct two sources of uncertainty shocks using the cross section of asset returns. The model suggests that the cash-flow uncertainty and the growth uncertainty are identified by the idiosyncratic volatilities of equity returns on assets in place (Equation (2.33)) and on growth options (Equation (2.38)), respectively. However, there is no equivalent to an equity on pure growth options or an equity on pure assets in place in the data. So, I appeal to stock returns of firms with low and high book-to-market ratios to approximate the equity returns on growth options and assets in place, respectively. An advantage of these measures is that they are available at high frequencies since they are based on financial data.

To be more precise, I sort firms into 10 portfolios on the basis of book-to-market ratios.[25] The basic idea is to use a firm's book-to-market ratio as an inverse measure of growth opportunities held by the firm. This idea follows the conventional wisdom that the market value of growth options is accounted in the market value of the firm, but not in the book value of assets. As a result, a firm's book-to-market ratio should be negatively associated with the firm's value of growth options relative to its total value. In a recent paper by Kogan and Papanikolaou (2014), the authors empirically validate the book-to-market ratio by comparing it with an alternative empirical measure of growth opportunities based on return sensitivity to investment-specific technology shocks. The break points for sorting firms are based on New York Stock Exchange deciles of book-to-market ratios, provided on Ken French's web site.

I first extract the idiosyncratic component of log returns. For each firm, the idiosyncratic component is constructed for every month. More precisely, to obtain the idiosyncratic component of firm f within the month t_m,[26] I appeal to the following

[25] The book-to-market ratio is computed using book equity and market capitalization constructed from Compustat items. I strictly following Fama and French's methodology.

[26] A firm-month return observation is included if (i) the stock has CRSP share code 10 or 11, and (ii)

factor model

$$r_{f,t_d} - r_{t_d} = a_{f,t_m} + \beta_{f,t_m}^T \mathbf{F}_{t_d} + \epsilon_{f,t_d}, \tag{2.39}$$

where $\mathbf{F}_{t_d}$ denotes the vector of factors and t_d denotes a daily observation in the past L_m months indexed by $t_m, t_m - 1, \cdots, t_m - L_m + 1$. The idiosyncratic component of the log return is estimated by the regression residual ϵ_{f,t_d} in (2.39). In the benchmark case, I choose $L_m = 3$.[27] Also, I use Fama and French (1993) three factors for the factor structure, following the standard in literature. Robustness check shows that the main results are not altered if using market returns, four factors in Carhart (1997), or principal components similar to Herskovic, Kelly, Lustig, and Nieuwerburgh (2014).

The idiosyncratic volatility of firm f's stock return in month t_m is then calculated as the standard deviation of residuals ϵ_{f,t_d} within that month, not including residuals in months $t_m - L_m + 1, \cdots, t_m - 1$. However, the standard deviation of ϵ_{f,t_d} is the idiosyncratic volatility of leveraged stock returns instead of all-equity returns as in the model; in other words, it is an amplified volatility by firm's financial leverage. More precisely, the underlying idiosyncratic shock to the value of firm's assets is amplified by a factor of the leverage ratio and pass through to the idiosyncratic equity returns. So the idiosyncratic volatility of stock returns is amplified by a factor of the leverage ratio. The leverage ratio is constructed by the sum of the book value of debt and the market value of equity divided by the market value of equity, similar to Welch (2004).[28] To construct the idiosyncratic volatility of the all-equity firm's returns, I need to adjust $\text{std}(\epsilon_f)$ by the leverage ratio. As a result, a panel of firm-month idiosyncratic volatility estimates for all-equity firms are obtained. Building on this firm-month panel, I construct two monthly time series, denoted by $\sigma_{\mathbf{k},t_m}$ and $\sigma_{\mathbf{s},t_m}$.

the firm has at least 17 return observations within the month, and (iii) the firm has no missing returns for the past 36 months.

[27]The empirical results are robust to alternative choices $L_m = 1, 2, 4, 6$. The way I constructed idiosyncratic returns are similar to the approach used in Herskovic, Kelly, Lustig, and Nieuwerburgh (2014) except several divergences. First, we have different frequencies. I focus on monthly idiosyncratic volatilities, whereas they construct annually idiosyncratic volatilities. Second, my factor regressions have overlapping rolling windows, while theirs rolling windows for factor regressions do not overlap.

[28]If the leverage ratio is missing from Compustat, I use the overall average leverage ratio in the same category as an approximation.

For constructing the series $\sigma_{\mathbf{k},t_m}$, I use the equal-weighted average of the idiosyncratic volatilities of those firms whose book-to-market ratios lie in the top 30% quantiles.[29] On the other hand, for the series $\sigma_{\mathbf{s},t_m}$, I use the idiosyncratic volatilities of those firms whose book-to-market ratios lie within the bottom three deciles. I first compute the average idiosyncratic volatilities within each category. I then regress each of the three series of logged average idiosyncratic volatilities onto the series $\log(\sigma_{\mathbf{k},t_m})$. The first principle component of the three residual series accounts for over 76% of their total variation. I use the first principle component to construct $\log(\sigma_{\mathbf{s},t_m})$.

Eventually, I construct an index tracking time-variation of the cash-flow uncertainty ν_{c,t_m} by the series $\log(\sigma_{\mathbf{k},t_m})$ with the constant level and linear trend taken out.[30] An index of tracking the growth uncertainty ν_{g,t_m} can be constructed by using $\log(\sigma_{\mathbf{s},t_m})$ with the constant level and linear trend taken out. It is quite intuitive why we need to take out the effect of cash-flow uncertainty from the average idiosyncratic volatility of low book-to-market firm stock returns. It is simply because their idiosyncratic volatilities are inevitably affected by cash-flow uncertainty shocks.

Results. The quarterly uncertainty indices, denoted by ν_{g,t_q} and ν_{c,t_q}, and annually uncertainty indices, denoted by ν_{g,t_y} and ν_{c,t_y}, are simply defined as the average of monthly uncertainty indices within each quarter and each year, respectively. Figure 2-6 illustrates the time variation of the cash-flow uncertainty annual index and the growth uncertainty annual index. The horizontal segments represent estimated high/medium/low regimes of uncertainties. The levels of the segments in the plots are the estimated average levels of uncertainty within each regime. The regimes and levels are estimated using regime-switching models (e.g., Hamilton, 1989, 1994; Tim-

[29] I also use the first principle component of the three average idiosyncratic volatilities for the firms with the highest 10%, with the second highest 10%, and with the third highest 10% book-to-market ratios. The first principle component accounts for over 73% of the total variation. It leads to almost the same result for the approximation of $\nu_{c,t}$.

[30] In order to focus on business-cycle behaviors, it is important to get rid of the long-run increasing trend in firm-level idiosyncratic volatilities of stock returns. There are indeed significant long-run increasing trends in both σ_{g,t_m} and σ_{c,t_m}, consistent with empirical results in Campbell, Lettau, Malkiel, and Xu (2001).

mermann, 2000).[31] Here, I employ the simplest regime-switching model specification for the cash-flow uncertainty:

$$\nu_{c,t_y} = a(\omega_{c,t_y}) + \epsilon_{c,t_y} \tag{2.40}$$

where ω_{c,t_y} is a latent state variable that follows a Markov chain jumping between discrete states over time and the constant $a(\omega_{c,t_y})$ characterizes the average uncertainty level within each regime. The residual ϵ_{c,t_y} is assumed to be uncorrelated with the latent state variable ω_{c,t_y}. This simplest regime-switching model basically provides a time-series clustering analysis, in which a particular year is classified as a high/medium/low regime when its likelihood of being in that regime is larger than 50%. This simple clustering analysis helps us with a better understanding of the uncertainty dynamics. The estimated regimes and their transition probabilities are useful in calibrating of the model's transition matrices of the Markov chains of uncertainty levels. The regime-switching model is estimated using the EM algorithm which maximizes the marginal likelihood of observable variables. For the growth uncertainty, its regime shifts are estimated similarly as in the specification (2.40). The point estimation of Markov transition probabilities are summarized in Table 2.6.

It is observed from Table 2.6 that the high growth uncertainty regime is more persistent compared with the high regime of cash-flow uncertainty. The conditional probability of staying in high state is 81.9% for the growth uncertainty and is 67.5% for the cash-flow uncertainty. Also, in the long run, the growth uncertainty stays in the high state much more often than the cash-flow uncertainty (47.6% versus 24.9%). The growth uncertainty on average lasts longer in the high state, because it is usually associated with political unstable periods, technological revolutions, and energy supply shifts; the resolution of the uncertainty about those events typically takes a long pe-

[31] The regime-switching model has not only proved its success in macroeconomics, but also been widely adopted in asset pricing and financial portfolio research (e.g., Ang and Bekaert, 2002; Dai, Singleton, and Yang, 2007).

riod.[32] Table 2.6 provides targets for the calibrations of $\mathcal{Q}_c$ and $\mathcal{Q}_g$ in Table 4.4.[33]

2.4.2 Over-identifications of Uncertainty: Idiosyncratic Dispersions

Because the idiosyncratic volatilities of value firm returns and growth firm returns proposed above are new measures for the cash-flow uncertainty and the growth uncertainty, it is important to first establish the validity of these measures. Specifically, by the definitions of two kinds of uncertainties, I derive more direct measures for the two kinds of uncertainties based on the cross-sectional distribution of idiosyncratic shocks in firm-level sales and capital expenditures. Particularly, the idiosyncratic dispersion in log sales growth rates should provide an ideal measure for the cash-flow uncertainty, while the idiosyncratic dispersion in log investment rates should provide an ideal measure for the growth uncertainty. The two measures based on the idiosyncratic dispersions strictly follow the formal definitions of the two sources of uncertainties; they are also consistent with the model's implications. However, the idiosyncratic-dispersion-based measures are only available at low frequencies (annual or quarterly) for a period of fifty years.[34] Now, I statistically verify whether the shocks constructed using the idiosyncratic volatility of stock returns validly serve as proxies for the uncertainty shocks

[32] The episodes of high growth uncertainty in the 1950s are mainly due to the fact that the 1950s was the first decade of post-war era and the starting decade of the Cold War. The international and domestic political uncertainty stayed very high for U.S. over the period. The episode of high growth uncertainty around 1970 is due to a major technological revolution in history (e.g., Perez, 2002). As the time approached the end of the 1960s, the old industries of oil, automobiles and mass production became matured, and new industries of information technology and telecommunications began to take the place from 1971. The episode of high growth uncertainty starting from the end of 1970s and lasting until the mid-1980s is mainly due to the long-lasting high oil price volatility (e.g., Peter Ferderer, 1996; Jo, 2012). The high oil price volatility was triggered by Iranian revolution from late 1978 to 1979. The Iranian Revolution, which began in late 1978, resulted in a drop of 3.9 million barrels per day of crude oil production from Iran and a large drop of oil supply from OPEC from 1978 to 1981. In early 1981, the U.S. Government responded to the oil crisis by removing price and allocation controls on the oil industry, which made oil prices more volatile. The episode of high growth uncertainty in the late 1990s is the result of the internet revolution. In the mid-1990s, the civilian Internet was transformed from a military safety net. At that time, the enormous potential of the internet to change all other industries and businesses aroused great growth uncertainty.

[33] The details for the calibration of $\mathcal{Q}_c$ and $\mathcal{Q}_g$ based on the estimated transition probabilities of uncertainty states can be found in the online appendix.

[34] I construct them from Compustat datasets. For quarterly frequency idiosyncratic dispersions, the time series are only available as early as 1984 which gives about 30 year data.

from two different origins.

Idiosyncratic dispersions of sales growth rates. The sales intensity of firm f over $[t, t + \mathrm{d}t)$ is $y_{f,t}$. According to Proposition 1, the sales is linear in firm's assets in place: $y_{f,t} = y(w_t)k_{f,t}$. Because the equilibrium wage w_t only depends on the aggregate state variables, it readily leads to the dynamics of log sales growth rates:

$$\mathrm{d}\log(y_{f,t}) = \underbrace{\mathrm{d}\log\left(y(w_t)\right) - \delta \mathrm{d}t + \sigma \mathrm{d}Z_t}_{\text{only depending on aggregate shocks}} + \underbrace{\nu_{c,t}\mathrm{d}W_{f,t}.}_{\text{idiosyncratic shocks}} \tag{2.41}$$

The only source of the cross-sectional heterogeneity comes from the idiosyncratic shocks $\nu_{c,t}\mathrm{d}W_{f,t}$. Thus, the interdecile range (IDR) in the cross section of log sales growth rates implied by the model is

$$\mathrm{IDR}\left[\mathrm{d}\log(y_{f,t})\right] = \aleph \nu_{c,t}, \tag{2.42}$$

where $\aleph$ is a universal constant that is approximately $\aleph \approx 2.5633$. Therefore, the cross-sectional dispersion in sales growth rates naturally identifies the cash-flow uncertainty, which basically justifies the name of such kind of uncertainty.

Idiosyncratic dispersions of investment rates. The firm-level investment rate, normalized by the aggregate investment rate, has the following expression in the model:

$$\frac{\tau_t g_{f,t}}{q_t k_{f,t}} / \frac{\tau_t g_t}{q_t k_t} = \lambda^{-1} \overline{\Gamma}_\alpha(\xi_t/\nu_{g,t})^{-1} \left(\frac{\varepsilon_{f,t}}{\nu_{g,t}}\right)^{\frac{1}{1-\alpha}} \mathbf{1}_{(\varepsilon_{f,t} \geq \xi_t)},$$

where $\frac{\tau_t g_{f,t}}{q_t k_{f,t}}$ is the firm-level capital expenditure normalized by tangible capital stock (the firm-level investment rate) and $\frac{\tau_t g_t}{q_t k_t}$ is the aggregate investment rate. The source of the cross-sectional heterogeneity comes from the idiosyncratic IST shock $\varepsilon_{f,t}$. The cross-sectional standard deviation (CSD) of idiosyncratic shocks in investment rates is characterized by a strictly increasing function of $\nu_{g,t}/\xi_t$. This is formally summarized in the following proposition with proofs given in the online appendix.

Proposition 8 (Growth Uncertainty versus Dispersions of Investment Rates)**.** *In equilibrium, the dispersion of idiosyncratic shocks in investment rates depends on $\nu_{g,t}$ positively. That is,*

$$CSD\left[\frac{\tau_t g_{f,t}}{q_t k_{f,t}} / \frac{\tau_t g_t}{q_t k_t}\right] = \lambda^{-1}\mathcal{J}_\alpha(\nu_{g,t}/\xi_t),$$

where $\mathcal{J}_\alpha(\cdot)$ is a deterministic strictly increasing function.

Similarly, the exercising boundary ξ_t is relatively much stable compared to $\nu_{g,t}$. As a result, the cross-sectional standard deviation CSD $\left[\frac{\tau_t g_{f,t}}{q_t k_{f,t}} / \frac{\tau_t g_t}{q_t k_t}\right]$ also provides an (approximate) identification for the growth uncertainty $\nu_{g,t}$.

Methodology. Now, I extract idiosyncratic shocks in sales growth rates and in investment rates. Once that is done, I compute their dispersions in the cross section of firms. In this empirical exercise of extracting the idiosyncratic unexpected component, I adopt the method similar to Purnanandam and Rajan (2014) in which the predictable component, the aggregate unexpected component, and the idiosyncratic unexpected component are statistically separated and estimated by using dynamic panel regression models (e.g., Holtz-Eakin, Newey, and Rosen, 1988; Arellano and Bond, 1991).

I first measure the idiosyncratic unexpected component of firm's investment rates. For each firm f, the investment rate of year t_y, denoted by IoK_{f,t_y}, is computed as the capital expenditure $CapEx_{f,t_y}$ deflated by capital stock of tangible assets K_{f,t_y-1} in the previous year, and then normalized by the aggregate investment-to-capital ratio

IoK_{t_y-1}.[35] That is,

$$IoK_{f,t_y} \equiv \left(CapEx_{f,t_y}/K_{f,t_y-1}\right)/IoK_{t_y-1}.$$

Here, the capital expenditure of firm f within year t_y is measured by the Compustat item capx, and the capital stock of tangible assets is measured by the Compustat item ppent. In this empirical exercise, I use the following regression model to extract the idiosyncratic shock in IoK_{f,t_y}:

$$\begin{aligned} IoK_{f,t_y} = & \underbrace{\lambda_{capx,t_y}}_{\text{agg. component}} \\ & + \underbrace{a_{capx,f} + \beta_{capx,1} IoK_{f,t_y-1} + \beta_{capx,2} CoK_{f,t_y-1} + \beta_{capx,3} MoB_{f,t_y-1}}_{\text{firm-level expected component}} \\ & + \underbrace{\epsilon_{capx,f,t_y}}_{\text{idio. shock}}, \end{aligned}$$

where $a_{capx,f}$ is the fixed effect capturing the firm-level predictability, λ_{capx,t_y} is the year effect capturing aggregate time-varying effect (can be caused by some latent aggregate factors) and captures the aggregate shock, CoK_{f,t_y} is the cash flow deflated by capital stock of tangible assets in the previous year, and MoB_{f,t_y} is the market-to-book ratio of assets capturing the investment opportunity of firm f in year t_y. The variables CoK_{f,t_y} and MoB_{f,t_y} are needed, particularly because the literature of the cash-flow-sensitivity of investment argues that cash flows can have impact on investment decisions. Though there are different ways to measure CoK_{f,t_y} and MoB_{f,t_y} in the data, my measures fol-

[35] In order to extract the idiosyncratic volatility of investment rates that only caused by the growth uncertainty $\nu_{g,t}$ (i.e. the idiosyncratic volatility), I need to remove the scaling effect time-varying volatility of aggregate investment rates. More precisely, the regression needs to make sure that the heteroskedasticity in the aggregate volatility of investment rate shocks does not alter the idiosyncratic shock ϵ_{capx,f,t_y} specified in the econometric model. Bachmann, Caballero, and Engel (2006) show that the volatility in the aggregate investment rate (IoK) is high when the past aggregate investment rate is high. So, I normalize the firm-level investment rate by the aggregate one; it serves as the simplest way to guarantee the idiosyncratic shock ϵ_{capx,f,t_y} not to be affected by the past aggregate investment rates through current aggregate volatility. This normalization is also consistent with the implications of the model in Proposition 8. I follow Bachmann, Caballero, and Engel (2006) and Favilukis and Lin (2013) to construct the aggregate investment rate using nominal annual private fixed nonresidential investment and the annual private nonresidential capital stock at year-end prices from the Bureau of Economic Analysis (BEA).

low the cash-flow-sensitivity of investment literature (e.g., Fazzari, Hubbard, and Petersen, 1988; Kaplan and Zingales, 1997).[36]

The residuals ϵ_{capx,f,t_y} in the regression model above capture the idiosyncratic shock in investment rates. The result of this procedure is a firm-year panel of idiosyncratic shocks ϵ_{capx,f,t_y}. I estimate the dynamic panel regression model using the GMM estimator proposed by Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991), with the first lagged value of capital expenditure rate as a GMM instrument variable.

Second, I measure the idiosyncratic unexpected component in sales growth rates. Following the literature (e.g., Bloom, 2009; Bachmann and Bayer, 2014; Herskovic, Kelly, Lustig, and Nieuwerburgh, 2014), the sales growth rate is measured as follows:

$$GoS_{f,t_y} \equiv \log\left(Sales_{f,t_y}/Sales_{f,t_y-1}\right).$$

Here, the variable $Sales_{f,t_y}$ is the sales of firm f within year t_y and I use the Compustat item sale for its values. I focus on the following regression model to extract the unexpected idiosyncratic component of GoS_{f,t_y}:

$$GoS_{f,t_y} = \underbrace{\lambda_{sale,t_y}}_{\text{agg. component}} + \underbrace{a_{sale,f} + \beta_{sale,1} GoS_{f,t_y-1}}_{\text{firm-level expected component}} + \underbrace{\epsilon_{sale,f,t_y}}_{\text{idio. shock}} \tag{2.43}$$

where $a_{sale,f}$ is the fixed effect capturing the firm-level predictability, and λ_{sale,t_y} is the year effect capturing the aggregate component (even there are latent factors) which includes the aggregate shock. The residuals ϵ_{sale,f,t_y} captures the idiosyncratic shocks in sales growth rates. The result of this procedure is a firm-year panel of idiosyncratic shocks. Similarly, I estimate the dynamic panel regression model using the GMM esti-

[36]More precisely, to construct CoK_{f,t_y}, the cash flow is measured by the sum of income before extraordinary items (Compustat item ib) and depreciation (Compustat item dp), and the capital stock of tangible assets is measured by net property, plant and equipment (Compustat item ppent). To construct MoB_{f,t_y}, the market value of assets is measured by the book value of assets (Compustat item at) plus the market value of common stock (Compustat item prccf $\times$ Compustat item csho) less the sum of book value of common stock (Compustat item ceq) and balance sheet deferred taxes (Compustat item txdb).

mator with the first lagged value of cash flow rate as a GMM instrument variable.

Now, after obtaining these two firm-year panels of idiosyncratic shocks, I construct two annual time series of cross-sectional dispersions, denoted by σ_{capx,t_y} and σ_{sale,t_y}. The series σ_{capx,t_y} are the cross-sectional standard deviations (CSD) of idiosyncratic shocks in investment rates across all firms within year t_y following Bachmann and Bayer (2014), while the series σ_{sale,t_y} are the interdecile ranges (IDR) of idiosyncratic shocks in sales growth rates across all firms within year t_y. Like the indices based on idiosyncratic stock volatilities, I focus on linearly-detrended series. In particular, I denote ν_{capx,t_y} and ν_{sale,t_y} the annual time series $\log\left(\sigma_{capx,t_y}\right)$ and $\log\left(\sigma_{sale,t_y}\right)$ with linear trends removed, respectively.

Results. Consistent with the predictions of my model (Equations (2.33) and (2.38)), the underlying shocks that drive the idiosyncratic sales dispersions ($\Delta\nu_{sale,t_y} \equiv \nu_{sale,t_y} - \nu_{sale,t_y-1}$) are particularly associated with the cash-flow uncertainty shocks ($\Delta\nu_{c,t_y} \equiv \nu_{c,t_y} - \nu_{c,t_y-1}$), but not with the growth uncertainty shocks ($\Delta\nu_{g,t_y} \equiv \nu_{g,t_y} - \nu_{g,t_y-1}$). On the contrary, as predicted by the model (Propositions 7, and 8), the shocks that drive the idiosyncratic investment dispersions ($\Delta\nu_{capx,t_y} \equiv \nu_{capx,t_y} - \nu_{capx,t_y-1}$) are particularly associated with the growth uncertainty shocks $\Delta\nu_{g,t_y}$, but not the cash-flow uncertainty shocks $\Delta\nu_{c,t_y}$. More precisely, Panel A of Figure 2-7 shows that the shocks of the idiosyncratic investments dispersions $\Delta\nu_{capx,t_y}$ can be statistically explained by the growth uncertainty shocks $\Delta\nu_{g,t_y}$ with the estimated slope 1.84 and the t-statistic 2.93. Panel B of Figure 2-7 shows that the shocks of idiosyncratic sales dispersions $\Delta\nu_{sale,t_y}$ cannot be statistically explained by the growth uncertainty shocks $\Delta\nu_{g,t_y}$, because the estimated slope -0.016 is not statistically different from zero. Its t-statistic is -0.028. Panel C of Figure 2-7 shows that the shocks of idiosyncratic investment dispersions $\Delta\nu_{capx,t_y}$ cannot be statistically explained by the cash-flow uncertainty shocks $\Delta\nu_{c,t_y}$. The slope is estimated to be -0.31 with t-statistic -0.62. Panel D of Figure 2-7 shows that the shocks of idiosyncratic sales dispersions $\Delta\nu_{sale,t_y}$ can be statistically explained by the cash-flow uncertainty shocks $\Delta\nu_{c,t_y}$ with the slope estimated to be 1.11 and its t-statistic to be 2.75.

Therefore, empirical evidence supports using the idiosyncratic volatilities of equity returns on assets in place as a measure of the cash-flow uncertainty and on growth options as a measure of the growth uncertainty. In other words, the uncertainty shocks of different origins, as fundamental macroeconomic shocks, can be identified and measured using panels of asset returns. Importantly, the asset pricing data allows for high frequency proxies for these underlying macroeconomic shocks. From the asset pricing perspective, the results show that the macroeconomic uncertainty shocks can have direct and significant impacts on the cross-sectional behavior of asset returns.

Discussion: cyclicality of cross-sectional dispersions. The cyclicality of the idiosyncratic investment dispersion ν_{capx,t_y} and the idiosyncratic sales dispersion ν_{sales,t_y}, as well as the growth uncertainty ν_{g,t_y} and the cash-flow uncertainty ν_{c,t_y}, are reported in Table 2.7. There, the cyclical component of real GDP per capita is estimated by using the one-sided HP filter. There are three points which worth mentioning about the statistics in Table 2.7. First, consistent with the main findings of Bachmann and Bayer (2014) as reproduced in Table 2.4, the cross-sectional dispersion of investment rates is statistically significantly pro-cyclical. In fact, the results reported here (the Pearson and Kendall correlations are 0.31 and 0.20, respectively) reinforce theirs. This is because the firm-specific predictable component, the aggregate predictable component, and the potential scaling effect have been all removed when the idiosyncratic sale dispersion ν_{sale,t_y} and the idiosyncratic investment dispersion ν_{capx,t_y} are constructed. Second, the dispersion of idiosyncratic shocks in sales growth rates ν_{sale,t_y} is countercyclical (the Pearson and Kendall correlations are -0.16 and -0.10, respectively), though annual estimated correlations are not significant (the p-values for Pearson and Kendall correlations are 0.26 and 0.32, respectively).[37] Third, the growth uncertainty $\nu_{g,t}$ is procyclical (the Pearson and Kendall correlations are 0.13 and 0.05, respectively), while the cash-flow uncertainty $\nu_{c,t}$ is strongly countercyclical (the Pearson and Kendall cor-

[37]The insignificance results can be a result of short sample length and high persistency in the annual level time series.

relations are -0.31 and -0.21, respectively).

Importantly, my theoretical and empirical results provide a natural and robust reconciliation for the so-called investment dispersion puzzle in the macroeconomics literature (e.g., Bachmann and Bayer, 2014). Basically, I show what has been missing in the macroeconomic models with heterogeneous firms is the growth uncertainty shocks. It's been a substantial literature documenting that the cross-sectional dispersion of micro-level fundamentals vary dramatically over time. In particular, it's been a consensus that the underlying shocks driving dispersions of sales have strongly adverse macroeconomic effects (e.g., Bloom, 2009; Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry, 2013; Herskovic, Kelly, Lustig, and Nieuwerburgh, 2014). However, in a recent work by Bachmann and Bayer (2014), the authors find pro-cyclical dispersion of firm-level investment in Germany, the United States, and the United Kingdoms. Quantitatively, they examine whether shocks in the dispersion of sales growth rates can generate pro-cyclical investment dispersions; they build their quantitative exercise upon the framework of Khan and Thomas (2008), Bloom (2009) and Bachmann, Caballero, and Engel (2013). They show that only very small shocks to sales growth dispersion can generate pro-cyclical investment dispersion, and shocks with such small scales fail to generate observed business cycles. These empirical patterns impose additional cross-equation restrictions on the properties of uncertainty shocks used in macroeconomic models; in particular, they pose quantitative challenges to the uncertainty-driven business cycle models, such as Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2013).

In my model with growth uncertainty shocks, as suggested theoretically and verified empirically, the dispersion of investment rates is mainly driven by the growth uncertainty, but not by the cash-flow uncertainty; on the contrary, the dispersion of sales growth rates is driven by the cash-flow uncertainty, but not by the growth uncertainty. Therefore, my model naturally reproduces the empirical patterns for the dynamics of sales dispersions and investment dispersions.

2.4.3 Inspecting the Mechanism: the Role of Risk Sharing Condition

In this section, I explore the empirical tests of the basic mechanism. The basic mechanism of the model is that the effects of growth uncertainty shocks on asset prices in the cross section are determined by the risk sharing condition in the economy. In particular, I focus on testing two most direct implications of the basic mechanism, which are summarized as follows. The first direct implication (Basic Implication I) is that in response to a positive growth uncertainty shock, the value of growth options increases relative to assets in place when the risk sharing condition is good, but decreases otherwise; the cash-flow uncertainty always tends to suppress the value of assets in place relative to the value of growth options, regardless of the risk sharing condition. The second direct implication (Basic Implication II) is that the growth uncertainty shock carries a positive market price of risk when risk sharing condition is good, but carries a negative one otherwise; the cash-flow uncertainty shock always carries a negative market price.

Testing basic implication I. I use the value spread, high minus low book-to-market portfolio returns, to approximate the relative value change of assets in place to growth options in the data. To add robustness of the testing results, I set up three tests using different econometric tools, and I also use three different measures of risk sharing conditions in the economy.

Regime-switching models. My model implies that the betas of value spreads with respect to growth uncertainty shocks are informative about the underlying state of risk sharing conditions. More precisely, when the beta of value spreads to growth uncertainty shocks is negative, the underlying risk sharing condition is likely to be good; alternatively, the underlying risk sharing condition is likely to be poor. In order to provide a direct test on this implication, I appeal to the regime-switching econometric model studied by Hamilton (1989, 1994) and Timmermann (2000). In my monthly regime-switching econometric specification, the underlying risk sharing condition is the latent state variable, denoted by ω_{x,t_m}. The latent state variable ω_{x,t_m} is to be uncovered from the data. It is assumed that ω_{x,t_m} follows a two-state Markov chain process; it's transition probabilities are to be estimated using the observables in the model. The observables include monthly market excess returns $r_{M,t_m} - r_{f,t_m}$, uncertainty shocks $\Delta\nu_{g,t_m}$ and $\Delta\nu_{c,t_m}$, and monthly value spreads $r_{H,t_m} - r_{L,t_m}$ with r_{H,t_m} and r_{L,t_m} to be returns of high and low book-to-market portfolio, respectively. More precisely, the econometric model is specified as follows:

$$\begin{aligned} r_{H,t_m} - r_{L,t_m} = a_{v,t_m} + \beta_{v,z,t_m}\left(r_{M,t_m} - r_{f,t_m}\right) \\ + \beta_{v,g,t_m}\Delta\nu_{g,t_m} + \beta_{v,c,t_m}\Delta\nu_{c,t_m} + \epsilon_{v,t_m} \end{aligned} \tag{2.44}$$

where the coefficients are time-varying and depend on the latent state $a_{v,t_m} \equiv a_v\left(\omega_{x,t_m}\right)$ and $\beta_{v,\iota,t_m} \equiv \beta_{v,\iota}\left(\omega_{x,t_m}\right)$ for $\iota \in \{z, g, c\}$. The latent state variable ω_{x,t_m} takes values in $\{\mathsf{Good}, \mathsf{Bad}\}$. Here, Good (Bad) stands for the state in which the risk sharing condition is good (bad). The state of risk sharing condition is unobservable in the econometric model and the identification implied by the theory is that

$$\beta_{v,g}\left(\mathsf{Good}\right) < \beta_{v,g}\left(\mathsf{Bad}\right). \tag{2.45}$$

Moreover, statistically, it is assumed that the residual term ϵ_{v,t_m} is not only uncorrelated with the input variables but also independent of the latent state variable ω_{x,t_m}.

I estimate the regime-switching model (2.44) using the EM algorithm that maximizes the marginal likelihood function of observables. The estimation results of the regime-switching model consist of two parts: one is the statistical inference about the coefficients which are summarized in Table 2.8; the other is the estimated likelihood of the risk sharing condition being Bad for every month. The estimated likelihoods are displayed in Figure 2-8.

In Table 2.8, Column (3) shows that the loadings of value spreads on growth uncertainty shocks change from negative ($\beta_{v,g}$ is estimated to be -1.76) to positive ($\beta_{v,g}$ is estimated to be 6.03) as the underlying state moves from Good to Bad. The signs are statistically significant at 75% confidence level. In the econometric analysis, the only restriction used for identifying the Good state is the inequality (2.45). There is no restriction imposed on the sign of growth uncertainty beta $\beta_{v,g}$ in the estimation. As a result, the sign switching itself empirically supports prediction of the theoretical model. It should be noted that the significance level of the coefficients tend to be understated compared to the econometric model in which the risk sharing condition is assumed to be known. This is because a large amount of randomness about the latent states have to be taken into account when drawing statistical inferences about the regression coefficients in the regime-switching model. Moreover, Column (4) of Table 2.8 verifies another prediction of the theory: the growth options always offer a hedge against the cash-flow uncertainty shock. More precisely, the coefficient $\beta_{v,c}$ is estimated to be negative in both states (-13.08 in Bad versus -1.00 in Good). In particular, the sign is significant at 95% confidence level in Bad and 75% confidence level in Good. However, it is still unclear whether the state Bad in the model truly corresponds to the state of poor risk sharing in the data. Thus, I need to compare the estimated Bad state with the measures of risk sharing conditions in the data.

In fact, the regime-switching econometric model does not offer an exact answer to the question which state the economy is in. Instead, it allows one to estimate the likelihood of the economy being in certain state.[38] In Figure 2-8, the estimated likelihood of being in Bad state is plotted in Panel D and is compared with three measures of risk sharing conditions in the data. The first empirical measure (in Panel A) is the Reinhart-Rogoff financial crisis index.[39] The second empirical measure (in Panel B) is the financial condition index based on broker-dealer leverages. The third empirical measure (in Panel C) is the credit spread index. Gilchrist and Zakrajsek (2012), Krishnamurthy and

[38] Of course, the state of the economy can be estimated based on the estimated likelihood. In practice, the economy is labeled by a particular state when the estimated likelihood of being that state is higher than a predetermined threshold. For example, 50% is used as the threshold, like in Figure 2-6.

[39] It is constructed based on U.S. banking/currency crisis, U.S. stock market crashes, U.K. banking/currency crisis, German banking/currency crisis, and France banking/currency crisis. I use a simplest nonlinear filter to form U.S. investors' expectation about financial sector conditions. If there are two or more crisis, investors have a bad outlook for financial conditions; if there is zero crisis, investors form a promising outlook for financial conditions; otherwise, they form a medium outlook.

Muir (2015), and Ivashina and Scharfstein (2010) show that credit spreads can serve as a crucial gauge of the degree of strains in the financial system. The basic idea is that fluctuations in credit spreads reflect shifts in the effective supply of funds offered by financial intermediaries. They found that an adverse shock to the equity valuations of the highly-leveraged financial intermediaries, relative to the market return, leads to an immediate and persistent increase in credit spreads.

My estimated likelihood of being in Bad state is plotted in Panel D. To interpret the levels of the time series in Figure 2-8, I set zero as the benchmark state in which the risk sharing condition is at its medium level. According to their definitions, positive index values indicate worse financial conditions than the medium state; negative index values indicate better financial conditions than the medium state. The three indices in Panels A–C capture the periods of stressed financial sector. Figure 2-8 shows that the Bad state is actually associated with poor financial conditions. Comparing the estimated financial condition (in Panel D) with the Reinhart-Rogoff financial condition index (in Panel A), the broker-dealer leverage index (in Panel B), and the credit spread index (in Panel C), the estimation results (in Panel D) are clearly consistent with the observations in the data (in Panels A, B, and C). More precisely, the four time series capture the major periods of financial stress in the history of the United States; at the same time, they also agree with each other upon the major periods of excellent financial conditions for U.S. economy.[40]

Most importantly, the estimation of financial condition (in Panel D) only depends on stock returns and the model's prediction about the cross-sectional impacts of growth uncertainty shocks. In other words, the estimation has almost zero prior information about financial conditions, which reinforces the power of the empirical result as a support for my theory. Now, I formally quantify the statistical association between the empirical measures of risk sharing conditions and the estimated likelihood of Bad state,

[40] My estimation, together with the three empirical measures, capture the financial crises around 1976, around 1990, around 2003, and around 2008; they also capture the periods of excellent financial conditions including the late 1990s, the periods around 2005, and the periods after 2014. In the online appendix, I also compare my estimation with other empirical measures of financial conditions including the financial condition index proposed by Brave and Butters (2011).

which are reported in Table 2.9. I use both the Pearson correlation and the Kendall rank correlation to quantify the associations. As reported in Columns (1) and (2) of Table 2.9, the credit spread index is used as the benchmark, and it is significantly correlated with both the Reinhart-Rogoff financial condition index, the broker-dealer leverage index, and the estimated likelihood of Bad state. In Columns (3) and (4) of Table 2.9, I also report the corresponding statistical associations of the financial condition indices in my model based on simulated samples. Because there is no corporate bond in my model, I use the equity premium as the proxy for credit spread. The risk sharing condition in the simulated data is measured by using Θ_t in (2.32).[41] The likelihood of Bad is estimated for the simulated data in the same way as for the real data. The associations between the simulated indices are comparable to those in the real data.

Uncertainty betas of book-to-market sorted portfolios. I also verify the theoretical implication by looking into the loadings of book-to-market sorted portfolios on uncertainty shocks $\Delta\nu_{c,t_m}$ and $\Delta\nu_{g,t_m}$ in subsamples corresponding to the periods of good or bad financial conditions. I first use the Reinhart-Rogoff index as the measure of risk sharing condition in the economy to construct subsamples. The betas of the book-to-market portfolios with respect to the market excess return $r_{M,t_m} - r_{t_m}$, the growth uncertainty shock $\Delta\nu_{g,t_m}$, and the cash-flow uncertainty shock $\Delta\nu_{c,t_m}$ are estimated within each of the two subsamples: one subsample includes the periods of good financial conditions; the other subsample includes periods of financial stress. The estimated betas are reported in Table 2.10. Panel A reports the beta estimates when the risk sharing condition is poor, while Panel B reports the beta estimates when the risk sharing condition is good. Comparing Columns (2) and (3) with Columns (5) and (6) in Table 2.10, the empirical results are almost perfectly in line with the theoretical prediction about the loadings on two sources of uncertainty shocks. More precisely, the beta on $\Delta\nu_{g,t}$ increases from -3.11 to 0.21 for the stock returns of the firms with the lowest 10%

[41] I can also use the endogenous state variable x_t or the consumption share dispersion Θ_t^e quantify risk sharing condition in the simulated data, because they are equally valid as the measure of risk sharing condition in my model.

book-to-market ratios (growth firms) versus those with highest 10% book-to-market ratios (value firms) when risk sharing condition is bad, as shown in Panel A with the sorting scheme #1; however, the growth uncertainty beta decreases from 2.40 to −12.66 for growth firms versus value firms when risk sharing condition is good, as shown in Panel B with the sorting scheme #1. Moreover, according to the sorting scheme #1, the beta on $\Delta \nu_{c,t_m}$, for the stock returns of growth firms versus value firms, decreases from 2.56 to −17.85 when the risk sharing condition is bad and decreases from 14.10 to −8.36 when the risk sharing condition is good. Importantly, as shown in Figure 2.10, the empirical findings are robust to various sorted book-to-market portfolios (e.g. the sorting schemes #2 and #3).

I then use the financial condition index based on broker-dealer leverages as the measure of risk sharing condition in the economy. The estimated betas are reported in Table 2.11. According to the sorting scheme #1, the beta on $\Delta \nu_{g,t_m}$ for growth firms versus value firms increases from −2.81 to 1.10 when the risk sharing condition is bad (in Panel A), while it decreases from −1.61 to −3.74 when the risk sharing condition is good (in Panel B). Moreover, under the sorting scheme #1, the beta on $\Delta \nu_{c,t_m}$ for growth firms versus value firms decreases from 6.80 (11.27) to −21.37 (−1.34) when the risk sharing condition is bad (good). The empirical results are robust across various sorting schemes (#2 and #3). Therefore, the results in Table 2.11 show that the empirical findings in Table 2.10 are quite robust against other measures of risk sharing conditions.

At last, I use the credit spread index as the measure of risk sharing condition in the economy. I fit the credit spread into a simplest three-state regime-switching model like in (2.40). The estimation results using 50% to be the threshold is to cluster each quarter into three categories: high/median/low credit spread levels.[42] The estimated betas are reported in Table 2.12. According to the sorting scheme #1, the beta on $\Delta \nu_{g,t_m}$

[42]The periods of low risk sharing condition (i.e. high credit spread) include 1974Q4 – 1976Q3, 1980Q2 – 1983Q3, 1988Q4 – 1992Q2, 2002Q1 – 2003Q4, 2008Q1 – 2009Q4, 2010Q2 – 2010Q4, and 2011Q4 – 2012Q4. On the other hand, the periods of high risk sharing condition (i.e. low credit spread) include 1973Q1 – 1974Q3, 1977Q3 – 1979Q3, 1987Q3 – 1988Q3, 2000Q1 – 2000Q3, 2004Q4 – 2005Q2, 2004Q4 – 2005Q2, and 2013Q4 – 2014Q3.

for growth firms versus value firms increases from −2.81 to 8.08 when the risk sharing condition is bad (in Panel A), while it decreases from 2.37 to −1.34 when the risk sharing condition is good (in Panel B). Moreover, under the sorting scheme #1, the beta on $\Delta\nu_{c,t_m}$ for growth firms versus value firms decreases from 0.78 (1.61) to −10.13 (−9.45) when the risk sharing condition is bad (good). The empirical results are robust across various sorting schemes (#2 and #3). Therefore, the results in Table 2.12 reinforce that the empirical findings in Table 2.10 are quite robust against other measures of risk sharing conditions.

However, statistically, it is still unclear how significantly the role of risk sharing conditions in altering the impact of growth uncertainty on the value of growth options relative to assets in place. To investigate the statistical significance, I compute the t-statistics for the estimated betas of extreme book-to-market-sorted portfolios. In Table 2.13, it shows that the sign changes (Column (1) versus Column (3)) are significant based on one-sample statistical tests; the statistical result is particularly strong when using the credit spread index as the measure for risk sharing conditions (in Panel C).

Linear models with interaction terms. Now, I set up a linear regression model in which the dependent variable is the value spread and the independent variables include the interaction terms between the uncertainty shocks and the risk sharing condition. The risk sharing condition is measured by the Reinhart-Rogoff financial condition index (reported in Columns (5) and (6)) or by the financial condition index based on broker-dealer leverages (reported in Columns (7) and (8)) or by the credit spread index (reported in Columns (3) and (4)). The regression model with interaction terms is specified as follows:

$$\begin{aligned} r_{H,t_y} - r_{L,t_y} = a_{vi} + \beta_{vi,z}\left(r_{M,t_y} - r_{t_y}\right) + \beta_{vi,g}\Delta\nu_{g,t_y} + \beta_{vi,c}\Delta\nu_{c,t_y} \\ + \beta_{vi,x}\text{regime-x}_{t_y} + \gamma_{vi,g}\left[\Delta\nu_{g,t_y} \times \text{regime-x}_{t_y}\right] \\ + \gamma_{vi,c}\left[\Delta\nu_{c,t_y} \times \text{regime-x}_{t_y}\right] + \epsilon_{vi,t_y} \end{aligned} \tag{2.46}$$

where vi in the subscript of coefficients means that they are coefficients for the value spread regression with interactions. Here, $r_{M,t_y} - r_{t_y}$ is the market excess return, $\Delta\nu_{g,t_y}$ and $\Delta\nu_{c,t_y}$ are uncertainty shocks, and $r_{H,t_y} - r_{L,t_y}$ is the value spread with r_{H,t_y} and r_{L,t_y} to be the returns of high and low book-to-market-portfolio returns, respectively. The variable regime-x_{t_y} is an aggregate state variable characterizing the condition of risk sharing in the economy.

The focus of this test has two folds: one is to test whether the coefficient $\gamma_{vi,g}$ in (2.46) is significantly positive; the other is to test whether the coefficient $\beta_{vi,c}$ is significantly negative. The regression (2.46) provides a formal statistical framework for testing whether the switching signs between Column (2) and Column (5) in Tables 2.10, 2.11, and 2.12 are statistically significant. In computing the t-statistics of coefficients, I appeal to Newey and West (1987a, 1994) for the robust covariance matrix estimation with one year lag.

In Table 2.14, Column (1) shows that the value premium exists; it is about 5.65% and statistically significant. More importantly, Column (1) shows that the market excess return fails to explain the value premium, because the intercept term is significantly nonzero and the $F-$statistic is insignificant. Column (2) shows that the uncertainty shocks have large explanatory power for value spreads, since the $F-$statistic for the regression (2) has significance less than 0.5%. It also shows that the impact of cash-flow uncertainty on value spreads is significantly negative, which is consistent with the theoretical prediction of my model. Regression (3) shows that the risk sharing condition helps explain the value spread when it interacts with growth uncertainty shocks. Perfectly in line with the prediction of the model, the coefficient of the interaction term $\Delta\nu_{g,t_y} \times \text{regime-x}_{t_y}$ is significantly positive, with estimate 407.44 and t-statistic 2.61. Moreover, the adjusted R^2 increases from 19.13% to 22.75% from the regression (2), and the intercept term becomes insignificantly positive. In Column (4), I further add in the interaction term between the risk sharing condition and the cash-flow uncertainty shock. The regression results of Column (3) are almost unaffected. The coefficient of the extra interaction term $\Delta\nu_{c,t_y} \times \text{regime-x}_{t_y}$ is insignificantly positive, with estimate

66.93 and t-statistic 0.72. The regression results in Columns (5)–(6) and the regression results in Columns (7)–(8) show the robustness of the regression results in Columns (3) and (4) when the measure of risk sharing conditions changes to the Reinhart-Rogoff financial condition index and the broker-dealer leverage index displayed in Figure 2-8, respectively. Furthermore, as shown in Table 2.14, the loadings of value spreads on cash flow shocks, denoted as $\beta_{vi,c}$ in (2.46), are significantly negative across all regressions and different measures for risk sharing conditions.[43]

Testing basic implication II. To verify the model predictions on stochastic discount factors, I explore the possibility of the cross section of stock returns. Because the prediction is specifically on the market price of risk for the growth uncertainty shock and the cash-flow uncertainty shock, I focus on portfolios of firms' stocks sorted based on the two uncertainty shocks, separately. As long as the loadings of firm stock returns on the uncertainty shocks are fairly persistent, the ex ante differential sensitivity to uncertainty shocks will lead to the ex post differential sensitivity. The differential average returns of sorted portfolios then are informative about these market price of risk associated with the uncertainty shocks.

In Table 2.15, the average returns for uncertainty-sorted portfolios are reported for the full sample (in Columns (5) and (6)) and two subsamples (in Columns (1) – (4)). One subsample corresponds to the periods of poor risk sharing conditions (reported in Columns (1) and (2)), while the other subsample corresponds to the periods of good risk sharing conditions (reported in Columns (3) and (4)). In Panel A, the risk sharing condition is measured by the Reinhart-Rogoff financial condition index; in Panel B, the risk sharing condition is gauged by the broker-dealer leverage index; in Panel C, the risk sharing condition is gauged by the credit spread index. Across all columns in Table 2.15, it shows that the firms with a higher exposure to the cash-flow uncertainty shock, on average, gain lower returns; it thus implies that cash-flow uncertainty shocks tend to carry a negative market price of risk, no matter whether the risk sharing condition

[43] They are all significant except regressions in Columns (7)–(8) in which the risk sharing condition is measured by the financial condition index based on broker-dealer leverages.

is good or bad. In particular, over the whole sample, the valuation spread between the firms with a high exposure to the cash-flow uncertainty shock versus those with a low exposure is statistically significantly negative; the spread is -5.11% with the t-statistic equal to -3.02.

However, the firms with a higher exposure to the growth uncertainty shock, on average, gain lower returns when risk sharing is limited (see Columns (1)); in contrast, they gain higher average returns otherwise (see Columns (3)). More precisely, if the Reinhart-Rogoff financial condition index is used as the measure for risk sharing conditions (in Panel A), the spread between high versus low ν_g-sorted portfolios changes from -2.30 (with t-statistic -1.24) to 1.96 (with t-statistic 0.91) when the risk sharing condition improves. This empirical pattern is robust against different choices of measures of risk sharing conditions. Particularly, if the broker-dealer leverage index is used to construct the regimes of risk sharing conditions (in Panel B), the spread between high versus low ν_g-sorted portfolios changes from -3.00 (with t-statistic -2.444) to 3.73 (with t-statistic 3.33) as the risk sharing condition improves; if the credit spread index is used to construct the regimes of risk sharing conditions (in Panel C), the spread between high versus low ν_g-sorted portfolios changes from -5.34 (with t-statistic -1.14) to 4.36 (with t-statistic 0.96) as the risk sharing condition improves. This suggests that the growth uncertainty shock tends to carry a negative market price of risk when the risk sharing condition is bad and a positive market price of risk otherwise.

2.5 Conclusion

I have studied an investment-based general equilibrium model with two sources of uncertainty shocks and endogenous imperfect risk sharing. The model provides a fundamental mechanism which can help reconcile seemingly contradictory empirical findings in asset pricing and macroeconomics under a unified framework.

There are two main new insights provided by this paper. First, the source of uncertainty shocks matters, since they affect the economy through different asset classes.

The characteristics of the assets determine the impact of uncertainty shocks from certain origin on asset prices and investment. In particular, the growth uncertainty shocks can increase asset prices and investment because of the option feature embedded in growth options. Second, the risk sharing condition plays a vital role in shaping the impact of uncertainty shocks. When risk sharing is severely limited, a rise in uncertainty distorts agents' real investment decisions and portfolio allocations in an inefficient manner. If agents' preference over smoothing consumption across time (governed by the elasticity of intertemporal substitution) is not very strong, even the growth uncertainty shock can suppress asset prices, decrease investment, deteriorate risk sharing conditions, and hence carry a negative market price of risk.

This paper, moreover, discovers the linkage between the cross section of asset returns and uncertainty shocks from different origins. Because different sources of uncertainty shocks do not affect firms symmetrically, then the cross section of asset returns can help identify the source of uncertainty shocks. Financial data with a larger cross section and a higher frequency can serve well for uncovering the uncertainty shocks used in macroeconomic models. Moreover, as shown theoretically and empirically, the cross sectional exposures of asset returns to growth uncertainty shocks are largely driven by the risk sharing condition; hence the time-varying cross-sectional exposures to the growth uncertainty shock are informative about the underlying economy state of risk sharing.

Table 2.1: BASELINE PARAMETRIZATION

PARAMETER	SYMBOL	VALUE
A. PREFERENCES		
Subjective discount rate	ρ	0.0111
Relative risk aversion	γ	6
EIS coefficient	ψ	2
B. ASSETS IN PLACE IN CONSUMPTION GOODS SECTOR		
Capital share in production function	φ	0.3
Assets in place depreciation rate	δ	15%
Aggregate volatility	σ	10%
Cash-flow uncertainty	ν_c^L/ν_c^H	10%/50%
Transition of cash-flow uncertainty	$\lambda^{(\nu_c^L,\nu_c^H)}/\lambda^{(\nu_c^H,\nu_c^L)}$	0.111/0.39
C. GROWTH OPTIONS IN CONSUMPTION GOODS SECTOR		
Investment goods share in production function	α	0.9
Growth uncertainty	ν_g^L/ν_g^H	10%/49%
Investment opportunity arrival rate	λ	3.33
Transition of growth uncertainty	$\lambda^{(\nu_g^L,\nu_g^H)}/\lambda^{(\nu_g^H,\nu_g^L)}$	0.1/0.44
Fixed adjustment cost rate	ϖ	0.0083
Aggregate growth options	$\bar{s}$	1
D. INVESTMENT GOODS SECTOR		
Average productivity level	z_ι	1.03
E. LABOR MARKET		
Population share	$\varkappa$	2.04%
Average lifespan	μ	1/40
F. FINANCIAL MARKET		
Severity of agency problem	ϕ	0.4
Pledgeability of human capital	ϱ	5%

NOTE: This table reports the calibrated parameters of the model. The annualized values are used in the table for the dynamic parameters. When choosing the values of the parameters, both inside and outside-model data are employed.

Table 2.2: MODEL VERSUS DATA: UNCONDITIONAL MOMENTS OF MACROECONOMIC CYCLES

	MEAN	STDEV	AC(1)	CORRELATION		
				$\Delta\log(c_{t+1})$	$\Delta\log(c_t)$	$\Delta\log(y_t)$
	(1)	(2)	(3)	(4)	(5)	(6)
			A. DATA			
$\Delta\log(c_t)$	1.46 [0.98, 1.93]	3.77 [3.31, 4.17]	0.36 [-0.13, 0.17]			
$\Delta\log(y_t)$	**1.67** [0.65, 2.63]	**5.78** [3.43, 7.49]	0.28 [0.00, 0.42]	0.34 [-0.03, 0.50]	0.92 [0.89, 0.94]	
$\Delta\log(i_t)$	1.35 [-2.58, 5.20]	**36.00** [10.92, 49.55]	0.43 [0.29, 0.56]	0.38 [0.09, 0.45]	0.83 [0.78, 0.87]	0.87 [0.85, 0.93]
			B. MODEL			
$\Delta\log(c_t)$	1.92 [0.74, 3.09]	3.96 [2.96, 4.33]	0.32 [0.11, 0.50]			
$\Delta\log(y_t)$	**1.92** [0.73, 3.06]	**4.01** [3.35, 4.70]	0.50 [0.34, 0.65]	0.62 [0.49, 0.73]	0.52 [0.33, 0.70]	
$\Delta\log(i_t)$	2.36 [-0.47, 5.73]	**55.38** [32.41, 77.63]	0.30 [-0.00, 0.49]	0.23 [0.07, 0.39]	-0.44 [-0.26, -0.61]	0.30 [0.17, 0.43]

NOTES: The table compares unconditional moments of the data to their simulated analogies in the model. Panel A reports the mean, standard deviation, and autocorrelation of U.S. output (y), consumption (c), and investment (i) log growth rates, as well as their cross-correlation coefficients. All variables are real (adjusted by CPI) and scaled by U.S. population. The 95% confidence intervals are reported in brackets; they are obtained by applying stationary block bootstrap method in which the block size is random (see Politis and Romano, 1994a,b). The average block size is determined by the adaptive block length selection procedure of Politis and White (2004) and Patton, Politis, and White (2009). Data are sampled at the annual frequency. Their sources and construction details are explained in the appendix. All variables are reported in percentage points, except for the autocorrelation and cross-correlation coefficients. The moments of the consumption growth and the output growth are from the extended long sample of Barro and Ursúa (2008) with sample period 1790 – 2014. The sample periods of net payout growth and investment are 1929 – 2014, and the labor supply growth is only available during 1948 – 2014. Panel B reports simulated moments in the model. I simulate the model at the weekly frequency and then time-aggregate the simulated data to construct annual observations. In brackets, they are the 5% and 95% quantiles across 1,000 independent simulations, each with a length of 80 years.

Table 2.3: MODEL VERSUS DATA: UNCONDITIONAL MOMENTS OF MACROECONOMIC RATIOS

RATIOS (%)	A. DATA MEAN (1)	A. DATA STDEV (2)	A. DATA AC(1) (3)	B. MODEL MEAN (4)	B. MODEL STDEV (5)	B. MODEL AC(1) (6)
INVESTMENT TO OUTPUT	**16.47**	5.23	0.90	**16.60**	6.41	0.70
	[13.85, 18.79]	[2.84, 6.23]	[0.60, 0.87]	[14.74, 18.28]	[4.83, 7.89]	[0.58, 0.83]
NET PAYOUT TO CONSUMPTION	**5.46**	3.07	0.85	**6.30**	4.03	0.70
	[3.93, 7.00]	[2.22, 3.42]	[0.59, 0.83]	[5.03, 7.49]	[2.42, 5.28]	[0.55, 0.84]
WAGE INCOME TO OUTPUT	**75.26**	4.02	0.96	**75.25**	2.04	0.71
	[72.94, 77.69]	[2.71, 4.48]	[0.74, 0.94]	[74.55, 75.89]	[1.55, 2.49]	[0.58, 0.83]
CAPITAL TO OUTPUT	**169.24**	46.96	0.90	**196.20**	48.35	0.71
	[144.40, 195.55]	[29.85, 53.13]	[0.66, 0.89]	[186.77, 204.51]	[33.96, 65.33]	[0.48, 0.82]

NOTES: The table compares unconditional moments of the data to their simulated analogies in the model. Panel A reports the mean, standard deviation, and autocorrelation of U.S. investment/output ratio, net payout/consumption ratio, wage income/output ratio, and capital/output ratio. The 95% confidence intervals are reported in brackets; they are obtained by applying stationary block bootstrap method in which the block size is random (see Politis and Romano, 1994a,b). The average block size is determined by the adaptive block length selection procedure of Politis and White (2004) and Patton, Politis, and White (2009). Data are sampled at the annual frequency. Their sources and construction details are explained in the appendix. All variables are reported in percentage points, except for the autocorrelation coefficients. The sample period is 1929 – 2014. Panel B reports simulated moments in the model. I simulate the model at the weekly frequency and then time-aggregate the simulated data to form annual observations. In brackets, they are the 5% and 95% quantiles across 1,000 independent simulations, each with a length of 80 years.

Table 2.4: MODEL VERSUS DATA: FUNDAMENTAL DISPERSIONS

DISPERSIONS (%)	MEAN (1)	STDEV (2)	AC(1) (3)	CORR $\Delta \log(y_t)$ (4)
	A. DATA			
IDR SALES GROWTH	**49.02**	**12.32**	0.80	-17.32
	[42.20, 55.90]	[8.39, 13.71]	[0.64, 0.89]	[-36.85, 3.80]
CSD INVESTMENT RATE	**40.85**	**7.25**	0.66	43.28
	[37.03, 44.71]	[5.22, 7.93]	[0.48, 0.77]	[30.45, 59.99]
	B. MODEL			
IDR SALES GROWTH	**53.02**	**16.03**	0.69	-27.66
	[42.20, 61.90]	[9.84, 23.19]	[0.57, 0.80]	[-45.51, -13.83]
CSD INVESTMENT RATE	**45.12**	**13.50**	0.71	23.82
	[39.13, 49.98]	[10.31, 16.37]	[0.43, 0.79]	[1.89, 40.35]

NOTES: The table compares unconditional moments of the data to their simulated analogies in the model. Panel A reports, in the data, the mean, standard deviation, autocorrelation, and cyclicality of Compustat sales dispersion measured by the cross-sectional interdecile range (IDR) and Compustat capital expenditures dispersion measured by the cross-sectional standard deviation (CSD). The sales are deflated by the sales in the previous year, and capital expenditure is deflated by capital stock in the previous year. Sales is constructed using item `sales`, capital expenditure is constructed using item `capx`, and capital stock is constructed using item `ppent`. The 95% confidence intervals are reported in brackets; they are obtained by applying stationary block bootstrap method in which the block size is random (see Politis and Romano, 1994a,b). The average block size is determined by the adaptive block length selection procedure of Politis and White (2004) and Patton, Politis, and White (2009). Data are sampled at the annual frequency. Their sources and construction details are explained in the appendix. All variables are reported in percentage points, except for the autocorrelation coefficients. All variables have the sample period of 1966 – 2014. Panel B reports simulated moments in the model. I simulate the model at the weekly frequency and then time-aggregate the simulated data to form annual observations. In brackets, they are the 5% and 95% quantiles across 1,000 independent simulations, each with a length of 80 years.

Table 2.5: MODEL VERSUS DATA: UNCONDITIONAL ASSET PRICING MOMENTS

	A. DATA			B. MODEL		
RETURNS (%)	MEAN (1)	STDEV (2)	AC(1) (3)	MEAN (4)	STDEV (5)	AC(1) (6)
EQUITY PREMIUM	4.47	20.83	0.03	4.95	19.25	0.12
	[0.85, 8.07]	[18.08, 22.99]	[-0.19, 0.16]	[3.17, 6.58]	[11.37, 27.91]	[-0.10, 0.33]
INTEREST RATE	**1.31**	2.71	0.62	**1.53**	2.92	0.71
	[0.63, 2.13]	[2.07, 3.19]	[0.39, 0.92]	[0.49, 2.66]	[2.41, 3.43]	[0.58, 0.83]
NET PAYOUT YIELD	2.25	3.60	0.79	3.67	4.14	0.68
	[1.78, 2.71]	[2.73, 3.91]	[0.56, 0.88]	[2.34, 4.71]	[3.31, 4.92]	[0.55, 0.81]
VALUE SPREAD	5.05	25.21	0.11	7.57	16.84	-0.02
	[0.57, 9.57]	[21.14, 28.54]	[-0.16, 0.20]	[5.88, 9.57]	[12.39, 21.39]	[-0.29, 0.26]

NOTES: The table compares unconditional moments of the data to their simulated analogies in the model. Panel A reports the mean, standard deviation, and autocorrelation of U.S. equity premium, the real interest rate, the net payout yield, and the value spread which is the return spread between two portfolios of firms with the top and bottom decile of book-to-market ratios. The 95% confidence intervals are reported in brackets; they are obtained by applying stationary block bootstrap method in which the block size is random (see Politis and Romano, 1994a,b). The average block size is determined by the adaptive block length selection procedure of Politis and White (2004) and Patton, Politis, and White (2009). Data are sampled at the annual frequency. Their sources and construction details are explained in the appendix. All variables are reported in percentage points, except for the autocorrelation coefficients. All variables have the sample period of 1929 – 2014. Panel B reports simulated moments in the model. I simulate the model at the weekly frequency and then time-aggregate the simulated data to form annual observations. In brackets, they are the 2.5% and 97.5% quantiles across 1,000 independent simulations, each with a length of 80 years.

Table 2.6: ESTIMATED TRANSITION: UNCERTAINTY REGIMES

	MARKOV TRANSITION PROBABILITIES (%)					
	G-Uncert			C-Uncert		
(states)	HIGH	MEDIUM	LOW	HIGH	MEDIUM	LOW
HIGH	81.9	16.3	1.8	67.5	20.1	12.4
MEDIUM	21.6	70.5	7.9	13.9	80.5	5.6
LOW	6.6	13.3	80.1	8.6	2.7	88.8
STATIONARY DIST.	47.6	34.4	18.0	24.9	31.7	43.4

NOTES: This table reports point estimation of Markov transition probabilities of the latent states for two kinds of uncertainty, respectively. G-Uncert stands for growth uncertainty, and C-Uncert stands for cash-flow uncertainty. The numbers are estimates of the regime-switching model in (2.40) using the EM algorithm. The estimation is based on annual sample from 1953 to 2014.

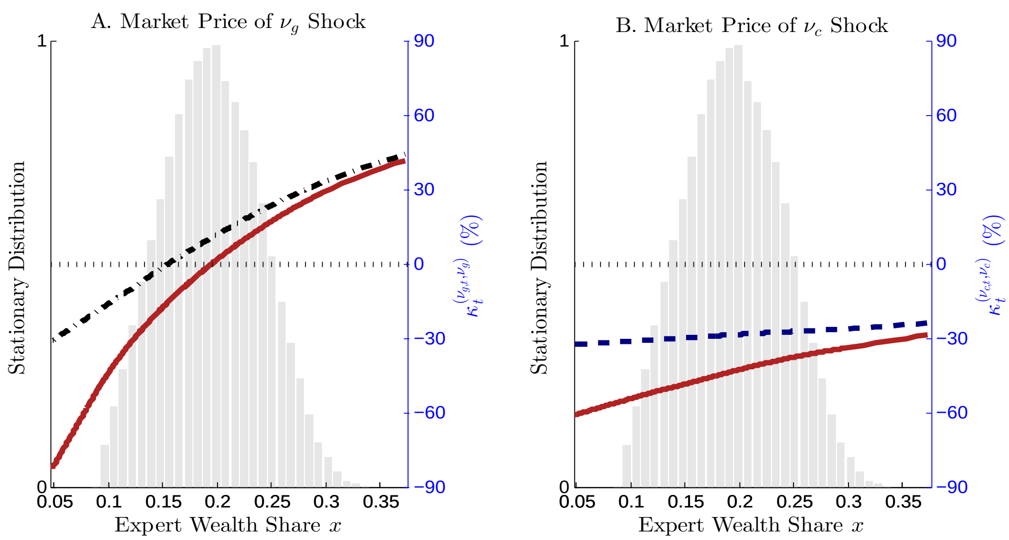

Figure 2-1: Policy Functions: market price of risk

NOTES: This figure illustrates the market price of risk for two uncertainty shocks under the calibration summarized in Table 4.4. Panel A shows the market price of risk for growth uncertainty shocks; Panel B is about the market price of risk for cash-flow uncertainty shocks. The red solid curve corresponds to the normal state of the world where both uncertainties are at low levels; the blue dashed curve corresponds to state of high growth uncertainty; and, the black dashed-dotted curve corresponds to the state of high cash-flow uncertainty. The grey distribution in the background is the stationary distribution of the endogenous state variable x_t.

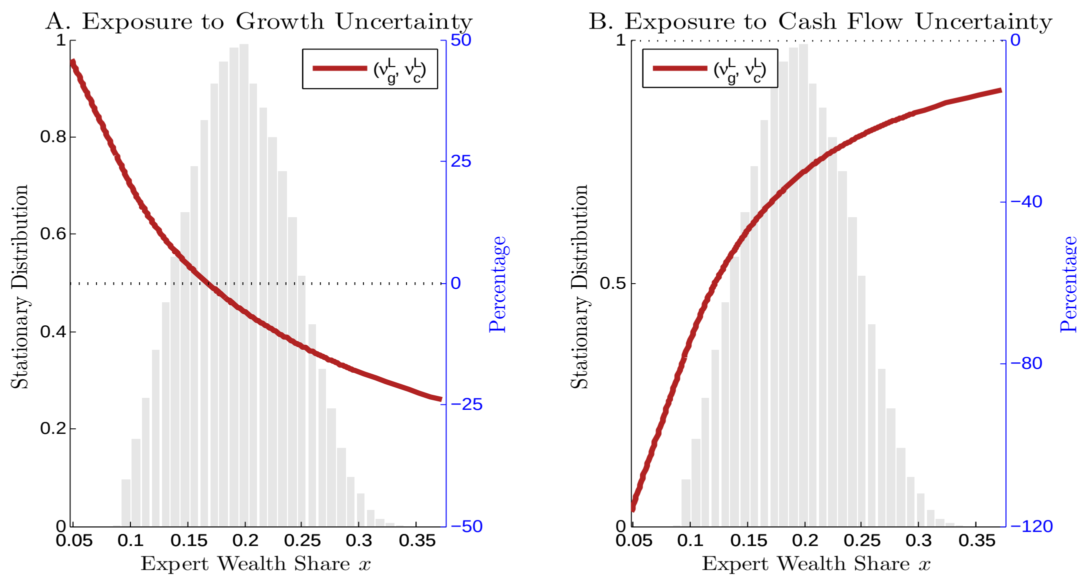

Figure 2-2: Value Spread's Uncertainty Exposures

NOTES: This figure illustrates the uncertainty exposure of value spreads under the calibration summarized in Table 4.4. Panel A shows the exposure of value spreads to growth uncertainty shocks; Panel B shows the exposure of value spreads to cash-flow uncertainty shocks. The red solid curve corresponds to the normal state of the world where both uncertainties are at low levels. The grey distribution in the background is the stationary distribution of the endogenous state variable x_t.

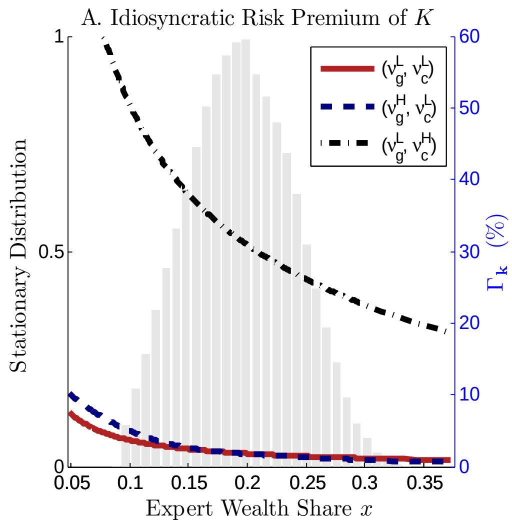

Figure 2-3: Idiosyncratic Risk Premia

NOTES: This figure illustrates the idiosyncratic risk premia under the calibration summarized in Table 4.4. Panel A is about the premium on the idiosyncratic cash flow risk $dW_{f,t}$, and Panel B is about the premium on the idiosyncratic investment risk associated with $\varepsilon_{f,t}$ and $dN_{f,t}$. The red solid curve corresponds to the normal state of the world where both uncertainties are at low levels; the blue dashed curve corresponds to state of high growth uncertainty; and, the black dashed-dotted curve corresponds to the state of high cash-flow uncertainty. The grey distribution in the background is the stationary distribution of the endogenous state variable x_t.

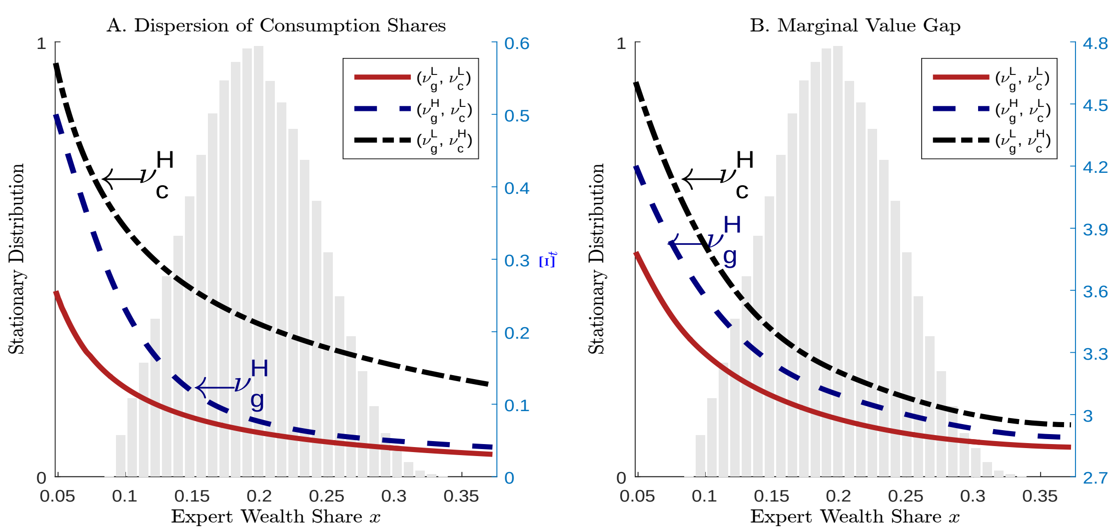

Figure 2-4: Limited Risk Sharing

NOTES: This figure illustrates the consumption dispersion and marginal value gap under the calibration summarized in Table 4.4. Panel A is about the premium on the idiosyncratic cash flow risk $dW_{f,t}$, and Panel B is about the premium on the idiosyncratic investment risk associated with $\varepsilon_{f,t}$ and $dN_{f,t}$. The red solid curve corresponds to the normal state of the world where both uncertainties are at low levels; the blue dashed curve corresponds to state of high growth uncertainty; and, the black dashed-dotted curve corresponds to the state of high cash-flow uncertainty. The grey distribution in the background is the stationary distribution of the endogenous state variable x_t.

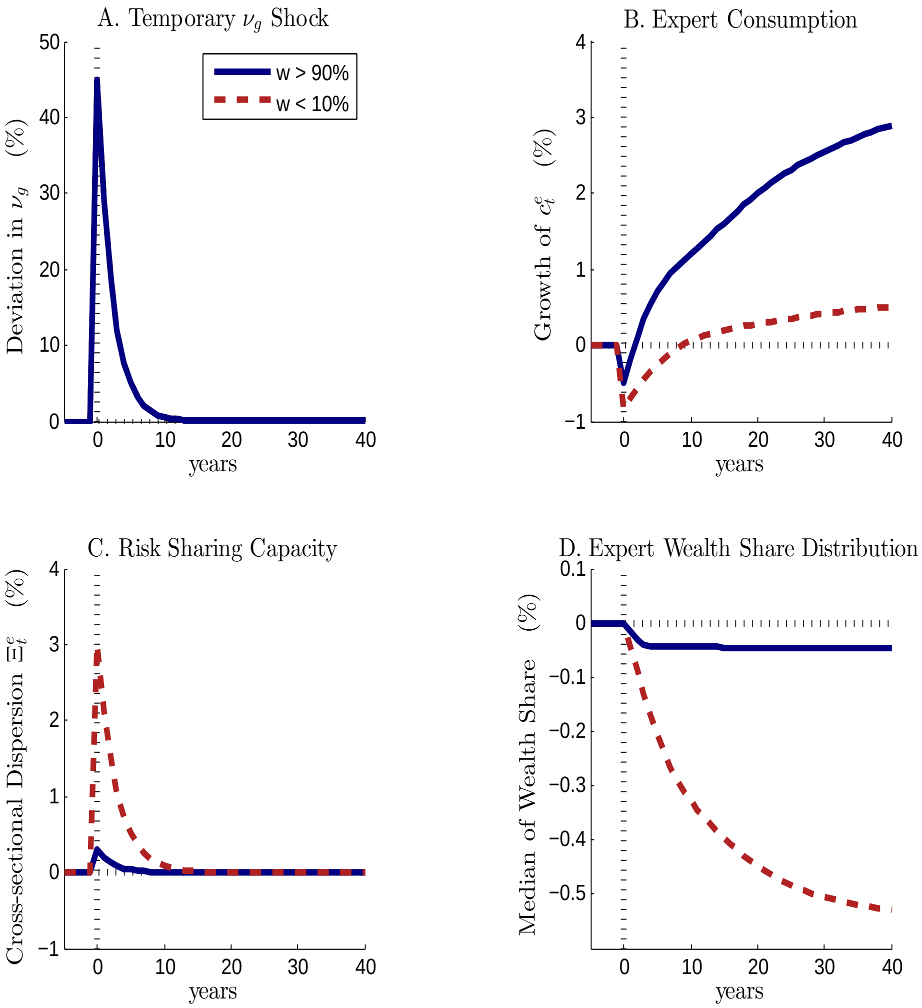

Figure 2-5: Impulse Responses to Growth Uncertainty Shocks

NOTES: This figure illustrates the impulse-response to a temporary growth uncertainty shock under the calibration summarized in Table 4.4. Panel A shows the temporal shock as an impulse; Panel B is about the responses of experts' aggregate consumption; Panel C is about the responses of conditional cross-sectional variance of consumption share growth, and Panel D is about median of cross-section of experts' consumption shares. The blue solid curve corresponds to the states of good risk sharing conditions; the red dashed curve corresponds to states of high poor risk sharing conditions. The grey distribution in the background is the stationary distribution of the endogenous state variable x_t.

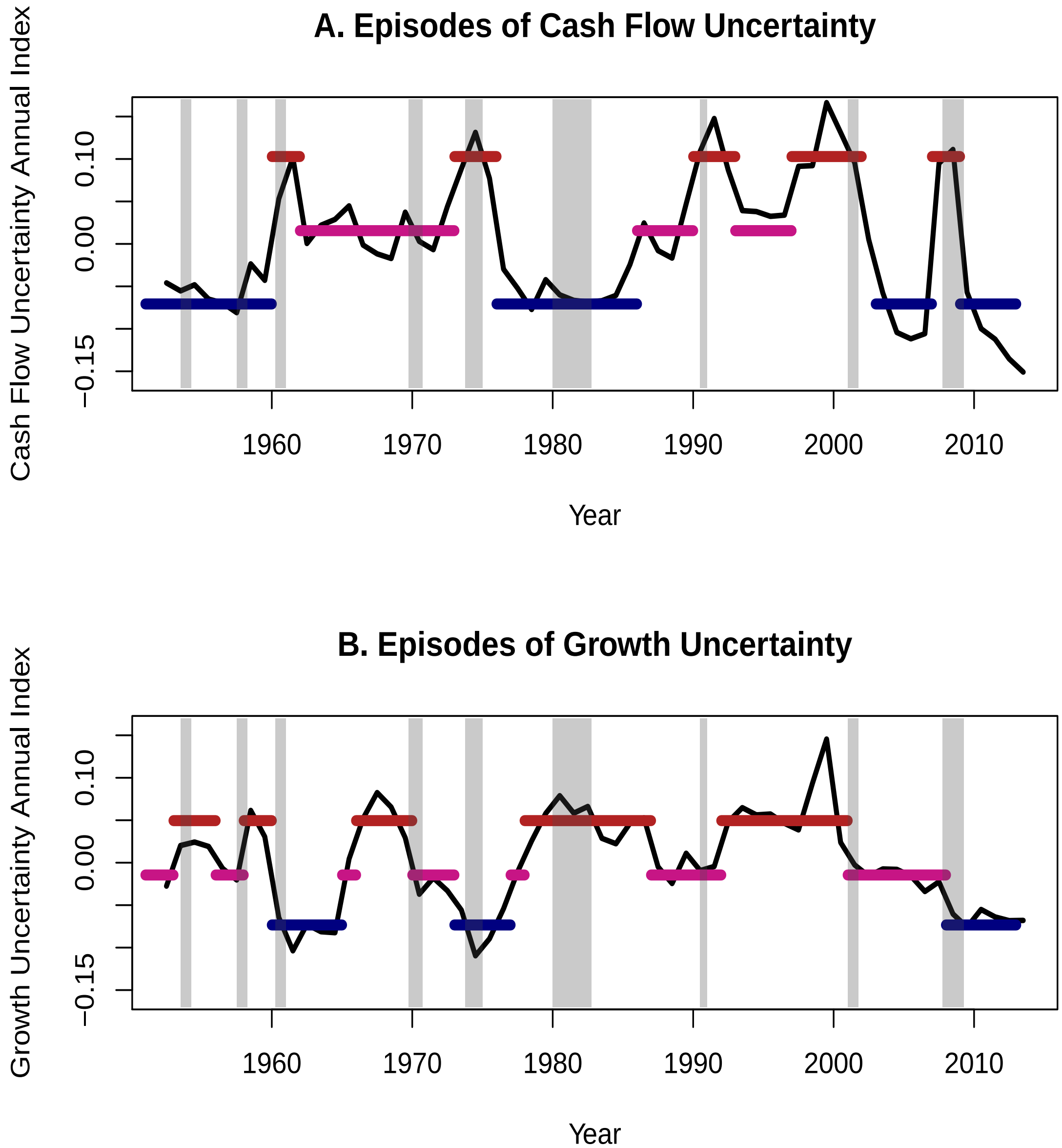

Figure 2-6: Annual Uncertainty indices, Estimated Uncertainty Regimes and U.S. Recessions

NOTES: This figure plots the annual indices of the cash-flow uncertainty and the growth uncertainty. The annual index is defined as the average of twelve monthly index values within each year. The monthly indices of uncertainty are constructed as described in Section 2.4.1.

Table 2.7: CYCLICALITY: UNCERTAINTY AND IDIOSYNCRATIC DISPERSION

CORRELATIONS WITH THE CYCLICAL COMPONENT OF OUTPUT:

	A. Dispersion		B. Uncertainty	
	PEARSON	KENDALL	PEARSON	KENDALL
	(1)	(2)	(3)	(4)
INVESTMENT	0.31 (0.03)	0.20 (0.04)	0.13 (0.39)	0.05 (0.62)
CASH FLOW	−0.16 (0.26)	−0.10 (0.32)	−0.31 (0.03)	−0.21 (0.03)

NOTES: This table reports correlations of uncertainty indices and idiosyncratic dispersions with the cyclical component of U.S. real GDP per capita. The cyclical component is extracted from the log real GDP per capita by using the one-sided Hodrick-Prescott (HP) filter. Panel A reports the cyclicality of the dispersion of idiosyncratic shocks in sales growth rates ν_{sales,t_y} (CASH FLOW row) and the idiosyncratic shocks in investment rates ν_{capx,t_y} (INVESTMENT row). Panel B reports the cyclicality of the cash-flow uncertainty ν_{c,t_y} (CASH FLOW row) and the growth uncertainty ν_{g,t_y} (INVESTMENT row). Column (1) and Column (3) report the Pearson correlations, while Column (3) and Column (4) report the Kendall rank correlations. The Kendall rank correlation measures the similarity of the orderings of the data when ranked by each of the quantities. Thus, it provides a non-parametric measure of the association of two time series. The validity of the Pearson correlation is more dependent on the parametric gaussian assumption. The p values are reported inside the parentheses. The sample is annual from 1966 to 2014. The reliable dispersion estimates are only available after 1966 in annual Compustat fundamentals.

Table 2.8: THE REGIME-SWITCHING MODEL: ESTIMATED COEFFICIENTS

ESTIMATED COEFFICIENTS IN THE REGIME-SWITCHING MODEL (2.44):

	a_v	$\beta_{v,z}$	$\beta_{v,g}$	$\beta_{v,c}$
	(1)	(2)	(3)	(4)
Bad state	0.47	0.41	6.03	−13.08
95% CI	[0.09, 1.07]	[−0.27, 0.55]	[0.29, 11.21]	[−19.92, −0.16]
75% CI	[0.28, 0.67]	[0.27, 0.48]	[2.60, 6.84]	[−13.88, −5.47]
Good state	0.46	−0.41	−1.76	−1.00
95% CI	[0.11, 1.99]	[−0.57, 0.89]	[−10.42, 1.07]	[−7.44, 0.73]
75% CI	[0.26, 0.59]	[−0.49, −0.26]	[−4.19, −0.99]	[−5.27, −0.80]

NOTES: This table reports the estimation results of the regime-switching model (2.44). The model is estimated using the EM algorithm which maximizes the marginal likelihood of observables. The 95% and 75% confidence intervals are reported in brackets; they are obtained by applying stationary block bootstrap method in which the block size is random (see Politis and Romano, 1994a,b). The average block size is determined by the adaptive block length selection procedure of Politis and White (2004) and Patton, Politis, and White (2009). Data are sampled at the monthly frequency from January 1953 to December 2014.

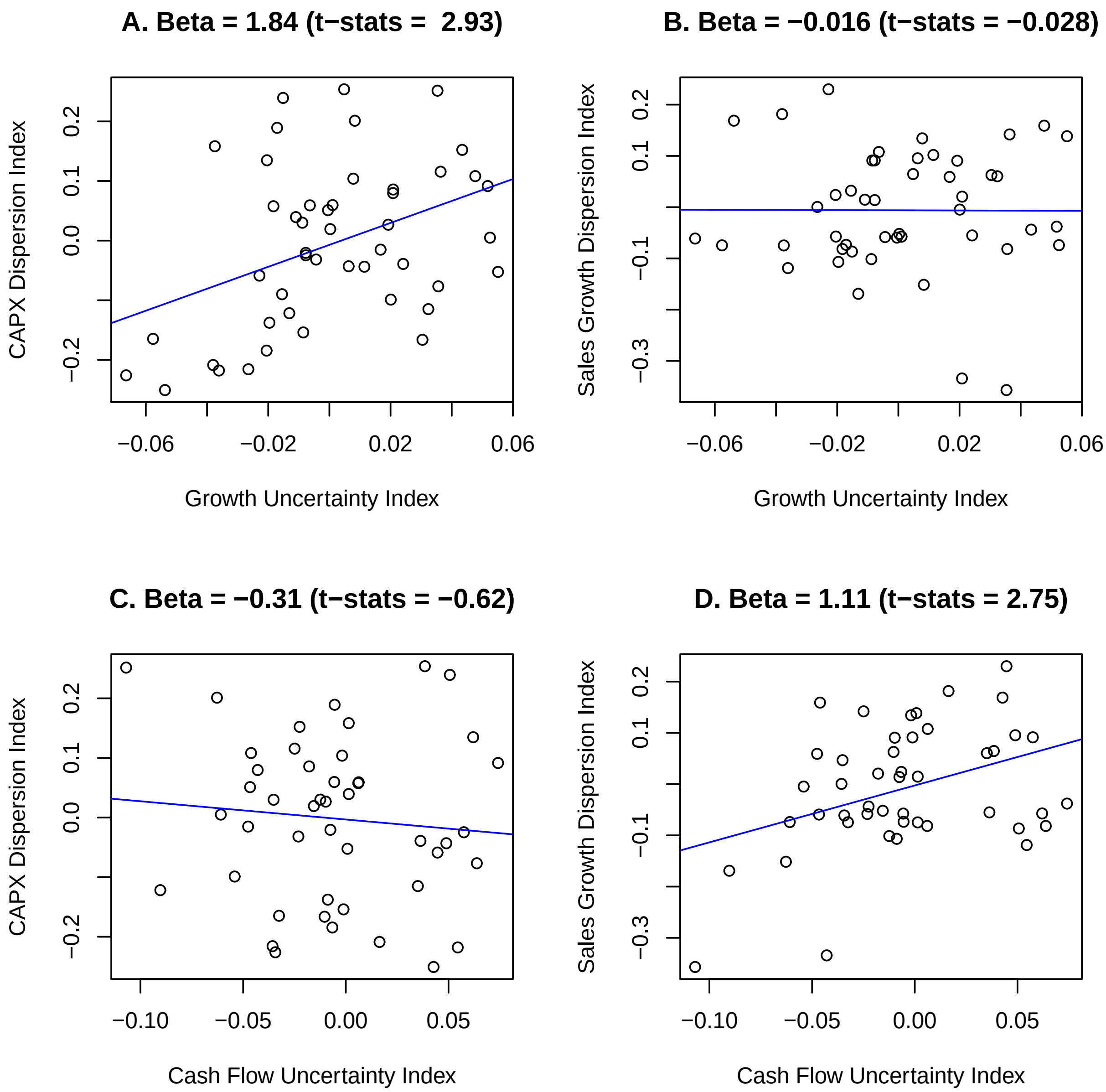

Figure 2-7: STATISTICAL ASSOCIATIONS: UNCERTAINTY VERSUS CROSS-SECTION DISPERSION

NOTES: This figure plot the annual changes of idiosyncratic sales dispersions and idiosyncratic investment dispersions against annual changes of uncertainty indices $\Delta\nu_{c,t_y}$ and $\Delta\nu_{g,t_y}$.

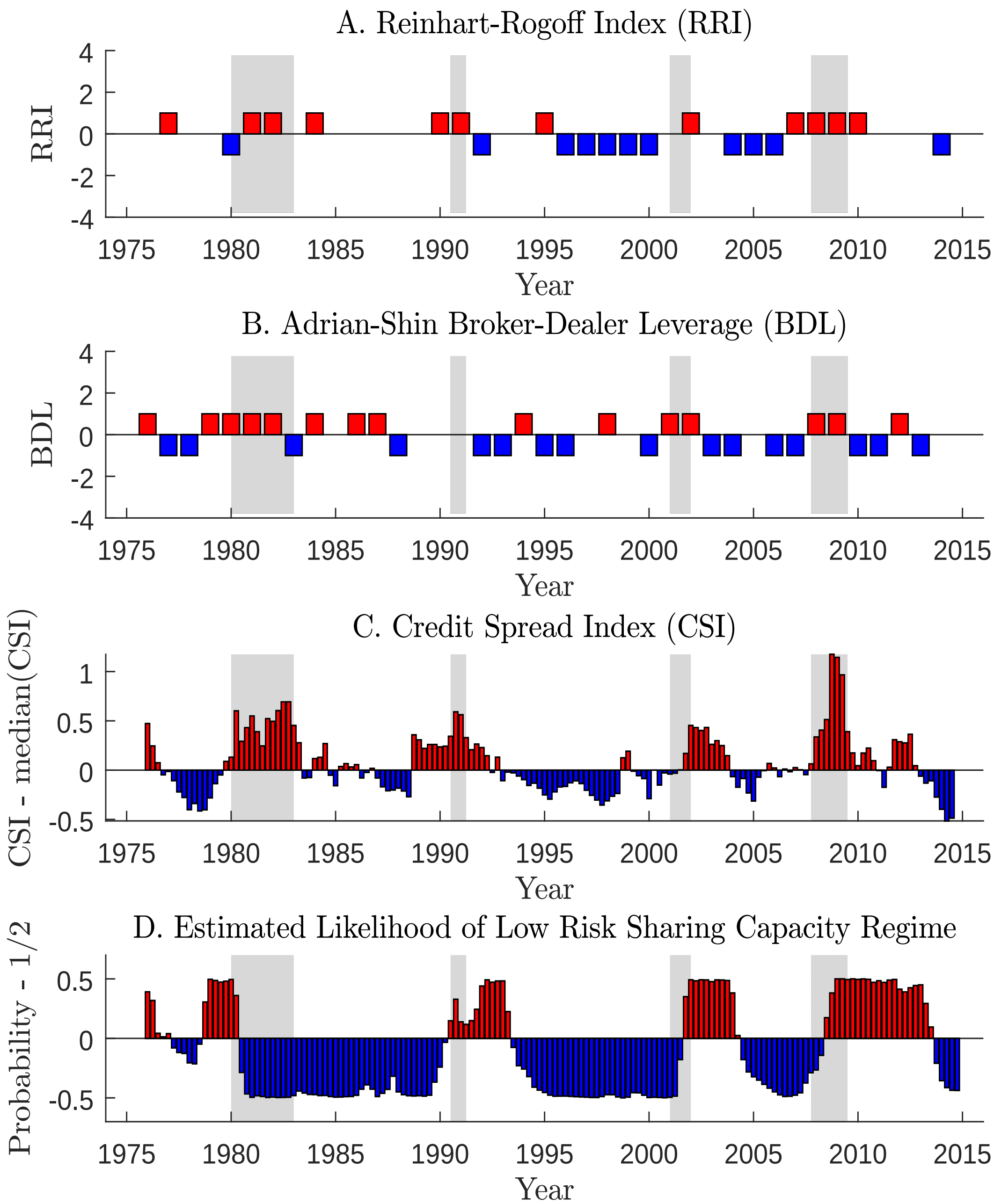

Figure 2-8: Measures and Estimation of Risk Sharing Condition of U.S. Economy

NOTES: The figure presents three measures of risk sharing conditions in the data (in Panels A, B, and C) and the estimated likelihood of being in the Bad state (in Panel D).

Table 2.9: STATISTICAL ASSOCIATIONS: ESTIMATION AND MEASURES OF RISK SHARING CONDITIONS

CORRELATION OF CREDIT SPREAD INDEX WITH:	A. DATA		B. MODEL	
	PEARSON	KENDALL	PEARSON	KENDALL
	(1)	(2)	(3)	(4)
ESTIMATED LIKELIHOOD OF BEING IN STATE Bad	0.38	0.23	0.36	0.29
	(0.00)	(0.00)	(0.00)	(0.00)
REINHART-ROGOFF FINANCIAL INDEX	0.52	0.39	0.27	0.18
	(0.00)	(0.01)	(0.00)	(0.00)
BROKER-DEALER LEVERAGE INDEX	0.42	0.29	−−	−−
	(0.00)	(0.07)	−−	−−

NOTES: This table reports the statistical association between the credit spread index, the Reinhart-Rogoff financial index, the Broker-Dealer leverage index, and the estimated likelihood of being in the Bad state plotted in Figure 2-8. Panel A shows the Pearson correlation (in Column (1)) and the Kendall rank correlation (in Column (2)) of the credit spread index with estimated likelihood of Bad state. At the same time, Panel B reports the corresponding statistical moments in the simulated data based on my model. Because there is no corporate bond in my model, I use the equity risk premium as the proxy for credit spread. The risk sharing condition in the simulated data is measured by using Θ_t in (2.32). The likelihood of Bad is estimated for the simulated data in the same way as for the real data. The Kendall rank correlation measures the similarity of the orderings of the data when ranked by each of the quantities. Thus, it provides a non-parametric measure of the association of two time series. The validity of the Pearson correlation is more dependent on the parametric gaussian assumption. The p values are reported inside the parentheses. The sample of indices are annual. They are time-aggregated from monthly or quarterly sample by averaging within each year. The simulated data are monthly and time-aggregated into quarterly frequency in the same way.

Table 2.10: UNCERTAINTY BETAS: THE REINHART-ROGOFF FINANCIAL INDEX

	A. BAD RISK SHARING CONDITION (REINHART-ROGOFF INDEX IS HIGH)			B. GOOD RISK SHARING CONDITION (REINHART-ROGOFF INDEX IS LOW)		
	MKT EX-RET	$\Delta\nu_g$	$\Delta\nu_c$	MKT EX-RET	$\Delta\nu_g$	$\Delta\nu_c$
BOOK-TO-MARKET SORT	(1)	(2)	(3)	(4)	(5)	(6)
SORT #1: SIX PORTFOLIOS						
LOWEST 10% (GROWTH)	1.06	−3.11	2.56	1.18	2.40	14.10
10% − 30%	0.96	−0.81	0.91	0.90	−1.85	−7.85
30% − 50%	0.99	0.91	0.68	0.79	−11.21	−15.98
50% − 70%	0.99	2.55	−0.66	0.72	−11.56	−10.97
70% − 90%	0.98	0.77	−4.48	0.62	−12.15	−14.29
HIGHEST 10% (VALUE)	1.22	0.21	−17.85	0.64	−12.66	−8.36
SORT #2: FIVE PORTFOLIOS						
LOWEST 20% (GROWTH)	1.02	−2.79	0.89	1.10	2.55	8.23
20% − 40%	0.98	0.50	2.04	0.87	−8.39	−12.18
40% − 60%	0.97	2.03	−0.46	0.78	−9.80	−14.50
60% − 80%	0.97	1.94	−0.27	0.60	−11.37	−15.55
HIGHEST 20% (VALUE)	1.05	−0.98	−9.69	0.63	−11.40	−12.74
SORT #3: THREE PORTFOLIOS						
LOWEST 30% (GROWTH)	1.00	−2.39	1.97	1.06	1.91	4.38
30% − 70%	0.99	1.78	0.32	0.76	−11.49	−14.37
HIGHEST 30% (VALUE)	1.97	0.32	−4.98	0.62	−12.59	−13.49

NOTES: The table reports estimated betas of book-to-market sorted portfolios with respect to the market excess return, the growth uncertainty shock, and the cash-flow uncertainty shock. In particular, it compares the estimation results of two subsamples. One subsample consists of the periods in which the risk sharing condition is good, while the other subsample consists of the periods in which the risk sharing condition is poor. The periods of good or poor risk sharing conditions are estimated using the Reinhart-Rogoff index shown in Panel A of Figure 2-8. The regression model for estimating the betas is $r_{BM,t_m} = a_{BM} + \beta_{BM,z}(r_{M,t} - r_{f,t}) + \beta_{BM,g}\Delta\nu_{g,t_m} + \beta_{BM,c}\Delta\nu_{c,t_m} + \epsilon_{BM,t_m}$, where BM stands for a book-to-market portfolio and r_{BM,t_m} is the return of the book-to-market portfolio labeled by BM. The reported estimates are obtained by using the ordinary-least-squares method. To account for the heteroskedasticity in stock returns, I also use the weighted-least-squares method with inverse market variance to be the weights. The estimation results are quite similar, because the regressions are totally separated for different subsamples. And, the heteroskedasticity does not show up dramatically and hence not bias the estimation within each subsample. The data are monthly from January of 1976 to December 2014.

Table 2.11: UNCERTAINTY BETAS: THE BROKER-DEALER LEVERAGE INDEX

	A. BAD RISK SHARING CONDITION (BD LEVERAGE INDEX IS HIGH)			B. GOOD RISK SHARING CONDITION (BD LEVERAGE INDEX IS LOW)		
	MKT EX-RET	$\Delta\nu_g$	$\Delta\nu_c$	MKT EX-RET	$\Delta\nu_g$	$\Delta\nu_c$
BOOK-TO-MARKET SORT	(1)	(2)	(3)	(4)	(5)	(6)
SORT #1: SIX PORTFOLIOS						
LOWEST 10% (GROWTH)	1.09	−2.81	6.80	1.03	−1.61	11.27
10% − 30%	0.98	2.41	−0.99	0.97	−2.00	0.98
30% − 50%	0.98	3.82	−3.35	0.91	−5.75	−3.75
50% − 70%	0.89	1.66	−6.59	0.87	−4.79	−3.67
70% − 90%	0.88	1.22	−9.79	0.84	−4.30	−1.40
HIGHEST 10% (VALUE)	1.03	1.10	−21.37	1.06	−3.74	−1.34
SORT #2: FIVE PORTFOLIOS						
LOWEST 20% (GROWTH)	1.05	−1.94	3.81	1.02	−1.35	6.99
20% − 40%	0.98	4.86	−1.52	0.95	−3.49	−0.72
40% − 60%	0.93	2.90	−5.48	0.89	−7.02	−3.67
60% − 80%	0.86	0.78	−6.20	0.79	−3.30	−3.69
HIGHEST 20% (VALUE)	0.92	0.65	−14.44	0.95	−3.82	−2.83
SORT #3: THREE PORTFOLIOS						
LOWEST 30% (GROWTH)	1.03	−0.77	3.62	1.01	−1.21	5.41
30% − 70%	0.94	2.45	−4.26	0.89	−4.89	−4.30
HIGHEST 30% (VALUE)	0.91	1.71	−10.30	0.88	−3.86	−1.95

NOTES: The table reports estimated betas of book-to-market sorted portfolios with respect to the market excess return, the growth uncertainty shock, and the cash-flow uncertainty shock. In particular, it compares the estimation results of two subsamples. One subsample consists of the periods in which the risk sharing condition is good, while the other subsample consists of the periods in which the risk sharing condition is poor. The periods of good or poor risk sharing conditions are estimated using the Broker-Dealer Leverage Index shown in Panel B of Figure 2-8. The regression model for estimating the betas is $r_{BM,t_m} = a_{BM} + \beta_{BM,z}(r_{M,t} - r_{f,t}) + \beta_{BM,g}\Delta\nu_{g,t_m} + \beta_{BM,c}\Delta\nu_{c,t_m} + \epsilon_{BM,t_m}$, where BM stands for a book-to-market portfolio and r_{BM,t_m} is the return of the book-to-market portfolio labeled by BM. The reported estimates are obtained by using the ordinary-least-squares method. To account for the heteroskedasticity in stock returns, I also use the weighted-least-squares method with inverse market variance to be the weights. The estimation results are quite similar, because the regressions are totally separated for different subsamples. And, the heteroskedasticity does not show up dramatically and hence not bias the estimation within each subsample. The data are monthly from January of 1976 to December 2014.

Table 2.12: UNCERTAINTY BETAS: THE CREDIT SPREAD INDEX

	A. BAD RISK SHARING CONDITION (CREDIT SPREAD IS HIGH)			B. GOOD RISK SHARING CONDITION (CREDIT SPREAD IS LOW)		
	MKT EX-RET	$\Delta\nu_g$	$\Delta\nu_c$	MKT EX-RET	$\Delta\nu_g$	$\Delta\nu_c$
BOOK-TO-MARKET SORT	(1)	(2)	(3)	(4)	(5)	(6)
SORT #1: SIX PORTFOLIOS						
LOWEST 10% (GROWTH)	1.07	−2.81	0.78	1.03	2.37	1.61
10% − 30%	0.97	0.41	0.24	1.06	0.92	−0.15
30% − 50%	0.98	0.80	0.59	1.00	−3.92	0.27
50% − 70%	0.98	2.56	−3.22	0.86	−2.23	−1.82
70% − 90%	0.98	2.43	0.07	0.86	−4.51	−0.35
HIGHEST 10% (VALUE)	1.24	8.08	−10.13	0.98	−1.34	−9.45
SORT #2: FIVE PORTFOLIOS						
LOWEST 20% (GROWTH)	1.03	−0.97	1.73	1.04	0.56	2.71
20% − 40%	0.96	0.20	−0.67	1.04	0.88	−2.12
40% − 60%	0.97	0.76	−0.92	0.94	−2.13	−2.10
60% − 80%	0.96	2.77	−1.01	0.82	−5.82	−0.67
HIGHEST 20% (VALUE)	1.06	4.69	−3.81	0.90	−2.00	−0.45
SORT #3: THREE PORTFOLIOS						
LOWEST 30% (GROWTH)	1.01	−1.53	0.53	1.05	1.02	0.96
30% − 70%	0.98	1.26	−0.84	0.93	−2.83	−0.55
HIGHEST 30% (VALUE)	1.03	3.08	−0.77	0.88	−3.83	−2.96

NOTES: The table reports estimated betas of book-to-market sorted portfolios with respect to the market excess return, the growth uncertainty shock, and the cash-flow uncertainty shock. In particular, it compares the estimation results of two subsamples. One subsample consists of the periods in which the risk sharing condition is good, while the other subsample consists of the periods in which the risk sharing condition is bad. The periods of good or bad risk sharing conditions are estimated by using the simplest three-state regime-switching model of the credit spread index. The periods of bad risk sharing conditions are those estimated to have high credit spread index level, while the periods of good risk sharing conditions are those estimated to have low credit spread index level. The regression model for estimating the betas is $r_{BM,t_m} = a_{BM} + \beta_{BM,z}(r_{M,t} - r_{f,t}) + \beta_{BM,g}\Delta\nu_{g,t_m} + \beta_{BM,c}\Delta\nu_{c,t_m} + \epsilon_{BM,t_m}$, where BM stands for a book-to-market portfolio and r_{BM,t_m} is the return of the book-to-market portfolio labeled by BM. The reported estimates are obtained by using the ordinary-least-squares method. To account for the heteroskedasticity in stock returns, I also use the weighted-least-squares method with inverse market variance to be the weights. The estimation results are quite similar, because the regressions are totally separated for different subsamples. The heteroskedasticity does not show up significantly and hence not bias the estimation within each subsample. The data are monthly from January of 1976 to December 2014.

Table 2.13: MODEL VERSUS DATA: UNCERTAINTY EXPOSURES

	LOW RISK SHARING CONDITION		HIGH RISK SHARING CONDITION		ALL	
BOOK-TO-MARKET SORT	$\Delta\nu_g$ (1)	$\Delta\nu_c$ (2)	$\Delta\nu_g$ (3)	$\Delta\nu_c$ (4)	$\Delta\nu_g$ (5)	$\Delta\nu_c$ (6)
	A. DATA: REINHART-ROGOFF INDEX					
LOW 10%	−3.11	2.56	2.40	14.10	−0.47	3.84
	(−0.75)	(0.85)	(0.47)	(3.51)	(-0.08)	(0.62)
HIGH 10%	0.21	−17.85	−12.66	−8.36	3.96	−30.01
	(0.04)	(−3.35)	(−2.42)	(−2.53)	(0.64)	(−3.11)
	B. DATA: BROKER-DEALER LEVERAGE INDEX					
LOW 10%	−2.81	6.80	−1.61	11.27	−0.47	3.84
	(−0.75)	(2.21)	(−0.34)	(3.32)	(−0.08)	(0.62)
HIGH 10%	1.10	−21.37	−3.74	−1.34	3.96	−30.01
	(0.23)	(−3.73)	(−0.36)	(−0.24)	(0.64)	(−3.11)
	C. DATA: CREDIT SPREAD INDEX					
LOW 10%	−1.25	3.33	18.82	26.38	−0.47	3.84
	(−0.15)	(0.37)	(2.33)	(3.12)	(−0.08)	(0.62)
HIGH 10%	25.19	−41.17	−7.10	−32.40	3.96	−30.01
	(2.05)	(−6.78)	(−0.90)	(−3.36)	(0.64)	(−3.11)
	D. MODEL					
GROWTH	−0.88	1.14	0.31	0.11	−0.09	0.43
	(−3.66)	(5.42)	(1.99)	(2.10)	(−0.16)	(3.03)
VALUE	−0.31	−1.06	−0.37	−0.13	−0.34	−0.47
	(−1.91)	(−5.13)	(−2.34)	(−2.22)	(−2.01)	(−3.49)

NOTES: The table compares unconditional moments of the data to their simulated analogies in the model. It reports boot-to-market sorted portfolios' uncertainty betas for the whole sample and two subsamples. The t-statistics are reported in the parentheses. In computing the t-statistics, the standard errors are estimated using Newey and West (1987a, 1994) method with one lag. Data are sampled at the monthly frequency. Their sources and construction details are explained in the appendix. The sample period is 1976 – 2014. The risk sharing regimes are measured by using the Reinhart-Rogoff Index (the Broker-Dealer Leverage Index) in the Panel A (Panel B), while the risk sharing regimes are measured by using the credit spread index in Panel C. Panel D reports the simulated results based on the model. I simulate at the weekly frequency and then time-aggregate the simulated data to form monthly observations. In parentheses, they are t-statistics computed using 1,000 independent simulations, each with a length of 400 years.

Table 2.14: INTERACTIONS: RISK SHARING CONDITIONS AND UNCERTAINTY SHOCKS

RETURN SPREADS OF HIGHEST 10% AND LOWEST 10% BOOK-TO-MARKET PORTFOLIO (VALUE SPREAD) ARE REGRESSED ON	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RISKS (INPUT VARIABLES)								
INTERCEPT (α_{vi})	5.65	5.82	6.03	5.72	6.56	6.53	5.54	5.52
	(2.45)	(2.12)	(1.39)	(1.27)	(2.37)	(2.59)	(3.95)	(6.39)
MKT EX-RET ($\beta_{vi,z}$)	0.21	0.22	0.18	0.18	0.21	0.22	0.03	0.04
	(1.91)	(1.60)	(1.54)	(1.46)	(3.06)	(2.55)	(0.22)	(0.89)
$\Delta\nu_g$ ($\beta_{vi,g}$)		59.45	−93.55	−85.98	8.39	4.98	−28.12	−37.90
		(1.26)	(−1.29)	(−1.14)	(0.29)	(0.20)	(−0.74)	(−1.61)
$\Delta\nu_c$ ($\beta_{vi,c}$)		−183.33	−181.78	−151.40	−212.63	−213.63	−55.35	−19.12
		(−2.31)	(−3.29)	(−3.39)	(−4.63)	(−3.01)	(−1.27)	(−0.73)
regime-x ($\beta_{vi,x}$)			−3.59	−3.15	−2.35	−2.64	−11.40	−11.22
			(−0.53)	(−0.45)	(−0.74)	(−1.01)	(−3.86)	(−7.57)
$\Delta\nu_g \times$ regime-x ($\gamma_{vi,g}$)			407.44	390.06	90.89	100.56	80.48	88.64
			(2.61)	(2.38)	(2.27)	(3.24)	(1.82)	(6.25)
$\Delta\nu_c \times$ regime-x ($\gamma_{vi,c}$)				66.39		37.45		63.17
				(0.72)		(0.67)		(2.67)
ADJ-R^2 (%)	1.50	19.13	22.75	21.66	18.42	17.32	16.63	14.27
F-STATISTIC	1.92	5.73	4.53	3.76	3.66	3.06	3.29	2.75
	(0.17)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.04)

NOTES: The table reports the results of regressions for value spreads. Column (1) reports the results of regressing value spreads on the constant term and the excess market return. Column (2) the uncertainty shocks into the regression. In Columns (3) and (4), the risk sharing condition (regime-x_{t_y}) is measured by the credit spread index; in Columns (5) and (6), the risk sharing condition (regime-x_{t_y}) is measured by the Reinhart-Rogoff Index; and, in Columns (7) and (8), the risk sharing condition (regime-x_{t_y}) is measured based on the Broker-Dealer Leverage Index. In Columns (3), (5), and (7), an extra independent variable regime-x_{t_y} and its interaction term with the growth uncertainty shock $\Delta\nu_{g,t_y} \times$ regime-x_{t_y}. In Columns (4), (6), and (8), the interaction terms between the state of risk sharing condition and the cash-flow uncertainty shock are added on the top of the regression (3), (5) and (7), respectively. The regressions are annual, because the state variable regime-x_{t_y} is quite slow moving and monthly returns cause too much unnecessary noise for the inference about the slow moving state variable. The annual indices are constructed by averaging monthly or quarterly indices within each year. The coefficients are estimated based on weighted-least-square estimation where weights are inverse market return variance. The weighted-least-squares method is necessary, since heteroskedasticity shows up largely in this unified regression and it is correlated with the explanatory state variable regime-x_{t_y}. The data are from 1953 to 2014 for regressions in (1), (2), (5), and (6); due to restrictions of availability, the data are from 1976 to 2014 for regressions in (3), (4), (6), and (8).

Table 2.15: MODEL VERSUS DATA: UNCERTAINTY-BETA SORTED PORTFOLIOS

UNCERTAINTY-BETA SORT	BAD RISK SHARING CONDITION $\Delta \nu_g$ (1)	$\Delta \nu_c$ (2)	GOOD RISK SHARING CONDITION $\Delta \nu_g$ (3)	$\Delta \nu_c$ (4)	ALL $\Delta \nu_g$ (5)	$\Delta \nu_c$ (6)
	A. DATA: REINHART-ROGOFF INDEX					
LOW 20%	3.77	6.89	13.79	16.79	12.38	17.45
	(0.87)	(1.11)	(3.23)	(2.86)	(6.49)	(8.18)
HIGH 20%	1.46	0.21	15.75	14.11	12.34	12.34
	(0.37)	(0.03)	(3.98)	(2.26)	(7.98)	(6.62)
HIGH − LOW	−2.30	−6.68	1.96	−2.27	−0.04	−5.11
	(−1.24)	(−1.69)	(0.91)	(−0.42)	(−0.02)	(−3.02)
	B. DATA: BROKER-DEALER LEVERAGE INDEX					
LOW 20%	7.99	10.19	12.76	14.91	12.38	17.45
	(1.31)	(1.69)	(4.13)	(3.60)	(6.49)	(8.18)
HIGH 20%	4.98	5.21	16.49	14.12	12.34	12.34
	(1.21)	(0.81)	(5.92)	(3.33)	(7.98)	(6.62)
HIGH − LOW	−3.00	−4.99	3.73	−0.80	−0.04	−5.11
	(−2.44)	(−1.19)	(3.33)	(−0.27)	(−0.02)	(−3.02)
	C. DATA: CREDIT SPREAD INDEX					
LOW 20%	14.32	13.04	13.41	19.09	12.38	17.45
	(2.31)	(1.64)	(2.07)	(2.59)	(6.49)	(8.18)
HIGH 20%	8.98	9.45	17.77	15.37	12.34	12.34
	(1.31)	(1.16)	(2.43)	(3.10)	(7.98)	(6.62)
HIGH − LOW	−5.34	−3.59	4.36	−3.72	−0.04	−5.11
	(−1.13)	(−0.60)	(0.96)	(−1.23)	(−0.02)	(−3.02)
	D. MODEL					
LOW	16.97	18.33	11.22	16.03	13.73	16.33
	(6.66)	(6.84)	(4.49)	(6.42)	(8.11)	(9.65)
HIGH	7.79	7.34	15.79	11.39	13.34	11.76
	(2.60)	(2.76)	(6.28)	(4.56)	(7.88)	(6.95)
HIGH − LOW	−9.18	−10.99	4.57	−4.64	−0.39	−4.57
	(−3.06)	(−3.98)	(1.79)	(−1.86)	(−0.23)	(−2.70)

NOTES: The table compares unconditional moments of the data to their simulated correspondences in the model. Within each Panel, it reports the average returns of uncertainty-beta sorted portfolios for the whole sample and two subsamples. The difference is that Panel A (Panel B) uses the Reinhart-Rogoff Index (Broker-Dealer Leverage Index) to measure risk sharing conditions, while Panel C uses the credit spread index to measure risk sharing conditions. The t-statistics are reported in the parentheses. Data are sampled at the monthly frequency. Their sources and construction details are explained in the online appendix. The sample period is 1976 – 2014. Panel D reports simulated average returns based on uncertainty-beta sorted portfolios in the model. I simulate the model at the weekly frequency and then time-aggregate the simulated data to form monthly observations. In parentheses, the numbers are t-statistics computed using 1,000 independent simulations, each with a length of 400 years.

Chapter 3

Measuring the "Dark Matter" in Asset Pricing Models

3.1 Measuring Model Fragility

In this section, we first introduce a formal measure of model fragility. Then we derive asymptotic properties of the fragility measure.

3.1.1 A Generic Model Structure

Consider a baseline model $\mathbb{P}$, which is a part of the full structural model $\mathcal{Q}$. The baseline model $\mathbb{P}$ has a $D_\Theta \times 1$ parameter vector $\theta \in \Theta$ and it specifies the dynamics of a vector of variables $\mathbf{x_t}$. In comparison, the full structural model $\mathcal{Q}$ has extra parameters as summarized by the vector ψ in addition to θ. The model $\mathcal{Q}$ aims to capture certain features of the distribution $\mathbb{Q}$ that governs the joint dynamics of $\mathbf{x_t}$ and additional variables $\mathbf{y_t}$.

We assume that the stochastic process $\{\mathbf{x_t}\}$ is strictly stationary and has a stationary distribution $\mathbb{P}_0 \equiv \mathbb{P}_{\theta_0}$. The true joint distribution for $\mathbf{x^n} \equiv (\mathbf{x_1}, \cdots, \mathbf{x_n})$ is $\mathbb{P}_{0,n} \equiv \mathbb{P}_{\theta_0,n}$, with the corresponding parameter vector θ_0, which is unknown to the econometrician. The density function for $\mathbf{x^n}$ is $\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)$. Without much loss of generality, we assume the process $\mathbf{x}$ is first-order Markov.[1] Similarly, we assume that the stochastic process $(\mathbf{x_t}, \mathbf{y_t})$ is strictly stationary and has a stationary distribution $\mathbb{Q}_0$. The econometrician does not need to specify the full functional form of the joint distribution of $(\mathbf{x^n}, \mathbf{y^n}) \equiv \{(\mathbf{x_t}, \mathbf{y_t}) : \ t = 1, \cdots, n\}$, which we denote by $\mathbb{Q}_{0,n}$. The corresponding joint density is $q_0(\mathbf{x^n}, \mathbf{y^n})$.

We evaluate the performance of a structural model under the Generalized Method of Moments (GMM) framework. Specifically, we assume that the model builder is concerned with the model's in-sample and out-of-sample performance as represented by a set of moment conditions, based on a $D_g \times 1$ vector of functions $g(\mathbf{x}, \mathbf{y}; \theta, \psi)$ of data observations $(\mathbf{x}_t, \mathbf{y}_t)$ and the parameter vector (θ, ψ) satisfying the following conditions

$$\mathbb{E}_{\mathbb{Q}_0}\left[g(\mathbf{x_t}, \mathbf{y_t}; \theta_0, \psi_0)\right] = 0. \tag{3.1}$$

We require the conditional score functions for the likelihood of $\mathbb{P}$, $\partial \ln \pi_{\mathbb{P}}(\mathbf{x}_t; \theta)/\partial\theta$,[2] to be included in the vector of moment conditions $g(\mathbf{x_t}, \mathbf{y_t}, \theta, \psi)$.

Denote

$$g_n(\mathbf{x^n}, \mathbf{y^n}; \theta, \psi) \equiv \frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}; \theta, \psi). \tag{3.2}$$

Then, the (efficient) GMM minimizes

$$J_{n,S_0}(\theta, \psi; \mathbf{x^n}, \mathbf{y^n}) \equiv n g_n(\mathbf{x^n}, \mathbf{y^n}; \theta, \psi)^T S_0^{-1} g_n(\mathbf{x^n}, \mathbf{y^n}; \theta, \psi), \tag{3.3}$$

where J_{n,S_0} is often referred to as the J-statistic, and S_0 has the following explicit formula

$$S_0 \equiv \sum_{\ell=-\infty}^{+\infty} \mathbb{E}_{\mathbb{Q}_0}\left[g(\mathbf{x_t}, \mathbf{y_t}; \theta_0, \psi_0) g(\mathbf{x_{t-\ell}}, \mathbf{y_{t-\ell}}; \theta_0, \psi_0)^T\right]. \tag{3.4}$$

[1] If the original random variable x_t is a m_0-th order Markov with $m_0 > 1$, we can construct a new random vector $\mathbf{x}_t$ stacking variables x_t with a sufficient number of lags so that $\mathbf{x_t}$ is first-order Markov. More precisely, we assume that the underlying time series x_t with $t = 1, \cdots, n$ is m-dependent process and the conditional density is $\pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})$ for some positive integer constant m_0. For the stacked vector $\mathbf{x_t} = (x_t, \cdots, x_{t-K_0})^T$ with $K_0 \geq m_0$, the conditional density for x_t under $\mathbb{P}$ can be rewritten as $\pi_{\mathbb{P}}(\mathbf{x_t}; \theta) = \pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})$.

[2] $\pi_{\mathbb{P}}(\mathbf{x_t}; \theta)$ is the conditional density for the m_0-th order Markovian underlying process x_t, i.e. $\pi_{\mathbb{P}}(\mathbf{x_t}; \theta) = \pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})$.

The matrix S_0 is the covariance matrix of the moment conditions at the true θ_0. In practice, when S_0 is unknown, we can replace it with a consistent estimator $\widehat{S}_n$. The consistent estimators of S_0 are provided by Newey and West (1987a), Andrews (1991), and Andrews and Monahan (1992).

We use GMM to evaluate model performance because of the concern of model misspecification. The GMM approach gives the model builder flexibility to choose which aspects of the model to emphasize when estimating model parameters. This is in contrast to the likelihood approach, which relies on the full probability distribution implied by the structural model.[3]

Finally, we introduce some further notation. We denote the expected Fisher information matrix for the baseline model as $\mathbf{I}_{\mathbb{P}}(\theta)$,

$$\mathbf{I}_{\mathbb{P}}(\theta) \equiv \mathbb{E}_{\mathbb{P}_0}\left[\frac{\partial}{\partial\theta}\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)\frac{\partial}{\partial\theta^T}\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)\right]. \tag{3.5}$$

We denote the GMM analog for the structural model as $\mathbf{I}_{\mathbb{Q}}(\theta)$, (see Hansen, 1982; Hahn, Newey, and Smith, 2011)

$$\mathbf{I}_{\mathbb{Q}}(\theta) \equiv G_0(\theta)^T S_0^{-1} G_0(\theta), \tag{3.6}$$

where

$$G_0(\theta) \equiv \mathbb{E}_{\mathbb{Q}_0}\left[\frac{\partial g(\mathbf{x_t}, \mathbf{y_t}; \theta, \psi_0)}{\partial\theta^T}\right]. \tag{3.7}$$

When evaluated at the true θ_0, the two matrices $\mathbf{I}_{\mathbb{P}}(\theta_0)$ and $\mathbf{I}_{\mathbb{Q}}(\theta_0)$ are the Fisher information matrix for the parametric family of the baseline model and the GMM information matrix for the structural model at the true θ_0, respectively. Computing the expectation in (3.7) requires knowing the distribution $\mathbb{Q}_0$. In cases when $\mathbb{Q}_0$ is unknown, $G_0(\theta)$ in (3.6) can be replaced by its consistent estimator $n^{-1}\sum_{t=1}^{n}\partial g(\mathbf{x_t}, \mathbf{y_t}; \theta, \psi_0)/\partial\theta^T$.

[3]Our assumptions of the knowledge of the likelihood function for the baseline model $\mathbb{P}$ and the GMM approach for the structural model $\mathcal{Q}$ are not restrictive. As a special case, we recover the MLE when we include the full score function for model $\mathcal{Q}$ as the moments. There are also ways to construct our fragility measure without relying on the likelihood function for $\mathbb{P}$, e.g., by using the limited information likelihood.

3.1.2 Model Fragility

We now define our measure of model fragility.

Definition 3 (Fragility Measure)**.** *Let $\pi(\theta)$ be a prior distribution on θ and let $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$ be the posterior on θ based on $\mathbb{P}$ and $\mathbf{x^n}$. We define the fragility measure for the structural model $\mathbb{Q}$ relative to the baseline model $\mathbb{P}$ as*

$$\varrho(\mathbf{x^n}, \mathbf{y^n}) \equiv \int d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, \theta\}\pi_{\mathbb{P}}(\theta|\mathbf{x^n})d\theta, \tag{3.8}$$

where

$$d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, \theta\} = J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \theta) - J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \widehat{\theta}^{\mathbb{Q}}), \tag{3.9}$$

$J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \theta)$ is the J-statistic defined in (3.3), and $\widehat{\theta}^{\mathbb{Q}}$ is the GMM estimator.[4]

The idea of our fragility measure is to quantify the in-sample over-fitting by a structural model. In Equation (3.9), $d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, \theta\}$ is the J-distance statistic (the GMM analog of the log likelihood ratio) of the model with parameter vector $\widehat{\theta}^{\mathbb{Q}}$, which provides the best in-sample fit of the data based on the GMM criterion, against an alternative model with the parameter vector θ. Assuming true parameter is θ instead of $\widehat{\theta}^{\mathbb{Q}}$, the fact that the J-statistic based on $\widehat{\theta}^{\mathbb{Q}}$ is smaller is a sign of over-fitting.

The weights attached to alternative models are essential for our definition of model fragility. We consider alternative values of θ while holding the rest of the structural parameters ψ fixed, i.e., we assume $\psi = \psi_0$. Starting with a prior $\pi(\theta)$, we weigh the various alternative models using $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$ – the posterior for θ based on the baseline model $\mathbb{P}$ and data $\mathbf{x^n}$. The weighted average of $d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, \theta\}$ over the entire set of alternative models represents the average degree of model over-fitting.

[4]As emphasized by Newey and West (1987b), it is crucial to define the GMM likelihood ratio test statistics based on the optimal GMM estimator, for which $W = S_0^{-1}$. This is because another choice of the weighting matrix W will destroy the asymptotic property of having chi-squared distribution as limiting distribution and will break the asymptotic equivalence between GMM likelihood ratio test statistics and other GMM test statistics such as Wald and LM test statistics. Similarly, as highlighted in Kim (2002), it is important to use efficient weighting matrix for developing the theory of limited information likelihood based on GMM.

The weights on various alternative models depend on $\mathbb{P}$, therefore $\varrho(\mathbf{x^n}, \mathbf{y^n})$ is a measure of fragility of $\mathcal{Q}$ *relative to the baseline model* $\mathbb{P}$. Thus, our measure of model fragility depends on the choice of the baseline model. Many structural models involve both a statistical model of exogenous variables and restrictions on the endogenous variables which are derived from the economic model. For such models, a natural choice may be to take $\mathbb{P}$ to be the statistical model, with $\mathbf{x_t}$ being the exogenous variables. $\mathbf{y_t}$ would then be the vector of endogenous variables in the structural model. In this context, our fragility measure quantifies the fragility of the structural model relative to the statistical model.

Alternatively, a structural model could be taken as the baseline model. Then, the fragility measure applies to the over-fitting caused by the additional economic restrictions imposed by $\mathcal{Q}$ relative to the baseline model.

The distribution over the alternative models also depends on the choice of the prior $\pi(\theta)$. If the econometrician does not have any information about θ beyond the baseline model and the data $\mathbf{x^n}$, an "uninformative" prior would be a desirable choice, one candidate being the Jeffreys prior. In many cases a truly uninformative prior is difficult to define, especially in the presence of constraints. If the econometrician has additional information about θ outside the model (e.g., from additional data or other models), such information can be incorporated through an informative prior.

Our definition of model fragility builds upon Spiegelhalter, Best, Carlin, and van der Linde (2002), who propose a related measure of model complexity for statistical models. Our measure differs from theirs in two respects. First, we adopt the GMM framework as opposed to the likelihood framework to address the issue of stochastic singularities that arise in structural models and to give the econometrician the flexibility to focus on specific features of a model. Second, Spiegelhalter, Best, Carlin, and van der Linde (2002) do not explicitly specify the weighting of alternative models.

3.1.3 Alternative Interpretations

The measure $\varrho(\mathbf{x^n}, \mathbf{y^n})$ has three related interpretations. It can be viewed as a measure of model complexity. It is also a measure of out-of-sample performance for the structural model. Finally, it can be seen as a measure of sensitivity of the structural model to small perturbations in its parameters.

Model complexity

It is well known that adding degrees of freedom to a model can increase the risk of over-fitting the data in sample with poor out-of-sample performance (see, for example, Fisher, 1922). The in-sample over-fitting is reflected by a large gap, on average, between $J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \widehat{\theta}^{\mathbb{Q}})$ (the global minimum of the J-statistic in sample) and $J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \theta)$ for alternative values of θ.

Out-of-sample performance

In statistics, the concept of model complexity is tightly linked to out-of-sample performance.[5] Let $(\tilde{\mathbf{x}}^{\mathbf{n}}, \tilde{\mathbf{y}}^{\mathbf{n}})$ denote a new data sample generated from the true distribution $\mathbb{Q}_{0,n}$. Let the J statistic corresponding to the moment conditions, $J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \widehat{\theta}^{\mathbb{Q}})$, serve as the loss function for evaluating the out-of-sample performance of model $\mathcal{Q}$.[6] Then, the out-of-sample performance of the model can be summarized by a risk function based on the expected J statistic,

$$\mathbb{E}_{\mathbb{Q}_{0,n}}^{(\tilde{\mathbf{x}}^{\mathbf{n}}, \tilde{\mathbf{y}}^{\mathbf{n}})} \left[J_{n,S_0}(\tilde{\mathbf{x}}^{\mathbf{n}}, \tilde{\mathbf{y}}^{\mathbf{n}}; \widehat{\theta}^{\mathbb{Q}}) \right]. \tag{3.10}$$

[5] Spiegelhalter, Best, Carlin, and van der Linde (2002), Ando (2007) and Gelman, Hwang, and Vehtari (2013), among others, argue that the effective number of parameters is tightly connected to the out-of-sample predictive accuracy of a model.

[6] Under the limited information likelihood interpretation of GMM, which we outline in Section 3.2.2, the loss function (3.10) can be rewritten as $-2\ln \pi_{\mathbb{Q}}(\cdot|\widehat{\theta}^{\mathbb{Q}})$, where $\pi_{\mathbb{Q}}(\cdot|\widehat{\theta}^{\mathbb{Q}})$ is the limited information likelihood function for the moment conditions. The expected limited information likelihood is a standard measure for out-of-sample performance. In linear Gaussian models it reduces to an expected sum of squared errors, which has been extensively used as a measure of out-of-sample fit in applied statistics. Similar risk functions for out-of-sample performance are also adopted in the macroeconomics literature (see, e.g. Smets and Wouters, 2007).

The lower the value of this risk function, the better the expected out-of-sample performance of the model.

An obvious challenge in computing the risk function in (3.10) is that the true distribution $\mathbb{Q}_{0,n}$ is unknown. One solution is to replace $\mathbb{Q}_{0,n}$ with the Bayesian predictive distribution (see, e.g. Spiegelhalter, Best, Carlin, and van der Linde, 2002; Gelfand and Ghosh, 1998; Ando, 2007).[7]

When n is large, the sample $(\mathbf{x^n}, \mathbf{y^n})$ reflects the population distribution of the data, and the expected loss function can be approximated as

$$\mathbb{E}^{(\tilde{\mathbf{x}}^{\mathbf{n}},\tilde{\mathbf{y}}^{\mathbf{n}})|\mathbf{x^n}}\left[J_{n,S_0}(\tilde{\mathbf{x}}^{\mathbf{n}}, \tilde{\mathbf{y}}^{\mathbf{n}}; \widehat{\theta}^{\mathbb{Q}})\right] \approx J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \widehat{\theta}^{\mathbb{Q}}) + 2\varrho(\mathbf{x^n}, \mathbf{y^n}). \tag{3.11}$$

Thus, the higher the value of $\varrho(\mathbf{x^n}, \mathbf{y^n})$, the worse the out-of-sample performance of the structural model.

Model sensitivity

It is common practice to examine robustness of a model by conducting sensitivity analysis. Intuitively, the goal is to check whether model's key implications are sensitive to small perturbations in the parameter values. To formalize this approach, one needs to define what constitutes "small perturbations" in parameter values and how the evaluate sensitivity of the model performance to such perturbations.

Our definition of $\varrho(\mathbf{x^n}, \mathbf{y^n})$ can be viewed as a multivariate model sensitivity measure. The GMM J-distance statistic $d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, \theta\}$ in (3.9) measure how model performance is affected by a change in the parameter vector from $\widehat{\theta}^{\mathbb{Q}}$ to an alternative value θ. $\varrho(\mathbf{x^n}, \mathbf{y^n})$ evaluates changes in the moments relative to their covariance matrix. It shows on average how sensitive the model performance is with respect to all possible perturbations in θ, where the notion of "small perturbations" is captured by assigning higher weights to the values of θ deemed more likely based on the posterior distribu-

[7]The Bayesian predictive distribution is given by $\int \pi_{\mathbb{Q}}(\tilde{\mathbf{x}}^{\mathbf{n}}, \tilde{\mathbf{y}}^{\mathbf{n}}|\theta, \psi_0)\pi_{\mathbb{P}}(\theta|\mathbf{x^n})d\theta$, where $\pi_{\mathbb{Q}}(\tilde{\mathbf{x}}^{\mathbf{n}}, \tilde{\mathbf{y}}^{\mathbf{n}}|\theta, \psi_0)$ is the limited information likelihood function for the moment conditions defined in Section 3.2.2.

tion from the baseline model, $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$.

3.1.4 A Generalized Fragility Measure

Next, we generalize the fragility measure $\varrho(\mathbf{x^n}, \mathbf{y^n})$ to allow for transformations of the parameter vector θ.

Definition 4 (Fragility Measure with Feature Functions)**.** *Let f be a $\mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_f}$ continuous differentiable mapping with $1 \leq D_f \leq D_\Theta$. Then, we define*

$$\varrho^f(\mathbf{x^n}, \mathbf{y^n}) \equiv \int d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\}\pi_{\mathbb{P}}(\theta|\mathbf{x^n})d\theta, \tag{3.12}$$

$$\text{where} \quad d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\} = \inf_{\tilde{\theta}: f(\tilde{\theta}) = f(\theta)} J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \tilde{\theta}) - J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \widehat{\theta}^{\mathbb{Q}}). \tag{3.13}$$

$J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \theta)$ is the J-statistic defined in (3.3), and $\widehat{\theta}^{\mathbb{Q}}$ is the GMM estimator.

Transforming the original parameter vector is useful, for example, if one wants to measure model's robustness with respect to a low-dimensional subset in the parameter space. For instance, to measure model robustness with respect to the first D_0 elements of θ ($D_0 < D_\Theta$), we set $f(\theta) = \mathbf{F}\theta$, where $\mathbf{F} = [\mathbf{I}_{D_0}, \mathbf{0}]$. In the special case of $f(\theta) = \mathbf{F}\theta$, with $\mathbf{F}$ being an arbitrary full-rank $D_\Theta \times D_\Theta$ matrix, we recover the original fragility measure, $\varrho^f(\mathbf{x^n}, \mathbf{y^n}) = \varrho(\mathbf{x^n}, \mathbf{y^n})$.

3.1.5 Asymptotic Fragility Measure

In practice, computation of the fragility measures $\varrho(\mathbf{x^n}, \mathbf{y^n})$ and $\varrho^f(\mathbf{x^n}, \mathbf{y^n})$ may be complicated by the complex form of the likelihood function of the baseline model, the curse of dimensionality induced by high-dimensional parameter spaces, and the additional minimization problem involved in the definition of the generalized measure. In this section, we derive an asymptotic approximation for the fragility measures and an eigen-decomposition of the asymptotic approximation.

Definition 5 (Asymptotic Fragility Measure)**.** *The asymptotic fragility measure corresponding to a full-rank $D_{\mathbf{v}} \times D_{\Theta}$ matrix $\mathbf{v}$ is defined as*

$$\varrho_a^{\mathbf{v}}(\theta_0) \equiv \mathbf{tr}\left[\left(\mathbf{v}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}^T\right)^{-1}\left(\mathbf{v}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}^T\right)\right], \tag{3.14}$$

where $\mathbf{I}_{\mathbb{P}}(\theta_0)$ and $\mathbf{I}_{\mathbb{Q}}(\theta_0)$ are the information matrices defined in (3.5) and (3.6).

In the special case where $\mathbf{v}$ is a full-rank $D_\Theta \times D_\Theta$ matrix, $\varrho_a^{\mathbf{v}}(\theta_0)$ is independent of the choice of $\mathbf{v}$. In that case we denote the asymptotic fragility measure as $\varrho_a(\theta_0)$.

The asymptotic fragility measure is connected to the original fragility measure defined in Section 3.1.2, as we show in the following theorem.

Theorem 1. *Assume the regularity conditions in Section A.3.3 hold. Consider a feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_f}$ with $\mathbf{v} = \partial f(\theta_0)/\partial\theta^T$ being the $D_f \times D_\Theta$ Jacobian matrix. Then, $\varrho_a^{\mathbf{v}}(\theta_0)$ and $\varrho^f(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$ are asymptotically related as*

$$\varrho^f(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n}) \rightsquigarrow 2\varrho_a^{\mathbf{v}}(\theta_0) - D_f + \varepsilon, \quad \mathbb{E}_{\mathbb{Q}_0}[\varepsilon] = 0, \tag{3.15}$$

where the distribution of the random variable ε depends on the feature vector $\mathbf{v}$ and the information matrices $\mathbf{I}_{\mathbb{P}}(\theta_0)$ and $\mathbf{I}_{\mathbb{Q}}(\theta_0)$, and $\rightsquigarrow$ denotes convergence in distribution. Moreover, convergence of expectations is also guaranteed:

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}_{0,n}}\left[\varrho^f(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})\right] = 2\varrho_a^{\mathbf{v}}(\theta_0) - D_f. \tag{3.16}$$

Proof. See Appendix A.3.3. □

Convergence of expectations shows that the quantity $2\varrho_a^{\mathbf{v}}(\theta_0) - D_f$ indeed provides a valid asymptotic approximation to the average fragility measure, defined as $\mathbb{E}_{\mathbb{Q}_{0,n}}\left[\varrho^f(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})\right]$. In practice, with a large sample size, the econometrician can use re-sampling methods such as bootstrap based on the sample $(\mathbf{x}^\mathbf{n}, \mathbf{y}^\mathbf{n})$ to estimate the average fragility measure. The limiting result in (3.18) guarantees that $2\varrho_a^{\mathbf{v}}(\theta_0) - D_f$ provides a reasonable approximation for such average fragility measure when sample size is large.

The asymptotic measure $\varrho_a^{\mathbf{v}}(\theta_0)$ is determined entirely by the properties of the model and does not depend on the data sample. This measure focuses on local departures from the true parameter vector θ_0. Theorem 1 shows that $\varrho_a^{\mathbf{v}}(\theta_0)$ characterizes average model fragility as sample size approaches infinity.

The following result follows immediately from Theorem 1 and establishes the asymptotic equivalence of the fragility measures (without feature functions):

Corollary 3. *Assume the regularity conditions in Section A.3.3 hold.*

$$\varrho(\mathbf{x^n}, \mathbf{y^n}) \rightsquigarrow 2\varrho_a(\theta_0) - D_\Theta + \varepsilon, \quad \mathbb{E}_{\mathbb{Q}_0}[\varepsilon] = 0, \tag{3.17}$$

where the random variable ε has distribution that only depends on the information matrices $\mathbf{I}_\mathbb{P}(\theta_0)$ and $\mathbf{I}_\mathbb{Q}(\theta_0)$, and $\rightsquigarrow$ denotes convergence in distribution. Moreover, convergence of expectations is also guaranteed:

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}_{0,n}}\left[\varrho(\mathbf{x^n}, \mathbf{y^n})\right] = 2\varrho_a(\theta_0) - D_\Theta. \tag{3.18}$$

The measure $\varrho_a^{\mathbf{v}}(\theta_0)$ is defined for a specific feature function f. One might be interested in searching among a class of feature functions to find the worse-case configuration. While this is generally difficult to do for the finite-sample fragility measure $\varrho^f(\mathbf{x^n}, \mathbf{y^n})$, it is actually quite straightforward for the asymptotic measure, because the original infinite-dimensional optimization problem is reduced to a finite-dimensional one. This leads us to define the following worst-case asymptotic fragility measure.

Definition 6. *The worst-case asymptotic fragility measure for the class of D-dimensional feature functions ($D \leq D_\Theta$) is defined as:*

$$\varrho_a^D(\theta_0) = \max_{v \in \mathbb{R}^{D\times D_\Theta}, \mathbf{Rank}(v)=D} \mathbf{tr}\left[\left(v\mathbf{I}_\mathbb{Q}(\theta_0)^{-1}v^T\right)^{-1}\left(v\mathbf{I}_\mathbb{P}(\theta_0)^{-1}v^T\right)\right]. \tag{3.19}$$

The problem in (3.19) is a generalized eigenvalue problem. The following proposition summarizes its solution.

Proposition 9. *Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{D_\Theta}$ be the eigenvalues of $\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}$, with the corresponding eigenvectors $e_1, e_2, \cdots, e_{D_\Theta}$ in $\mathbb{R}^{D_\Theta}$. Then, the D-dimensional worst-case asymptotic model fragility measure is equal to*

$$\varrho_a^D(\theta_0) = \lambda_1 + \lambda_2 + \cdots + \lambda_D, \tag{3.20}$$

with the worst-case D-dimensional linear subspace of the parameter space characterized by the matrix $\mathbf{v}_D^* = [v_1^* \; v_2^* \; \cdots \; v_D^*]$,

$$v_i^* = \mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}} e_i \Big/ \left| \mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}} e_i \right|. \tag{3.21}$$

As a special case, the overall asymptotic fragility measure is given by

$$\varrho_a(\theta_0) = \lambda_1 + \lambda_2 + \cdots + \lambda_{D_\Theta}. \tag{3.22}$$

Proof. See Appendix A.3.1. □

The intuition behind the worst-case asymptotic model fragility measure is as follows. Through the matrix $\mathbf{v}$, we search over all D-dimensional linear subspaces of the parameter space to find the maximum discrepancy between the inverses of the two information matrices, $\mathbf{I}_{\mathbb{P}}(\theta_0)$ and $\mathbf{I}_{\mathbb{Q}}(\theta_0)$. In the context of MLE and GMM estimation, the inverse of the information matrix is linked to the asymptotic covariance matrices of the estimators. Since we require the conditional score functions for the baseline model to be included in the moment conditions for the GMM, the asymptotic efficiency of the GMM estimator for the structural model dominates that of the baseline model. The asymptotic fragility measure effectively compares the asymptotic covariance matrices of these two estimators to isolate the information provided by the structural model restrictions.

We can view Proposition 9 as a decomposition of the overall fragility of a model into D_Θ 1-dimensional linear subspaces. The i-th largest eigenvalue λ_i ($1 \leq i \leq D_\Theta$) of $\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}$ gives the marginal contribution of the 1-dimensional linear subspace associated with v_i^* to the overall fragility measure. In the language of sensitivity analysis, such a decomposition reveals the directions in which small perturbations of parameters can have the largest impact on the model output.

The asymptotic fragility measure has a natural sample-size interpretation. Consider the case of $D = 1$. In this case, we ask what is the minimum sample size required for the estimator of the baseline model to match or exceed the precision of the estimator for the full structural model in all 1-dimensional linear subspaces of the parameter space. Because the asymptotic covariance of the estimator is proportional to the sample size n, the required sample size for the baseline model is ϱ_a^1 times the sample size for the structural model to achieve at least the same estimation accuracy.

From Proposition 9, it is easy to see that the worst-case asymptotic fragility measure $\varrho_a^D(\theta_0)$ is monotonically increasing in the dimension of the subpsace D.[8]

Proposition 10. (Monotonicity) *For $D_1 \leq D_2 \leq D_\Theta$,*

$$\varrho_a^{D_1}(\theta) \leq \varrho_a^{D_2}(\theta). \tag{3.23}$$

3.2 Model Fragility and Informativeness of Cross-Equation Restrictions

In this section we formalize the intuition that excessive informativeness of cross-equation restrictions tends to be associated with model fragility.

3.2.1 Chernoff Information

Our asymptotic fragility measures are based on the information matrices (the likelihood-based Fisher information or the generalized Fisher information for GMM) from the baseline model and the full structural model. We show that by comparing the infor-

[8] A similar monotonicity property applies to $\varrho^f(\mathbf{x^n}, \mathbf{y^n})$. Let $\widetilde{f} = [f, f_1]'$, where f and f_1 are continuously differentiable and $D_{\widetilde{f}} \leq D_\Theta$. Then $\varrho^f(\mathbf{x^n}, \mathbf{y^n}) < \varrho_c^{\widetilde{f}}(\mathbf{x^n}, \mathbf{y^n})$.

mation matrices, the asymptotic fragility measure quantifies the informativeness of the cross-equation restrictions in the structural model. We define the latter notion precisely below.

We start by introducing the concept of Chernoff information. Chernoff information gives the asymptotic geometric rate (Chernoff rate) at which the detection error probability (the weighted average of the error probabilities in selection between two alternative models) decays as the sample size increases. Intuitively, Chernoff information measures the difficulty of discriminating among alternative models.[9]

Consider a model with density $p(x|\theta_0)$ and an alternative model with density $p(x|\theta)$. Assume the densities are absolutely continuous relative to each other. The Chernoff information between the two models is defined as (see, e.g., Cover and Thomas (1991)):

$$C^*(p(x|\theta) \ : \ p(x|\theta_0)) \equiv -\ln \min_{\alpha\in[0,1]} \int_{\mathcal{X}} p(x|\theta_0)^{\alpha} p(x|\theta)^{1-\alpha} \mathrm{d}x. \tag{3.24}$$

The cross-equation restrictions imposed by the structural model increase efficiency of parameter estimation, which makes it is easier to distinguish model $\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)$ from local alternatives, $\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0 + n^{-\frac{1}{2}}u)$ (u is a vector), compared to distinguishing $\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)$ from $\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0 + n^{-\frac{1}{2}}u)$. Informativeness of cross-equation restrictions for discrimination between alternative models can be captured asymptotically by the ratio of two Chernoff rates, computed with and without imposing the cross-equation restrictions. The following proposition connects such ratio to the asymptotic fragility measure $\varrho_a(\theta_0)$.

Proposition 11. *Assume the regularity conditions in Section A.3.3 hold. Then, there exist D_Θ linearly independent $D_\Theta-$dimensional vectors $u_1, \cdots, u_{D_\Theta}$ such that*

$$\varrho_a(\theta_0) = \lim_{n\to\infty} \sum_{i=1}^{D_\Theta} \frac{C^*(\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_{u_i}) \ : \ \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) \ : \ \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0))}, \tag{3.25}$$

where $\theta_{u_i} = \theta_0 + n^{-\frac{1}{2}} u_i$ and n is the sample size.

[9] Anderson, Hansen, and Sargent (2003) use Chernoff rate to motivate a measure of model misspecification in their analysis of robust decision making.

Proof. See Appendix A.3.2. □

3.2.2 Relative Entropy and Effective Sample Size

In Section 3.2.1, we have shown that the asymptotic fragility measure can be interpreted as an asymptotic measure of informativeness of the cross-equation restrictions in the structural model via the Chernoff rates. In finite samples, we use the Bayesian method to measure the informativeness of cross-equation restrictions. Starting with a prior on θ, $\pi(\theta)$, we obtain posterior distributions through the baseline model and the structural model, respectively. Then, the discrepancy between the two posteriors shows how the cross-equation restrictions affect the inference about θ. In this section, we establish the connection between this finite-sample measure of the informativeness of cross-equation restrictions and the asymptotic fragility measure.

Given the prior $\pi(\theta)$ and data $\mathbf{x^n}$, the posterior density of θ in the baseline model is $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$. To derive the posterior density in the structural model $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$, we first introduce the Limited Information Likelihood (LIL) for GMM. Given the GMM J-statistic $J_{n,S_0}(\theta, \psi_0; \mathbf{x^n}, \mathbf{y^n})$ and the true joint density $q_0(\mathbf{x^n}, \mathbf{y^n})$, the Limited Information Likelihood is

$$\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta, \psi_0) = \mathcal{C}_n \exp\left\{-\frac{J_{n,S_0}(\theta, \psi_0; \mathbf{x^n}, \mathbf{y^n})}{2}\right\} q_0(\mathbf{x^n}, \mathbf{y^n}), \tag{3.26}$$

where $\mathcal{C}_n$ are constants.[10] The corresponding limited information posterior is

$$\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n}) = \frac{\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta)\pi(\theta)}{\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n})}. \tag{3.27}$$

We quantify the discrepancy between probability distribution using a standard statistical measure, the relative entropy (also known as the Kullback-Leibler divergence). The relative entropy between $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$ and $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$ is

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta|\mathbf{x^n})) = \int \ln\left(\frac{\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})}{\pi_{\mathbb{P}}(\theta|\mathbf{x^n})}\right)\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})d\theta\,. \tag{3.28}$$

Intuitively, we can think of the log posterior ratio $\ln(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})/\pi_{\mathbb{P}}(\theta|\mathbf{x^n}))$ as a measure of the discrepancy between the two posteriors at a given θ. Then the relative entropy is the average discrepancy between the two posteriors over all possible θ, where the average is computed under the constrained posterior. $\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta|\mathbf{x^n}))$ is finite if and only if the support of the posterior $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$ is a subset of the support of the posterior $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$, that is, Assumption PQ in Appendix A.3.3 holds.

The magnitude of relative entropy is difficult to interpret directly, and we propose an intuitive "effective sample size" interpretation. Instead of imposing the cross-equation restrictions from the structural model, one can gain extra information about θ within the baseline model with additional data. We evaluate the amount of additional data under the baseline model needed to match the informativeness of cross-equation restrictions.

Suppose we draw additional data $\mathbf{\tilde{x}^m}$ of sample size m from the Bayesian predictive

[10] Kim (2002) derives the LIL from the I-projection theory based on relation entropy (see, e.g. Csiszár, 1975; Jaynes, 1982; Jones, 1989). He chooses the probability measure that minimizes the relative entropy distance from the true probability measure $\mathbb{Q}_0$, out of a set of probability measures satisfying the same moment conditions. Kim (2002) extends Bayesian methods of moments of Zellner (1996, 1998) to the general case of GMM for deriving a limited information posterior and derive a limited information likelihood that is not considered by Zellner (1996, 1998). This method is also considered by Chernozhukov and Hong (2003) as a special case of their "Quasi-Bayesian method" and studied empirically by Yin (2009) as "Bayesian GMM". The limited information likelihood formulation reconciles the efficient GMM estimation with the maximum likelihood estimation approach. For example, the over-identification test (i.e., the J-distance test), the Wald test and Lagrange multiplier test in the GMM framework (see, e.g. **?**Newey, 1985; Newey and West, 1987b) have formal likelihood-based counterparts under the limited information likelihood formulation.

distribution

$$\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}}) \equiv \int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})\mathrm{d}\theta. \tag{3.29}$$

Again, we measure the gain in information from this additional sample $\tilde{\mathbf{x}}^{\mathbf{m}}$ using relative entropy,

$$\mathbf{D}_{KL}(\pi_{\mathbb{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})) = \int \ln\left(\frac{\pi_{\mathbb{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})}{\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})}\right)\pi_{\mathbb{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})\mathrm{d}\theta\,. \tag{3.30}$$

$\mathbf{D}_{KL}(\pi_{\mathbb{P}}(\theta|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}))$ depends on the realization of the additional sample of data $\tilde{\mathbf{x}}^{\mathbf{m}}$. The average relative entropy (information gain) over possible future samples $\{\tilde{\mathbf{x}}^{\mathbf{m}}\}$ according to the Bayesian predictive distribution $\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}})$ equals the *mutual information* between $\tilde{\mathbf{x}}^{\mathbf{m}}$ and θ given $\mathbf{x}^{\mathbf{n}}$:

$$\begin{aligned}\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}}) &\equiv \mathbb{E}^{\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x}^{\mathbf{n}}}\left[\mathbf{D}_{KL}(\pi_{\mathbb{P}}(\theta'|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta'|\mathbf{x}^{\mathbf{n}}))\right] \\ &= \int\int \mathbf{D}_{KL}(\pi_{\mathbb{P}}(\theta'|\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta'|\mathbf{x}^{\mathbf{n}}))\,\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)\,\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}})\,\mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\,\mathrm{d}\theta.\end{aligned} \tag{3.31}$$

Like the relative entropy, the mutual information is always positive. It is easy to check that $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}}) = 0$ when $m = 0$. Under the assumption that the prior distribution is nonsingular and the parameters in the likelihood function are well identified, and additional general regularity conditions, $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}})$ is monotonically increasing in m and converges to infinity as m increases. These properties ensure that we can find an extra sample size m that equates (approximately, due to the fact that m is an integer) $\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}})||\pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}))$ with $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}};\theta|\mathbf{x}^{\mathbf{n}})$.

Definition 7 (Effective-Sample Size Information Measure)**.** *For a feature function vector $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_f}$, we define the effective-sample measure of the informativeness of the cross-equation restrictions as*

$$\varrho^{f}_{KL}(\mathbf{x}^{\mathbf{n}},\mathbf{y}^{\mathbf{n}}) = \frac{n+m^*_f}{n}, \tag{3.32}$$

where m_f^* satisfies

$$\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}_f^*}; f(\theta)|\mathbf{x}^{\mathbf{n}}) \leq \mathbf{D}_{KL}(\pi_{\mathbb{Q}}(f(\theta)|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathbb{P}}(f(\theta)|\mathbf{x}^{\mathbf{n}})) < \mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}_f^*+1}; f(\theta)|\mathbf{x}^{\mathbf{n}}), \quad (3.33)$$

with $\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(f(\theta)|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathbb{P}}(f(\theta)|\mathbf{x}^{\mathbf{n}}))$ being the relative entropy between the constrained and unconstrained posteriors of $f(\theta)$ and $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; f(\theta)|\mathbf{x}^{\mathbf{n}})$ being the conditional mutual information between the additional sample of data $\mathbf{x}^{\mathbf{m}}$ and the transformed parameter $f(\theta)$ given the existing sample of data $\mathbf{x}^{\mathbf{n}}$.

For scalar-valued feature functions, there exists a direct connection between our asymptotic fragility measure and the relative-entropy based informativeness measure. We stablish the result in Theorem 2 below using the approximation results summarized in Propositions 12 and Proposition 13.

Proposition 12 (Relative Entropy). *Consider a feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \partial f(\theta_0)/\partial\theta^T$. Let the MLE from the baseline model $\mathbb{P}_\theta$ be $\widehat{\theta}^{\mathbb{P}}$, and the GMM estimator from the structural model $\mathbb{Q}_\theta$ be $\widehat{\theta}^{\mathbb{Q}}$. Under the regularity conditions stated in Appendix A.3.3,*

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(f(\theta)|\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})||\pi_{\mathbb{P}}(f(\theta)|\mathbf{x}^{\mathbf{n}})) - \frac{n}{2\mathbf{v}'\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}(f(\widehat{\theta}^{\mathbb{P}}) - f(\widehat{\theta}^{\mathbb{Q}}))^2$$
$$\to \frac{1}{2}\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} + \frac{1}{2}\frac{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} - 1/2, \quad (3.34)$$

where convergence is in probability under $\mathbb{Q}_{0,n}$.

Proof. See Appendix A.3.3. ☐

Proposition 12 generalizes the results in Lin, Pittman, and Clarke (2007) (Theorem 3) in two important aspects. First, our results extend the traditional results to the more general GMM framework, building on the Bayesian GMM formulation (see e.g. Kim, 2002; Chernozhukov and Hong, 2003). Second, our results allow for general weak dependence among the observations, which makes our results applicable to time series models in finance and economics.

Proposition 13 (Mutual Information)**.** *Under the assumptions in Subsection A.3.3, suppose that $m/n \to \varsigma \in (0,\infty)$ as both m and n approach infinity,*

$$\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; f(\theta)|\mathbf{x}^{\mathbf{n}}) - \frac{1}{2}\ln\left(\frac{m+n}{n}\right) \to 0, \tag{3.35}$$

where convergence is in probability under $\mathbb{Q}_{0,n}$.[11]

Proof. See Appendix A.3.3. □

The following theorem establishes asymptotic equivalence between the fragility measure $\varrho_a^{\mathbf{v}}(\theta_0)$ and the effective sample size information measure $\varrho_{KL}^{f}(\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}})$.

Theorem 2. *Consider a feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \partial f(\theta_0)/\partial\theta^T$. Under the regularity conditions stated in Appendix A.3.3, it must hold that*

$$\ln \varrho_{KL}^{f}(\mathbf{x}^{\mathbf{n}}, \mathbf{y}^{\mathbf{n}}) \rightsquigarrow \ln\left[\varrho_a^{\mathbf{v}}(\theta_0)\right] + \left[1 - \varrho_a^{\mathbf{v}}(\theta_0)^{-1}\right](\chi_1^2 - 1), \tag{3.36}$$

where χ_1^2 is a chi-square random variable with degrees of freedom 1 and $\rightsquigarrow$ denotes convergence in distribution.

Proof. See Appendix A.3.3. □

3.3 Applications

In this section, we implement our fragility measure in the context of two widely studied asset pricing models. The first example is a rare disaster model, for which we compute the asymptotic fragility measure analytically. The second example is a long-run risk model. We use this example to illustrate how one can diagnose the sources of fragility in a more complex model.

[11] There exist related approximation results for mutual information $\mathbf{I}(\tilde{\mathbf{x}}^{\mathbf{m}}; \theta|\mathbf{x}^{\mathbf{n}})$, which consider large m while holding the observed sample size n fixed. For more details, see Clarke and Barron (1990, 1994) and references therein. See also the case of non-identically distributed observations by Polson (1992), among others. Our results differ in that we allow both m and n to grow. Ours is a technically nontrivial extension of the existing results because, as n increases, the posterior distribution of θ given $\mathbf{x}^{\mathbf{n}}$ must be updated according to the Bayes rule.

3.3.1 Disaster Risk Model

Rare economic disasters are a natural source of "dark matter" in asset pricing models. It is difficult to evaluate the likelihood of rare events statistically. Yet, agents' aversion to large disasters can have an economically large effect on asset prices.[12]

We consider a disaster risk model similar to Barro (2006b). The structural model describes the log growth rate of aggregate consumption g_t and the excess log return on the market portfolio r_t, which jointly follow the process

$$\begin{pmatrix} g_t \\ r_t \end{pmatrix} = (1 - z_t)u_t - z_t \begin{pmatrix} v_t \\ bv_t + \epsilon_t \end{pmatrix}, \tag{3.37}$$

where z_t has an IID Bernoulli distribution, is independent of the other random variables, and takes the value of 1 and 0 with probability p and $1 - p$. We assume that realizations of z_t are observable. Outside of disasters ($z_t = 0$), g_t and r_t are jointly normal with mean (μ, η). Their covariance in the non-disaster state is

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}.$$

In a disaster state ($z_t = 1$), the log of decline in consumption v_t follows a truncated exponential distribution, $v_t \sim 1_{\{v \geq \underline{v}\}} \lambda e^{-\lambda(v-\underline{v})}$, with the lower bound for disaster size equal to $\underline{v}$. Conditional on a disaster, the average disaster size is $\underline{v} + 1/\lambda$. The excess log return in a disaster is linked to the decline in consumption with a leverage factor b. In addition, we add an independent shock $\epsilon_t \sim N(0, \nu^2)$ to r_t so that r_t and g_t are imperfectly correlated in a disaster state.

The representative agent has a separable, constant relative risk aversion utility function $\sum_0^\infty \delta^t c_t^{1-\gamma}/(1-\gamma)$, where $\gamma > 0$ is the coefficient of relative risk aversion. The equity premium $\eta = \mathbb{E}[r_t]$ can be derived from the consumption Euler equation, which

[12] See the early work by Rietz (1988), and recent developments by Barro (2006b), Gabaix (2012), Martin (2012), Wachter (2013), and Collin-Dufresne, Johannes, and Lochstoer (2013), among others.

is approximately[13]

$$\eta \approx \gamma\rho\sigma\tau - \frac{\tau^2}{2} + e^{\gamma\mu - \frac{\gamma^2\sigma^2}{2}}\Delta(\lambda)\frac{p}{1-p}, \tag{3.38}$$

where

$$\Delta(\lambda) = \lambda\left(\frac{e^{\gamma\underline{v}}}{\lambda-\gamma} - \frac{e^{\frac{\nu^2}{2}+(\gamma-b)\underline{v}}}{\lambda+b-\gamma}\right). \tag{3.39}$$

Equation (3.38) provides a cross-equation restriction among the processes of consumption growth g_t, the disaster state z_t, and the excess log return of the market portfolio r_t. The first two terms on the right hand side give the market risk premium due to Gaussian consumption shocks. The third term is the disaster risk premium. We need $\lambda > \gamma$ for the risk premium to be finite, which sets an upper bound for the average disaster size and dictates how heavy the tail of the disaster size distribution can be.

The fact that the equity premium η explodes as λ approaches γ is a crucial feature for our analysis. Even when we consider extremely rare disasters (very small p), we can still generate an arbitrarily large risk premium η by making the average disaster size sufficiently large (lowering λ towards γ). Extremely rare and large disasters are difficult to rule out based on standard statistical tests. Below we illustrate how our fragility measure can detect fragility in models with such features.

Asymptotic fragility measure

Equations (3.37) and (3.38) together specify the full structural model $\mathbb{Q}$. TWe set the baseline model $\mathbb{P}$ to be the model for consumption growth g_t. To focus our discussion on the rare disasters, we treat the parameters μ, σ, $\underline{v}$, τ, ρ, b and ν as known. This simplifying assumption allows us to obtain a simple closed-form expression for the asymptotic fragility measure. Thus, $\theta = (p, \lambda)$, while the structural parameters of $\mathbb{Q}$ include $\psi = \gamma$.

[13] The approximation we make here is $e^{\eta + \frac{\tau^2}{2} - \gamma\rho\sigma\tau} \approx 1 + \eta + \frac{\tau^2}{2} - \gamma\rho\sigma\tau$.

The asymptotic fragility measure is (see Appendix A.1.1 for details)

$$\varrho_a(p,\lambda) = 2 + \frac{p\Delta(\lambda)^2 + p(1-p)\lambda^2\dot{\Delta}(\lambda)^2}{(1-\rho^2)\tau^2(1-p)^2}e^{2\gamma\mu-\gamma^2\sigma^2}, \tag{3.40}$$

where $\dot{\Delta}(\lambda)$ is the first derivative of $\Delta(\lambda)$,

$$\dot{\Delta}(\lambda) = -\frac{e^{\gamma\underline{v}}\gamma}{(\lambda-\gamma)^2} + \frac{e^{(\gamma-b)\underline{v}}(\gamma-b)}{(\lambda-\gamma+b)^2}e^{\nu^2/2}. \tag{3.41}$$

The one-dimensional worst-case asymptotic fragility measure is $\varrho_a^1(p,\lambda) = \varrho_a(p,\lambda) - 1$.

As Equation (3.38) shows, $\Delta(\lambda)$ and $\dot{\Delta}(\lambda)$ are related to the sensitivity of the equity premium to the disaster probability p and disaster size parameter λ, respectively. When λ approaches γ, both $\Delta(\lambda)$ and $\dot{\Delta}(\lambda)$ approach infinity. Thus, disaster risk models with high average disaster size are fragile according to our measure.

Quantitative analysis

In our quantitative analysis, we use annual real per-capita consumption growth (nondurables and services) from the NIPA and returns on the CRSP value-weighted market portfolio for the period of 1929 to 2011. We fix the parameters μ, σ, ν, τ and ρ at the values of the corresponding moments of the empirical distribution of consumption growth and excess stock returns: $\mu = 1.87\%$, $\sigma = 1.95\%$, $\tau = 19.14\%$, $\nu = 34.89\%$ and $\rho = 59.36\%$. The lower bound for disaster size is $\underline{v} = 7\%$. The leverage parameter b is 3. In Figure 3-1, we plot the 95% and 99% confidence regions for (p,λ) based on the baseline model.

The 95% confidence region for (p,λ) is quite wide. For low values of the disaster probability p, the baseline model has little power to reject models with a wide range of average disaster size values (λ). Figure 3-1 also shows the equity premium isoquants for different levels of relative risk aversion: lines with the combinations of p and λ required to match the average equity premium of 5.89% for a given value of γ. The

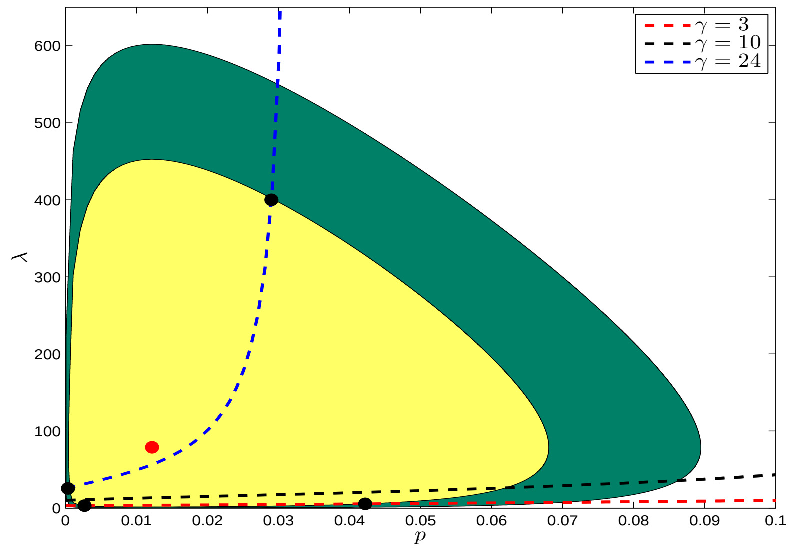

Figure 3-1: The 95% and 99% confidence regions of (p, λ) for the unconstrained model and the equity premium isoquants implied by the asset pricing constraint for $\gamma = 3, 10, 24$. The maximum likelihood estimates are $(\widehat{p}_{MLE}, \widehat{\lambda}_{MLE}) = (0.0122, 78.7922)$.

fact that these isoquants all intersect with the 95% confidence region implies that even for low risk aversion ($\gamma = 3$), there exist many combinations of (p, λ) that not only match the observed equity premium, but also are "consistent with the macro data" in a sense that they cannot be rejected by the macro data based on standard statistical tests. In the remainder of this section, we refer to a calibration of (p, λ) that is within the 95% confidence region as an "acceptable calibration."[14]

While it is difficult to distinguish among a wide range of calibrations using standard statistical tools based on the macro data, these calibrated models differ significantly based on our fragility measures. We focus on four alternative calibrations, as denoted by the four points located at the intersections of the equity premium isoquants ($\gamma = 3, 24$) and the boundary of the 95% confidence region in Figure 3-1. For $\gamma = 3$, the two points are ($p = 4.22\%, \lambda = 5.46$) and ($p = 0.27\%, \lambda = 3.14$). For $\gamma = 24$, the two points are ($p = 2.9\%, \lambda = 396.7$) and ($p = 0.037\%, \lambda = 25.49$).

[14] Julliard and Ghosh (2012) estimate the consumption Euler equation using the empirical likelihood method and show that the model requires a high level of relative risk aversion to match the equity premium. Their empirical likelihood criterion rules out any large disasters that have not occurred in the historical sample, hence requiring the model to generate high equity premium using moderate disasters.

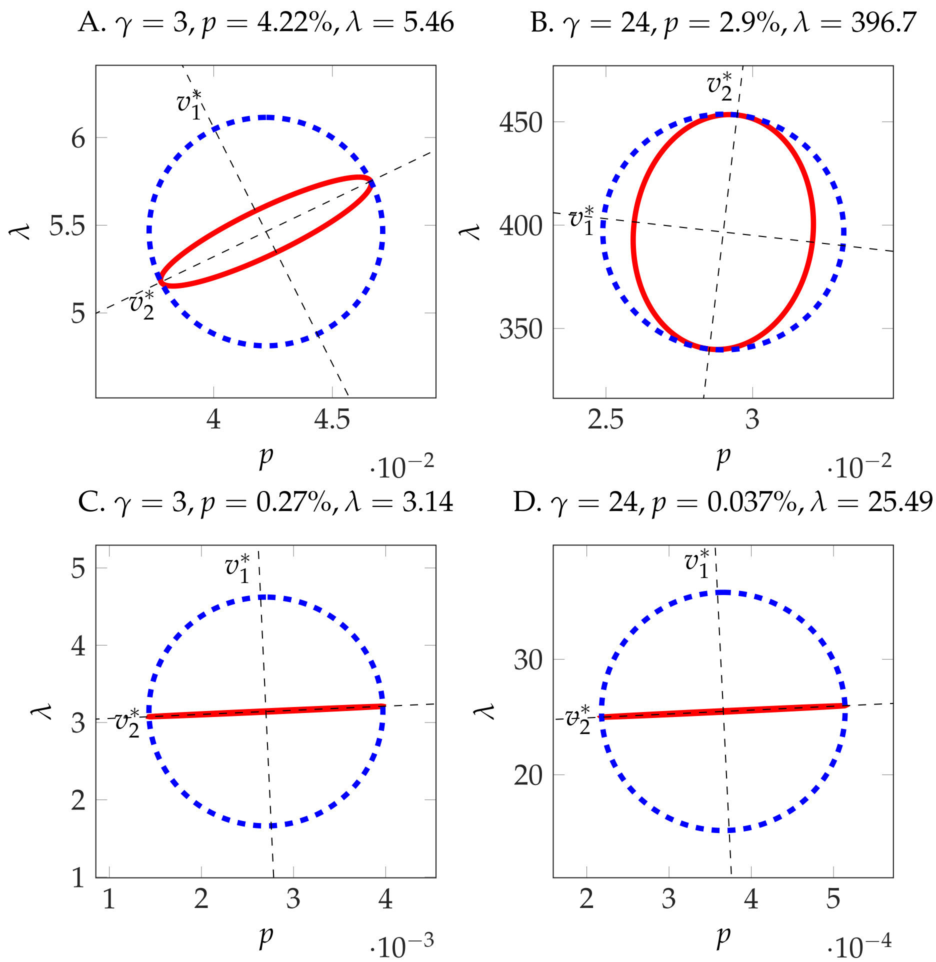

Figure 3-2: 95% confidence regions for the asymptotic distribution of the MLEs for four "acceptable calibrations." In Panels A through D, the asymptotic fragility measures are $\varrho_a = 25.47, 2.81, 3.7 \times 10^4$, and 2.9×10^4 respectively.

With only two parameters in θ, we illustrate the worst-case asymptotic fragility measure by plotting the asymptotic confidence regions for (p, λ) in the baseline model and the structural model, as determined by the respective information matrices $\mathbf{I}_{\mathbb{P}}(\theta)$ and $\mathbf{I}_{\mathbb{Q}}(\theta)$.[15] In each panel of Figure 3-2, the largest dash-line circle is the 95% confidence region for (p, λ) under the baseline model. The smaller solid-line ellipse is the 95% confidence region for (p, λ) under the structural model. The reason that the confidence region under the structural model is smaller than that under the baseline model

[15] In fact, we use all the score functions of likelihoods of $\mathbb{P}$ and $\mathbb{Q}$ to construct the moments, so the optimal GMM estimation is asymptotically equivalent to the MLE in this case.

is that the GMM moments in the structural model contain both the score function of the likelihood for the baseline model and the cross-equation restrictions. In this example, the two confidence regions coincide[16] in the direction of v_2^* and differ the most in the direction of v_1^*. Moreover, with enough extra data, the confidence region for the unconstrained estimator can be made small enough to reside within the confidence region of the constrained estimator.

In Panel A, with $\gamma = 3, p = 4.22\%, \lambda = 5.46$, $\varrho_a(p,\lambda) = 25.47$ and $\varrho_a^1(p,\lambda) = 24.47$. This means that under the baseline model, we need to increase the amount of consumption data by a factor of 24.47 to match or exceed the precision in estimation of any linear combination of p and λ afforded by the equity premium constraint. Panels C and D of Figure 3-2 correspond to the calibrations with "extra rare and large disasters." For $\gamma = 3$ and 24, $\varrho_a^1(p,\lambda)$ rises to 3.7×10^4 and 2.9×10^4, respectively. If we raise γ to 24 while changing the annual disaster probability to 2.9% and lowering the average disaster size to 7.25%, $\varrho_a^1(p,\lambda)$ drops to 1.81. The reason is that by raising the risk aversion coefficient we are able to reduce the average disaster size.

So far, we have been examining the fragility of a specific calibrated structural model. We can also assess the fragility of a general class of models by plotting the distribution of $\varrho_a(\theta)$ based on a particular distribution of θ. For example, if econometricians are interested in fragility of a class of disaster risk models where the risk aversion γ is fixed at a given level and the uncertainty of the parameters in θ is explicitly taken into account, the posterior for (p,λ) under the structural model (i.e., constrained posterior distribution) denoted by $\pi_Q(\theta|\mathbf{g^n}, \mathbf{z^n}, \mathbf{r^n}; \gamma)$ is proposed to be used as the distribution of θ. Since the constrained posterior updates the prior $\pi(\theta)$ based on information from the data and the asset pricing constraint, it can be viewed as summarizing our knowledge of the distribution of θ assuming the model constraint is valid.

We implement this idea in Figure 3-3. For each value of γ, the boxplot shows the $1, 25, 50, 75$, and 99-th percentile of the distribution of $\varrho_a(\theta)$ based on the posterior

[16]This is not true in general. When localized, the cross-equation restriction from the equity premium in this model is a linear constraint. Thus, the parameter estimates are not affected along the direction of the constraint.

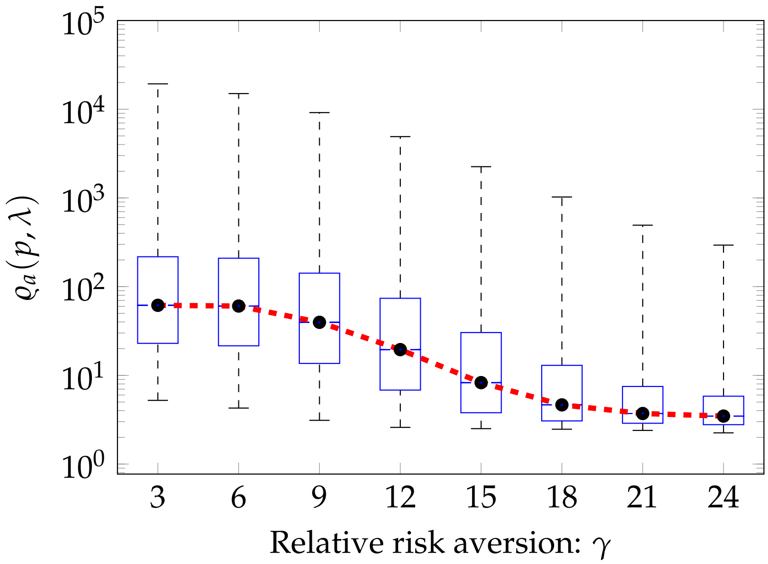

Figure 3-3: Distribution of the asymptotic fragility measure $\varrho_a(p, \lambda)$ for different levels of risk aversion. For each γ, the boxplot shows the $1, 25, 50, 75$, and 99-th percentile of the distribution of $\varrho_a(p, \lambda)$ based on the constrained posterior for (p, λ).

$\pi_{\mathbb{Q}}(\theta|\mathbf{g^n}, \mathbf{z^n}, \mathbf{r^n}; \gamma)$. The asymptotic fragility measures are higher when the levels of risk aversion are low. For example, for $\gamma = 3$, the $25, 50$, and 75-th percentile of the distribution of $\varrho_a(p, \lambda)$ are 23.0, 61.6, and 217.4, respectively. This is because a small value of γ forces the constrained posterior for θ to place more weight on "extra rare and large" disasters, which imposes particularly strong restrictions on the parameters (p, λ). As γ rises, the mass of the constrained posterior shifts towards smaller disasters, which imply lower information ratios. For $\gamma = 24$, the 25, 50, and 75-th percentile of the distribution of $\varrho_a(p, \lambda)$ drop to 2.8, 3.5, and 5.8, respectively.

Uncertainty about γ

So far in this example we have computed the fragility measure for θ conditional on specific values of the structural parameters ψ (specifically, the risk aversion coefficient γ). The econometrician could be interested in assessing the fragility of a more general class of models, which not only takes into account his uncertainty about the baseline model parameters θ, but also the uncertainty about the structural parameters ψ. We

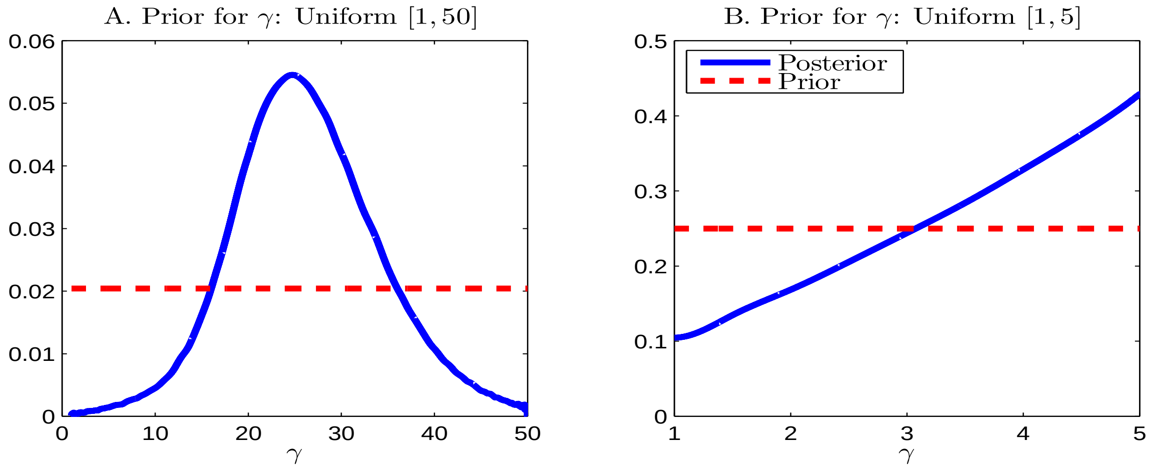

Figure 3-4: Constrained posterior distribution for risk-aversion coefficient γ. Panel A displays the constrained posterior and prior densities for γ when the prior is uniform on $[1, 5]$. Panel B displays the results when the prior is uniform on [1,50].

now illustrate how to deal with such uncertainty in a Bayesian framework.

We consider two ways to set up the econometrician's prior belief on γ. The first is an "uninformative prior" that allows for a wide range of possible values on γ. Specifically, we assume γ is uniformly distributed between 1 and 50. Alternatively, the econometrician might prefer models with low levels of risk aversion. For example, Barro (2006b) states that the usual view in the finance literature is that γ is in the range of 2 to 5. Thus, we also consider an "informative prior" on γ that is uniform between 1 and 5. In addition to the prior on γ, we adopt the Jeffreys priors for p, λ (see Appendix A.1.2 for details). Using the macro and return data, we then obtain the constrained posterior $\pi_{\mathbb{Q}}(p, \lambda, \gamma | \mathbf{g^n}, \mathbf{z^n}, \mathbf{r^n})$.

We plot the constrained posterior marginal density for γ from the two different priors in Figure 3-4. In the case of "uninformative prior" (Panel A), the median value for γ in the constrained posterior is 25.8, and the probability that γ is less than 10 is 3.9%. The posterior is clearly very different from the prior, suggesting that asset prices convey significant information about γ to the econometrician in this case. In contrast, with an informative prior on γ (Panel B), the constrained posterior on γ is concentrated on low values and is relatively close to the prior. The median value for γ

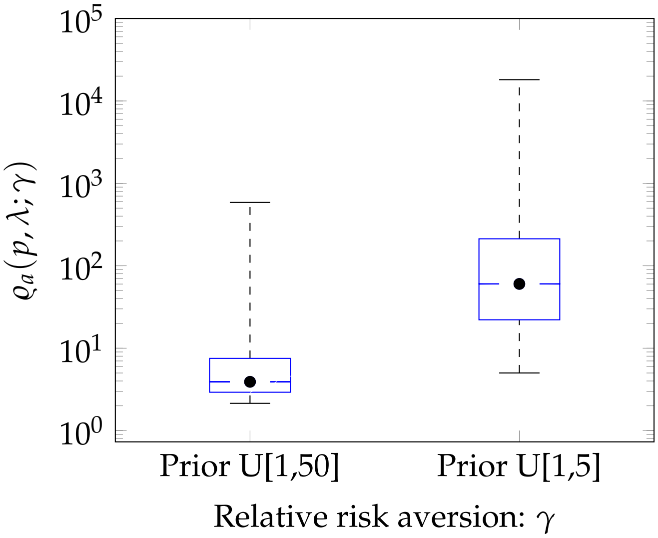

Figure 3-5: Distribution of the asymptotic fragility measure $\varrho_a(p,\lambda;\gamma)$ based on two different priors for γ. For each prior, the boxplot shows the $1, 25, 50, 75$, and 99-th percentile of the distribution of $\varrho_a(p,\lambda;\gamma)$ based on the constrained posterior for (p,λ,γ).

in the constrained posterior is 3.60 in this case.

Next, we plot in Figure 3-5 the distributions of the asymptotic fragility measure $\varrho_a(p,\lambda;\gamma)$ based on the two constrained posteriors $\pi_{\mathbb{Q}}(p,\lambda,\gamma|\mathbf{g^n},\mathbf{z^n},\mathbf{r^n})$. In the case with an uninformative prior for γ, the $25, 50$, and 75-th percentile of the distribution of $\varrho_a(p,\lambda;\gamma)$ are 2.9, 3.9, and 7.5, respectively. In contrast, the asymptotic fragility measure is significantly higher in the case of an informative prior that favors low values for γ. The $25, 50$, and 75-th percentile of $\varrho_a(p,\lambda;\gamma)$ are 22.1, 60.2, and 212.5, respectively. This finding is consistent with the result in Figure 3-3. Under the high values for γ, we do not need large and rare disasters to match the observed equity premium, which reduces the sensitivity of the cross-equation restrictions, hence lowering the asymptotic fragility measure.

3.3.2 Long-run risk model

In the second example, we consider a long-run risk model similar to Bansal and Yaron (2004b) and Bansal, Kiku, and Yaron (2012b). In the model, the representative agent has recursive preferences as in Epstein and Zin (1989a) and Weil (1989) and maximizes his lifetime utility,

$$V_t = \left[(1-\delta) C_t^{1-1/\varpi} + \delta \left(\mathbb{E}_t \left[V_{t+1}^{1-\gamma} \right] \right)^{\frac{1-1/\varpi}{1-\gamma}} \right]^{\frac{1}{1-1/\varpi}}, \tag{3.42}$$

where C_t is consumption at time t, δ is the rate of time preference, γ is the coefficient of risk aversion for timeless gambles, and ϖ is the elasticity of intertemporal substitution when there is perfect certainty.

The log growth rate of consumption Δc_t, the conditional mean of consumption growth x_t, and the conditional volatility of consumption growth σ_t follow the process

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \epsilon_{c,t+1} \tag{3.43a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_t \epsilon_{x,t+1} \tag{3.43b}$$

$$\sigma_{t+1}^2 = \overline{\sigma}^2 + \nu(\sigma_t^2 - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1} \tag{3.43c}$$

where the shocks $\epsilon_{c,t}, \epsilon_{x,t}$, and $\epsilon_{\sigma,t}$ are *i.i.d.* $N(0,1)$ and mutually independent.[17]

Next, the log dividend growth Δd_t follows the processes

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c} \sigma_t \epsilon_{c,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1} + \sigma_{d,u} \epsilon_{d,t+1}^u, \tag{3.44}$$

where the shocks $\epsilon_{d,t}$ and $\epsilon_{d,t}^u$ are *i.i.d.* $N(0,1)$ and mutually independent with the other shocks in (3.43a)-(3.43c). Compared to the dividend process in Bansal, Kiku, and Yaron (2012b), we have added an extra shock $\epsilon_{d,t}^u$ to avoid stochastic singularities.[18]

From the consumption Euler equation, one can derive a linear approximation of the stochastic discount factor,

$$m_{t+1} = \Gamma_0 + \Gamma_1 x_t + \Gamma_2 \sigma_t^2 - \lambda_c \sigma_t \epsilon_{c,t+1} - \lambda_x \varphi_x \sigma_t \epsilon_{x,t+1} - \lambda_\sigma \sigma_w \epsilon_{\sigma,t+1}. \tag{3.45}$$

[17] The volatility process (3.43c) potentially allows for negative values of σ_t^2. Following the literature, we impose a small positive lower bound $\underline{\sigma}^2$ for σ_t^2 in simulations.

[18] For example, without the additional shock, $r_{m,t+1}^e$ is a deterministic function of $\Delta c_{t+1}, x_{t+1}, \Delta d_{t+1}$ and σ_{t+1}^2 conditional on the information up to time t, which poses a stochastic singularity.

Table 3.1: Benchmark Calibration for the Long-Run Risk Model

Preferences	δ	γ	ϖ			
	0.9989	10	1.5			
Consumption	μ_c	ρ	φ_x	$\overline{\sigma}$	ν	σ_w
	0.0015	0.975	0.038	0.0072	0.999	$2.8e-6$
Dividends	μ_d	ϕ_d	$\varphi_{d,c}$	$\varphi_{d,d}$	$\sigma_{d,u}$	
	0.0015	2.5	2.6	5.96	0.005	

The formulae for the coefficients Γ_0, Γ_1, Γ_2, λ_c, λ_x, and λ_σ are given in Appendix A.2. Moreover, the equilibrium excess (log) return follows

$$r^e_{m,t+1} = \mu^e_{r,t} + \beta_c \sigma_t \epsilon_{c,t+1} + \beta_x \sigma_t \epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} + \varphi_{d,d} \sigma_t \epsilon_{d,t+1} + \sigma_{d,u} \varepsilon^u_{d,t+1}, \quad (3.46)$$

where the conditional average (log) excess return is

$$\mu^e_{r,t} = \lambda_c \beta_c \sigma_t^2 + \lambda_x \beta_x \varphi_x \sigma_t^2 + \lambda_\sigma \beta_\sigma \sigma_w^2 - \frac{1}{2} \sigma^2_{r_m,t}. \quad (3.47)$$

The expressions for β_c, β_x, β_σ, and $\sigma_{r_m,t}$ are given in Appendix A.2.

Quantitative Analysis

We choose the model of consumption as the baseline model $\mathbb{P}$. We assume that the econometrician observes the process for consumption, the latent variables x_t and σ_t^2, and the process for asset returns. We make the latent variables observable to be consistent with the postulated process for asset returns, which is derived assuming that these variables are observable.

Accordingly, $\theta = (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w)$. By measuring the fragility of the long-run risk model relative to this particular benchmark, we can interpret the fragility measure as quantifying the additional information that asset pricing restrictions provide for the consumption dynamics (in particular, the long-run risk components) relative to information contained in consumption data.

Table 3.2: Fragility Measures for the Long-Run Risk Models

	ϱ_a	ϱ_a^1	$\varrho_a^{\mathbf{v}}$					
			μ_c	ρ	φ_x	$\overline{\sigma}^2$	ν	σ_w
Benchmark Model	2016	2000	1.013	4.591	1.005	19.809	6.565	1.011
$\nu = 0.98, \gamma = 27$	19	12	1.133	5.714	1.015	1.115	3.688	1.001

Note: The direction corresponding to the worst-case 1-dimensional fragility measure ϱ_a^1 is given by $v_1^* = [0.0021, -0.0004, -0.0003, 0.1286, -0.0053, 0.9917]$.

The benchmark calibration of the model follows Bansal, Kiku, and Yaron (2012b) and is summarized in Table 3.1. As Bansal, Kiku, and Yaron (2012b) (Table 2, p. 194) show, the simulated moments match the set of key asset pricing moments in the data reasonably well.

The first row of Table 3.2 reports the fragility measures for the benchmark calibration. The asymptotic fragility measure is $\varrho_a = 2016$, indicating a high level of model fragility. The worst-case 1-dimensional asymptotic fragility measure is also high, $\varrho_a^1 = 2000$, which implies that the sample size needs to be 2000 times longer in order for the baseline model estimator to match the precision of the estimator for the full structural model in all 1-D directions.

The large size of ϱ_a^1 implies that the model under the benchmark calibration is highly sensitive to perturbations in the parameters in a single direction, as identified by v_1^* (i.e. the worst direction). However, this does not mean that one can discover the full scope of the fragility issue by examining individual parameters one at a time. We demonstrate this point by computing the individual parameter-based fragility measure $\varrho_a^{\mathbf{v}}$, where $\mathbf{v}$ is the appropriate standard basis vector $\mathbf{e}_i$ whose i-th element is one and other elements are zeros. As Table 3.2 shows, the fragility measures for all the individual parameters are relatively small. While the measure is somewhat larger in magnitude for ρ (the persistence of conditional mean consumption growth), $\overline{\sigma}^2$ (long-run variance of consumption growth), and ν (the persistence of conditional variance of consumption growth), all of the univariate measures are much smaller than ϱ_a and

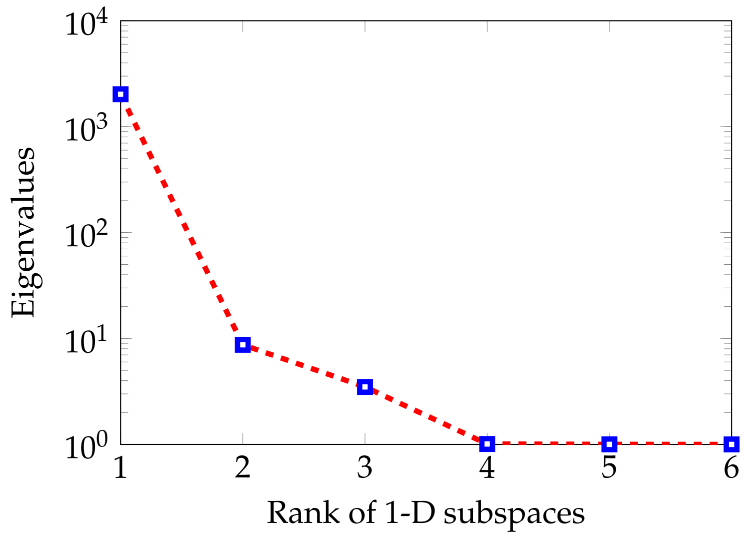

Figure 3-6: Eigenvalues for the matrix $\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}$ under the benchmark calibration.

ϱ_a^1. Had we focused only on the sensitivity of the model's properties to individual parameters in θ, we would have missed the high fragility measure for the full model.

Diagnosing the sources of fragility Besides measuring the fragility of the model, our asymptotic results have provided a set of tools to diagnose the sources of fragility. First, the rankings of the eigenvalues of $\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{\frac{1}{2}}$ are informative. Each eigenvalue denotes the marginal contribution of a 1-dimensional subspace to the overall fragility measure (see Definition 6 and Proposition 9). As Figure 3-6 shows, there are large differences between the eigenvalues. Model fragility along the worst direction in 1-dimensional subspaces, as captured by the leading eigenvalue, is 2000, which accounts for over 99% of the total fragility. This result means that one can dramatically reduce the dimensionality (from 6 to 1) when analyzing the fragility of this model.

Second, the worst-case direction (i.e., the worst-case 1-dimensional subspace) is v_1^*. Knowing that the majority of the model fragility is concentrated in this direction, we can conveniently search for the fragile moments in the model by examining which moments are the most sensitive to the change in θ along the direction of v_1^*. For illustration, we focus on four moments from the long-run risk model, the risk loading and

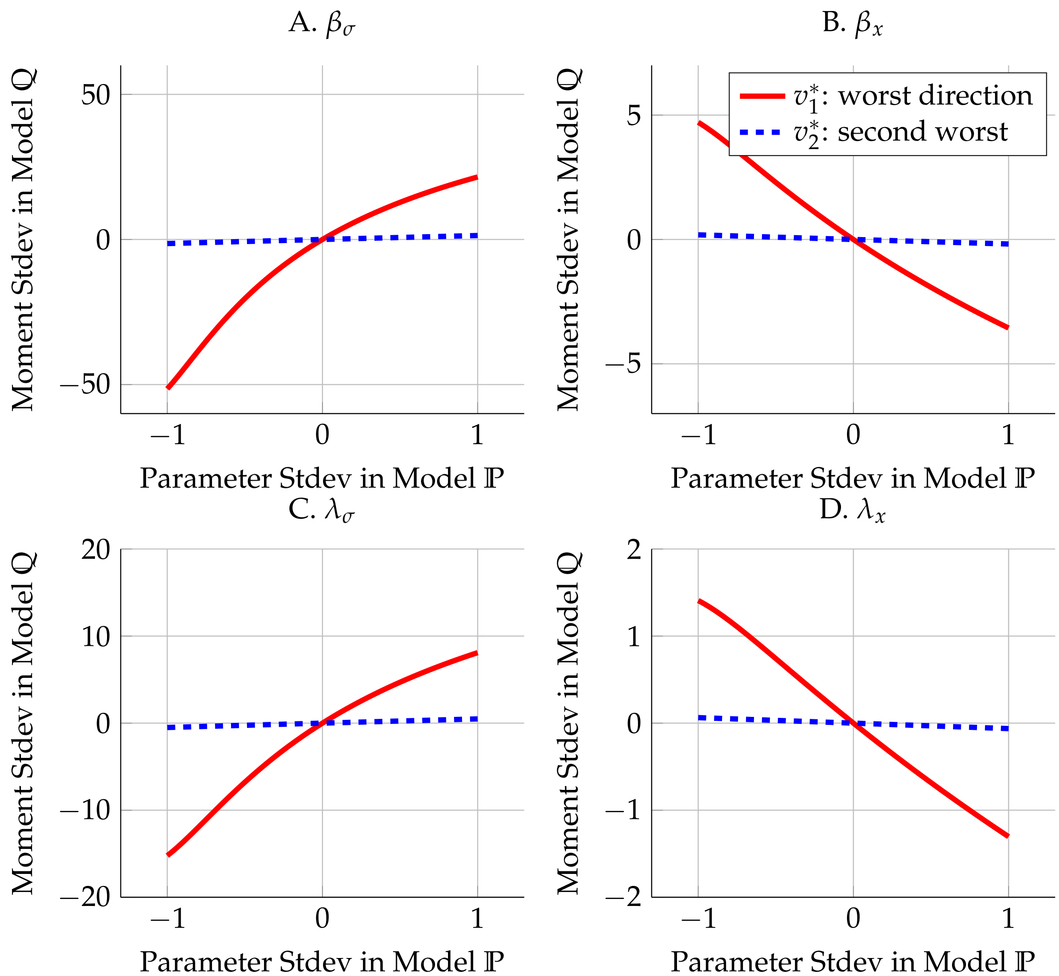

Figure 3-7: Sensitivity of return betas and risk prices with respect to the perturbation along the worst-case direction in the benchmark calibration.

price of risk for volatility shocks ($\beta_\sigma \sigma_w, \lambda_\sigma \sigma_w$), and for growth shocks ($\beta_x \sigma_t, \lambda_x \varphi_x \sigma_t$). The conditional market excess return depends crucially on these moments (see Equation (3.47)).

In Figure 3-7, we plot the sensitivities of $\beta_\sigma, \lambda_\sigma, \beta_x$ and λ_x with respect to perturbations of θ along the worst direction v_1^* (solid line) and compare them to the sensitivities of the same set of moments to perturbations of θ along the second-worst direction v_2^* (dash line). We measure the size of a perturbation of θ relative to the standard deviation of θ in the baseline model $\mathbb{P}$. We measure sensitivity of a moment as the change in the moment normalized by the moment's standard deviation in the structural model

Q.

The risk loading and the price of risk for volatility shocks are both highly sensitive to changes in θ along the direction of v_1^*, while the corresponding sensitivities to changes in θ along the direction of v_2^* are all very low in comparison. For example, a one standard deviation change in θ along the direction of v_1^* can lead to a 51-standard deviation change in β_σ under the structural model. Thus, an important source of fragility of the long-run risk model based on the benchmark calibration is in the risk exposure of the market portfolio to volatility shocks. If the true value of θ is slightly different from the benchmark calibration along the direction v_1^*, this version of the long-run risk model will perform poorly at explaining the relation between asset returns and volatility shocks out of sample.

Finally, we can further trace the sources of fragility by examining how λ_σ and β_σ are determined (see (A.23) and (A.43)). For example, the fact that the persistence parameter for the conditional variance of consumption growth, ν, is close to 1, makes both β_σ and λ_σ sensitive to changes in θ. This motivates us to consider an alternative calibration with a smaller value for ν. Specifically, we change ν from 0.999 to 0.98, and simultaneously raise the coefficient of relative risk aversion γ from 10 to 27 in order to match the unconditional equity premium as in the benchmark calibration. The rest of the parameters are unchanged. This alternative calibration produces asset pricing moments largely similar to those in the benchmark calibration. However, based on our fragility measures, the alternative calibration is much less fragile compared to the benchmark calibration. As Table 3.2 shows, under the alternative calibration, ϱ_a drops from 2016 to 19, and ϱ_a^1 drops from 2000 to 12.

3.4 Conclusion

In this paper, we propose a new measure of model fragility by quantifying a model's tendency of in-sample over-fitting. We formally connect the fragility of structural models to the informativeness of the cross-equation restrictions imposed on the parameters.

We also provide a tractable asymptotic approximation to the fragility measure, which helps with diagnosing sources of model fragility.

Our methodology has a broad range of applications. In addition to the examples of applications in asset pricing that we consider in this paper, our measure can be used to assess robustness of structural models in many other areas of economics, such as structural IO and structural corporate finance.

Chapter 4

The Volatility of International Capital Flows and Foreign Assets

4.1 Key Facts on U.S. International Capital Flows and Current Accounts

In this section, we review key facts on the U.S. current account and net foreign assets and then turn to the U.S. international capital flows.

4.1.1 Current Accounts and Net Foreign Assets

The current account is the sum of the trade balance (exports minus imports), the net dividend payments, and the net interest payments. In all but one of the last thirty years, the U.S. current account has been consistently negative, mostly because the U.S. imports more than it exports. The sum of the past cumulated current accounts is now close to 60% of GDP.

This alarming level contrasts with the net foreign asset position of the U.S. Consistent with a stream of negative current accounts, the net foreign asset position of the U.S. declined, reaching -20% of the U.S. GDP at the end of the sample. There is considerable uncertainty in the measure of the net foreign asset position. Yet, it appears much smaller than the cumulated past current accounts. As Gourinchas and Rey (2007, 2010, 2013) argue, this discrepancy suggests large valuation effects: the U.S. receives on average larger returns on their assets than they pay on their liabilities. While there is some uncertainty in the magnitude of the returns and their difference, it appears likely that the difference in returns at least partly compensates the deficit in the current account.

In this view, the sustainability of the current account relies on the ability of the U.S. to pocket large returns on its foreign investments. Such large returns in the past may have been unexpected and thus pure luck, or expected and thus reflecting differences in risk premia. As Gourinchas and Rey (2013) note, a difference in expected returns between U.S. assets and liabilities is consistent with the broad asset allocation of the country, since the U.S. is short domestic debt and long foreign equity. The U.S. may thus receive large expected returns on its levered equity investments, as a compensation for their risk, while paying low returns on its debt.

4.1.2 Equity and Bond Flows

The levered position of the U.S. economy has clear implications for the dynamics of its net foreign assets. In theory, the foreign asset positions can change either because their unit values change, a pure valuation effect, or because their quantities change, as a result of capital reallocation and thus international capital flows. In practice, a statistical gap exists between the changes in foreign assets on the one hand and the sum of the valuation effects and capital flows. Even after taking into account this statistical gap, a clear difference emerges between the dynamics of the U.S. foreign assets and liabilities.

According to the changes in U.S. equity and bond assets and liabilities over the last twenty years, three key patterns appear: (i) the volatility in foreign equity holdings is mostly due to valuation changes, not net capital flows; (ii) the volatility of foreign equity assets is much larger than the volatility of U.S equity liabilities; (iii) but the volatility of bond liabilities is mostly due to net capital flows, not valuation changes.

The last recession illustrates these patterns vividly: the value of foreign equity held by U.S. investors plummeted, and so did the value of the foreign equity holdings in the U.S. But the magnitudes are different: in the worst quarter of the crisis, the foreign investors lost $600 billions in U.S. equity wealth, while the U.S. investors lost $1 trillion in foreign equity wealth, amounting to a wealth transfer of $400 billions from the U.S. to the ROW in just one quarter. By comparison, bond values remain relatively stable. These patterns are intuitive — stocks tend to be more volatile than bonds — but they highlight the key difficulties in modeling international capital assets flows: the volatilities of holdings and flows are country- and asset-specific.

We turn now to a model that can potentially assess the volatility of equity and bond holdings and flows. The model features both expected and unexpected valuation changes, as well as portfolio rebalancing.

4.2 Model

In this section, we describe the model, starting with the endowment processes and the preferences, before turning to the market frictions.

4.2.1 Endowments

The model features two endowment economies. In each country, the endowment has a world and a country-specific component.

World Endowment The world endowment, denoted e_t, is described by a Lucas tree whose stochastic growth follows a time-homogeneous Markov process. In the absence of disasters, the growth rate of the global component is g_t, which takes values in a discrete set $\mathcal{S}_g$ and is governed by a Markov transition matrix Π_g. But growth switches from "normal" times, denoted $\xi_t = 0$, to "disaster" times, denoted $\xi_t = -1$, with some probability p_t. The disaster probability p_t follows a homogeneous Markov process with values in $\mathcal{S}_p$ and transition matrix Π_p. Once the economy is in its disaster state, it remains there the next period with probability p_d. The global endowment growth is thus:

$$\log \frac{e_{t+1}}{e_t} = g_{t+1} + \varphi_d \xi_{t+1},$$

where φ_d denotes the size of the world disaster. Three state variables therefore describe the world endowment: the growth rate in normal times, g_t, the occurrence of a disaster, ξ_t, and the probability of a disaster, p_t.

Country-specific Endowments The country-specific endowments, $e_{i,t}$, follow independent time-homogeneous Markov processes, denoted $a_{1,t}$ and $a_{2,t}$. Both take values in the set $\mathcal{S}_a$ and share the same transition matrix Π_a. Thus, the exogenous state of the economy is summarized by $s_t = (a_{1,t}, a_{2,t}, p_t, g_t, \xi_t)$. The total endowment in each country is:

$$\log e_{i,t} = \log e_t + a_{i,t}, \quad \text{for } i = 1, 2,$$

and the log endowment growth of country i is equal to:

$$\log \frac{e_{i,t+1}}{e_{i,t}} = \underbrace{[g_{t+1} + \varphi_d \xi_{t+1}]}_{\text{Global Component}} + \underbrace{\Delta a_{i,t+1}.}_{\text{Country-specific Component}}$$

Note that the model features permanent shocks to the level of endowments.[1] This feature is key as Alvarez and Jermann (2005) and Hansen and Scheinkman (2014) show, in a preference-free setting, that permanent shocks account for most of the variance of the pricing kernel. Lustig, Stathopoulos, and Verdelhan (2015), however, find that bond markets behave as if exchange rates are mostly driven by temporary components as if the permanent components were similar across countries. Our model features both a global permanent and two transitory components in the endowments. In other

[1] In many Markov economies used to study portfolio choices, such as Judd, Kubler, and Schmedders (2003), Kubler and Schmedders (2003), and Stepanchuk and Tsyrennikov (2015), endowments, dividends and labor income depend on the current exogenous shock alone, i.e. $e^i : \mathcal{S} \to \mathbb{R}_{++}$ is a time-invariant function. In our model, because the shocks to the world component e_t are permanent, the endowments, dividends and labor income depend on both the current shock and the world component e_t. Heaton and Lucas (1996) and Brumm, Grill, Kubler, and Schmedders (2013) also present models with stochastic growth and permanent shocks to study their asset pricing implications.

words, our economy is a Lucas-type economy with stochastic growth: the economy fluctuates around the stochastic trend governed by the world endowment e_t, whose sample path is driven by permanent shocks.

4.2.2 Preferences

In each country, there are two groups of agents: workers and investors. Both groups of agents in both countries maximize their utility over consumption. The utility function is recursive, following Kreps and Porteus (1978a) and Epstein and Zin (1989b). It is defined over a final consumption good that aggregates, with a constant elasticity of substitution (CES), the domestic and foreign goods. The value function of an agent in country i takes the following recursive form:

$$V_{i,t} = \left\{ C_{i,t}^{\frac{1-\gamma_i}{\theta_i}} + \beta \left[\mathbb{E}_t V_{i,t+1}^{1-\gamma_i} \right]^{\frac{1}{\theta_i}} \right\}^{\frac{\theta_i}{1-\gamma_i}}, \tag{4.1}$$

$$\text{where } C_{1,t} = \left[s \left(c_{1,t}^1 \right)^{\rho} + (1-s) \left(c_{1,t}^2 \right)^{\rho} \right]^{1/\rho} \tag{4.2}$$

$$\text{and } C_{2,t} = \left[(1-s) \left(c_{2,t}^1 \right)^{\rho} + s \left(c_{2,t}^2 \right)^{\rho} \right]^{1/\rho}. \tag{4.3}$$

The time discount factor is β, the risk aversion parameter is $\gamma_i \geq 0$, and the intertemporal elasticity of substitution (EIS) is $\psi_i \geq 0$. The parameter θ_i is defined by $\theta_i \equiv (1-\gamma_i)/(1-\frac{1}{\psi_i})$. The consumption home bias parameter s is between 0.5 and 1, and the elasticity of substitution between the domestic and foreign goods is $\epsilon = 1/[1-\rho]$. The aggregate consumption of an agent in country 1 is denoted $C_{1,t}$: it includes the consumption of goods produced in country 1, denoted $c_{1,t}^1$, as well as the consumption of goods produced in country 2, denoted $c_{1,t}^2$. More generally, $c_{i,t}^j$ denotes the consumption of good j by agent i at time t.

The CES consumption aggregators immediately imply the following price indices:

$$P_{1,t} = \left[s^{\epsilon} p_{1,t}^{1-\epsilon} + (1-s)^{\epsilon} p_{2,t}^{1-\epsilon} \right]^{1/(1-\epsilon)} \text{ and } P_{2,t} = \left[(1-s)^{\epsilon} p_{1,t}^{1-\epsilon} + s^{\epsilon} p_{2,t}^{1-\epsilon} \right]^{1/(1-\epsilon)}, \tag{4.4}$$

where $p_{1,t}$ and $p_{2,t}$ are the prices for goods produced by country 1 and country 2 respectively.[2] We normalize the price system assuming that:

$$p_{1,t} + p_{2,t} = 1.$$

Our calibration assumes a preference for an early resolution of uncertainty: for each agent $i \in \{1,2\}$, the EIS and risk-aversion parameters are above one ($\psi_i > 1$, $\gamma_i > 1$, and $\theta_i < 0$ for $i = 1,2$). After transformation, $U_i \equiv \dfrac{V_i^{1-\psi_i^{-1}}}{1-\psi_i^{-1}}$, the utility function can be re-written as:

$$U_{i,t} = \frac{C_{i,t}^{1-\psi_i^{-1}}}{1-\psi_i^{-1}} + \beta \mathbb{E}_t \left[U_{i,t+1}^{\theta_i} \right]^{1/\theta_i}.$$

As the notation above suggests, we assume that countries differ in their EIS and risk-aversion preference parameters: $\psi_1 > \psi_2$ and $\gamma_1 < \gamma_2$. Cross-country differences in risk-aversion are key in Gourinchas, Rey, and Govillot (2010): in their model, the relatively less risk-averse U.S. agent insures the ROW agent by taking a levered position in ROW equity. The risky position of the U.S. accounts for the difference between the returns on its assets and liabilities. Differences in EIS have received some recent empirical support. Vissing-Jorgensen (2002b) shows that the values of the EIS are larger for the U.S. households with larger financial positions; a similar reasoning at the aggregate level would suggest that the U.S. may have a higher EIS than the ROW. Likewise, Havranek, Horvath, Irsova, and Rusnak (2013) find that households in richer countries and countries with higher asset market participation have higher values of EIS. Differences in preference parameters are also shortcuts for differences in financial sectors' sizes and skills as modeled in Mendoza, Quadrini, and Rios-Rull (2009) and inMaggiori (2015).

[2] The terms of trade is $q \equiv p_2/p_1$, and hence the real exchange rate is:

$$Q \equiv \frac{P_2}{P_1} = \left[\frac{(1-s)^{\epsilon} q^{1-\epsilon} + s^{\epsilon}}{s^{\epsilon} q^{1-\epsilon} + (1-s)^{\epsilon}} \right]^{1/(1-\epsilon)}.$$

4.2.3 Limited Market Participation

Both workers and investors are characterized by the same preferences, but workers are hand-to-mouth, i.e. they do not have access to financial markets and consume their labor income every period, whereas investors participate in financial markets.

Financial Income Investors trade three assets: one stock in each country, as well as an international bond. The stocks are long-term assets, while the bond is one-period. The net supply of each stock is one, while the net supply of the bond is zero.

The international bond, bought at price q_t^b at date t, is a claim on e_{t+1} units of a composite good, which is a bundle of α goods from country 1 and $1-\alpha$ goods from country 2, with $\alpha = 1/2$. The price of the composite good at date $t+1$ is equal to: $p_{\alpha,t+1} = \alpha p_{1,t+1} + (1-\alpha) p_{2,t+1}$. We model only one instead of two bonds for computational reasons: an equilibrium with two bonds is more difficult to determine. Note that adding a second bond would not be enough for the markets to be complete, and our simplification thus appears innocuous.

In each country, a stock is a claim to a stream of dividends $d_{i,t} > 0$ measured in units of good i. Stocks are traded at the ex-dividend prices, denoted $q_{1,t}$ and $q_{2,t}$. The dividends are leveraged payoffs of endowments:

$$d_{i,t} = e_t \left[\overline{d} + s_\xi \left(\exp(\varphi_d \xi_t) - 1 \right) + s_g \left(\exp(g_t) - 1 \right) + s_a \left(\exp(a_{i,t}) - 1 \right) \right].$$

The leverage is time-varying, as in Longstaff and Piazzesi (2004b). As a result, the dividend growth rate is not perfectly correlated to the endowment growth rate.

Labor Income Labor income in country i, denoted $\omega_{i,t}$, is the fraction of the total endowment not distributed as dividends:

$$\omega_{i,t} = e_t \left[1 - \overline{d} - s_\xi \left(\exp(\varphi_d \xi_t) - 1 \right) - s_g \left(\exp(g_t) - 1 \right) - s_a \left(\exp(a_{i,t}) - 1 \right) \right].$$

In the model, since leverage is time-varying, the income share is also time-varying. Unlike the dividend cash flow that can be traded by buying and selling long-lived equities, the future labor income cash flow cannot be traded: potential reasons include financial frictions, capital income taxation, or poor enforcement of property rights. Workers thus face a hard constraint: they cannot participate in financial markets and cannot work around this constraint.

Since workers are hand-to-mouth, their consumption can be easily obtained. Let I_I denote the share of labor income received by investors in each country. The workers in country i receive a total income of $(1 - I_I)\omega_{i,t}$ in terms of their domestic goods. Their budget constraint implies that $(1 - I_I)\omega_{i,t}p_{i,t} = P_{i,t}C_{w,i,t}$, and their consumption levels are:

$$\begin{aligned}
c^1_{w,1,t} &= s^{\epsilon}\left[\frac{p_{1,t}}{P_{1,t}}\right]^{-\epsilon}\frac{(1-I_I)\omega_1 p_{1,t}}{P_{1,t}}, \quad \text{and} \quad c^2_{w,1,t} = (1-s)^{\epsilon}\left[\frac{p_2}{P_{1,t}}\right]^{-\epsilon}\frac{(1-I_I)\omega_1 p_{1,t}}{P_{1,t}}, \\
c^1_{w,2,t} &= (1-s)^{\epsilon}\left[\frac{p_{1,t}}{P_{2,t}}\right]^{-\epsilon}\frac{(1-I_I)\omega_{2,t} p_{2,t}}{P_{2,t}}, \quad \text{and} \quad c^2_{w,2,t} = s^{\epsilon}\left[\frac{p_{2,t}}{P_{2,t}}\right]^{-\epsilon}\frac{(1-I_I)\omega_{2,t} p_{2,t}}{P_{2,t}},
\end{aligned}$$

where, again, $c^j_{i,t}$ denotes the consumption of good j by agent i at time t. The investors' optimal consumption solves a more complicated optimal portfolio problem.

4.2.4 Borrowing and Short-Selling Constraints

In the model, investors face two specific constraints: (i) they cannot short equity and (ii) their borrowing ability is limited.

The short-selling constraint on equity positions and the presence of labor income together imply that some risk cannot be hedged. This plays a crucial role in determining the portfolio position of the agents since the perfect conditional correlation between non-tradable income and dividends gives investors an incentive to short their own equity. Let $\vartheta^j_{i,t}$ denote the holding of stock j by agent i at date t: the subscript characterizes the country holder and the superscript characterizes the goods in which the asset is denominated. Formally, the short-selling constraint is:

$$\vartheta_{i,t}^{j} \geq 0, \quad \text{for } i, j = 1, 2. \tag{4.5}$$

The borrowing constraint is such that debt can always be repaid since the amount due is always above or equal to the financial wealth of the borrower in the worst state of the world next period:

$$b_{i,t} \geq -B_{i,t}, \quad \text{for } i, j = 1, 2, \tag{4.6}$$

$$\text{where } B_{1,t} \equiv \min_{s^{t+1} \succcurlyeq s^t} \left\{ w_{1,t+1} \frac{p_{1,t+1}}{p_{\alpha,t+1}} + \sum_{j=1}^{2} \vartheta_{1,t}^{j} \frac{q_{j,+1} + p_{j,t+1} d_{j,t+1}}{p_{\alpha,t+1}} \right\}, \tag{4.7}$$

$$B_{2,t} \equiv \min_{s^{t+1} \succcurlyeq s^t} \left\{ w_{2,t+1} \frac{p_{2,t+1}}{p_{\alpha,t+1}} + \sum_{j=1}^{2} \vartheta_{2,t}^{j} \frac{q_{j,t+1} + p_{j,t+1} d_{j,t+1}}{p_{\alpha,t+1}} \right\}, \tag{4.8}$$

where the minimum is taken on all possible states the next period: the symbol $\succcurlyeq$ denotes the partial order on the tree $\mathbb{S}$ such that node $s^{t_1} \succcurlyeq s^{t_2}$ if s^{t_1} is a descendant of s^{t_2}. The right hand side of Equations (4.7) and (4.8) describe the lowest possible sum of labor income and equity wealth for investors in countries 1 and 2 respectively next period. Labor income and equity wealth are thus collateral, securing international debt. Bonds cannot be used as collateral as there is a unique bond in the model: if one country lends, the other must borrow. As a result, the country that borrows has no bond to post as collateral. The borrowing constraint remains potentially binding even in the long run because investors cannot become rich enough to forget it: the non-participation of workers to financial markets prevents investors from lending money to workers, accumulating wealth up to the point when the borrowing constraints are no longer relevant.

The short-selling and borrowing constraints are key: they rule out defaults and address the survivorship or degenerated stationary distribution issue highlighted in Lucas and Stokey (1984) and Anderson (2005). In our model, despite the heterogeneity in agents' preferences, both agents survive in the long run because the collateral and short-sale constraints prohibit them from assuming more and more debt over time. The consumption of investors satisfy the following budget constraint:

$$\begin{aligned}\sum_{j=1}^{2} p_{j,t} c_{i,t}^{j} \quad &+ \quad \sum_{j=1}^{2} q_{j,t} \vartheta_{i,t}^{j} + q_t^b b_{i,t} \\ &= \quad p_{i,t}\omega_{i,t} + \sum_{j=1}^{2} \left[q_{j,t} + p_{j,t} d_{j,t} \right] \vartheta_{i,t-1}^{j} + p_{\alpha,t} b_{i,t-1}. \end{aligned} \tag{4.9}$$

In the next section, we define the competitive equilibrium in the model and prove that a wealth-recursive equilibrium exists. This proof is not purely formal: as pointed out by Kubler and Polemarchakis (2004), the approximate equilibria obtained by numerical methods may exist even when no exact equilibrium exists. The following section guarantees that the wealth-recursive Markov equilibrium exists. The reader mostly interested by the simulation results can skip this section.

4.3 Equilibrium

Before characterizing the equilibrium, we formulate the country's optimization Bellman equation into a compact and manageable form.

4.3.1 Time-Shift

We appeal to the "time shift" proposed by Dumas and Lyasoff (2012). We translate the combined borrowing constraints in Equations (4.6), (4.7), and (4.8) into a group of separate constraints as follows, for each date t and $t+1$:

$$\mathbb{C}_1(t, t+1) \equiv p_{1,t+1}\omega_{2,t+1} + \sum_{j=1}^{2} \vartheta_{1,t}^{j} \left[q_{j,t+1} + p_{j,t+1} d_{j,t+1} \right] + b_{1,t} p_{\alpha,1,t+1} \geq 0,$$

$$\mathbb{C}_2(t, t+1) \equiv p_{2,t+1}\omega_{2,t+1} + \sum_{j=1}^{2} \vartheta_{2,t}^{j} \left[q_{j,t+1} + p_{j,t+1} d_{j,t+1} \right] + b_{2,2} p_{\alpha,2,t+1} \geq 0.$$

The Lagrangian multiplier for each of the $|\mathcal{S}|$ borrowing constraints is $\mu^b_{i,t,s_{t+1}}$. The $|\mathcal{S}|$ Lagrangian multipliers are endogenous variables in period t. Likewise, each short-selling constraint is associated with a multiplier $\mu^j_{i,t}$. The recursive form of the value function leads to the following Bellman equation with Lagrangian multipliers, for every $t \geq 0$:

$$
\begin{gathered}
U_i(W_{i,t}; s^t) = \\
\min_{\mu^j_{i,t} \geq 0,\ \mu^b_{i,t,s_{t+1}} \geq 0} \ \max_{c^j_{i,t}, \theta^j_{i,t}, b_{i,t}} \ \frac{C_{i,t}^{1-\psi_i^{-1}}}{1-\psi_i^{-1}} \\
\beta \mathbb{E}_t \left[U_i(W_{i,t+1}; s^{t+1})^{\theta_i} \right]^{1/\theta_i} + \textstyle\sum_{j=1}^{2} \mu^j_{i,t} \vartheta^j_{i,t} + \sum_{s_{t+1} \in \mathcal{S}} \mu^b_{i,t,s_{t+1}} \mathbb{C}_i(t, t+1),
\end{gathered}
$$

subject to the inter-temporal budget constraints:

$$
\begin{aligned}
W_{i,t} &= p_{1,t} c^1_{i,t} + p_{2,t} c^2_{i,t} + \vartheta^1_{i,t} q_{1,t} + \vartheta^2_{i,t} q_{2,t} + b_{i,t} q_{\alpha,t}, \\
\text{and } W_{i,t+1} &= p_{i,t+1} \omega_{i,t+1} I_I + \sum_{j=1}^{2} \vartheta^j_{i,t} \left(q_{j,t+1} + p_{j,t+1} d_{j,t+1} \right) + b_{i,t} p_{\alpha,t+1}.
\end{aligned}
$$

4.3.2 Definitions

Let us now define formally the competitive equilibrium.

Definition 8. *A competitive equilibrium with initial asset holdings* $\{\vartheta_i(s^{-1}), b_i(s^{-1})\}_{i=1,2}$ *and initial shock* s_0 *is a collection of prices* $\mathcal{P}^{\mathbb{S}} = \left\{ \left(p_i(s^t), q_i(s^t), q^b(s^t) \right)_{i=1,2} \right\}_{s^t \in \mathbb{S}}$*, consumption allocations* $\mathcal{C}^{\mathbb{S}} = \left\{ \left(c^1_i(s^t), c^2_i(s^t) \right)_{i=1,2} \right\}_{s^t \in \mathbb{S}}$*, and international asset holdings* $\mathcal{A}^{\mathbb{S}} = \left\{ \left(\vartheta^1_i(s^t), \vartheta^2_i(s^t), b_i(s^t) \right)_{i=1,2} \right\}_{s^t \in \mathbb{S}}$ *such that*

(i) given the price system $\mathcal{P}^{\mathbb{S}}$*, each investor in country* $i \in \{1,2\}$ *solves the optimization problem* $U_i(\mathcal{C}^{\mathbb{S}}_i)$ *with the consumption plan* $\mathcal{C}^{\mathbb{S}}_i$ *and the asset holdings* $\mathcal{A}^{\mathbb{S}}_i$ *lying in the sequential budget set* $\mathbb{B}_{\mathbb{S}}\left(\mathcal{P}^{\mathbb{S}}\right)$ *described in Equation (4.9) under the short-selling constraint described in Equation (4.5) and the borrowing constraints described in Equations (4.6), (4.7), and (4.8);*

(ii) given the same price system $\mathcal{P}^{\mathbb{S}}$*, each worker in country* $i \in \{1,2\}$ *maximizes her utility*

under her budget constraint;

(iii) equity markets and bond markets clear, i.e. for $j = 1,\ 2$ *and for all dates* t*:*

$$\begin{aligned} \vartheta^j_{1,t} + \vartheta^j_{2,t} &= 1, \\ b_{1,t} + b_{2,t} &= 0. \end{aligned}$$

(iv) goods markets clear, i.e. for $j = 1,\ 2$ *and for all dates* t

$$c^j_{1,t} + c^j_{2,t} = e_{j,t}.$$

The borrowing constraints in the agents' optimization problem not only constitute a market imperfection but also ensure the existence of a solution to the agents' optimization problem (see e.g., Levine and Zame, 1996; Magill and Quinzii, 1996; and Hernandez and Santos, 1996). Although the proof of the existence of a competitive equilibrium in Lucas-type infinite-horizon exchange economies with heterogeneous agents and incomplete markets exists, it is impossible to compute the equilibrium in general because it is not unique and the equilibria are mathematically equivalent to an infinite number of equilibrium prices – a infinite dimensional problem. Duffie, Geanakoplos, Mas-Colell, and McLennan (1994) show that if the exogenous shocks' dynamics can be characterized by a finite-valued time-homogeneous Markov process, then there exists a competitive equilibrium in which the endogenous variables can be summarized by a finite number of endogenous state variables as well as the exogenous state variables. The endogenous state variables follow a time-homogeneous Markov process having a time invariant transition with an ergodic measure. This type of equilibrium is called recursive Markov equilibria. A recursive Markov equilibrium in which the wealth distribution summarizes all the endogenous state variables is called a wealth-recursive Markov equilibrium. Duffie, Geanakoplos, Mas-Colell, and McLennan (1994) show that a recursive Markov equilibrium is a competitive equilibrium under general regularity conditions. Under mild regularity conditions, Kubler and Schmedders (2003) in their Lemma 2 show that a wealth-recursive Markov equilibrium is a competitive equilibrium. Their proof does not apply to our model, but we show how to extend their result. In order to do so, let us first rigorously define the wealth-recursive Markov equilibrium.

Because we have two heterogeneous representative investors in the economy, the wealth portion of the agent 1 fully characterizes the wealth distribution. The wealth share of country 1 is denoted w:

$$w_t \equiv \frac{W_{1,t}}{W_{1,t} + W_{2,t}},$$

where the total wealth in the economy is $W_{1,t} + W_{2,t} = \sum_{j=1}^{2} \left[p_{j,t} e_{j,t} + q_{j,t} \right]$. Let $\mathcal{Y}$ denote the space of all possible endogenous variables that occur in the economy at some node s^t. That is, $\mathcal{Y}$ consists of all vectors:

$$\left\{ \left(c_i^1, c_i^2 \right)_{i=1,2}, \left(\vartheta_i^1, \vartheta_i^2, b_i \right)_{i=1,2}, \left(p_i, q_i, q_i^b \right)_{i=1,2}, \left(\mu_i^1, \mu_i^2, \mu_{i,\tilde{s}}^b \right)_{i=1,2;\tilde{s}\in\mathcal{S}} \right\} \tag{4.10}$$

such that, for $i, j \in \{1, 2\}$:

$$c_i^j, p_j, q_j, q^b, \mu_i^j, \mu_{i,\tilde{s}}^b \in \overline{\mathbb{R}}_+, \quad \text{and} \quad \vartheta_i^j, b_i^j \in \mathbb{R}_+,$$
$$p_1 + p_2 = 1, \quad \text{and} \quad \vartheta_i^j \mu_i^j = 0, \quad \text{and} \quad \vartheta_1^j + \vartheta_2^j = 1, \quad \text{and} \quad b_1 + b_2 = 0.$$

The Lagrangian multiplier μ_i^j corresponds to the short-selling constraint of the agent in the country i on the stock j, for $i, j \in \{1, 2\}$, while the Lagrangian multiplier $\mu_{i,\tilde{s}}^b$ corresponds to agent i's borrowing constraint. The space of endogenous variables $\mathcal{Z}$ is a closed subset of $\overline{\mathbb{R}}^{2\times(11+|\mathcal{S}|)}$. The space of both exogenous and endogenous variables is $\mathcal{Z} \equiv \mathcal{Y} \times \mathcal{S}$. Let $\widehat{\mathcal{Z}} \equiv [0, 1] \times \mathcal{Y} \times \mathcal{S} \times \overline{\mathbb{R}}_+$.

The expectation correspondence maps the variables $\hat{z} \in \widehat{\mathcal{Z}}$ in the current period to a subset of the space of endogenous variables in next period $([0, 1] \times \mathcal{Y})^{|\mathcal{S}|}$, where $([0, 1] \times \mathcal{Y})^{|\mathcal{S}|}$ is the Cartesian product of $|\mathcal{S}|$ copies of $[0, 1] \times \mathcal{Y}$. More precisely, the expectation correspondence is denoted by

$$\Phi \; : \; \widehat{\mathcal{Z}} \rightrightarrows ([0,1] \times \mathcal{Y})^{|\mathcal{S}|} \, ,$$

such that for a given state in current period $\widehat{z} \equiv (w, y, s, e) \in \widehat{\mathcal{Z}}$, the country 1's wealth share $\{w(\tilde{s}) \; : \; \tilde{s} \in \mathcal{S}\}$ of next period and the vector of endogenous variables $\{\tilde{y}(\tilde{s}) \; : \; \tilde{s} \in \mathcal{S}\}$ in the next period lies in the set $\Phi(\widehat{z})$ if and only if they are consistent with the intertemporal budget constraints, the first-order conditions and market clearing conditions.

Definition 9. *A wealth-recursive Markov equilibrium consists of a (nonempty valued) "policy correspondence"* $\Pi : [0,1] \times \mathcal{S} \times \overline{\mathbb{R}}_+ \rightrightarrows \mathcal{Y}$*, where* $\mathcal{Y}$ *is the space of endogenous policy variables defined in (4.10) - (4.11) and a "transition map"* $\Omega \, : \, [0,1] \times \mathcal{S} \to [0,1]^{|\mathcal{S}|}$ *such that for any given* $(w, s, e) \in [0,1] \times \mathcal{S} \times \overline{\mathbb{R}}_+$ *with* $(\tilde{w}(\tilde{s}))_{\tilde{s}\in\mathcal{S}} = \Omega(w, s)$*, it holds that* $\forall \; y \in \Pi(w, s, e)$ *and* $\forall \; \tilde{y}(\tilde{s}) \in \Pi(\tilde{w}(\tilde{s}), \tilde{s}, \tilde{e}))$ *with* $\tilde{e} \equiv e \times \zeta(\tilde{s})$ *and* $\tilde{s} \in \mathcal{S}$*,*

$$(\tilde{w}(\tilde{s}), \tilde{y}(\tilde{s}))_{\tilde{s}\in\mathcal{S}} \in \Phi(w, y, s, e).$$

For notational simplicity, we denote $\tilde{w}(\tilde{s}) = \Omega(w, s; \tilde{s})$*.*

We now turn to our main theorem.

4.3.3 Existence of a Wealth-Recursive Markov Equilibrium

Theorem 3. *Assuming that there exists* $d_m > 0$ *and* $\omega_m > 0$ *such that* $d_i(s_t)/e(s_t) > d_m$ *and* $\omega_i(s_t)/e(s_t) > \omega_m$ *for all* $i = 1, 2$ *and* $s_t \in \mathcal{S}$*, there exists a wealth-recursive Markov equilibrium in the economy with heterogenous agents with recursive utility described in Section 4.2.*

Proof. The assumption guarantees that the dividend and wage incomes, as percentages of world GDP, are bounded from below. The proof of the theorem is reported in Appendix B.3. It consists of three main steps. First, we show that for any T-truncated economy, the competitive equilibrium's policy functions are uniformly bounded if a competitive equilibrium exists.[3] In this step, we generalize the results of Kubler and Schmedders (2003) and Duffie, Geanakoplos, Mas-Colell, and McLennan (1994) to allow stochastic growth in the economy, lower-bounded utility functions and Epstein-Zin-Weil preferences. Second, we show the existence of competitive equilibrium for each T-truncated economy. Third, we show the existence of wealth-recursive Markov equilibrium exists for the infinite-horizon economy by backward induction. □

Theorem 3 extends the results of Kubler and Schmedders (2003) to a large class of preferences and to stochastic growth. Duffie, Geanakoplos, Mas-Colell, and McLennan (1994) and Kubler and Schmedders (2003) crucially assume that the utility is not bounded from below, which guarantees that the equilibrium variables are all uniformly bounded. They focus on the time-separable CRRA utility function whose coefficient of relative risk aversion is not smaller than one. However, for the Epstein-Zin-Weil preferences with an EIS parameter bigger than one, the utility function is not bounded from below, and thus their arguments do not go through. We use the results inGeanakoplos and Zame (2013), who show the existence of a competitive equilibrium for a two-period incomplete-market model, and combine them with the proofs in Kubler and Schmedders (2003) in order to extend their results.

The wealth recursive formulation of the agent's optimization problem makes it natural to consider wealth-recursive Markov equilibrium of the economy. The intuition is that the wealth distribution among agents at the beginning of each period presumably influences prices and allocations in that period. Intuitively, one would expect that the wealth distribution constitutes a sufficient endogenous state space. The argument would be that the initial distribution of wealth is the only endogenous variable that influences the equilibrium behavior of the economy. However, as pointed by Kubler and Schmedders (2002), the wealth distribution alone does not always constitute a sufficient endogenous state space, mainly because the equilibrium decisions at time t also

[3]The T-truncated economy is defined to be a finite-horizon economy built on an event tree, denoted by $\mathbb{S}^T$, which consists of all the nodes and edges along the path $s^T = (s_0, s_1, \cdots, s_T)$ in the original event tree $\mathbb{S}$. The endowments and asset payoffs at the nodes of the truncated tree, as well as agents' preferences and portfolio constraints at these nodes, are the identical to the original infinite-horizon economy.

must be consistent with expectations at time $t-1$ and that these expectations at time $t-1$ cannot always be summarized in the wealth distribution alone. Our existence result allow us to proceed further in simulating the model. Theorem 3, however, does not guarantee the uniqueness of the equilibrium or the existence of non-degenerate ergodic measure. But Theorem 3 offers a key characteristic of the solution method.

Corollary 4. *Under the same assumptions as in Theorem 3, the policy correspondence* Π *and value functions* U_i *in a wealth-recursive Markov equilibrium have the following forms, for* $i, j \in \{1, 2\}$,

$$c_i^j(w,s,e) \equiv c_i^j(w,s)e, \quad \vartheta_i^j(w,s,e) \equiv \vartheta_i^j(w,s), \quad b_i^j(w,s,e) \equiv b_i^j(w,s), \tag{4.11}$$

$$p_i(w,s,e) \equiv p_i(w,s), \quad q_i(w,s,e) \equiv q_i(w,s)e, \quad q_i^b(w,s,e) \equiv q_i^b(w,s)e, \tag{4.12}$$

$$\mu_i^j(w,s,e) \equiv \mu_i^j(w,s)e^{1-\psi_i^{-1}}, \ \mu_{i,\tilde{s}}^b(w,s,e) \equiv \mu_{i,\tilde{s}}^b(w,s)e^{-\psi_i^{-1}}, \tag{4.13}$$

$$\text{and } U_i(w,s,e) = U_i(w,s)e^{1-\psi_i^{-1}}. \tag{4.14}$$

Proof. The proof is in Appendix B.2. □

Corollary 4 suggests that the components of the policy correspondence in equilibrium are homogeneous in terms of the size of the global economy e to different degrees, because the level of the global tree e controls the scale of the economy and shocks on the size of global tree are permanent shocks. For example, the consumption, the bond holdings and the equity prices are degree-one homogeneous in the size of the economy, which is intuitive because only the consumption shares between agents and the debt ratios of each agent matter for the economy and the size of the economy is proportional to the amount of commodities attached to equity. Furthermore, the equity shares and the bond prices are invariant to the scale of the economy, because the total amount of the equity is normalized to be one and by definition the claim of a unit of bond is always assumed to be one unit of commodity. As a standard property, the Epstein-Zin-Weil preference U_i is homogeneous in $1-\psi_i^{-1}$ degrees in term of wealth that is proportional to the size of the economy. The shadow values are also homogeneous in term of economy scale according to the value functions. Thus, without loss of generality, we can assume that the endowment level of the global tree in current period is one, i.e. $e = 1$. Therefore, when solving for the equilibrium, we only need to focus on the wealth share w and the exogenous shock s .

4.4 Calibration

This section describes our data set and the key statistics on GDP, consumption, international trade, and asset prices that define our calibration.

4.4.1 Data

Our data come from different sources. At the quarterly frequency, GDP, consumption and international trade series are from the OECD, while international capital stocks and flows are from the International Monetary Fund (IMF). International capital flows come from Bluedorn, Duttagupta, Guajardo, and Topalova (2013); the balance of payments of each country is the primary source of the data. Foreign equity return indices are built by Datastream; for the U.S., the equity return series come from CRSP. Interest rates correspond to Treasury Bills or money market rates from the IMF. At the annual frequency, long time-series of capital stocks come from Lane and Milesi-Ferretti (2007).

This dataset is used to characterize two countries, the U.S. and the ROW. The ROW is defined as the aggregate of the G10 countries, excluding the U.S. (i.e., Belgium, Canada, Japan, France, Germany, Italy, Netherlands, Sweden, Switzerland, and U.K.). Each period, the ROW GDP and consumption growth rates are obtained by weighting each country-specific real growth rates by the share of its real GDP (measured at purchasing power parity) in total GDP. Indices are built from the growth rates and HP-filtered with a smoothing coefficient of 1600, as it is usual for quarterly series (Hodrick and Prescott, 1997). The sample period is 1973.1–2010.4.

4.4.2 Macroeconomic and Financial Variables

Let us now rapidly review the properties of macroeconomic and financial variables in the U.S. and ROW.

Production, Consumption, and International Trade Table 4.1 reports the mean, standard deviation, and autocorrelation of U.S. GDP and consumption growth rates, as well as their rest-of-the-world (ROW) counterparts. The table also reports similar summary statistics on the U.S. net exports and trade openness. Net exports are obtained as the difference between exports and imports, both scaled by GDP. Trade openness corresponds to the average of imports and exports, also scaled by GDP.

The macroeconomic data exhibit classic features of real business cycles. In both the US and the ROW, consumption appears less volatile than GDP, a common finding among developed countries. GDP and consumption are less volatile in the ROW than in the US as some of the foreign shocks average out across foreign countries. GDP growth rates are more correlated across countries than consumption growth rates. These characteristics appear on growth rates as well as on HP-filtered series. Trade openness is around 10%, while net exports are on average -2%; both measures are very persistent.

Interest Rates, Equity, and Currency Returns Panel A of Table 4.2 reports the mean, standard deviation, and autocorrelation of U.S. and rest-of-the-world (ROW) real interest rates, dividend yields, real equity returns and excess returns, as well as their cross-country correlation coefficients. Over the last forty years, the average real equity returns in the U.S. and ROW are respectively equal to 8.4% and 4.7% per year, leading to average equity excess returns respectively equal to 6.4% and 2.7%.[4] The dividend yields are 3.1% and 2.8% in the U.S. and ROW, implying price dividend ratios of 32 and 37. The price-dividend ratios are volatile, and thus either future dividend growth

[4]The Datastream series understate the aggregate equity return: for the U.S., the difference between the CRSP and Datastream estimates is equal to 2.7% on average over our sample period. The discrepancy is certainly related to the Datastream focus on only a subset of large firms. The equity premium for the ROW is thus likely much higher than reported here.

Table 4.1: GDP, Consumption, and International Trade

	Data				Model			
	Mean	Std	AC(1)	Corr(RW,US)	Mean	Std	AC(1)	Corr(RW,US)
				Panel A: Raw Series (Growth Rates and Ratios)				
US GDP	0.68 (0.09)	0.83 (0.07)	0.39 (0.08)		0.68	0.87	0.44	
US Consumption	0.74 (0.07)	0.68 (0.06)	0.34 (0.08)		0.68	0.87	0.46	
ROW GDP	0.53 (0.07)	0.62 (0.09)	0.48 (0.08)	0.45 (0.10)	0.68	0.87	0.44	0.47
ROW Consumption	0.54 (0.04)	0.51 (0.05)	0.04 (0.11)	0.34 (0.09)	0.68	0.85	0.44	0.50
US Net Exports/GDP	-2.13 (0.24)	1.71 (0.13)	0.98 (0.05)		-1.74	0.41	0.96	
US Trade Openness	10.44 (0.27)	1.92 (0.16)	0.98 (0.05)		8.28	0.23	0.96	
				Panel B: HP-Filtered				
US GDP		1.53 (0.15)	0.87 (0.06)			1.08	0.82	
US Consumption		1.21 (0.10)	0.88 (0.06)			1.07	0.82	
ROW GDP		1.13 (0.13)	0.88 (0.05)	0.65 (0.06)		1.08	0.82	0.43
ROW Consumption		0.72 (0.07)	0.80 (0.07)	0.47 (0.09)		1.05	0.82	0.47
US Net Exports/GDP		0.46 (0.04)	0.77 (0.06)			0.14	0.69	
US Trade Openness		0.53 (0.07)	0.81 (0.06)			0.08	0.69	

Notes: Panel A reports the mean, standard deviation, and autocorrelation of U.S. rest-of-the-world (ROW) GDP and consumption growth rates, as well as their cross-country correlation coefficients. It also reports the mean, standard deviation, and autocorrelation of U.S. net exports and trade openness. Net exports are obtained as the difference between exports and imports, both scaled by GDP. Trade openness corresponds to the average of imports and exports, also scaled by GDP. Panel B reports the same test statistics (except for the mean) for HP-filtered series in levels. Standard errors are reported in parentheses; they are obtained by block-boostrapping. Data are quarterly, from the OECD database. All variables are reported in percentage points, except for the autocorrelation and cross-country correlation coefficients. The sample period is 1973.1–2010.4. The simulated moments correspond to samples without disasters.

or future equity excess returns must be predictable (Campbell and Shiller, 1988). Equity returns are volatile both in the U.S. and in the ROW aggregate (18% on an annual basis) but appear largely correlated (0.8) among the most developed countries. In the model, the wealth consumption ratio is large and volatile, as it is in the data (Lustig, et al., 2013).

Predictability regressions show that equity excess returns are predictable over long horizons. Panel B of Table 4.2 reports the slope coefficients (β_{pd} or β_{cay}) and the R^2 obtained in predictability tests of equity excess returns over 5 years on dividend yields or, for the U.S., the consumption-wealth ratio of Lettau and Ludvigson (2001). The slope coefficients are statistically significant and the R^2 range from 10% to 30%. The model matches particularly well the amount of predictability implied by the wealth-consumption ratio. Panel C of Table 4.2 reports the mean, standard deviations, and autocorrelations of expected equity excess returns in the U.S. obtained using either the price-dividend ratio or the wealth consumption ratio as predictors. Expected equity excess returns, i.e. risk premia, are clearly time-varying.

Table 4.3 focuses on exchange rates. The real exchange rate between the U.S. and the ROW has an annualized volatility of 8.9% and a small and insignificant autocorrelation. Carry trade excess returns are obtained by building three portfolios of currencies sorted by their interest rates: carry trades then correspond to strategies long the last portfolio of high interest rate currencies and short the first portfolio of low interest rate currencies. The carry trade offers an average excess return of 2.5% in the sample and a Sharpe ratio of 0.28, higher than the Sharpe ratios on U.S. and ROW aggregate equity markets. Carry trade excess returns tend to be low when global equity volatility surges: the correlation between the two is significantly negative. The exchange rate of low interest rate countries tend to appreciate while the exchange rate of high interest rate countries tend to depreciate when global volatility increases, leading in both cases to carry trade losses. This pattern is at the root of a risk-based explanation of the large average carry trade excess returns. Risk-averse investors expecting losses in bad times require a risk premium as a compensation for bearing the exchange rate risk.

Table 4.2: Dividend Yields, Equity Returns, and Interest Rates

	Data				Model			
	Panel A: Moments							
	Mean	Std	AC(1)	Corr(ROW,US)	Mean	Std	AC(1)	Corr(ROW,US)
US Dividend Yield	4.36	1.35	0.94		5.37	0.95	0.90	
	(0.20)	(0.11)	(0.05)					
ROW Dividend Yield	2.76	0.88	0.95	0.72	3.12	0.43	0.86	0.96
	(0.12)	(0.07)	(0.05)	(0.07)				
US Real Equity Returns	8.37	17.03	0.12		9.36	16.55	-0.05	
	(2.84)	(1.55)	(0.10)					
ROW Real Equity Returns	4.73	17.76	0.13	0.08	6.62	12.78	-0.04	0.95
	(3.12)	(1.34)	(0.08)	(0.09)				
US Real Money Market	1.87	2.61	0.82		0.86	3.38	0.88	
	(0.34)	(0.20)	(0.06)					
ROW Real Money Market	2.07	2.39	0.96	0.63	-1.46	5.76	0.88	1.00
	(0.32)	(0.22)	(0.05)	(0.07)				
US Equity Excess Returns	6.39	16.99	0.12		8.50	17.72	-0.01	
	(2.83)	(1.60)	(0.10)					
ROW Equity Excess Returns	2.69	17.34	0.12	0.08	8.09	15.19	0.09	0.95
	(2.88)	(1.48)	(0.07)	(0.09)				
	Panel B: Predictability Tests							
	β_{pd}	R^2	β_{cay}	R^2	β_{pd}	R^2	β_{cay}	R^2
US Pred.	0.37	0.09	0.54	0.23	1.44	0.44	0.52	0.36
	(0.18)	(0.03)	(0.14)	(0.03)				
ROW Pred.	1.24	0.31			2.84	0.39		
	(0.34)	(0.03)						
	Panel C: Expected Equity Excess Returns							
	Mean	Std	AC(1)		Mean	Std	AC(1)	
US Exp. ER (D/P)	4.28	1.28	0.98		8.50	7.14	0.90	
	(2.68)	(0.82)	(0.05)					
US Exp. ER (cay)	4.28	2.66	0.93		8.50	3.21	0.94	
	(2.68)	(1.20)	(0.05)					

Notes: Panel A of the table reports the mean, standard deviation, and autocorrelation of U.S. and rest-of-the-world (ROW) real interest rates, dividend yields, real equity returns and excess returns, as well as their cross-country correlation coefficients. Real equity returns are obtained by subtracting three-month realized inflation to nominal equity returns. Real interest rates correspond to nominal interest rates minus 12-month inflation. Panel B reports the slope coefficients (β_{pd} or β_{cay}) and the R^2 in predictability tests of equity excess returns over 5 years on dividend yields or, for the U.S., the consumption-wealth ratio of Lettau and Ludvigson (2001). Panel C report the mean, standard deviation, and autocorrelation of the expected U.S. equity excess returns. Expected excess returns over the next quarter are obtained using either the dividend yield or the wealth-consumption ratio. Standard errors are reported in parentheses; they are obtained by block-boostrapping. Data are quarterly, from the Datastream (equity indices and dividend yields) and IMF (money market rates) databases. All variables are reported in percentage points, except for the autocorrelation and cross-country correlation coefficients.

Table 4.3: Exchange Rates

	Data				Model			
	Panel A: Exchange Rates and Currency Excess Returns							
	Mean	Std	AC(1)	Corr(ER, WV.)	Mean	Std	AC(1)	Corr(ER, WV)
ROW Real FX chge	-3.99	8.94	0.07		-0.00	4.09	-0.04	-0.72
	(14.50)	(0.50)	(0.08)					
Carry ER	3.67	8.95	0.11	-0.45	4.61	4.76	0.16	-0.37
	(1.42)	(1.31)	(0.08)	(0.11)				
	Panel B: Backus-Smith Correlations							
	$\frac{C_I^{US}}{C_I^{RoW}}, Q$	$\frac{C_W^{US}}{C_W^{RoW}}, Q$	$\frac{C^{US}}{C^{RoW}}, Q$		$\frac{C_I^{US}}{C_I^{RoW}}, Q$	$\frac{C_W^{US}}{C_W^{RoW}}, Q$	$\frac{C^{US}}{C^{RoW}}, Q$	
Growth			-0.11		-0.44	0.13	-0.11	
			(0.09)					
HP filter			0.01		-0.40	0.19	-0.06	
			(0.10)					

Notes: Panel A of the table reports the mean, standard deviation, and autocorrelation of the real exchange rate change between the U.S. and the ROW, as well as the same moments for the currency carry trade excess returns, along with its correlation with world equity volatility. Currencies are sorted by the level other short-term interest rates into three portfolios as in Lustig and Verdelhan (2007). Carry trade excess returns correspond to the returns on the high interest rate portfolios minus the returns on the low interest rate portfolio. Panel B of the table reports the Backus-Smith correlation between exchange rates and the relative consumption in the U.S. and ROW. Consumption and exchange rates are either measured on growth rates or H.P.-filtered. Consumption corresponds to workers' (denoted C_W) or investors' (denoted C_I) or aggregate (C) consumption. Standard errors are reported in parentheses; they are obtained by block-boostrapping. Data are quarterly, from the Datastream (exchange rates) and IMF (money market rates) databases. All variables are reported in percentage points, except for the autocorrelation and cross-country correlation coefficients. Exchange rate changes and currency excess returns are annualized (i.e., average obtained on quarterly returns are multiplied by 4 and the standard deviations are multiplied by 2). The sample period is 1973.1–2010.4. The simulated moments correspond to samples without disasters.

Table 4.4: Parameters

Parameter (Quarterly)	Symbol	Home/Foreign
Panel A: Preferences		
Subjective discount rate	β	0.99
Relative risk aversion	γ_1/γ_2	3.8/4
EIS coefficient	ψ_1/ψ_2	2.4/1.1
Consumption ES coefficient	ϵ	0.885
Consumption share coefficient	s	0.93
Panel B: Endowment		
Country-spec. volatility	σ_c	2%
Global volatility	σ_g	0.6%
Average growth	μ_g	0.675%
Panel C: Dividend and Wage Income		
Wage income share of investors	W_I	10%
Dividend share of output	$\overline{d}$	5%
Dividend leverage on country-spec. shock	s_d	0.19
Dividend leverage on global. shock	s_g	0.19
Dividend leverage on disaster shock	s_{gd}	0.9
Panel D: Disasters		
Disaster size	φ_d	9.7%
Disaster escaping prob.	$1 - p_d$	11.1%
Average log prob.	$\log(\overline{p})$	log(0.314%)
Std. log prob.	σ_p	4.9%
Autocorr. log prob.	ρ_p	0.9

Notes: This table reports the parameters used in the benchmark simulation of the model. The two countries share the same parameters, except for their risk-aversion and elasticity of substitution.

4.4.3 Parameters

We use data on macroeconomic variables and asset returns to calibrate our model, starting with the endowment processes. Table 4.4 reports all the parameters of the model.

In the simulation, the two countries differ in their risk-aversion (3.3 for the U.S. vs. 4 for the ROW) and their IES (2.4 for the U.S. vs 1.4 for the ROW). The other preference parameters are the same in both countries. The subjective discount factor is 0.99. The domestic consumption share is 0.95, and the elasticity of substitution between the domestic and foreign goods is 0.885.

We follow Rouhenworst (1995) to calibrate the Markov processes such that they replicate the GDP series. The dividend share of total endowment is assumed to be ten times more volatile than the labor income share.

The average probability of a disaster is low, equal to 0.3%, but the disaster size is large: when it occurs, it entails a GDP decrease of 9.7%. The probability of leaving the disaster state the next period is 11.1%. The log probability of a disaster is persistent, with an autocorrelation of 0.9, and volatile, with a standard deviation of 4.9%. As the disaster probability is not directly observed, its parameters are subject to a large uncertainty. The model parameters are in line with those suggested by Barro (2006a) and Gourio (2012).

Going back to Tables 4.1, 4.2, and 4.3, we check that the model reproduces the basic features of GDP, consumption, interest rates, equity prices and returns, and exchange rates. The attentive reader can compare moment by moment, series by series, the actual to the simulated data. The main discrepancy is the volatility of net exports and trade openness, which are more volatile in the data than in the model.

The model delivers a large equity premium. It also delivers time-variation in equity returns that is in line with the data. In the data, price-dividend and wealth-consumption ratios predict future equity returns. The model reproduces these findings. The volatility of the expected excess return obtained using the price-dividend ratio is higher in the model than in the data, but the volatility of the expected excess return obtained using the wealth-consumption ratio is the same in the model and the data. The current calibration, however, implies dividend yields that are more correlated than in the data. Likewise, the realized returns are more correlated in the model than in the data. As a result, the simulated cross-country correlation of realized and expected returns is counterfactually high. The model also misses the level of the ROW risk-free rate, calling for an adjustment in the EIS parameter.

The model delivers exchange rates that are less volatile than in the data, but the currency risk premium is the same in the model and the data. While frictionless complete markets where agents are characterized by constant relative risk-aversion imply a perfect correlation between the exchange rate changes and relative consumption growth

(Backus and Smith, 1993), our model implies a negative correlation, closer to its empirical counterpart.

Overall, the model delivers its premises: large and time-varying risk premia with reasonable endowment and preference assumptions. We turn now to the simulation results obtained with this calibration.

4.5 Benchmark Simulation

We start by describing the policy functions and then turn to the key result of the paper: the comparison between the volatility of foreign assets and capital flows in the model and in the data.

4.5.1 Policy Functions

Symmetric Countries To build intuition on the model, let us start with the case of symmetric countries: both countries share the same preference parameters ($\gamma_1 = \gamma_2 = 4$ and $\psi_1 = \psi_2 = 2$), and all the other parameters are the same. Figure 4-1 reports the distribution of relative wealth along with policy functions that describe the asset holdings.

The upper left panel shows that the distribution of relative wealth, defined as $w_t \equiv W_{1,t}/[W_{1,t} + W_{2,t}]$, is symmetric, centered around 0.5 as expected. The lower right panel shows the amount of lending and borrowing chosen by country 1 (the U.S.). When the U.S. is relatively poor, the U.S. borrows from the ROW; when the U.S. is relatively rich, the U.S. lends to the ROW. The policy function is perfectly symmetric around the 0.5 relative wealth. On average, the U.S. does not have any debt. The role of the borrowing constraint appears when one country is much poorer than the other. For example, when the ROW is relatively poor (on the right hand side of the graph) and the U.S. holds more than 70% of total wealth, then any additional increase in the U.S. wealth decreases its lending to the ROW. The ROW would like to borrow but is not rich enough to post collateral. The borrowing constraint becomes binding. The

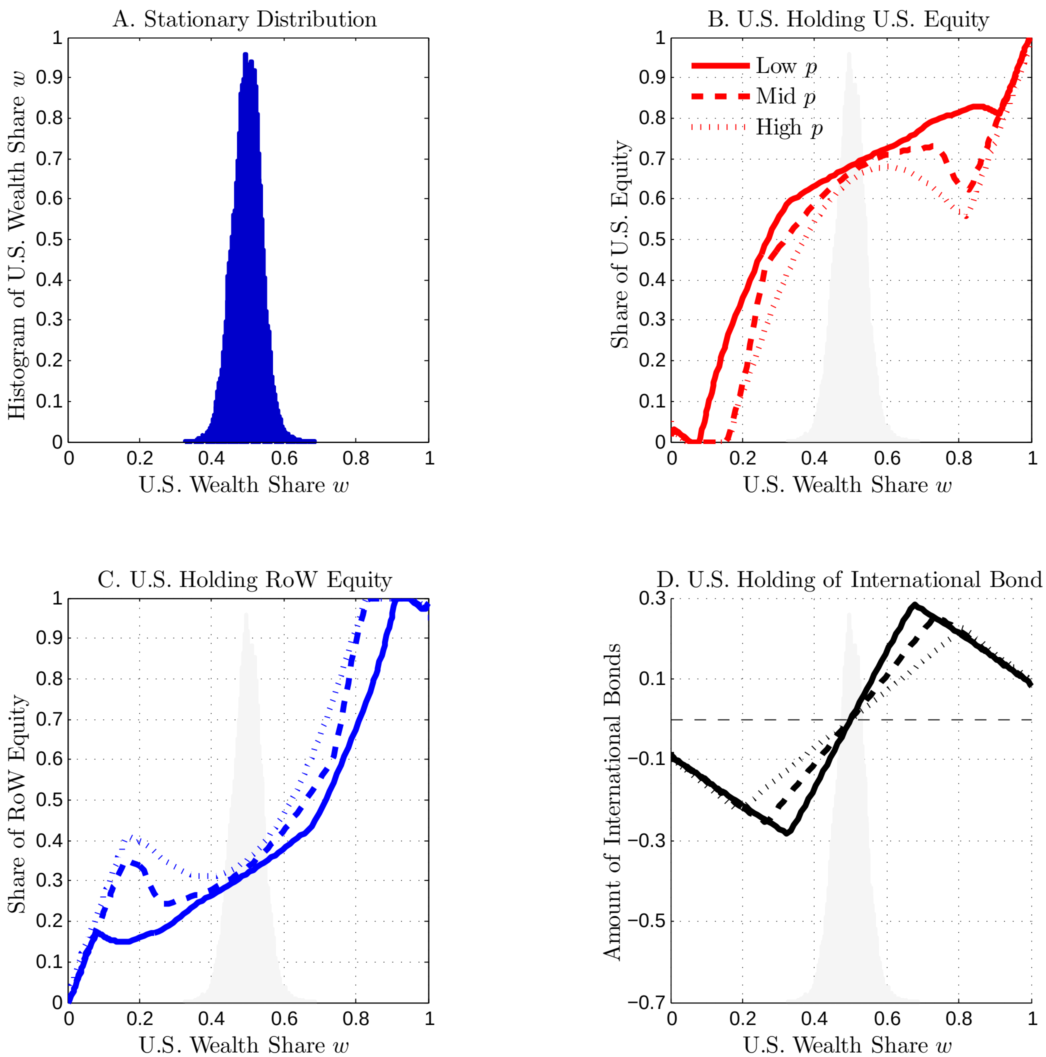

Figure 4-1: Relative Wealth and Asset Holdings in the Symmetric Case

This is the symmetric case with $\gamma_1 = \gamma_2 = 4$ and $\psi_1 = \psi_2 = 2$. The Panel A of this figure reports the stationary distribution of relative wealth, defined as $w_t \equiv W_{1,t}/[W_{1,t} + W_{2,t}]$, where the country 1 corresponds to the U.S. and country 2 the ROW.

distribution of relative wealth shows that this state of the world happens rarely in the model.

The upper right panel describes the U.S. holdings of U.S. equity. The home bias in consumption implies that the U.S. holds more than half of U.S. equity even when the two countries share the same wealth level. When the U.S. become relatively richer, they invest more in their own equity. The increase in their equity holdings is not monotone. At high wealth level, the binding borrowing constraint of the ROW impacts the U.S. equity choice. Because the U.S. cannot lend as much as they would like, they adjust their equity position downwards. This mechanism is particularly strong when the disaster probability is high, and thus equity prices are low: in that case, the ROW has less collateral and borrows less, thus affecting more the equity holdings of the U.S. At the other extreme, when the U.S. is relatively very poor, the U.S. would like to short their own equity, but the short-selling constraint on equity binds, and the U.S. simply stop holding equity. The lower left panel describes the U.S. holdings of the ROW equity. Since equity is either held by the U.S. or the ROW, the set of policy functions in that panel mirrors the previous one.

Asymmetric Countries We turn now to the asymmetric case. Figure 4-2 reports the distribution of relative wealth and the policy functions in that model. As Panel A shows, the simulation delivers again a stationary distribution of relative wealth. The U.S., which is less risk-averse, tends to be wealthier on average than the ROW.

The three other panels describe the U.S. holdings of the U.S. equity, ROW equity, and international bonds. At the mode of relative wealth, the U.S. holds a large share of U.S. equity (Panel B of Figure 4-2), again in line with the well-known home equity bias, but also a large share of foreign equity (Panel C of Figure 4-2). To do so, the U.S. tends to borrow from the ROW (Panel D of Figure 4-2) and thus exhibits a levered position in equity markets: borrowing on average from the ROW in order to buy U.S. and ROW equity. Only when the U.S. is much much wealthier than the ROW does the U.S. lend to the ROW. As in the symmetric case, the U.S. lending increases with U.S. relative

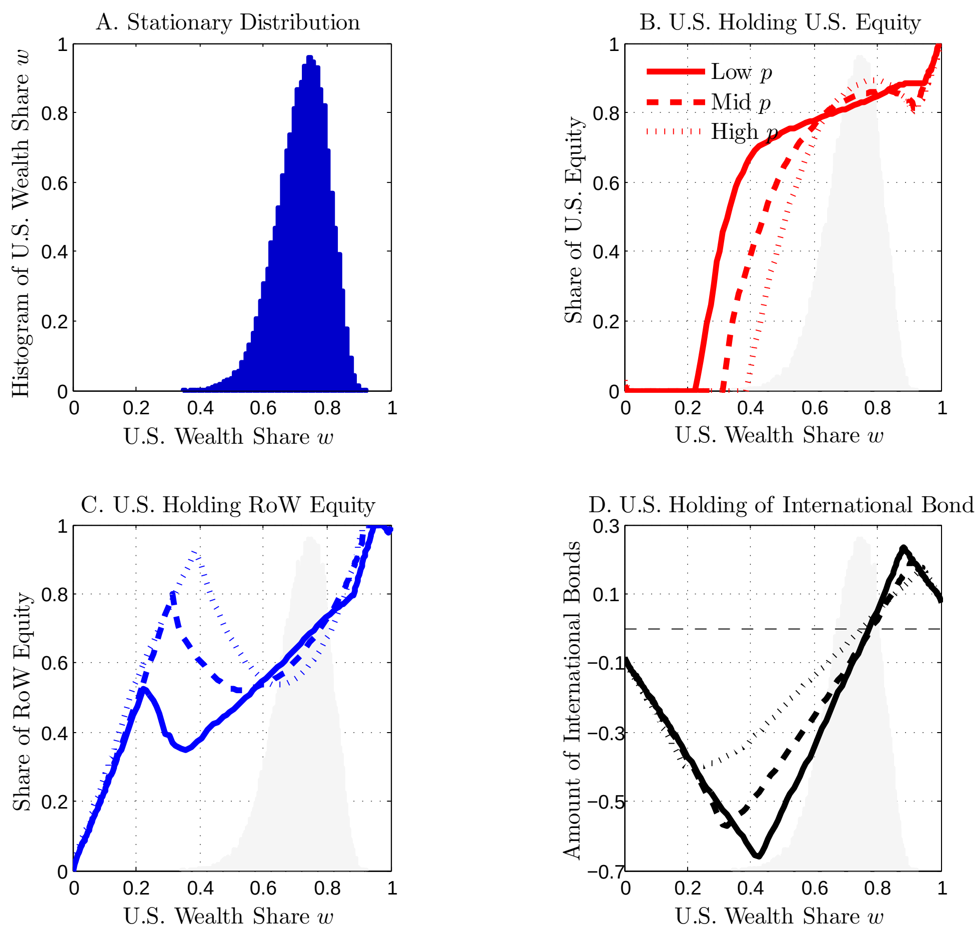

Figure 4-2: Relative Wealth and Asset Holdings

The Panel A of this figure reports the stationary distribution of relative wealth, defined as $w_t \equiv W_{1,t}/[W_{1,t} + W_{2,t}]$, where the country 1 corresponds to the U.S. and country 2 the ROW.

wealth up to a point, where the borrowing constraint binds for the ROW: the ROW is then so poor that it can no longer collateralize its borrowing. After that point, the U.S. lending decreases with the U.S. relative wealth.

4.5.2 International Capital Stocks and Flows

We turn now to the comparison between actual and simulated foreign capital stocks and flows.

Stocks Over the last forty years, the total stocks of U.S. foreign assets and liabilities (even scaled by U.S. GDP) has increased tremendously from less than 10% to more than 160%. The large increase in international positions occurs across all four categories of investments reported in the balance of payments and international investments statistics: debt, equity, FDI, and other investments. It follows an increase in the financial openness of the US and ROW, as encoded for example from the restrictions on cross-border financial transactions reported in the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions. To parallel the model, we report statistics on two asset categories, equity vs debt, both built from the the Lane and Milesi-Ferretti dataset (2007). All "equity" stocks correspond to the sum of equity, foreign direct investment, and other investments. For debt, we focus on net debt holdings because the model features only one international bond. Net debt assets correspond to the difference between debt portfolio assets and liabilities.

Table 4.5 reports basic summary statistics on U.S. international stocks. Because of the trend in foreign holdings, we report statistics on raw data as well as on HP-filtered series.

While the average level of debt is slightly higher in the data than in the model, the average equity position is much higher in the model than in the data. The current calibration offers expected equity excess returns that are too large, inducing large foreign holdings. The foreign capital holdings are also more volatile in the model than in the data, particularly for equity assets. They are also too persistent compared to their ac-

Table 4.5: U.S. International Capital Stocks

	Raw Data			HP-Filtered Series			
	Min	Mean	Max	Std	AC(1)	Corr US GDP	Corr ROW GDP
				Panel I: Data			
US All "Equity" assets	13.62	45.63	108.86	6.25	0.23	0.40	0.04
US All "Equity" liabilities	8.93	39.12	84.24	4.20	0.39	0.44	0.04
US Net All "Equity" assets	-3.90	6.51	24.61	3.29	0.36	0.20	0.03
US Net Debt assets	-41.82	-14.02	-2.36	1.09	0.54	-0.32	-0.03
US Net Foreign assets	-29.54	-6.73	4.56	3.20	0.20	0.08	-0.03
				Panel II: Model			
US All "Equity" assets	118.41	356.45	742.15	39.64	0.69	0.19	-0.04
US All "Equity" liabilities	0.00	69.12	444.46	13.70	0.69	0.27	0.00
US Net All "Equity" assets	77.05	287.33	484.13	29.03	0.69	0.13	0.05
US Net Debt assets	-71.43	-9.72	20.17	2.23	0.69	-0.11	-0.04
US Net Foreign assets	43.74	277.61	500.19	27.41	0.69	0.13	0.04

Notes: This table reports the min, mean, max, standard deviation, autocorrelation, and cross-country correlation of U.S. international capital stocks in different asset classes. All "equity" stocks correspond to the sum of equity, foreign direct investment, and other investments. Net all "equity" assets correspond to the difference between all "equity" assets and liabilities. Net debt assets correspond to the difference between debt portfolio assets and liabilities. The last two columns correspond to the cross-country correlation coefficients between international capital flows and U.S. or rest-of-the-world (ROW) HP-filtered GDP series. All series are scaled by GDP. The min, mean, and max statistics are computed on raw data, while the standard deviation, autocorrelation, and correlations are computed on HP-filtered series. Standard errors are reported in parentheses; they are obtained by block-boostrapping. Data are annual, from the Lane and Milesi-Ferretti dataset and the OECD. All variables are reported in percentage points, except for the autocorrelation and cross-country correlation coefficients. The sample period is 1973–2010.

tual counterparts. The model, however, captures the cyclicality of U.S. equity assets and liabilities with respect to the U.S. GDP, as well as the counter-cyclicality of the net U.S. debt position.

Flows In the data, the large increase in total assets and liabilities is accompanied by a large increase in the size and volatility of all categories of international capital flows. Balance of payments record international capital flows at the quarterly frequency, distinguishing between foreign direct investment, portfolio flows, and the remainder, denoted "other flows."[5] To quantify the volatility of the capital flows, Table 4.6 reports

[5]Gross outflows are defined as net purchases of foreign financial instruments by domestic residents. Gross inflows are defined as net sales of domestic financial instruments to foreign residents. By convention, negative outflows mean that residents are buying more foreign assets than they are selling,

some simple summary statistics. Total U.S. equity outflows attain more than 13% of GDP, and sometimes even reverse sign. The total equity inflows amount to close to 12% GDP at their maximum.

Turning to HP-filtered series to eliminate the trends, both equity inflows and outflows exhibit a low but significant autocorrelation of around 0.2. The autocorrelation of net equity flows is only 0.1, much lower than the autocorrelation of net debt inflows (0.3). The total net inflows (debt and equity) are essentially uncorrelated. Total gross inflows and outflows tend to increase (more capital flowing abroad and in the U.S.) when US and ROW GDP are high, delivering significant correlation coefficients between capital flows and GDP series.

Table 4.6 also shows that capital flows tend to shrink in times of high aggregate volatility. We measure aggregate volatility as the cross-country average of the realized standard deviations of daily equity returns over each quarter. When aggregate volatility increases, capital outflows out of the U.S. become less negative, i.e. shrink in magnitude. likewise, capital inflows in the U.S. decrease. Such correlations appear clearly for equity and debt portfolio flows, as well as for the "other" flows and the total inflows and outflows. Foreign direct investment and net capital flows, however, do not exhibit any significant correlation with aggregate volatility. These correlation are best exemplified during the Great Recession. As already noted by several authors, the Great Recession is characterized by retrenchment: foreigners pull out their wealth out of U.S. equity and equity-like assets (equity inflows turn negative), while U.S. residents repatriate part of their foreign equity-like holdings (outflows turn positive). These unusual patterns coincide with large increases in world volatility, from pre-crisis levels of 20% to close to 60% (in annualized terms). Net debt inflows remain positive during the spike in volatility but turn negative when volatility recesses.

As in the data, our model produces volatile stock holdings because the value of the

contributing positively to negatively to net inflows. Intuitively, a negative outflow means than money is leaving the home country and flowing to the foreign country. Positive inflows means that foreigners are purchasing more domestic assets than they are selling, contributing positively to net inflows. Intuitively, a positive inflow means that money is flowing into the home country. Up to accounting errors, net inflows are then the sum of gross outflows and gross inflows.

Table 4.6: U.S. International Capital Flows

	Raw Data			HP-Filtered Series				
	Min	Mean	Max	Std	AC(1)	Corr US GDP	Corr ROW GDP	Corr World Vol.
				Panel I: Data				
US All "Equity" Outflows	-13.43	-2.94	5.82	2.55	0.21	-0.22	-0.28	-0.06
	(1.22)	(0.33)	(1.02)	(0.29)	(0.08)	(0.09)	(0.10)	(0.12)
US All "Equity" Inflows	-5.89	3.16	12.43	2.42	0.18	0.26	0.24	0.12
	(1.47)	(0.33)	(1.00)	(0.29)	(0.09)	(0.09)	(0.09)	(0.10)
US All "Equity" Net Inflows	-3.56	0.22	4.70	1.40	0.11	0.05	-0.10	0.10
	(0.19)	(0.15)	(0.27)	(0.10)	(0.07)	(0.09)	(0.09)	(0.10)
US Net Debt Inflows	-3.55	1.73	8.56	1.23	0.30	0.24	0.36	0.05
	(1.35)	(0.25)	(0.82)	(0.18)	(0.06)	(0.11)	(0.12)	(0.07)
US Net Capital Inflows	-2.31	1.96	9.04	1.31	0.03	0.25	0.20	0.14
	(0.26)	(0.27)	(0.88)	(0.16)	(0.06)	(0.10)	(0.10)	(0.07)
				Panel II: Model				
US All "Equity" Outflows	-348.34	-0.15	415.95	17.38	-0.09	-0.08	0.08	-0.04
US All "Equity" Inflows	-280.42	-0.01	237.17	10.52	-0.09	0.08	-0.08	0.03
US All "Equity" Net Inflows	-116.75	-0.17	146.91	7.54	-0.09	-0.07	0.07	-0.06
US Net Debt Inflows	-34.58	-0.00	32.18	1.79	-0.09	-0.01	-0.04	0.14
US Net Capital Inflows	-128.24	-0.17	157.13	6.91	-0.09	-0.07	0.07	-0.03

Notes: This table reports the min, mean, max, standard deviation, autocorrelation, and cross-country correlation of U.S. international capital flows in different asset classes. All "equity" flows correspond to the sum of equity, foreign direct investment, and other investments. Net debt flows correspond to the sum of debt portfolio inflows and outflows. The next two columns correspond to the cross-country correlation coefficients between international capital flows and U.S. or rest-of-the-world (ROW) HP-filtered GDP series. The last column corresponds to the cross-country correlation coefficients between international capital flows and the change in world equity volatility. All series are scaled by GDP. The min, mean, and max statistics are computed on raw data, while the standard deviation, autocorrelation, and correlations are computed on HP-filtered series. Standard errors are reported in parentheses; they are obtained by block-boostrapping. Data are quarterly, from the Bluedorn, Duttagupta, Guajardo, and Topalova (2013) dataset, Datastream, and the OECD. All variables are reported in percentage points, except for the autocorrelation and cross-country correlation coefficients. The sample period is 1973.4–2010.4.

stock holdings move a lot. Recall that in the model as in the data, stock returns exhibit a 16% annualized volatility. The volatility of equity flows is much higher in the model, as it is in the data. Turning to bonds, the model reproduces the volatility of net debt flows, with little valuation effects.

Overall the model thus reproduces the start contrast between the volatility of the U.S. foreign assets and liabilities. Changes in debt liabilities are mostly due to changes in the amount of borrowing and thus international debt flows. To the contrary, changes in equity assets are mostly due to valuation changes. In the model, changes in equity prices and thus returns are either expected and unexpected. The large expected returns on ROW equity help the U.S. finance its negative trade balance and reimburse its debt. The model does not feature sovereign default and the negative trade balance is sustainable.

4.6 Conclusion

This paper presents a two-good, two-country real model that replicates basic stylized facts on equity excess returns and real interest rates. In the model, the U.S. borrows from the ROW and invests in ROW equity. The gross foreign asset positions are large and volatile. The changes in asset positions reflect both capital flows and changes in the value of the existing assets. The returns on existing assets feature an expected component that compensate investors for the risk of losing money in times of high marginal utility. Valuation effects appear key to understand the volatility of international asset holdings and the sustainability of current account imbalances.

Appendix A

Appendix: Measuring the "Dark Matter" in Asset Pricing Models

A.1 Disaster risk model

A.1.1 Asymptotic fragility measure

The probability density for (g, z) in the baseline model is

$$\pi_{\mathbb{P}}(g, z; \theta) = p^z (1-p)^{1-z} \left[\frac{1}{2\pi\sigma} \exp\left\{ -\frac{(g-\mu)^2}{2\sigma^2} \right\} \right]^{1-z} \times \left[\mathbf{1}_{\{-g>\underline{v}\}} \lambda \exp\left\{ -\lambda(-g-\underline{v}) \right\} \right]^z. \tag{A.1}$$

The Fisher information matrix for (p, λ) under the baseline model $\mathbb{P}$` is

$$\mathbf{I}_{\mathbb{P}}(p, \lambda) = \begin{bmatrix} \frac{1}{p(1-p)} & 0 \\ 0 & \frac{p}{\lambda^2} \end{bmatrix}. \tag{A.2}$$

Next, to derive the probability density function $\pi_{\mathbb{Q}}(g, r, z | \theta, \psi)$ in the structural model, we simply substitute the risk premium η in $\pi_{\mathbb{P}}(g, z; \theta)$ with the asset pricing constraint

(3.38) and add the indicator function for the restrictions on parameters:

$$\begin{aligned}
\pi_{\mathbb{Q}}(g,r,z|\theta,\psi) &= p^z(1-p)^{1-z} \\
&\times \left[\frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}}\right. \\
&\times \left.\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(g-\mu)^2}{\sigma^2}+\frac{(r-\eta(\theta,\psi))^2}{\tau^2}-\frac{2\rho(g-\mu)(r-\eta(\theta,\psi))}{\sigma\tau}\right]\right\}\right]^{1-z} \\
&\times \left[\mathbf{1}_{\{-g>\underline{v}\}}\lambda\exp\{-\lambda(-g-\underline{v})\}\frac{1}{\sqrt{2\pi}\nu}\exp\left\{-\frac{1}{2\nu^2}(r-bg)^2\right\}\right]^z \mathbf{1}_{\{\eta(\theta,\psi)>\underline{\eta}^*,\lambda>\gamma\}},
\end{aligned}$$

where $\underline{\eta}^*$ is a lower bound for the risk premium and

$$\eta(\theta,\psi) \equiv \gamma\rho\sigma\tau - \frac{\tau^2}{2} + e^{\gamma\mu-\frac{\gamma^2\sigma^2}{2}}\lambda\left(\frac{e^{\gamma\underline{v}}}{\lambda-\gamma} - e^{\frac{1}{2}\nu^2}\frac{e^{(\gamma-b)\underline{v}}}{\lambda+b-\gamma}\right)\frac{p}{1-p}. \tag{A.3}$$

Using the notation introduced by (3.39) and (3.41), we can express the Fisher information for (p,λ) under the constrained model $\mathbb{Q}_{\theta,\psi}$ as

$$\mathbf{I}_{\mathbb{Q}}(p,\lambda) = \begin{bmatrix} \frac{1}{p(1-p)} + \frac{\Delta(\lambda)^2}{(1-\rho^2)\tau^2}\frac{e^{2\gamma\mu-\gamma^2\sigma^2}}{(1-p)^3} & \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma\mu-\gamma^2\sigma^2}}{(1-p)^2}\Delta(\lambda)\dot{\Delta}(\lambda) \\ \frac{p}{(1-\rho^2)\tau^2}\frac{e^{2\gamma\mu-\gamma^2\sigma^2}}{(1-p)^2}\Delta(\lambda)\dot{\Delta}(\lambda) & \frac{p}{\lambda^2} + \frac{\dot{\Delta}(\lambda)^2}{(1-\rho^2)\tau^2}e^{2\gamma\mu-\gamma^2\sigma^2}\frac{p^2}{1-p} \end{bmatrix}. \tag{A.4}$$

Following the definition in (3.19) and Proposition 9, the asymptotic fragility measure is the sum of eigenvalues of matrix $\mathbf{I}_{\mathbb{Q}}(\theta)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta)^{1/2}$. This matrix has identical eigenvalues with the matrix $\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\mathbf{I}_{\mathbb{Q}}(\theta)\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}$. The latter eigenvalue problem is much easier to compute compared to the former one. In this case, the eigenvalues are available in simple closed form. This gives us the formula for $\varrho_a(\theta) = \varrho_a^2(\theta)$ in (3.40). The minimum information ratio is $\varrho_a^{v_2^*}(\theta) = 1$.

A.1.2 Posteriors

Next, we construct the posteriors of the parameters $\theta = (p,\lambda)$ and $\psi = \gamma$ under the baseline model and the structural model, respectively. We appeal to the Jeffreys prior for (p,λ) under the model without asset pricing constraint as the econometrician's

Table A.1: Independent Jeffreys/Reference priors for parameters

Parameters	Prior PDF (up to a constant)
p	$p^{-1/2}(1-p)^{-1/2}$
λ	$\lambda^{-1}\mathbf{1}_{(\lambda>0)}$

prior. The structural parameter γ has an independent prior $\pi(\gamma)$. The prior for γ can be delta distributions like in Figure 3-3 or the uniform priors as in Section 3.3.1. Given the likelihood function in (A.1), the parameters are mutually independent under the Jeffreys prior and their probability density functions (PDFs) are explicitly specified in Table A.1.

The constrained likelihood function (given by equation (**??**)) is "nonstandard" when we impose equality and inequality constraints on the parameters. Given the independent reference priors specified in Table A.1 and the "nonstandard" likelihood function, not only the analytical form of the posterior density function becomes inaccessible, but also the traditional Monte Carlo methods designed to draw i.i.d. samples from the posterior become inefficient. For simulations of posterior based on a "nonstandard" likelihood function, one of the general methods is the Approximate Bayesian Computation (ABC) method.[1] One issue concerning with applying the conventional ABC method to our disaster risk model is the lack of efficiency when the priors are flat. Given the specific structure of our problem, we propose a tilted ABC method to boost the speed of our simulation. The details of the procedure are in Appendix A.1.3.

A.1.3 ABC Method and Implementation

Given the special structure of our problem, we propose a tilted ABC method in the hope of boosting the speed of our simulation. The algorithm described here is for the case of joint estimation with the risk aversion coefficient γ. We illustrate the case where

[1] For general introduction to the ABC method, see Blum (2010) and Fearnhead and Prangle (2012), among others.

γ has the prior $\pi(\gamma)$. The algorithm can be adapted easily for the special case where the value of γ is fixed.

The posterior for (p, λ, γ) under the baseline model satisfies

$$
\begin{aligned}
p, \lambda, \gamma \mid \mathbf{r}, \mathbf{g}, \mathbf{z} \sim \text{Beta}\left(p|0.5 + n - \kappa_n, 0.5 + \kappa_n\right) \\
\otimes \text{Gamma}\left(\lambda|n - \kappa_n, \sum_{t=1}^{n} z_t(g_t - \underline{v})\right) \\
\otimes \pi(\gamma),
\end{aligned} \tag{A.5}
$$

where

$$
x_t = (g_t, r_t)^T, \quad \mu_n = \sum_{t=1}^{n}(1 - z_t)x_t / \sum_{t=1}^{n}(1 - z_t), \quad \kappa_n = \sum_{t=1}^{n}(1 - z_t), \quad \nu_n = \kappa_n - 1,
$$

$$
S_n = \sum_{t=1}^{n}(1 - z_t)(x_t - \mu_n)(x_t - \mu_n)^T, \quad s_n = \sum_{t=1}^{n} z_t(r_t - bg_t)^2.
$$

Define

$$
\overline{g} = \sum_{t=1}^{n}(1 - z_t)g_t/\kappa_n \quad \text{and} \quad \overline{r} = \sum_{t=1}^{n}(1 - z_t)r_t/\kappa_n.
$$

The posterior for (p, λ, γ) under the structural model satisfies:

$$
\begin{aligned}
\pi_Q(p, \lambda, \gamma|\mathbf{g^n}, \mathbf{r^n}, \mathbf{z^n}) \propto\ & p^{n-\kappa_n+1/2-1}(1 - p)^{\kappa_n+1/2-1} \\
& \times \mathbf{1}_{(\lambda>\gamma)}\lambda^{n-\kappa_n-1} \exp\left\{-\lambda \sum_{t=1}^{n} z_t(-g_t - \underline{v})\right\} \\
& \times \tau^{-1}(1 - \rho^2)^{-1/2} \\
& \times \exp\left\{-\frac{\kappa_n}{2(1 - \rho^2)\tau^2}\left[\eta(p, \lambda, \gamma) - \overline{r} - \rho\frac{\tau}{\sigma}(\mu - \overline{g})\right]^2\right\} \\
& \times \mathbf{1}_{\{\eta(p,\lambda,\gamma)>\underline{\eta}^*\}} \times \pi(\gamma).
\end{aligned} \tag{A.6}
$$

Then, the posterior distribution will not change if we view the model in a different way as follows:

$$\bar{r} \sim \mathrm{N}\left(\eta(p,\lambda,\gamma) + \rho\frac{\tau}{\sigma}(\bar{g}-\mu), \tau^2(1-\rho^2)\right) \text{ where } \eta(p,\lambda,\gamma) > \underline{\eta}^*,$$

with priors

$$\gamma \sim \pi(\gamma),$$

$$p \sim \text{Beta}(n-\kappa_n+1/2, \kappa_n+1/2),$$

$$\lambda \sim \text{Gamma}\left(\lambda | n-\kappa_n, \sum_{t=1}^{n} z_t(g_t-\underline{v}), \lambda > \gamma\right).$$

The tilted ABC method is implemented as follows.

Algorithm We illustrate the algorithm for simulating samples from the posterior (A.6) based ABC method. We choose the threshold in ABC algorithm as $\epsilon = \hat{\tau}/n/100$, where $\hat{\tau}$ is the sample standard deviation of the observations $r_1, \cdots, r_n$. Our tilted ABC algorithm can be summarized as follows:

For step $i = 1, \cdots, N$:

Repeat the following simulations and calculations:

(1) simulate $\tilde{\gamma} \sim \pi(\gamma)$,

(2) simulate $\tilde{p} \sim \text{Beta}(n-\kappa_n+1/2, \kappa_n+1/2)$,

(3) simulate $\tilde{\lambda} \sim \text{Gamma}\left(\lambda | n-\kappa_n, \sum_{t=1}^{n} z_t(g_t-\underline{v})\right)$,

(4) calculate $\tilde{\eta} = \eta(\tilde{\theta}, \tilde{\psi})$ with

$$\tilde{\theta} = (\tilde{p}, \tilde{\lambda}) \quad \text{and} \quad \tilde{\phi} = \tilde{\gamma},$$

(5) simulate $\tilde{r} \sim \mathrm{N}\left(\tilde{\eta} + \tilde{\rho}\frac{\tilde{\tau}}{\tilde{\sigma}}(\bar{g}-\tilde{\mu}), \tilde{\tau}^2(1-\tilde{\rho}^2)\right)$,

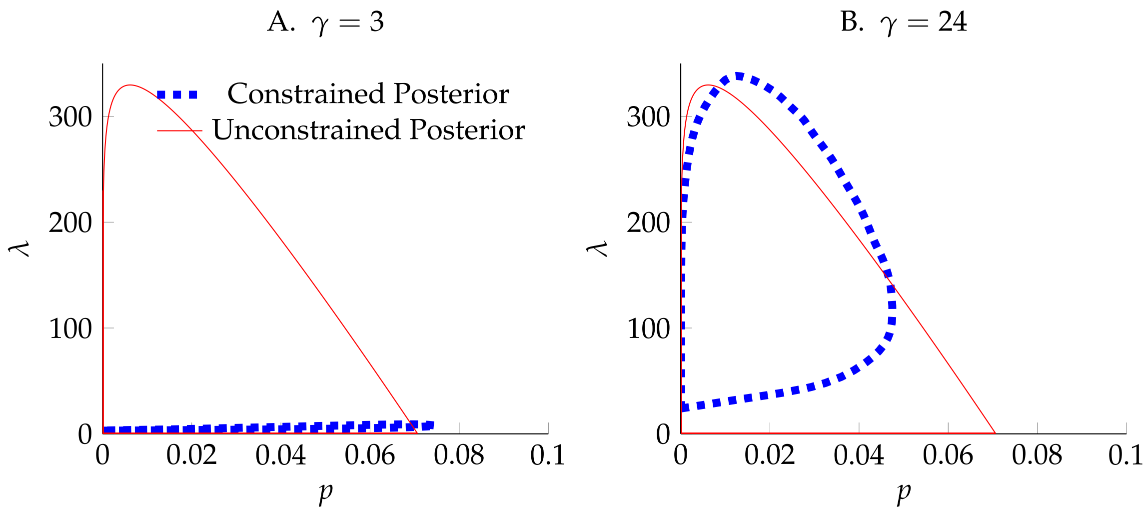

Figure A-1: The 95% Bayesian confidence regions for (p, λ). In the left panel, the posterior under the structural model (i.e. constrained posterior) sets $\gamma = 3$. In the right panel, the posterior under the structural model sets $\gamma = 24$. Both are compared with the posterior under the baseline model (i.e. unconstrained posterior).

Until (i) $|\tilde{r} - \bar{r}| < \epsilon$ and (ii) $\tilde{\eta} > \underline{\eta}^*$, we record

$$\theta^{(i)} = \tilde{\theta}$$
$$\phi^{(i)} = \tilde{\phi}$$

Set $i = i + 1$, if $i < N$; end the loop, if $i = N$.

Using this algorithm, we shall get simulated samples $\theta^{(1)}, \cdots, \theta^{(N)}$ from the posterior (A.6).

A.1.4 Results

Now, we show some examples of posteriors for (p, λ) when γ has delta priors. In Figure A-1 we illustrate their differences by plotting the 95% Bayesian confidence regions for (p, λ) according to the two posteriors. The 95% Bayesian region for (p, λ) under the baseline posterior distribution is similar to the 95% confidence region for (p, λ) under the baseline model (see Figure 3-1).

The shape of the 95% Bayesian region for the constrained posterior depends on the coefficient of relative risk aversion γ. When γ is high (e.g, $\gamma = 24$), the constrained posterior is largely similar to the unconstrained posterior (see Panel B), except that it assigns lower weight to the lower right region, because these relatively frequent and large disasters are inconsistent with the equity premium constraint. For a lower level of risk aversion, $\gamma = 3$, the constrained posterior is drastically different. The only parameter configurations consistent with the equity premium constraint are those with large average disaster size, with λ close to its lower limit γ.

A.2 Long-run Risk Model

We consider a long-run risk model similar to Bansal and Yaron (2004b) and Bansal, Kiku, and Yaron (2012b). The log growth rate of aggregate consumption Δc_t, the long-run risk component in consumption growth x_t, and stochastic volatility σ_t follow the joint processes

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_{p,t}\epsilon_{c,t+1} \tag{A.7a}$$

$$x_{t+1} = \rho x_t + \varphi_x \sigma_{p,t}\epsilon_{x,t+1} \tag{A.7b}$$

$$\sigma^2_{t+1} = \overline{\sigma}^2 + \nu(\sigma^2_t - \overline{\sigma}^2) + \sigma_w \epsilon_{\sigma,t+1} \tag{A.7c}$$

$$\sigma^2_{p,t+1} = \max(\underline{\sigma}^2, \sigma^2_{t+1}), \tag{A.7d}$$

where the shocks $\epsilon_{c,t}, \epsilon_{x,t}$, and $\epsilon_{\sigma,t}$ are i.i.d. standard normal variables and they are mutually independent. We formally denote the statistical model characterizing the three exogenous variables $(\Delta c_t, x_t, \sigma^2_t)$ with $t = 1, \cdots, n$ to be $\mathbb{P}$. The focus is the dynamic parameters $\theta \equiv (\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w)$.

The preference of the representative agent is assumed to be Epstein-Zin-Weil preference:

$$V_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\vartheta}} + \delta\left(\mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right]\right)^{\frac{1}{\vartheta}}\right]^{\frac{\vartheta}{1-\gamma}} \tag{A.8}$$

where $\vartheta = (1-\gamma)/(1-\psi^{-1})$. Define the wealth process and the gross return on consumption claims:

$$W_{t+1} = (W_t - C_t)R_{c,t+1}. \tag{A.9}$$

Therefore, the stochastic discount factor (SDF) can be expressed as follows:

$$M_{t+1} = \delta^{\vartheta} \left(\frac{C_{t+1}}{C_t}\right)^{-\vartheta/\psi} R_{c,t+1}^{\vartheta-1}. \tag{A.10}$$

The log SDF can be written as

$$m_{t+1} = \vartheta \log \delta - \frac{\vartheta}{\psi}\Delta c_{t+1} + (\vartheta - 1) r_{c,t+1}. \tag{A.11}$$

The state variables in long-run risk models are (x_t, σ_t^2). The log consumption growth rate Δc_{t+1} can be expressed in terms of x_t and σ_t^2 by assumption (3.43a). In contrast, the dependence of $r_{c,t+1}$ on the state variables are endogenous. To turn the system into an affine model, we first exploit the Campbell-Shiller log-linearization approximation:

$$r_{c,t+1} = \kappa_0 + \kappa_1 z_{t+1} + \Delta c_{t+1} - z_t, \tag{A.12}$$

where $z_t = \log(W_t/C_t)$ is log wealth-consumption ratio where wealth is the price of consumption claims. The log-linearization constants are determined by long-run steady state:

$$\kappa_0 = \log(1 + e^{\bar{z}}) - \kappa_1 \bar{z} \tag{A.13}$$

$$\kappa_1 = \frac{e^{\bar{z}}}{1 + e^{\bar{z}}}, \tag{A.14}$$

where $\bar{z}$ is the mean of the log price-consumption ratio.

Given the log-linearization approximation (A.12) - (A.14), we can search the equilibrium characterized by

$$z_t = A_0 + A_1 x_t + A_2 \sigma_t^2, \tag{A.15}$$

where the constants A_0, A_1 and A_2 are to be determined by the equilibrium conditions. Thus, the log return on consumption claim can be written as

$$r_{c,t+1} = \kappa_0 + \kappa_1 \left(A_0 + A_1 x_{t+1} + A_2 \sigma_{t+1}^2 \right) + \Delta c_{t+1} - \left(A_0 + A_1 x_t + A_2 \sigma_t^2 \right). \quad \text{(A.16)}$$

Therefore, the log SDF can be re-written in terms of state variables and exogenous shocks

$$m_{t+1} = \Gamma_0 + \Gamma_1 x_t + \Gamma_2 \sigma_t^2 - \lambda_c \sigma_t \epsilon_{c,t+1} - \lambda_x \sigma_t \varphi_x \epsilon_{x,t+1} - \lambda_\sigma \sigma_w \epsilon_{\sigma,t+1}, \quad \text{(A.17)}$$

where predictive coefficients are

$$\Gamma_0 = \log \delta - \psi^{-1} \mu_c - \frac{1}{2} \vartheta(\vartheta - 1) \left(\kappa_1 A_2 \sigma_w \right)^2 \quad \text{(A.18)}$$

$$\Gamma_1 = -\psi^{-1} \quad \text{(A.19)}$$

$$\Gamma_2 = (\vartheta - 1)(\kappa_1 \nu - 1) A_2 = \frac{1}{2} (\gamma - 1)(\psi^{-1} - \gamma) \left[1 + \left(\frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho} \right)^2 \right] \quad \text{(A.20)}$$

and the market price of risk coefficients are

$$\lambda_c = \gamma \quad \text{(A.21)}$$

$$\lambda_x = \left(\gamma - \psi^{-1} \right) \frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho} \quad \text{(A.22)}$$

$$\lambda_\lambda = -(\gamma - 1) \left(\gamma - \psi^{-1} \right) \frac{\kappa_1}{2(1 - \kappa_1 \nu)} \left[1 + \left(\frac{\kappa_1 \varphi_x}{1 - \kappa_1 \rho} \right)^2 \right] \quad \text{(A.23)}$$

It can be seen that as ρ or ν approaches to unit, the risk premium goes to infinity. The coefficients A_j's are determined by equilibrium condition (i.e. Euler Equation for price of consumption claim – pure intertemporal first-order condition of consumption decision), which is

$$1 = \mathbb{E}_t \left[M_{t+1} R_{c,t+1} \right] = \mathbb{E}_t \left[e^{m_{t+1} + r_{c,t+1}} \right] \quad \text{(A.24)}$$

It leads to the equilibrium conditions:

$$A_0 = \frac{1}{1-\kappa_1}\left[\log\delta + \kappa_0 + \left(1-\psi^{-1}\right)\mu_c + \kappa_1 A_2(1-\nu)\overline{\sigma}^2 + \frac{\vartheta}{2}\left(\kappa_1 A_2 \sigma_w\right)^2\right] \tag{A.25}$$

$$A_1 = \frac{1-\psi^{-1}}{1-\kappa_1\rho} \tag{A.26}$$

$$A_2 = -\frac{(\gamma-1)(1-\psi^{-1})}{2(1-\kappa_1\nu)}\left[1 + \left(\frac{\kappa_1\varphi_x}{1-\kappa_1\rho}\right)^2\right] \tag{A.27}$$

The long-run mean $\overline{z}$ is also determined endogenously in the equilibrium. More precisely, given all parameters fixed, we have $A_j = A_j(\overline{z})$ in Equations (A.25) – (A.27) because κ_0 and κ_1 are functions of $\overline{z}$. In the long-run steady state, we have

$$\overline{z} = A_0(\overline{z}) + A_2(\overline{z})\overline{\sigma}^2. \tag{A.28}$$

Thus, in the equilibrium, the long-run mean $\overline{z}$ is a function of all parameters in the model, according to (A.28) and Implicit Function Theorem,

$$\overline{z} = \overline{z}\left(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots\right). \tag{A.29}$$

And hence, we can also solve out κ_0 and κ_1 based on (A.29) as follows, whose explicit forms are usually not available

$$\kappa_0 = \kappa_0(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots) \;\text{ and }\; \kappa_1 = \kappa_1(\mu_c, \rho, \varphi_x, \overline{\sigma}^2, \nu, \sigma_w, \cdots). \tag{A.30}$$

The gradients κ_0 and κ_1 with respect to the parameters, such as ρ and ν, can be calculated using Implicit Function Theorem in (A.28).

Given the pricing kernel in the equilibrium, we can price assets. We specify the joint distribution of the exogenous state variables and the log dividend growth Δd_t, these joint distributional assumptions are part of the structural component of the model.

More precisely, we assume that the log dividend growth process is

$$\Delta d_{t+1} = \mu_d + \phi_d x_t + \varphi_{d,c}\sigma_{p,t}\epsilon_{c,t+1} + \varphi_{d,d}\sigma_{p,t}\epsilon_{d,t+1} + \sigma_{d,d}\epsilon^u_{d,t+1}. \tag{A.31}$$

Market Return

Using the Campbell-Shiller decomposition and linearization, we can represent the return in terms of log price-dividend ratio and log dividend growth:

$$r_{m,t+1} = \kappa_{m,0} + \kappa_{m,1} z_{m,t+1} + \Delta d_{t+1} - z_{m,t}, \tag{A.32}$$

where

$$\kappa_{m,0} = \log(1 + e^{\bar{z}_m}) - \kappa_{m,1}\bar{z}_m \tag{A.33}$$

and

$$\kappa_{m,1} = \frac{e^{\bar{z}_m}}{1 + e^{\bar{z}_m}} \tag{A.34}$$

and $\bar{z}_m$ is long-run mean of market log price-dividend ratio. We search for the equilibrium where the log market price-dividend ratio is a linear function of the states:

$$z_{m,t} = A_{m,0} + A_{m,1} x_t + A_{m,2}\sigma^2_t, \tag{A.35}$$

where the constants $A_{m,0}$, $A_{m,1}$ and $A_{m,2}$ are to be determined by equilibrium condition (i.e. Euler equation for market returns). Thus, we have

$$\begin{aligned} r_{m,t+1} - \mathbb{E}_t\left[r_{m,t+1}\right] = {} & \varphi_{d,c}\sigma_{p,t}\epsilon_{c,t+1} + \kappa_{m,1}A_{m,1}\varphi_x\sigma_{p,t}\epsilon_{x,t+1} \\ & + \kappa_{m,1}A_{m,2}\sigma_w\epsilon_{\sigma,t+1} + \varphi_{d,d}\sigma_{p,t}\epsilon_{d,t+1} + \sigma_{d,u}\epsilon^u_{d,t+1}, \end{aligned} \tag{A.36}$$

where

$$\mathbb{E}_t\left[r_{m,t+1}\right] = \mu_d + \kappa_{m,0} + (\kappa_{m,1} - 1)A_{m,0} + \kappa_{m,1}A_{m,2}(1-\nu)\bar{\sigma}^2 \tag{A.37}$$

$$+ \left[\phi_d + (\kappa_{m,1}\rho - 1)A_{m,1}\right] x_t + (\kappa_{m,1}\nu - 1)A_{m,2}\sigma^2_t. \tag{A.38}$$

Plugging the equation above into the Euler Equation

$$1 = \mathbb{E}_t \left[e^{m_{t+1} + r_{m,t+1}} \right], \tag{A.39}$$

we can derive the coefficients

$$A_{m,0} = \frac{1}{1 - \kappa_{m,1}} \times \left[\Gamma_0 + \kappa_{m,0} + \mu_d + \frac{1}{2}\sigma_{d,u}^2 + \kappa_{m,1} A_{m,2}(1-\nu)\overline{\sigma}^2 + \frac{1}{2}(\kappa_{m,1}A_{m,2} - \lambda_w)^2 \sigma_w^2 \right]$$

$$A_{m,1} = \frac{\phi_d - \psi^{-1}}{1 - \kappa_{m,1}\rho} \tag{A.40}$$

and

$$A_{m,2} = \frac{1}{1 - \kappa_{m,1}\nu} \left[\Gamma_2 + \frac{1}{2} \left(\varphi_{d,d}^2 + (\varphi_{d,c} - \lambda_c)^2 + (\kappa_{m,1} A_{m,1} \varphi_x - \lambda_x)^2 \right) \right] \tag{A.41}$$

In sum, according to (A.36), the market return can be re-written as the following beta representation for the priced aggregate shocks:

$$r_{m,t+1} - \mathbb{E}_t \left[r_{m,t+1} \right] = \beta_c \sigma_{p,t} \epsilon_{c,t+1} + \beta_x \sigma_{p,t} \epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\epsilon,t+1} + \varphi_{d,d} \sigma_{p,t} \epsilon_{d,t+1} + \sigma_{d,u} \epsilon_{d,t+1}^u. \tag{A.42}$$

where the betas are

$$\beta_c = \varphi_{d,c}, \ \ \beta_x = \kappa_{m,1} A_{m,1} \varphi_x, \ \text{ and } \ \beta_\sigma = \kappa_{m,1} A_{m,2} \tag{A.43}$$

Excess Market Return and Equity Premium

The Euler Equations for market return and riskfree rate can be written in one equation

$$\mathbb{E}_t \left[e^{m_{t+1}} \right] = \mathbb{E}_t \left[e^{m_{t+1} + r_{m,t+1}^e} \right]. \tag{A.44}$$

The risk premium is given by the beta pricing rule:

$$\mathbb{E}_t\left[r^e_{m,t+1}\right] = \lambda_c \sigma^2_{p,t}\beta_c + \lambda_x \sigma^2_{p,t}\beta_x + \lambda_\sigma \sigma^2_w \beta_\sigma - \frac{1}{2}\sigma^2_{r_m,t}, \tag{A.45}$$

$$\text{where} \quad \sigma^2_{r_m,t} = \beta^2_c \sigma^2_{p,t} + \beta^2_x \sigma^2_{p,t} + \beta^2_\sigma \sigma^2_w + \varphi^2_{d,d}\sigma^2_{p,t} + \sigma^2_{d,u}. \tag{A.46}$$

Similarly, the long-run mean of log market price-dividend ratio is

$$\bar{z}_m = A_{m,0}(\bar{z}_m) + A_{m,2}(\bar{z}_m)\bar{\sigma}^2. \tag{A.47}$$

Based on (A.42), the excess log return of market portfolio $r^e_{m,t+1} = r_{m,t+1} - r_{f,t}$ has the following expression:

$$\begin{aligned} r^e_{m,t+1} - \mathbb{E}_t\left[r^e_{m,t+1}\right] &= \beta_c \sigma_{p,t}\epsilon_{c,t+1} + \beta_x \sigma_{p,t}\epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\epsilon,t+1} \\ &\quad + \varphi_{d,d}\sigma_{p,t}\epsilon_{d,t+1} + \sigma_{d,u}\epsilon^u_{d,t+1}. \end{aligned} \tag{A.48}$$

In sum, the equilibrium excess return follows the dynamics:

$$\begin{aligned} r^e_{m,t+1} = \mu^e_{r,t} &+ \beta_c \sigma_{p,t}\epsilon_{c,t+1} + \beta_x \sigma_{p,t}\epsilon_{x,t+1} + \beta_\sigma \sigma_w \epsilon_{\sigma,t+1} \\ &+ \varphi_{d,d}\sigma_{p,t}\epsilon_{d,t+1} + \sigma_{d,u}\epsilon^u_{d,t+1}, \end{aligned}$$

where

$$\mu^e_{r,t} = \lambda_c \beta_c \sigma^2_{p,t} + \lambda_x \beta_x \sigma^2_{p,t} + \lambda_\sigma \beta_\sigma \sigma^2_w - \frac{1}{2}\left(\beta^2_c \sigma^2_{p,t} + \beta^2_x \sigma^2_{p,t} + \beta^2_\sigma \sigma^2_w + \varphi^2_{d,d}\sigma^2_{p,t} + \sigma^2_{d,u}\right).$$

Generalized Methods of Moments

The likelihood function of the baseline statistical model $\mathbb{P}_{\theta,n}$ can be seen clearly when re-arrange the terms

$$\frac{\Delta c_{t+1} - \mu_c - x_t}{\sigma_{p,t}} = \epsilon_{c,t+1} \tag{A.49a}$$

$$\frac{x_{t+1} - \rho x_t}{\varphi_x \sigma_{p,t}} = \epsilon_{x,t+1} \tag{A.49b}$$

$$\frac{(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)}{\sigma_w} = \epsilon_{\sigma,t+1} \tag{A.49c}$$

where $\epsilon_{c,t}$, $\epsilon_{x,t}$ and $\epsilon_{\sigma,t}$ are i.i.d. standard normal variables and they are mutually independent. We consider the GMM where the moments functions are identical to the score functions of the likelihood function. Thus, it is equivalent to the MLE. Denote the set of moment functions to be $g_{\mathbb{P}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2; \theta)$. More precisely, $g_{\mathbb{P}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma_{t+1}^2, \sigma_t^2; \theta)$ includes six moment conditions:

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\Delta c_{t+1} - \mu_c - x_t}{\sigma_{p,t}^2}$$

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(x_{t+1} - \rho x_t) x_t}{\varphi_x^2 \sigma_{p,t}^2}$$

$$1 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(x_{t+1} - \rho x_t)^2}{\varphi_x^2 \sigma_{p,t}^2}$$

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\left[(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)\right](\sigma_t^2 - \overline{\sigma}^2)}{\sigma_w^2}$$

$$1 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{\left[(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)\right]^2}{\sigma_w^2}$$

$$0 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{(\sigma_{t+1}^2 - \overline{\sigma}^2) - \nu(\sigma_t^2 - \overline{\sigma}^2)}{\sigma_w^2}$$

In the long-run risk model, the major focus is to understand the stock excess return's dynamics explained by the consumption process specified in (A.49a) - (A.49c). The joint distribution of the excess log return $r_{m,t+1}^e$ and the consumption variables can be seen clearly from the following formula:

$$\frac{r^e_{m,t+1} - \mu^e_{r,t} - \beta_c(\Delta c_{t+1} - \mu_{c,t} - x_t) - \beta_x \frac{x_{t+1} - \rho x_t}{\varphi_x} - \beta_\sigma \left[\widehat{\sigma}^2_{t+1} - \nu \widehat{\sigma}^2_t\right]}{\sqrt{\varphi^2_{d,d}\sigma^2_{p,t} + \sigma^2_{d,u}}} = \varepsilon_{t+1}$$

where $\widehat{\sigma}^2_t \equiv \sigma^2_t - \overline{\sigma}^2$ and ε_t is a standard normal variable defined as

$$\varepsilon_{t+1} \equiv \frac{\varphi_{d,d}\sigma_{p,t}\epsilon_{d,t+1} + \sigma_{d,u}\epsilon^u_{d,t+1}}{\sqrt{\varphi^2_{d,d}\sigma^2_{p,t} + \sigma^2_{d,u}}},$$

and

$$\mu^e_{r,t} = \lambda_\eta \beta_\eta \sigma^2_{p,t} + \lambda_e \beta_e \sigma^2_{p,t} + \lambda_w \beta_w \sigma^2_w - \frac{1}{2}\left(\beta^2_\eta \sigma^2_{p,t} + \beta^2_e \sigma^2_{p,t} + \beta^2_w \sigma^2_w + \varphi^2 \sigma^2_{p,t}\right).$$

We choose the over-identification moment constraints

$$g_{\mathbb{Q}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, r^e_{m,t+1}; \theta)$$

to include the score functions of the conditional likelihood of $r^e_{m,t+1}$ above. Thus, the moment conditions for the optimal GMM setup to assess the fragility of the benchmark version of long-run risk model are

$$\begin{aligned} & g(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, r^e_{t+1}; \theta) \\ & \quad \equiv \begin{bmatrix} g_{\mathbb{P}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t; \theta) \\ g_{\mathbb{Q}}(\Delta c_{t+1}, x_{t+1}, x_t, \sigma^2_{t+1}, \sigma^2_t, r^e_{m,t+1}; \theta) \end{bmatrix}. \end{aligned} \tag{A.50}$$

Intuitively, the over-identification moment conditions imposed by the long-run risk model on the dynamic parameter θ is through the cross-equation restrictions on the beta coefficients $\beta_c, \beta_x, \beta_\sigma$ and the pricing coefficients $\lambda_c, \lambda_x, \lambda_\sigma$. Because of the shocks $\varepsilon_{t+1}, \epsilon_{c,t+1}, \epsilon_{x,t+1}$, and $\epsilon_{\sigma,t+1}$) are mutually independent, the GMM setup in (A.50) is actually asymptotically equivalent to the MLE for the joint distribution of $(\Delta c_t, x_t, \sigma^2_t, r^e_{m,t})$.

It should be noted that the whole joint distribution of the economic variables, including $(\Delta c_t, x_t, \sigma_t^2, r_{m,t}^e)$ and many other variables such as dividend growth and price-dividend ratio, may have stochastic singularities and many features that are not the targets of the long-run risk model to explain at the first place. Following the spirits of GMM-based estimation and hypothesis testing for structural models, we focus on the moments targeted by the particular long-run risk model, while discarding most parts of the whole joint distribution of the model. The analytical formulas of the over-identification moment conditions are super complicated, since the dependence of the beta coefficients and pricing coefficients is extremely complicated in the long-run risk model. We ignore the formulas here and, in fact, we calculate them numerically to compute the fragility measures. Moreover, we compute the Fisher Information matrixes for the moment conditions $g_{\mathbb{P}}$ and g based simulated stationary time-series data generated according to the Monte Carlo method.

A.3 Proofs

A.3.1 Proof of Proposition 1

If we define $\mathbf{u} = \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1/2} v$, we can rewrite the $\varrho_a^D(\theta_0)$ as

$$\begin{aligned}
\varrho_a^D(\theta_0) &= \max_{\mathbf{u} \in \mathbb{R}^{D_\Theta \times D}, \mathbf{Rank}(\mathbf{u}) = D} \mathbf{tr}\left[\left(\mathbf{u}^T \mathbf{u}\right)^{-1} \left(\mathbf{u}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2} \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2} \mathbf{u}\right)\right] \\
&= \max_{\mathbf{u} \in \mathbb{R}^{D_\Theta \times D}, \mathbf{Rank}(\mathbf{u}) = D} \mathbf{tr}\left[\mathbf{u} \left(\mathbf{u}^T \mathbf{u}\right)^{-1} \mathbf{u}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2} \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\right]
\end{aligned}$$

The linear operator $\mathcal{P}_{\mathbf{u}} \equiv \mathbf{u}\left(\mathbf{u}^T \mathbf{u}\right)^{-1} \mathbf{u}^T$ is the projection operator onto the subspace spanned by the column vectors of $\mathbf{u}$. Therefore, we have

$$\varrho_a^D(\theta_0) = \max_{\mathbf{u} \in \mathbb{R}^{D_\Theta \times D}, \mathbf{Rank}(\mathbf{u}) = D} \mathbf{tr}\left[\mathcal{P}_{\mathbf{u}} \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2} \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\right].$$

The projection operator can be equivalently expressed in terms of the orthonormal column vectors lying in the subspace spanned by $\mathbf{u}$. Thus, without loss of any generality, we can assume that the column vectors of $\mathbf{u}$ are orthonormal vectors, i.e. $\mathbf{u}^T\mathbf{u}$ is a D-dimensional identity matrix. Therefore,

$$\begin{aligned}
\varrho_a^D(\theta_0) &= \max_{\mathbf{u}\in\mathbb{R}^{D_\theta\times D},\mathbf{Rank}(\mathbf{u})=D,\mathbf{u}^T\mathbf{u}=I} \mathbf{tr}\left[\mathbf{u}\mathbf{u}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\right] \\
&= \max_{\mathbf{u}\in\mathbb{R}^{D_\theta\times D},\mathbf{Rank}(\mathbf{u})=D,\mathbf{u}^T\mathbf{u}=I} \mathbf{tr}\left[\mathbf{u}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{u}\right] \\
&= \max_{\mathbf{u}\in\mathbb{R}^{D_\theta\times D},\mathbf{Rank}(\mathbf{u})=D,\mathbf{u}^T\mathbf{u}=I} \sum_{i=1}^{D} u_i^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}u_i \\
&= \lambda_1 + \lambda_2 + \cdots + \lambda_D.
\end{aligned}$$

The argmax matrix is $\mathbf{u}^* = [e_1^*, e_2^*, \cdots, e_D^*]$ whose column vectors are the corresponding eigenvectors. Thus, correspondingly, the worst-case matrix is $v^* = [v_1^*, v_2^*, \cdots, v_D^*]$ with $v_i^* = \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}e_i^*$. Moreover, for $i = 1, \cdots, D$, it holds that

$$\lambda_i = \frac{(v_i^*)^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}v_i^*}{(v_i^*)^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}v_i^*}. \tag{A.51}$$

A.3.2 Chernoff Rate and Detection Error Probability

Proof of Proposition 11

In the proof, we consider a mathematically more general case where the matrix $\mathbf{v}$ for $\varrho^{\mathbf{v}}(\theta_0)$ is not necessarily a $D_\Theta \times D_\Theta$ identity matrix. We assume that $\mathbf{v}$ is a full-rank $D_{\mathbf{v}} \times D_\Theta$ matrix with $1 \leq D_{\mathbf{v}} \leq D_\Theta$. Similar to the proof of Proposition 9, we can show that there exists a system of orthonormal basis $[\tilde{v}_1, \cdots, \tilde{v}_{D_{\mathbf{v}}}]$ of the linear space spanned by the column vectors of $\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1/2}\mathbf{v}$ such that

$$\varrho_a^{\mathbf{v}}(\theta_0) = \sum_{i=1}^{D_{\mathbf{v}}} \tilde{v}_i^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\tilde{v}_i = \sum_{i=1}^{D_{\mathbf{v}}} \frac{\hat{v}_i^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\hat{v}_i}{\hat{v}_i^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\hat{v}_i}, \tag{A.52}$$

where $\hat{v}_i = \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\tilde{v}_i$. Because $\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}$ has exactly the same eigenvalues as $\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathbb{Q}}(\theta_0)\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}$, there exist unit vectors $\tilde{u}_1, \cdots, \tilde{u}_{D_{\mathbf{v}}}$ such that

$$\tilde{v}_i^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{1/2}\tilde{v}_i = \tilde{u}_i^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathbb{Q}}(\theta_0)\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\tilde{u}_i. \tag{A.53}$$

Define $u_i = \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\tilde{u}_i / |\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\tilde{u}_i|$, we have

$$\varrho_a^{\mathbf{v}}(\theta_0) = \sum_{i=1}^{D_{\mathbf{v}}} \frac{u_i^T \mathbf{I}_{\mathbb{Q}}(\theta_0) u_i}{u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0) u_i}. \tag{A.54}$$

Now, let's consider the Chernoff rates for the "perturbed" parameters

$$\theta_{u_i} \equiv \theta_0 + n^{-1/2} u_i, \quad \text{for } i = 1, \cdots, D_{\mathbf{v}}.$$

First, we have the following identity

$$\int \left[\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)\right]^{1-\alpha} \left[\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i})\right]^{\alpha} \mathrm{d}\mathbf{x^n} = \int \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0) e^{\alpha\left[\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)\right]} \mathrm{d}\mathbf{x^n}. \tag{A.55}$$

According to Lemma 7.6 in van der Vaart (1998), we know that the condition of differentiability in quadratic mean holds for density functions in our case. Then, the Local Asymptotic Normality (LAN) condition holds, i.e.,

$$\begin{aligned} &\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0) \\ &\quad = u_i^T \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \ln \pi_{\mathbb{P}}(\mathbf{x_t}; \theta_0) \right] - \frac{1}{2} u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0) u_i + R_n, \end{aligned} \tag{A.56}$$

where $\mathbb{E}_{\mathbb{P}_0}|R_n|^2 \to 0$ because of the **Assumption F**. Under the regularity conditions, the following CLT holds, according to Theorem 2.4 of **?**,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \ln \pi_{\mathbb{P}}(\mathbf{x_t}; \theta_0) \rightsquigarrow N\left(0, \mathbf{I}_{\mathbb{P}}(\theta_0)\right). \tag{A.57}$$

Define the Moment Generating Function (MGF) of $\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)$:

$$\begin{aligned} M_n(\alpha) &\equiv \mathbb{E}_{\mathbb{P}_0}\left\{e^{\alpha\left[\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) - \ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)\right]}\right\} \\ &= e^{-\alpha\frac{1}{2}u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0)u_i}\mathbb{E}_{\mathbb{P}_0}\left\{e^{\alpha u_i^T\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta_0)\right]}\right\} + \epsilon_n(\alpha), \end{aligned}$$

where $\epsilon_n(\alpha) = o(1)$ for each $\alpha \in [0,1]$. Therefore, as n goes large,

$$M_n(\alpha, u_i) \to e^{-\frac{1}{2}\alpha(1-\alpha)u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0)u_i}, \quad \forall\, \alpha \in [0,1]. \tag{A.58}$$

We denote the Cumulant Generating Function (CGF) as

$$\Lambda_n(\alpha, u_i) = \ln M_n(\alpha, u_i).$$

The CGF $\Lambda_n(\alpha, u_i)$ is convex in α. Because the pointwise convergence for a sequence of convex functions implies their uniform convergence to a convex function(see e.g., Rockafellar, 1970), we know that

$$\Lambda_n(\alpha, u_i) \to -\frac{1}{2}\alpha(1-\alpha)u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0)u_i. \tag{A.59}$$

Based on the definition of Chernoff information in (3.24) and the identity in (A.55), we know that

$$\begin{aligned} &C^*(\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) \;:\; \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)) \\ &\equiv \max_{\alpha\in[0,1]} -\ln \int \left[\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0)\right]^{\alpha}\left[\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i})\right]^{1-\alpha} \mathrm{d}\mathbf{x^n} \\ &= \max_{\alpha\in[0,1]} -\Lambda_n(\alpha, u_i) \to \max_{\alpha\in[0,1]} \frac{1}{2}\alpha(1-\alpha)u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0)u_i = \frac{1}{8}u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0)u_i. \end{aligned} \tag{A.60}$$

In Equation (A.60) above, the convergence of maxima of $-\Lambda_n(\alpha, u_i)$ to the maximum of $\frac{1}{2}\alpha(1-\alpha)u_i^T \mathbf{I}(\theta_0)u_i$ is guaranteed by the uniform convergence of $\Lambda_n(\alpha, u_i)$.

Similarly, we can show that

$$C^*(\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_{u_i}) \; : \; \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)) = \frac{1}{8} u_i^T \mathbf{I}_{\mathbb{Q}}(\theta_0) u_i + o(1). \tag{A.61}$$

Therefore, we have

$$\lim_{n\to\infty} \frac{C^*(\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_{u_i}) \; : \; \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) \; : \; \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0))} = \frac{\frac{1}{8} u_i^T \mathbf{I}_{\mathbb{Q}}(\theta_0) u_i}{\frac{1}{8} u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0) u_i} = \frac{u_i^T \mathbf{I}_{\mathbb{Q}}(\theta_0) u_i}{u_i^T \mathbf{I}_{\mathbb{P}}(\theta_0) u_i}$$

Combining with (A.54), we obtain

$$\varrho_a^{\mathbf{v}}(\theta_0) = \lim_{n\to\infty} \sum_{i=1}^{D_{\mathbf{v}}} \frac{C^*(\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_{u_i}) \; : \; \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0))}{C^*(\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{u_i}) \; : \; \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_0))}. \tag{A.62}$$

Detection Error Probability

This subsection is mainly based on Section 12.9 in Cover and Thomas (1991). Assume $X_1, \cdots, X_n$ i.i.d. $\sim Q$. We have two hypothesis or classes: $Q = P_1$ with prior π_1 and $Q = P_2$ with prior π_2. The overall probability of error (detection error probability) is

$$P_e^n = \pi_1 E_1^{(n)} + \pi_2 E_2^{(n)},$$

where $E_1^{(n)}$ is the error probability when $Q = P_1$ and $E_2^{(n)}$ is the error probability when $Q = P_2$. Define the best achievable exponent in the detection error probability is

$$D^* = \lim_{n\to\infty} \min_{A_n \in \mathcal{X}^n} -\frac{1}{n} \log_2 P_e^{(n)}, \;\; \text{where } A_n \text{ is the acceptance region.}$$

The Chernoff's Theorem shows that $D^* = C^*(P_1 : P_2)$. More precisely, Chernoff's Theorem states that the best achievable exponent in the detection error probability is D^*, where

$$D^* = \mathbf{D}_{KL}(P_{\alpha^*}||P_1) = \mathbf{D}_{KL}(P_{\alpha^*}||P_2),$$

with

$$P_\alpha = \frac{P_1^\alpha(x) P_2^{1-\alpha}(x)}{\int_{\mathcal{X}} P_1^\alpha(x) P_2^{1-\alpha}(x) \mathrm{d}x}$$

and α^* is the value of α such that

$$\mathbf{D}_{KL}(P_{\alpha^*}||P_1) = \mathbf{D}_{KL}(P_{\alpha^*}||P_2) = C^*(P_1 : P_2).$$

According to the Chernoff's Theorem, intuitively, the best achievable exponent in the detection error probability is

$$P_e^{(n)} \doteq \pi_1 2^{-n\mathbf{D}_{KL}(P_{\alpha^*}||P_1)} + \pi_2 2^{-n\mathbf{D}_{KL}(P_{\alpha^*}||P_2)} = 2^{-nC^*(P_1:P_2)}. \tag{A.63}$$

A.3.3 Asymptotic Equivalence Theorems

In this section, we prove the main results in Section 3.2 of the paper. We first introduce necessary notations in Subsection A.3.3. Then, we introduce the standard regularity conditions in Subsection A.3.3. Thirdly, we prove the basic lemmas in Subsection A.3.3, which are themselves interesting and general. Fourthly, in Subsection A.3.3, we state and prove the sequence of propositions which serve as intermediate steps for the proof of the main results. Finally, the main results are proved in Subsection A.3.3 - Subsection A.3.3.

Generic Notations and Definitions

First, we introduce some notations for the matrices. For any real symmetric non-negative definite matrix A, we define $\lambda_M(A)$ to be the largest eigenvalue of A and define $\lambda_m(A)$ to be the smallest eigenvalue of A. For a matrix A, we define the spectral norm of A to be $||A||_{\mathcal{S}}$. By definition of spectral norm, we know that for any real matrix A,

$$||A||_{\mathcal{S}} \equiv \sqrt{\lambda_M(A^T A)}.$$

Denote $\overline{\lambda}(\theta)$ and $\underline{\lambda}(\theta)$ to be the largest eigenvalue and the smallest eigenvalue of $\mathbf{I}_{\mathbb{P}}(\theta)$, respectively. That is,

$$\overline{\lambda}(\theta) = \lambda_m(\mathbf{I}_{\mathbb{P}}(\theta)) \quad \text{and} \quad \underline{\lambda}(\theta) = \lambda_M(\mathbf{I}_{\mathbb{P}}(\theta)).$$

If the matrix $\mathbf{I}_{\mathbb{P}}(\theta)$ is continuous in θ, $\overline{\lambda}(\theta)$ and $\underline{\lambda}(\theta)$ are continuous in θ. We define upper bound and lower bound to be

$$\overline{\lambda} \equiv \sup_{\theta\in\Theta} \overline{\lambda}(\theta), \text{ and } \underline{\lambda} \equiv \inf_{\theta\in\Theta} \underline{\lambda}(\theta). \tag{A.64}$$

Second, we introduce some notations related to subsets in Euclidean spaces. We define the "Euclidean distance" between two sets $S_1,\ S_2 \subset \mathbb{R}^{D_\Theta}$ as follows

$$d_L(S_1, S_2) \equiv \inf\{|s_1 - s_2| :\ s_i \in S_i,\ i = 1,2\}. \tag{A.65}$$

For $\theta \in \mathbb{R}^{D_\Theta}$, we denote $\theta_{(1)}$ to be the first element of θ and denote $\theta_{(-1)}$ to be the $D_\Theta - 1$ dimensional vector containing all elements of θ other than $\theta_{(1)}$. Define the open ball centered at θ with radius r to be

$$\Omega(\theta, r) \equiv \{\vartheta :\ |\vartheta - \theta| < r\} \ \text{ and } \ \overline{\Omega}(\theta, r) \equiv \{\vartheta :\ |\vartheta - \theta| \leq r\}$$

Denote

$$\Omega_{(1)}(\theta, r) = \Omega(\theta_{(1)}, r) \subset \mathbb{R}^1, \ \text{ and } \ \Omega_{(-1)}(\theta, r) = \Omega(\theta_{(-1)}, r) \subset \mathbb{R}^{D_\Theta - 1}.$$

In addition, we define

$$\Theta_{-1}(\theta_{(1)}) \equiv \left\{ \theta_{(-1)} \in \mathbb{R}^{D_\Theta - 1} \mid \begin{pmatrix} \theta_{(1)} \\ \theta_{(-1)} \end{pmatrix} \in \Theta \right\}, \tag{A.66}$$

and we denote

$$V_\Theta \equiv \mathrm{Vol}(\Theta) < +\infty, \quad V_\Theta(\theta_{(1)}) \equiv \mathrm{Vol}(\Theta_{-1}(\theta_{(1)})) < +\infty,$$

$$\text{and } V_{\Theta,1} \equiv \sup_{\theta_{(1)} \in \Theta_{(1)}} V_\Theta(\theta_{(1)}) < +\infty.$$

Third, we introduce some notations on metrics of probability measures. Consider two probability measures P and Q with densities p and q with respect to Lebesgue measure, respectively. The Hellinger affinity between P and Q is denoted as

$$\alpha_H(P, Q) \equiv \int \sqrt{p(x)q(x)}\mathrm{d}x.$$

The total variation distance between P and Q is denoted as

$$||P - Q||_{TV} \equiv \int |p(x) - q(x)|\mathrm{d}x.$$

Fourth, we introduce notations for time series. The maximal correlation coefficient and the uniform mixing coefficient are defined as

$$\rho_{\max}(\mathcal{F}_1, \mathcal{F}_2) \equiv \sup_{f_1 \in L^2_{\text{real}}(\mathcal{F}_1), f_2 \in L^2_{\text{real}}(\mathcal{F}_2)} |\mathrm{Corr}(f_1, f_2)|,$$

and

$$\phi(\mathcal{F}_1, \mathcal{F}_2) \equiv \sup_{A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2} |P(A_2|A_1) - P(A_2)|,$$

where $L^2_{\text{real}}(\mathcal{F}_i)$ denote the space of square-integrable. $\mathcal{F}_i$-measurable, real-valued random variables, for any sub $\sigma-$fields $\mathcal{F}_i \subset \mathcal{F}$.

At last, we introduce some notations related to generalized method of moments. For the moment function $g(\mathbf{x_t}, \mathbf{y_t}, \theta)$ we define

$$g_0(\theta) \equiv \mathbb{E}_{\mathbb{Q}_0}\left[g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right] \text{ and } G_0(\theta) \equiv \mathbb{E}_{\mathbb{Q}_0}\left[\frac{\partial}{\partial\theta} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right]. \tag{A.67}$$

We simply denote $g_0(\theta_0)$ as g_0 and denote $G_0(\theta_0)$ as G_0. The J-statistic is

$$J_{n,S_0}(\theta, \psi; \mathbf{x^n}, \mathbf{y^n}) \equiv n g_n(\mathbf{x^n}, \mathbf{y^n}; \theta, \psi)^T S_0^{-1} g_n(\mathbf{x^n}, \mathbf{y^n}; \theta, \psi), \tag{A.68}$$

where S_0 has the following explicit formula

$$S_0 = \sum_{\ell=-\infty}^{+\infty} \mathbb{E}_{\mathbb{Q}_0} \left[g(\mathbf{x_t}, \mathbf{y_t}; \theta_0, \psi_0) g(\mathbf{x_{t-\ell}}, \mathbf{y_{t-\ell}}; \theta_0, \psi_0)^T \right]. \tag{A.69}$$

The GMM likelihood ratio statistic is

$$d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, \theta\} = J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \theta) - J_{n,S_0}(\mathbf{x^n}, \mathbf{y^n}; \widehat{\theta}^{\mathbb{Q}}), \tag{A.70}$$

where $\widehat{\theta}^{\mathbb{Q}}$ is the GMM estimator.

The Regularity Conditions

The regularity conditions we choose to impose on the behavior of the data are influenced by three major considerations. First, our assumptions are chosen to allow processes of sequential dependence. In particular, the processes allowed should be relevant to inter-temporal asset pricing models. Second, our assumptions are required to meet the analytical tractability. Third, our assumptions are sufficient conditions in the sense that we are not trying to provide the weakest conditions or high level conditions to guarantee the results work, but instead, we chose those which are relatively straightforward to check in practice.

Assumption S

The observed data $(\mathbf{x_t}, \mathbf{y_t})$ for $t = 1, 2, \cdots, n$ follow a strictly stationary process and have a joint stationary distribution $\mathbb{Q}_{0,n}$. Marginally, the sample $\mathbf{x_t}$ for $t = 1, 2, \cdots, n$ has a stationary distribution $\mathbb{P}_{0,n} \equiv \mathbb{P}_{\theta_0,n}$ within a parametric family $\{\mathbb{P}_{\theta,n} : \theta \in \Theta\}$.

Assumption MD

The time series $\mathbf{x_t}$ is first-order Markov. This is without loss of generality, because if the original random variable x_t is a m_0-th order Markov with $m_0 > 1$, we can construct a new random vector $\mathbf{x_t}$ stacking variables x_t with a sufficient number of lags so that $\mathbf{x_t}$ is first-order Markov. More precisely, we assume that the underlying time series x_t with $t = 1, \cdots, n$ is m-dependent process and the conditional density is $\pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})$ for some positive integer constant m_0. For the stacked vector $\mathbf{x_t} = (x_t, \cdots, x_{t-K_0})^T$ with $K_0 \geq m_0$, the conditional density for x_t under $\mathbb{P}$ can be rewritten as $\pi_{\mathbb{P}}(\mathbf{x_t}; \theta) = \pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})$.

Assumption M [2]

There exists constant $\lambda \geq 2d/(d-1)$, where d is the constant in **Assumption D**, such that $(\mathbf{x_t}, \mathbf{y_t})$ for $t = 1, 2, \cdots, n$ is uniform mixing and there exists a constant ϕ^* such that the uniform mixing coefficients satisfy

$$\phi(m; \theta) \leq \phi^* m^{-\lambda} \quad \text{for all } \theta \in \Theta,$$

where $\phi(m; \theta)$ is the uniform mixing coefficient under the probability measure $\mathbb{Q}_\theta$. Its definition is standard and can be found, for example, in White and Domowitz (1984) or Bradley (2005).

Assumption D The function $g(\mathbf{x}, \mathbf{y}, \theta)$ is twice continuously differentiable in θ almost everywhere. There exist dominating measurable functions $a_1(\mathbf{x}, \mathbf{y})$ and $a_2(\mathbf{x}, \mathbf{y})$, and

[2] Following the literature(see, e.g. White and Domowitz, 1984; Newey, 1985; Newey and West, 1987a), we adopt the mixing conditions as a convenient way of describing economic and financial data which allows time dependence and heterogeneity. The mixing conditions basically restrict the memory of a process to be weak so that large sample properties of the process are close to the case of assuming ergodicity for strictly stationary processes (see, e.g. Hansen, 1982). In particular, we employ the uniform mixing or $\phi-$mixing condition which is discussed in some detail by White and Domowitz (1984) where definition and its relationship with other type of mixing conditions can be found in the survey by Bradley (2005).

constant $d > 1$, such that almost everywhere

$$\begin{aligned}
&|g(\mathbf{x}, \mathbf{y}, \theta)|^2 \leq a_1(x, y), \;\; ||\partial g(\mathbf{x}, \mathbf{y}, \theta)/\partial\theta||_{\mathbb{S}}^2 \leq a_1(\mathbf{x}, \mathbf{y}), \\
&||\partial^2 g_{(i)}(\mathbf{x}, \mathbf{y}, \theta)/\partial\theta\partial\theta^T||_{\mathbb{S}}^2 \leq a_1(\mathbf{x}, \mathbf{y}), \;\; \text{for } i = 1, \cdots, D_g, \\
&|q_0(\mathbf{x}, \mathbf{y})| \leq a_2(\mathbf{x}, \mathbf{y}), \;\; |q_0(\mathbf{x_1}, \mathbf{y_1}, \mathbf{x_t}, \mathbf{y_t})| \leq a_2(\mathbf{x_1}, \mathbf{y_1}) a_2(\mathbf{x_t}, \mathbf{y_t}), \;\; \text{for } t \geq 2, \\
&\int [a_1(\mathbf{x}, \mathbf{y})]^d a_2(\mathbf{x}, \mathbf{y}) \mathrm{d}x\mathrm{d}y < +\infty, \;\; \int a_2(\mathbf{x}, \mathbf{y}) \mathrm{d}x\mathrm{d}y < +\infty.
\end{aligned}$$

We assume that the expected observed Fisher information matrix, defined as $\mathbf{I}_{\mathbb{Q}}(\theta) \equiv G_0(\theta)^T S_0^{-1} G_0(\theta)$, is positive definite for all $\theta \in \Theta$, where $G_0(\theta) \equiv \mathbb{E}_{\mathbb{Q}_0}[\partial g(\mathbf{x}, \mathbf{y}, \theta)/\partial\theta]$.

Assumption P

Suppose the parameter set is $\Theta \subset \mathbb{R}^{D_\Theta}$ with Θ compact[3]. And, the prior is absolutely continuous with respect to the Lebesgue measure with Radon-Nykodim density $\pi_{\mathbb{P}}(\theta)$, which is twice continuously differentiable and positive on Θ.[4] Denote $M_\pi \equiv \max_{\theta\in\Theta} \pi_{\mathbb{P}}(\theta)$ and $m_\pi \equiv \min_{\theta\in\Theta} \pi_{\mathbb{P}}(\theta)$.

Assumption F [5]

Suppose θ_0 is an interior point of parameter set Θ. The conditional densities $\pi_{\mathbb{P}}(\mathbf{x_t}; \theta)$ is twice continuously differentiable in parameter set, for almost every $\mathbf{x_t}$ under $\mathbb{P}_0$. For

[3] Compactness implies total boundness.

[4] In our diaster risk model, the parameter set is not compact due to the adoption of uninformative prior. However, in that numerical example, we can truncate the parameter set at very large values which will not affect the main numerical results.

[5] There are two information matrices that typically coincide (i.e. Information Equality) and have a basic role in the analysis. Define

$$\mathbf{J}_{\mathbb{P}}(\theta) \equiv -\mathbb{E}_{\mathbb{P}_0}\left[\frac{\partial^2}{\partial\theta\partial\theta^T} \ln \pi_{\mathbb{P}}(\mathbf{x_t}; \theta)\right].$$

When the **Assumption F** holds, we have $\mathbf{I}_{\mathbb{P}}(\theta_0) = \mathbf{J}_{\mathbb{P}}(\theta_0)$ (see e.g. Lehmann and Casella, 1998). Besides, it should be noted that conditions here are actually implied by **Assumption D**, because the conditional score function $\frac{\partial}{\partial\theta_j} \ln \pi_{\mathbb{P}}(\mathbf{x_t}; \theta)$ are assumed to be included in the moment condition $g(\mathbf{x_t}, \mathbf{y_t}, \theta)$. The dominating assumption, together with the uniform mixing assumption and stationarity assumption, imply the stochastic equicontinuity condition (i) in Proposition 3 of Chernozhukov and Hong (2003), which is, for some $\delta > 0$, the following function class is a Donsker class:

$$\left\{\frac{\ln \pi_{\mathbb{P}}(\mathbf{x}; \theta) - \ln \pi_{\mathbb{P}}(\mathbf{x}; \theta_0) - \partial \ln \pi_{\mathbb{P}}(\mathbf{x}; \theta_0)/\partial\theta(\theta - \theta_0)}{|\theta - \theta_0|} \; : \; |\theta - \theta_0| < \delta\right\}.$$

.

each pair of j and k, it holds that for some constant $\zeta > 0$ and large constant $C > 0$,

$$\mathbb{E}_{\mathbb{P}_0} \sup_{\theta\in\Theta} \left| \frac{\partial^2}{\partial\theta_j \partial\theta_k} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta) \right|^{2+\zeta} < C, \text{ and } \mathbb{E}_{\mathbb{P}_0} \sup_{\theta\in\Theta} \left| \frac{\partial}{\partial\theta_j} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta) \right|^{2+\zeta} < C,$$

and

$$\mathbb{E}_{\mathbb{P}_0} \sup_{\theta\in\Theta} |\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)|^{1+\zeta} < C. \tag{A.71}$$

We define the expected observed Fisher information matrix $\mathbf{I}_{\mathbb{P}}(\theta)$ to be

$$\mathbf{I}_{\mathbb{P}}(\theta) \equiv \mathbb{E}_{\mathbb{P}_0} \left[\frac{\partial}{\partial\theta} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta) \frac{\partial}{\partial\theta^T} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta) \right]$$

and assume that $\mathbf{I}_{\mathbb{P}}(\theta)$ is positive definite for all $\theta \in \Theta$.

Assumption PQ [6]

The probability measure defined by the density $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$ is dominated by the probability measure defined by the density $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$, for almost every $\mathbf{x^n}, \mathbf{y^n}$ under $\mathbb{Q}_0$.[7]

Assumption ID [8]

The parametric family of distributions $\mathbb{P}_\theta$ is sound, that is, the convergence of a sequence of parameter values is equivalent to the weak convergence of the distributions they index:

$$\theta \to \theta_0 \Leftrightarrow \mathbb{P}_\theta \to \mathbb{P}_{\theta_0}. \tag{A.72}$$

[6] This assumption is just to guarantee that $\mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n}) || \pi_{\mathbb{P}}(\theta|\mathbf{x^n})\right)$ to be well defined.

[7] The concept of dominating measure here is the one in measure theory. More precisely, this regularity condition requires that for any measurable set which has zero measure under $\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})$, it must also have zero measure under $\pi_{\mathbb{P}}(\theta|\mathbf{x^n})$.

[8] The identification assumption for the baseline model $\pi_{\mathbb{P}}(\mathbf{x};\theta)$ is a weak identifiability condition which implies that $\theta_1 \neq \theta_2 \Rightarrow \mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. More precisely, the true model θ_0 is identified in the sense that if $\theta \neq \theta_0$ and $\theta \in \Theta$, $\pi_{\mathbb{P}}(\mathbf{x};\theta) \neq \pi_{\mathbb{P}}(\mathbf{x};\theta_0)$ for on a measurable set with nonzero Lebesgue measure. By the continuity of $R_0(\theta) \equiv -\mathbb{E}_{\mathbb{P}_0} \ln(\pi_{\mathbb{P}}(\mathbf{x};\theta)/\pi_{\mathbb{P}}(\mathbf{x};\theta_0))$ and the compactness of Θ, together with the identification assumption for the baseline model, we know that for any $\delta > 0$, $\max_{\theta\in\Theta,|\theta-\theta_0|\geq\delta} R_0(\theta) - R_0(\theta_0) < 0$. With the uniformly weak convergence of $n^{-1}\sum_{t=1}^n \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)$ to $R_0(\theta)$, it follows that the Assumption 3 of Chernozhukov and Hong (2003) is satisfied according to the Lemma 1 in Chernozhukov and Hong (2003). By the continuity of $J_0(\theta) \equiv -g_0(\theta)^T S_0^{-1} g_0(\theta)$ and the compactness of Θ, together with the identification assumption for the moment condition, we know that for any $\delta > 0$, $\max_{\theta\in\Theta,|\theta-\theta_0|\geq\delta} J_0(\theta) - J_0(\theta_0) < 0$. With the uniformly weak convergence of $n^{-1}\sum_{t=1}^n g(\mathbf{x_t}, \mathbf{y_t}, \theta)$ to $g_0(\theta)$, it follows that the Assumption 3 of Chernozhukov and Hong (2003) is satisfied according to the Lemma 1 in Chernozhukov and Hong (2003).

And, the true parameter θ_0 is identified by the moment conditions in the sense that $\mathbb{E}_{\mathbb{Q}_0}[g(\mathbf{x}, \mathbf{y}, `)] = 0$ only if $\theta = \theta_0$.

Assumption FF [9]

The feature function $f : \Theta \to \mathbb{R}$ is twice continuously differentiable. We assume that there exist $D_\Theta - 1$ twice continuously differentiable functions $f_2, \cdots, f_{D_\theta}$ on Θ such that $F = (f, f_2, \cdots, f_{D_\theta}) : \Theta \to \mathbb{R}^{D_\Theta}$ is a one-to-one mapping (i.e. injection) and $F(\Theta)$ is a connected and compact D_Θ-dimensional subset of $\mathbb{R}^{D_\Theta}$.

Lemmas

As emphasized by **?**, the mixing conditions serve as an operating assumption for economic and financial processes, because mixing assumptions are difficult to verify or test. However, **?** argues that the restriction is not so restrictive in the sense that a wide class of transformations of mixing processes are themselves mixing.

Lemma 1. *Let* $\mathbf{z_t} = Z((\mathbf{x_{t+ø_0}}, \mathbf{y_{t+ø_0}}), \cdots, (\mathbf{x_{t+ø}}, \mathbf{y_{t+ø}}))$ *where* Z *is a measurable function onto* $\mathbb{R}^{D_z}$ *and two integers* $\tau_0 < \tau$. *If* $\{\mathbf{x_t}, \mathbf{y_t}\}$ *is uniform mixing with uniform mixing coefficients* $\phi(m) \le Cm^{-\lambda}$ *for some* $\lambda > 0$ *and* $C > 0$, *then* $\{\mathbf{z_t}\}$ *is uniform mixing with uniform mixing coefficients* $\phi_z(m) \le C_z m^{-\lambda}$ *for some* $C_z > 0$.

Proof. It can be derived directly from the definition of uniform mixing. □

The following classical result is put here for easy reference. And, it easily leads to a corollary which will be used repeatedly.

Lemma 2. *For any two* $\sigma-$*fields* $\mathcal{F}_1$ *and* $\mathcal{F}_2$, *it holds that*

$$\rho_{max}(\mathcal{F}_1, \mathcal{F}_2) \le 2\left[\phi(\mathcal{F}_1, \mathcal{F}_2)\right]^{1/2}.$$

[9] A simple sufficient condition for Assumption **FF** to hold is that f is a proper and twice continuously differentiable function on $\mathbb{R}^{D_\Theta}$ and $\frac{\partial f(\theta)}{\partial \theta_{(1)}} > 0$ at each $\theta \in \mathbb{R}^{D_\Theta}$. In this case, we can simply choose $f_k(\theta) \equiv \theta_{(k)}$ for $k = 2, \cdots, d$. Then, the Jacobian determinant of F is nonzero at each $\theta \in \mathbb{R}^{D_\Theta}$ and F is proper and twice differentiable mapping $\mathbb{R}^{D_\Theta} \to \mathbb{R}^{D_\Theta}$. According to the Hadamard's Global Inverse Function Theorem (see e.g. Krantz and Parks, 2013), we know that F is a one-to-one mapping and $F(\Theta)$ is a connected and compact D_Θ-dimensional subset of $\mathbb{R}^{D_\Theta}$.

Proof. The proof can be found in Ibragimov (1962) or Doob (1950, Lemma 7.1). □

Corollary 5. *Let $\{\mathbf{z_t}\}$ be strictly stationary process satisfying uniform mixing such that $\phi(m) \leq Cm^{-\lambda}$ for some $\lambda > 0$ and $C > 0$, then the autocorrelation function satisfies $\max_{i,j} \left|\mathrm{Corr}(\mathbf{z}_{(\mathbf{i}),\mathbf{t}}, \mathbf{z}_{(\mathbf{j}),\mathbf{t+m}})\right| \leq 2\sqrt{C}m^{-\lambda/2}$ where $\mathbf{z}_{(\mathbf{i}),\mathbf{t}}$ is the i-th element of $\mathbf{z_t}$.*

Proof. It directly follows from Lemma 1 and Lemma 2. □

Lemma 3. *Let $\{\mathbf{z_t}\}$ be a sequence of strictly stationary random vectors such that $\mathbb{E}_{\mathbb{P}_0}|\mathbf{z_t}|^2 < +\infty$ and it satisfies the uniform mixing condition with $\phi(m) \leq Cm^{-\lambda}$ for some $\lambda > 2$ and $C > 0$. Then, $\lim_{n\to\infty} n\mathrm{var}_0\left(n^{-1}\sum_{t=1}^{n}\mathbf{z_t}\right) = V_0 < +\infty$.*

Proof. Let $\mathbf{z}_{(\mathbf{i}),\mathbf{t}}$ be the i-th element of vector $\mathbf{z_t}$. Denote $\sigma_i^2 \equiv \mathrm{var}_0(\mathbf{z}_{(\mathbf{i}),\mathbf{t}})$ for each i. And, we denote the cross correlation to be $\rho_{i,j}(\tau) \equiv \mathrm{Corr}(\mathbf{z}_{(\mathbf{i}),\mathbf{t}}, \mathbf{z}_{(\mathbf{j}),\mathbf{t+ø}})$ for all t, τ, i and j. Then, we have, for each pair of i and j,

$$
\begin{aligned}
&n\mathrm{Cov}_0\left(n^{-1}\sum_{t=1}^{n}\mathbf{z}_{(\mathbf{i}),\mathbf{t}}, n^{-1}\sum_{t=1}^{n}\mathbf{z}_{(\mathbf{j}),\mathbf{t}}\right)\\
&\quad= \sigma_i\sigma_j\left[\rho_{i,j}(0) + 2\frac{n-1}{n}\rho_{i,j}(1) + \cdots + 2\frac{1}{n}\rho_{i,j}(n-1)\right].
\end{aligned}
$$

According to Corollary 5, we know that $\rho_{i,j}(m) = o(m^{-1})$. Thus, by verifying the Cauchy condition, we know that $\rho_{i,j}(0) + 2\frac{n-1}{n}\rho_{i,j}(1) + \cdots + 2\frac{1}{n}\rho_{i,j}(n-1)$ converges to a finite constant. □

The following two lemmas are extensions of Propositions 6.1 - 6.2 in Clarke and Barron (1990, Page 468-470). Lemma 4 shows that analogs of the soundness condition for certain metrics on probability measures imply the existence of **strongly uniformly exponentially consistent (SUEC)** hypothesis tests. A composite hypothesis test is called **uniformly exponentially consistency (UEC)** if its type-I and type-II errors are uniformly upper bounded by $e^{-\xi n}$ for some positive ξ, over all alternatively (see e.g. Barron, 1989). A **strongly uniformly exponentially consistent (SUEC)** test is a hypothesis test whose type-I and type-II errors are upper bounded by $e^{-\xi n}$ for some positive constant ξ, uniformly over all alternatives and all null parametric models over two subsets in the probability measure space. Lemma 5 shows that metrics with the desirable consistency property exists, which extends Proposition 6.2 in Clarke and Barron (1990).

Lemma 4. *Suppose d_G is a metric on the space of probability measures on X with the property that for any $\epsilon > 0$, there exists $\xi > 0$ and $C > 0$ such that*

$$\mathbb{P}_n \left\{ d_G(\widehat{\mathbb{P}}_n, \mathbb{P}) > \epsilon \right\} \le Ce^{-\xi n}, \tag{A.73}$$

uniformly over all probability measures $\mathbb{P}_n$, where $\widehat{\mathbb{P}}_n$ is the empirical distribution. And, the metric d_G also satisfies

$$d_G(\mathbb{P}_{\theta'}, \mathbb{P}_\theta) \to 0 \;\Rightarrow\; \theta' \to \theta. \tag{A.74}$$

Then, for any $\delta > \delta_1 \ge 0$ and for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$, there exists a SUEC hypothesis test of $\vartheta = \theta \in \overline{\Omega}(\theta_0, \delta_1)$ versus alternative $\Omega(\theta_0, \delta)^c$.

Proof. The proof is an extension based on that of Lemma 6.1 in Clarke and Barron (1990). From (A.74), for any given $\delta > \delta_1 \ge 0$, there exists $\epsilon_1 > 0$ such that it holds $d_L(\{\theta\}, \Omega(\theta_0, \delta)^c) > \delta - \delta_1 > 0$ implies that $d_G(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) > \epsilon_1$ for all $\theta' \in \Omega(\theta_0, \delta)^c$. Thus, for any $\delta > \delta_1 \ge 0$, there exists $\epsilon_1 > 0$ such that, $d_G(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) > \epsilon_1$ for all $\theta \in \Omega(\theta_0, \delta_1)$ and $\theta' \in \Omega(\theta_0, \delta)^c$. Therefore, if we have a SUEC test of

$$\mathbf{H}_0 : \; \mathbb{P} = \mathbb{P}_\theta \text{ versus } \mathbf{H}_A : \; \mathbb{P} \in \{\tilde{\mathbb{P}} : \; d_G(\tilde{\mathbb{P}}, \mathbb{P}_\theta) > \epsilon_1\}, \;\; \theta \in \overline{\Omega}(\theta_0, \delta_1).$$

then we have a SUEC test of

$$\mathbf{H}_0 : \; \theta' = \theta \text{ versus } \mathbf{H}_A : \; \mathbb{P} \in \{\tilde{\mathbb{P}} : \; |\theta' - \theta| > \delta - \delta_1\}, \;\; \theta \in \overline{\Omega}(\theta_0, \delta_1).$$

Let $\widehat{\mathbb{P}}_n$ be the empirical distribution, choose $\epsilon = \epsilon_1/2$ and let

$$\mathcal{A}_{\theta,n} \equiv \{\mathbf{x^n} : \; d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_\theta) \le \epsilon\}$$

be the acceptance region. By (A.73), we have that the probability of type-I error satisfies

$$\mathbb{P}_{\theta,n}\mathcal{A}^c_{\theta,n} \le Ce^{-\xi n}, \text{ for some } \xi > 0 \text{ and } C > 0, \text{ uniformly over } \theta \in \overline{\Omega}(\theta_0, \delta_1).$$

We want to show that the probability of type-II error $\mathbb{P}_{\theta'}\mathcal{A}_{\theta,n}$ is exponentially small uniformly over $\theta' \in \Omega(\theta_0, \delta)^c$ and $\theta \in \Omega(\theta_0, \delta_1)$.

Using triangular inequality, we have for any $\theta \in \overline{\Omega}(\theta_0, \delta_1)$ and $\theta' \in \Omega(\theta_0, \delta)^c$

$$2\epsilon \le d_G(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \le d_G(\mathbb{P}_\theta, \widehat{\mathbb{P}}_n) + d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}). \tag{A.75}$$

By definition and inequality (A.75), for each $\theta \in \Omega(\theta_0, \delta_1)$, on the acceptance region $\mathcal{A}_{\theta,n}$, we have

$$2\epsilon < \epsilon + d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}) \text{ for any } \theta' \in \Omega(\theta_0, \delta)^c. \tag{A.76}$$

and hence, for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$, on the acceptance region $\mathcal{A}_{\theta,n}$, it holds that

$$d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}) > \epsilon.$$

Therefore, for each $\theta \in \overline{\Omega}(\theta_0, \delta_1)$ and each $\theta' \in \Omega(\theta_0, \delta)^c$, we have

$$\mathbb{P}_{\theta',n}\mathcal{A}_{\theta,n} \le \mathbb{P}_{\theta',n}\left\{d_G(\widehat{\mathbb{P}}_n, \mathbb{P}_{\theta'}) > \epsilon\right\} \le Ce^{-\xi n}.$$

□

Lemma 5. *For a space of probability measures $\mathbb{P}$ on a separable metric space $\mathcal{X}$ such that the mixing coefficient $\phi(m) \le \Psi m^{-\lambda}$ with $\Psi > 0$ and $\lambda \ge 2$ being universal constant (independent of $\mathbb{P}$), denoted as $\mathcal{P}(\Psi, \lambda)$, there exists a metric $d_G(\mathbb{P}, \mathbb{Q})$ that satisfies the property (A.73) and such that convergence in d_G implies the weak convergence of the measures. Thus, in particular, for probability measures in a parametric family*

$$d_G(\mathbb{P}_{\theta'}, \mathbb{P}_\theta) \to 0 \Rightarrow \mathbb{P}_{\theta'} \to \mathbb{P}_\theta. \tag{A.77}$$

Proof. We extend the proof for the existence result in Proposition 6.2 in Clarke and Barron (1990) to allow for weak dependence. Let $\{F_i : i = 1, 2, \cdots\}$ be the countable field of sets generated by balls of the form $\{x : d_{\mathcal{X}}(x, s_{j_1}) \leq 1/j_2\}$ for $j_1, j_2 = 1, 2, \cdots$, where $d_{\mathcal{X}}$ denotes the metric for the space $\mathcal{X}$ and $s_1, s_2, \cdots$ is a countable dense sequence of points in $\mathcal{X}$. Define a metric on the space of probability measures as follows

$$d_G(\mathbb{P}, \mathbb{Q}) = \sum_{i=1}^{\infty} 2^{-i} |\mathbb{P}F_i - \mathbb{Q}F_i|.$$

According to Gray (1988, Page 251–253), if $d_G(\mathbb{P}_n, \mathbb{P}) \to 0$, then $\mathbb{P}_n$ converges weakly to $\mathbb{P}$. Now, for any $\epsilon > 0$, we choose $k \geq 1 - \ln(\epsilon) / \ln(2)$. Thus,

$$d_G(\widehat{\mathbb{P}}_n, \mathbb{P}) \leq \sum_{i=1}^{k} 2^{-i} \left| \widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i \right| + \sum_{i=k+1}^{\infty} 2^{-i} \leq \max_{1 \leq i \leq k} \left| \widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i \right| + \epsilon/2.$$

Then, we have

$$\mathbb{P}\left\{ d_G(\widehat{\mathbb{P}}, \mathbb{P}) > \epsilon \right\} \leq \mathbb{P}\left\{ \max_{1 \leq i \leq k} \left| \widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i \right| > \epsilon/2 \right\} \leq \sum_{i=1}^{k} \mathbb{P}\left\{ \left| \widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i \right| > \epsilon/2 \right\}$$

The Hoeffding-type inequality for uniform mixing process (see, e.g. Roussas, 1996, Theorem 4.1) guarantees that there exists $C_1 > 0$ and $\xi > 0$ such that

$$\sup_{\mathbb{P} \in \mathcal{P}(\Psi, \lambda)} \mathbb{P}\left\{ \left| \widehat{\mathbb{P}}_n F_i - \mathbb{P}F_i \right| > \epsilon/2 \right\} \leq C_1 e^{-\xi n}.$$

Thus,

$$\mathbb{P}\left\{ d_G(\widehat{\mathbb{P}}_n, \mathbb{P}) > \epsilon \right\} \leq C e^{-\xi n}, \tag{A.78}$$

with $C = kC_1$. □

Lemma 6 extends the large deviation result of Schwartz (1965, Lemma 6.1) to allow time dependence in the data process. Let $\mathbf{z^n} = (\mathbf{z_1}, \cdots, \mathbf{z_n})$ be strictly stationary and uniform mixing with $\phi(m) = O\left(m^{-\lambda}\right)$ for some positive λ. The process $\mathbf{z^n}$ have the joint density $\mathbb{P}_{\theta,n}$. Denote the mixture distribution of the parametric family $\mathbb{P}_{\theta,n}$

with respect to the conditional prior distribution $\pi_{\mathbb{P}}(\cdot|\mathcal{N}^c)$ to be $\mathbb{P}_{\mathcal{N}^c,n}$ with density $\pi_{\mathbb{P}}(\mathbf{z^n}|\mathcal{N}^c)$. More precisely, we define

$$\pi_{\mathbb{P}}(\mathbf{z^n}|\mathcal{N}^c) \equiv \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\mathbf{z^n}|\theta)\pi_{\mathbb{P}}(\theta|\mathcal{N}^c)\mathrm{d}\theta. \tag{A.79}$$

Lemma 6. *Assume that the mixing coefficient power $\lambda > 2$. Suppose there exist strongly uniformly exponentially consistent (SUEC) tests of hypothesis $\theta \in \mathcal{N}_0$ against the alternative $\theta \in \mathcal{N}^c$ such that $\mathcal{N}_0 \subset \mathcal{N}$ with $d_L(\mathcal{N}_0, \mathcal{N}^c) \geq \delta$ for some $\delta > 0$. Then, there exists $\xi > 0$ and a positive integer k such that, for all $\theta \in \mathcal{N}_0$,*

$$||\mathbb{P}_{\theta,n} - \mathbb{P}_{\mathcal{N}^c,n}||_{TV} \geq 2(1 - 2e^{-\xi m}), \;\; \textit{where } m + k \leq n.$$

Proof. We assume that there exists a sequence of SUEC tests, denoted by $\{\mathcal{A}_n\}$, for any sample with size n. Then, there exists a positive integer k such that for all $n \geq k$

$$\mathbb{P}_{\theta,n}\mathcal{A}_n < \frac{1}{8} \text{ for all } \theta \in \mathcal{N}_0 \quad \text{and} \tag{A.80}$$

$$\mathbb{P}_{\theta',n}\mathcal{A}_n > 1 - \frac{1}{8} \text{ for all } \theta' \in \mathcal{N}^c. \tag{A.81}$$

For each $j = 1, 2, \cdots$, we define

$$\mathcal{A}_{k,j} = \mathcal{A}_k\left(\mathbf{z_{j+1}}, \cdots, \mathbf{z_{j+k}}\right), \tag{A.82}$$

then, according to Lemma 1,

$$Y_m = \frac{1}{m}\sum_{j=1}^{m} \mathcal{A}_{k,j} \tag{A.83}$$

is an average of strictly stationary and uniform mixing such that $\phi(m) \leq \phi^* m^{-\lambda}$. The expectation of Y_m, under distribution $\mathbb{P}_{\theta,n}$, is $\mu(\theta)$ with

$$\mu(\theta) = \begin{cases} < \frac{1}{8} & \text{if } \theta \in \mathcal{N}_0 \\ > \frac{7}{8} & \text{if } \theta \in \mathcal{N}^c. \end{cases} \tag{A.84}$$

By Corollary 5 and the uniform mixing conditions with $\lambda > 2$, together with the fact that $\mathcal{A}_{k,j} \in [0,1]$, we know that the assumptions of Theorem 2.4 in **?** holds and hence the CLT for the time series holds, i.e.

$$m^{1/2} Y_m \xrightarrow{d} N(\mu(\theta), V(\theta)), \tag{A.85}$$

where $V(\theta) = \lim_{m \to +\infty} \mathbb{P}_{\theta,n} \left[m^{1/2} \sum_{j=1}^{m} (\mathcal{A}_{k,j} - \mu(\theta)) \right]^2$. From Lemma 3, we know that $V(\theta) \leq V^* < \infty$. Therefore, the moment generating functions converge

$$m^{-1} \ln \mathbb{P}_{\theta,n} e^{tmY_m} \to \frac{1}{2} t^2 V(\theta). \tag{A.86}$$

On the one hand, when $\theta \in \mathcal{N}^c$, we have $\mu(\theta) > \frac{1}{4}$, according to Theorem 8.1.1 of Taniguchi and Kakizawa (2000), we can achieve the following large deviation result

$$\lim_{m \to +\infty} m^{-1} \ln \mathbb{P}_{\theta,n} \left\{ Y_m \leq \frac{1}{4} \right\} = -\frac{1}{32V(\theta)} \leq -\frac{1}{32V^*}. \tag{A.87}$$

Therefore, there exists $\xi_1 > 0$ such that

$$\mathbb{P}_{\theta,n} \left\{ Y_m \leq \frac{1}{4} \right\} \leq e^{-\xi_1 m} \quad \text{for all } \theta \in \mathcal{N}^c. \tag{A.88}$$

Thus,

$$\mathbb{P}_{\mathcal{N}^c,n} \left\{ Y_m \leq \frac{1}{4} \right\} = \int_{\mathcal{N}^c} \mathbb{P}_{\theta,n} \left\{ Y_m \leq \frac{1}{4} \right\} \pi_{\mathbb{P}}(\theta | \mathcal{N}^c) \mathrm{d}\theta \leq e^{-\xi_1 m}, \quad \text{for } n \geq m + k. \tag{A.89}$$

On the other hand, when $\theta \in \mathcal{N}_0$, we have $\mu(\theta) < \frac{1}{4}$, according to Theorem 8.1.1 of Taniguchi and Kakizawa (2000), we obtain the large deviation result

$$\lim_{m \to +\infty} m^{-1} \ln \mathbb{P}_{\theta,n} \left\{ Y_m \geq \frac{1}{4} \right\} = -\frac{1}{32V(\theta)} \leq -\frac{1}{32V^*}. \tag{A.90}$$

Therefore, there exists $\xi_2 > 0$ such that

$$\mathbb{P}_{\theta,n}\left\{Y_m \geq \frac{1}{4}\right\} \leq e^{-\xi_2 m}. \tag{A.91}$$

For $\xi = \min(\xi_1, \xi_2)$, it follows that

$$||\mathbb{P}_{\theta,n} - \mathbb{P}_{\mathcal{N}^c,n}||_{TV} = 2\sup_{A\in\mathcal{F}} |\mathbb{P}_{\theta,n}A - \mathbb{P}_{\mathcal{N}^c,n}A| \geq 2(1-2e^{-\xi m}) \tag{A.92}$$

for $m + k \leq n$ by considering $A = \{Y_m \leq \frac{1}{4}\}$. □

The Intermediary Propositions

First, we introduce the generalized information equality. The information equality is known to hold for regular and correctly specified likelihood functions. The following proposition shows that the generalized information equality also holds for regular and correctly specified optimal generalized method of moments. This result is directly borrowed from Kim (2002, Theorem 2) and the proof can be found there. We introduce it here just for convenient references.

Proposition 14. *Suppose that the regularity assumptions in Subsection A.3.3 hold. Then, it is true that*

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}_0}\left[\frac{1}{n}\left(\frac{\partial \ln \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)}{\partial\theta^T}\right)\left(\frac{\partial \ln \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)}{\partial\theta^T}\right)^T\right] = \lim_{n\to\infty} -\mathbb{E}_{\mathbb{Q}_0}\left[\frac{1}{n}\frac{\partial^2 \ln \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0)}{\partial\theta\partial\theta^T}\right].$$

The following two propositions on asymptotic normality of conditional MLE and GMM estimators for time series are general and based on standard results. We introduce them mainly for easy references.

Proposition 15. *Suppose that the regularity assumptions in Subsection A.3.3 hold. Let $\widehat{\theta}^{\mathbb{P}}$ to be the conditional MLE estimator for the m-dependent sequence $\mathbf{x_t}$ with $t = 1, \cdots, n$. Then,* *we have*

$$\sqrt{n}(\widehat{\theta}^{\mathbb{P}} - \theta_0) \rightsquigarrow N(0, \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}).$$

Proof. The proof is based on the classical results in van der Vaart (1998, Theorem 5.41) and **?**, Theorem 2.3 & 2.4. Let

$$\psi_\theta(\mathbf{x_t}) = \frac{\partial}{\partial\theta} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta).$$

Then, we have $\dot{\psi}_\theta(\mathbf{x_t}) = \frac{\partial^2}{\partial\theta\partial\theta^T} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)$. We know that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_0}\psi_{\theta_0}(\mathbf{x_t}) = & \int \pi_{\mathbb{P}}(x_t, x_{t-1}, \cdots, x_{t-m_0}|\theta_0) \\
& \times \frac{\partial \pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})}{\pi_{\mathbb{P}}(x_t|\theta, x_{t-1}, \cdots, x_{t-m_0})} \mathrm{d}(x_t, x_{t-1}, \cdots, x_{t-m_0}) \\
= & \int \pi_{\mathbb{P}}(x_{t-1}, \cdots, x_{t-m_0}|\theta_0) \mathrm{d}(x_{t-1}, \cdots, x_{t-m_0}) \\
& \times \int \partial \pi_{\mathbb{P}}(x_t|\theta_0, x_{t-1}, \cdots, x_{t-m_0}) \mathrm{d}x_t
\end{aligned}$$

Because $\int \partial \pi_{\mathbb{P}}(x_t|\theta_0, x_{t-1}, \cdots, x_{t-m_0}) \mathrm{d}x_t = 0$ almost everywhere, we have the relationship $\mathbb{E}_{\mathbb{P}_0}\psi_{\theta_0}(\mathbf{x_t}) = 0$. Thus, the assumption A(i) of Theorem 2.4 in White and Domowitz (1984) is satisfied for $\psi_\theta(\mathbf{x_t})$.

Assumption F, together with **Assumption M** and Lemma 2, the assumptions A(ii) and A(iii) of Theorem 2.4 in White and Domowitz (1984) are satisfied for $\psi_\theta(\mathbf{x_t})$. It also guarantees that the dominating assumption in Theorem 2.3 of White and Domowitz (1984) for $\dot{\psi}_\theta(\mathbf{x_t})$. Thus, the uniform law of large numbers (ULLN) holds, i.e. $n^{-1}\sum_{t=1}^{n} \dot{\psi}_\theta(\mathbf{x_t}) \to \mathbb{E}_{\mathbb{P}_0}[\dot{\psi}_\theta(\mathbf{x_t})]$ uniformly on Θ. According to **Assumption F**, the matrix $\mathbf{I}_{\mathbb{P}}(\theta_0) \equiv \mathbb{E}_{\mathbb{P}_0}\left[\dot{\psi}_{\theta_0}(\mathbf{x_t})\right]$ is positive definite.

By definition, the conditional MLE $\widehat{\theta}_{\mathbb{P}}$ sets

$$\frac{1}{n}\sum_{t=1}^{n} \psi_{\widehat{\theta}_{\mathbb{P}}}(\mathbf{x_t}) = 0.$$

By Taylor expansion, we have

$$\sqrt{n}(\widehat{\theta}_{\mathbb{P}} - \theta_0) = -\left[n^{-1}\sum_{t=1}^{n}\dot{\psi}_{\tilde{\theta}}(\mathbf{x_t})\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_{\theta_0}(\mathbf{x_t})\right], \tag{A.93}$$

where $\tilde{\theta}$ is on the segment between $\widehat{\theta}_{\mathbb{P}}$ and θ_0. Now, let's first show that $\widehat{\theta}^{\mathbb{P}}$ converges to θ_0. Because θ_0 is identified according to **Assumption ID** and the condition $\mathbb{E}_{\mathbb{P}_0}[\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)] < \infty$ for all $\theta \in \Theta$ according to **Assumption F**, by Jensen's inequality, we have $L_{\mathbb{P}}(\theta) - L_{\mathbb{P}}(\theta_0) > 0$ with inequality holds for all $\theta \neq \theta_0$. We define $L_{\mathbb{P}}(\theta) \equiv \mathbb{E}_{\mathbb{P}_0}[-\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)]$. Besides, according to Theorem 2.3 in White and Domowitz (1984), we have $n^{-1}\sum_{t=1}^{n}\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta) \to L_{\mathbb{P}}(\theta)$ a.s. uniformly in $\theta \in \Theta$. The uniform convergence and identification of θ_0 guarantees that $\widehat{\theta}^{\mathbb{P}} \to \theta_0$ in $\mathbb{P}_{0,n}$. Therefore, we have

$$n^{-1}\sum_{t=1}^{n}\dot{\psi}_{\tilde{\theta}}(\mathbf{x_t}) \to \mathbf{I}_{\mathbb{P}}(\theta_0) \quad \text{in } \mathbb{P}_{0,n}.$$

In the end, by Theorem 2.4 of White and Domowitz (1984), we have the asymptotic normality result for the mixing process:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_{\theta_0}(\mathbf{x_t}) \rightsquigarrow N(0, \mathbf{I}_{\mathbb{P}}(\theta_0)).$$

Therefore, from Equation (A.93), we have

$$\sqrt{n}(\widehat{\theta}^{\mathbb{P}} - \theta_0) \rightsquigarrow N(0, \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}).$$

□

Proposition 16. *Suppose that the regularity assumptions in Subsection A.3.3 hold. Let $\widehat{\theta}^{\mathbb{Q}}$ to be the GMM estimator for the sequence $(\mathbf{x_t}, \mathbf{y_t})$ with $t = 1, \cdots, n$. Then, we have*

$$\sqrt{n}(\widehat{\theta}^{\mathbb{Q}} - \theta_0) \rightsquigarrow N(0, \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}).$$

Proof. All the assumptions of Lemma 1 in Newey (1985) are satisfied under the regularity conditions in Subsection A.3.3. Then, we have

$$\sqrt{n}(\widehat{\theta}^{\mathbb{Q}} - \theta_0) = -(G_0^T S_0^{-1} G_0)^{-1} G_0^T S_0^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta_0) + o_p(1)$$

where

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta_0) \rightsquigarrow N(0, S_0).$$

□

Proposition 17. *Denote $\theta_{(1)}$ to be the first element in θ. Then, we have*

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})) \leq \mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta|\mathbf{x^n})). \tag{A.94}$$

Proof. Denote $\theta_{(-1)}$ to be the vector containing all parameters other than $\theta_{(1)}$ in θ. Then, we have

$$\begin{aligned}
&\mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\right) \\
&\quad = \mathbb{E}^{\theta_{(1)}} \mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta_{(-1)}|\mathbf{x^n}, \mathbf{y^n}, \theta_{(1)})||\pi_{\mathbb{P}}(\theta_{(-1)}|\mathbf{x^n}, \theta_{(1)})\right) \qquad \text{(A.95)} \\
&\qquad + \mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})\right).
\end{aligned}$$

Because the term (A.95) is nonnegative, the result of the proposition is proved. □

Definition 10. *We define the following quantities which are expected negative log (limited information) likelihood will be used repeatedly in the proofs.*

$$H_{\mathbb{Q}}(\theta) \equiv \frac{1}{2} g_0(\theta)^T S_0^{-1} g_0(\theta), \ \ \textit{and} \ \ H_{\mathbb{P}}(\theta) \equiv -\int \pi_{\mathbb{P}}(\mathbf{x}|\theta_0) \ln \pi_{\mathbb{P}}(\mathbf{x}; \theta) d\mathbf{x}. \tag{A.96}$$

Define the sample correspondences as

$$\widehat{H}_{\mathbb{Q},n}(\theta) \equiv \frac{1}{2} \left[\frac{1}{n} \sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right] S_0^{-1} \left[\frac{1}{n} \sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right] \tag{A.97}$$

and

$$\widehat{H}_{\mathbb{P},n}(\theta) \equiv -\frac{1}{n}\sum_{t=1}^{n} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta). \tag{A.98}$$

Proposition 18. *Under the regularity conditions in Subsection A.3.3, we have*

$$\widehat{H}_{\mathbb{Q},n}(\theta) \to H_{\mathbb{Q}}(\theta) \text{ and } \widehat{H}_{\mathbb{P},n}(\theta) \to H_{\mathbb{P}}(\theta) \quad \text{a.s. uniformly in } \theta \in \Theta.$$

Proof. Simply follows the uniform law of large numbers (ULLN) in White and Domowitz (1984, Theorem 2.3). □

Proposition 19. *Under the assumptions in Subsection A.3.3, it holds that*

$$\mathbf{D}_{KL}(\pi_{\mathbb{P}}(\theta|\mathbf{x^n})||N(\widehat{\theta}^{\mathbb{P}}, n^{-1}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1})) \to 0 \text{ in } \mathbb{P}_{0,n}.$$

Proof. Our proof is mainly based on Theorem 2.1 in Clarke (1999) which is under the i.i.d. condition. In particular, we need to adjust two parts of the proof there, in order to extend the result to the case that the observations are time series with uniform mixing. First, we consider the quantity $\widehat{H}_{\mathbb{P},n}(\theta)$ defined in (A.98). When n is large enough, we obtain that

$$\sup_{\theta\in\Theta} |\widehat{H}_{\mathbb{P},n}(\theta)| \leq 1 + \frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta} |\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)|.$$

According to **Assumption F** and **Assumption M**, it follows from Theorem 2.3 of White and Domowitz (1984) that

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta} |\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)| \to \mathbb{E}_{\mathbb{P}_0} \sup_{\theta\in\Theta} |\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta)| \quad \text{a.s.}$$

Thus, we have $\sup_{\theta\in\Theta} |\widehat{H}_{\mathbb{P},n}(\theta)| = O_p(1)$.

The second part is to show that

$$\int u^T u \left| \pi_{\mathbb{P}}(\widehat{\theta}^{\mathbb{P}} + u/\sqrt{n}|\mathbf{x^n}) - \varphi_{\mathbb{P}}(u) \right| \mathrm{d}u \to 0 \quad \text{in } \mathbb{P}_{0,n} \tag{A.99}$$

where $\varphi_{\mathbb{P}}(u) = \sqrt{\frac{\det \mathbf{I}_{\mathbb{P}}(\theta_0)}{(2\pi)^{D_\Theta}}} \exp\left[-\frac{1}{2}u^T \mathbf{I}_{\mathbb{P}}(\theta_0)u\right]$.

In Clarke (1999), it shows that when $\mathbf{x_1}, \cdots, \mathbf{x_n}$ are i.i.d., the limit result (A.100) is satisfied under the regularity conditions in Subsection A.3.3. To extend this limit result to allow weak dependence, we appeal to Theorem 1 and Proposition 3 of Chernozhukov and Hong (2003) whose conditions are implied by the regularity assumptions in Subsection A.3.3. □

Proposition 20. *Under the assumptions in Subsection A.3.3, it holds that*

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(\theta|\mathbf{x^n}, \mathbf{y^n})||N(\widehat{\theta}^{\mathbb{Q}}, n^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1})) \to 0 \ \ in \ \ \mathbb{Q}_{0,n}.$$

Proof. Our proof is mainly based on Theorem 2.1 in Clarke (1999) which is under the i.i.d. condition and likelihood setting. In particular, we need to adjust two parts of the proof there, in order to extend the result to uniform mixing time series in GMM setting. First, we consider the quantity $\widehat{H}_{\mathbb{Q},n}(\theta)$ defined in (A.97). Then, it follows that

$$\sup_{\theta \in \Theta} |\widehat{H}_{\mathbb{Q},n}(\theta)| \leq \frac{1}{2\lambda_m(S_0)n} \sum_{t=1}^{n} a_1(\mathbf{x_t}, \mathbf{y_t}) = O_p(1),$$

where the inequality is due to the definition of the smallest eigenvalue $\lambda_m(S_0)$ and the regularity condition $\mathbf{D}$ in Section A.3.3, and the $O_p(1)$ probabilistic control is due to the Law of Large Numbers.

The second part is to show that

$$\int u^T u \left| \pi_{\mathbb{Q}}(\widehat{\theta}^{\mathbb{Q}} + u/\sqrt{n}|\mathbf{x^n}) - \varphi_{\mathbb{Q}}(u) \right| \mathrm{d}u \to 0 \ \ \text{in} \ \ \mathbb{Q}_{0,n}, \tag{A.100}$$

where $\varphi_{\mathbb{Q}}(u) = \sqrt{\frac{\det \mathbf{I}_{\mathbb{Q}}(\theta_0)}{(2\pi)^{D_\Theta}}} \exp\left[-\frac{1}{2}u^T \mathbf{I}_{\mathbb{Q}}(\theta_0)u\right]$.

In Clarke (1999), it shows that when $\mathbf{x_1}, \cdots, \mathbf{x_n}$ are i.i.d., the limit result (A.100) is satisfied under the regularity conditions in Subsection A.3.3. To extend this limit result to allow weak dependence, we appeal to Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003) whose conditions are implied by the regularity assumptions in Subsection A.3.3. □

Corollary 6. *Denote* $\mathbf{v} \equiv \nabla f(\theta_0)$. *Under the assumptions in Subsection A.3.3, we have*

$$\mathbf{D}_{KL}\left(\pi_{\mathbb{P}}(f(\theta)|\mathbf{x^n})||N(f(\widehat{\theta}^{\mathbb{P}}), n^{-1}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v})\right) \to 0 \ \ in \ \mathbb{P}_{0,n}, \tag{A.101}$$

and

$$\mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(f(\theta)|\mathbf{x^n}, \mathbf{y^n})||N(f(\widehat{\theta}^{\mathbb{Q}}), n^{-1}\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v})\right) \to 0 \ \ in \ \mathbb{Q}_{0,n}, \tag{A.102}$$

Proof. Because of Assumption **FF** and the assumptions in Subsection A.3.3 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta) = (\theta_{(1)}, \theta_{(2)}, \cdots, \theta_{(D_f)})$. Applying Proposition 17, Proposition 19 and Proposition 20, we know that the results hold. □

Definition 11. *We define the observed Fisher information matrices as*

$$\widehat{\mathbf{I}}_{\mathbb{P},n}(\theta) \equiv -\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^T}\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta) = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^T}\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta) + o_p(1), \tag{A.103}$$

and

$$\widehat{\mathbf{I}}_{\mathbb{Q},n}(\theta) \equiv \left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right]. \tag{A.104}$$

Definition 12. *The empirical score functions are*

$$s_{\mathbb{P},n}(\theta) \equiv \frac{1}{n}\frac{\partial}{\partial\theta}\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta),$$

and

$$
\begin{aligned}
s_{\mathbb{Q},n}(\theta) &\equiv \frac{1}{n}\frac{\partial}{\partial\theta}\ln \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta) \\
&= -\frac{1}{2}\frac{\partial}{\partial\theta}\left\{\left[\frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right]\right\} \\
&= -\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta)\right].
\end{aligned}
$$

The standardized empirical score functions are

$$
S_{\mathbb{P},n}(\theta) = \sqrt{n} s_{\mathbb{P},n}(\theta), \;\; \textit{and} \;\; S_{\mathbb{Q},n}(\theta) = \sqrt{n} s_{\mathbb{Q},n}(\theta).
$$

Proposition 21. *Under the assumptions in Subsection A.3.3, the uniform law of large numbers (ULLN) holds:*

$$
\sup_{\theta\in\Theta}\left|\left|\widehat{\mathbf{I}}_{\mathbb{Q},n}(\theta) - \mathbf{I}_{\mathbb{Q}}(\theta)\right|\right|_{\mathcal{S}} \to 0 \;\; \textit{in} \;\; \mathbb{Q}_{0,n}, \tag{A.105}
$$

where

$$
\mathbf{I}_{\mathbb{Q}}(\theta) \equiv G_0(\theta)^T S_0^{-1} G_0(\theta).
$$

Proof. Recall

$$
\widehat{\mathbf{I}}_{\mathbb{Q},n}(\theta) = \left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t}, \mathbf{y_t}, \theta)}{\partial\theta}\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t}, \mathbf{y_t}, \theta)}{\partial\theta}\right].
$$

Thus, we need to show that

$$
\sup_{\theta\in\Theta}\left|\left|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t}, \mathbf{y_t}, \theta)}{\partial\theta}\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t}, \mathbf{y_t}, \theta)}{\partial\theta}\right] - \mathbf{I}_{\mathbb{Q}}(\theta)\right|\right|_{\mathcal{S}} = o_p(1).
$$

We first consider the following decomposition

$$
\begin{aligned}
&\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right]-\mathbf{I}_{\mathrm{Q}}(\theta)\\
&=\frac{1}{2}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]\\
&\qquad+\frac{1}{2}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]
\end{aligned}
$$

By the triangular inequality of spectrum norm, it follows that

$$
\begin{aligned}
&\left\|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right]-\mathbf{I}_{\mathrm{Q}}(\theta)\right\|_{\mathcal{S}}\\
&\leq\frac{1}{2}\left\|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]\right\|_{\mathcal{S}}\\
&\qquad+\frac{1}{2}\left\|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]\right\|_{\mathcal{S}}\\
&=\left\|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]\right\|_{\mathcal{S}}
\end{aligned}
$$

Now, by using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left|\left|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right]-\mathbf{I}_{\mathbb{Q}}(\theta)\right|\right|_{\mathbb{S}} \\
&\leq\left|\left|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0\right]\right|\right|_{\mathbb{S}}^{1/2} \\
&\qquad\times\left|\left|\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0\right]\right|\right|_{\mathbb{S}}^{1/2} \\
&\leq\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}+G_0(\theta)\right|\right|_{\mathbb{S}}\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0(\theta)\right|\right|_{\mathbb{S}}/\lambda_m(S_0) \\
&\leq\left[\frac{1}{n}\sum_{t=1}^{n}\left|\left|\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}\right|\right|_{\mathbb{S}}+||G_0(\theta)||_{\mathbb{S}}\right]\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0(\theta)\right|\right|_{\mathbb{S}}/\lambda_m(S_0) \\
&\leq\left[\frac{1}{n}\sum_{t=1}^{n}a_1(\mathbf{x_t},\mathbf{y_t})^{1/2}+\mathbb{E}_{\mathbb{Q}_0}a_1(\mathbf{x},\mathbf{y})^{1/2}\right]\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0(\theta)\right|\right|_{\mathbb{S}}/\lambda_m(S_0)
\end{aligned}$$

By Theorem 2.3 of White and Domowitz (1984), the ULLN implies that

$$\sup_{\theta\in\Theta}\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}-G_0(\theta)\right|\right|_{\mathbb{S}}\to 0 \quad \text{in } \mathbb{Q}_{0,n}.$$

Thus, we complete the proof. □

Proposition 22. *Under the assumptions in Subsection A.3.3, the uniform law of large numbers (ULLN) holds:*

$$\sup_{\theta\in\Theta}\left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\theta)-\mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathbb{S}}\to 0 \;\; in \;\; \mathbb{P}_{0,n}, \tag{A.106}$$

where

$$\mathbf{I}_{\mathbb{P}}(\theta)\equiv-\mathbb{E}_{\mathbb{P}_0}\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\ln\pi_{\mathbb{P}}(\mathbf{x};\theta)\right].$$

Proof. By applying Theorem 2.3 of White and Domowitz (1984), the ULLN gives

$$-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^T}\ln\pi_{\mathbb{P}}(\mathbf{x_t};\theta)\to\mathbf{I}_{\mathbb{P}}(\theta) \quad \text{a.s. uniformly in } \Theta.$$

□

Corollary 7. *Under the assumptions in Subsection A.3.3, for any sequence of random variables* $\tilde{\theta}_n$ *such that* $\tilde{\theta}_n \to \theta_0$ *in* $\mathbb{Q}_{0,n}$*, we have*

$$\widehat{\mathbf{I}}_{\mathbb{Q},n}(\tilde{\theta}_n) \to \mathbf{I}_{\mathbb{Q}}(\theta_0) \ \text{ in } \ \mathbb{Q}_{0,n}, \ \text{ and } \ \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}_n) \to \mathbf{I}_{\mathbb{P}}(\theta_0) \ \text{ in } \ \mathbb{P}_{0,n},$$

where $\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})$ *and* $\widehat{\mathbf{I}}_{\mathbb{Q},n}(\tilde{\theta})$ *are observed Fisher Information matrixes defined in (A.103) and (A.104), respectively.*

Proposition 23. *Under the assumptions in Subsection A.3.3, it holds that the Hausman test statistic based on GMM estimation satisfies*

$$n(\widehat{\theta}^{\mathbb{P}} - \widehat{\theta}^{\mathbb{Q}})^T \left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\right]^{-1} (\widehat{\theta}^{\mathbb{P}} - \widehat{\theta}^{\mathbb{Q}}) \rightsquigarrow \chi^2_{D_\Theta}$$

Proof. The estimator $\widehat{\theta}^{\mathbb{Q}}$ is GMM estimator under constrained model and $\widehat{\theta}^{\mathbb{P}}$ is MLE under unconstrained model. Under the constrained model $\mathbb{Q}_0$, the estimator $\widehat{\theta}^{\mathbb{Q}}$ is asymptotic efficient with asymptotic variance $n^{-1}\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}$, while the estimator $\widehat{\theta}^{\mathbb{P}}$ is asymptotic normal but not asymptotic efficient. The estimator $\widehat{\theta}^{\mathbb{Q}}$ becomes inconsistent when $\mathbb{Q}_0$ is false, while the estimator $\widehat{\theta}^{\mathbb{P}}$ is always consistent since $\mathbb{P}_0$ is assumed always to be true. Thus, the statistic

$$n(\widehat{\theta}^{\mathbb{P}} - \widehat{\theta}^{\mathbb{Q}})^T \left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\right]^{-1} (\widehat{\theta}^{\mathbb{P}} - \widehat{\theta}^{\mathbb{Q}})$$

is effectively the Hausman specification test statistic based on GMM for subset of moments. In fact, the result directly follows from Theorem 3 of Newey (1985) with D_Θ to be the rank of $\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} - \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}$. □

Proposition 24. *Denote* $\mathbf{v} \equiv \nabla f(\theta_0)$*. Under the assumptions in Subsection A.3.3, it holds that the Hausman test statistic based on MLEs satisfies*

$$n(f(\widehat{\theta}^{\mathbb{P}}) - f(\widehat{\theta}^{\mathbb{Q}}))^T \left[\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v} - \mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1} (f(\widehat{\theta}^{\mathbb{P}}) - f(\widehat{\theta}^{\mathbb{Q}})) \rightsquigarrow \chi^2_{D_f}$$

Proof. By using the Delta method and Proposition 23, it follows that $\sqrt{n}(f(\widehat{\theta}^{\mathbb{P}}) - f(\widehat{\theta}^{\mathbb{Q}}))$ has asymptotic normal distribution with the asymptotic covariance matrix

$$\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v} - \mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}.$$

According to continuous mapping theorem, we know that the result holds. □

Proposition 25. *Suppose that the assumptions in Subsection A.3.3 are satisfied. Define the sets*

$$\mathcal{J}_{1,n}(\delta,\eta) \equiv \left\{ \left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}} \leq \eta \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1},\ \forall\, \theta \in \Omega(\theta_0,\delta) \text{ and } \tilde{\theta} \in \Omega(\theta,\delta) \right\},$$

and

$$\mathcal{J}_{2,n}(\theta,\delta,\eta) \equiv \left\{ \left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}} \leq \eta \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1},\ \forall\, \tilde{\theta} \in \Omega(\theta,\delta) \right\}.$$

Then, for any $\eta > 0$ there exists small enough positive constants δ_1 and δ such that

$$\mathbb{P}_{0,n}\mathcal{J}_{1,n}(\delta,\eta)^c = o\left(\frac{1}{n}\right) \text{ and } \sup_{\theta\in\Omega(\theta_0,\delta_1)} \mathbb{P}_{\theta,n}\mathcal{J}_{2,n}(\theta,\delta,\eta)^c = o\left(\frac{1}{n}\right).$$

Proof. Appealing to the fact that the Spectral norm and the Frobenius norm are equivalent for the $D_\Theta \times D_\Theta$ matrixes and following the argument on page 49 of Clarke and Barron (1994) or page 465 of Clarke and Barron (1990), we can prove the results very similarly. We omit the detailed proofs to avoid tedious repetition of the proofs in Clarke and Barron (1990) and Clarke and Barron (1994). □

Corollary 8. *Suppose that the assumptions in Subsection A.3.3 are satisfied. Define the sets*

$$\mathcal{J}_{3,n}(\delta,\eta) \equiv \Big\{ 1-\eta \leq \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \leq 1+\eta,$$
$$\forall\, \theta \in \Omega(\theta_0,\delta) \text{ and } \tilde{\theta} \in \Omega(\theta,\delta) \Big\},$$

and

$$\mathcal{J}_{4,n}(\theta,\delta,\eta) \equiv \left\{1-\eta \leq \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \leq 1+\eta,\ \ \forall\, \tilde{\theta} \in \Omega(\theta,\delta)\right\}.$$

Then, for any $\eta > 0$ there exists small enough positive constants δ_1 and δ such that

$$\mathbb{P}_{0,n}\mathcal{J}_{3,n}(\delta,\eta)^c = o\left(\frac{1}{n}\right),\ \ \textit{and}\ \ \sup_{\theta\in\Omega(\theta_0,\delta_1)} \mathbb{P}_{\theta,n}\mathcal{J}_{4,n}(\theta,\delta,\eta)^c = o\left(\frac{1}{n}\right).$$

Proof. We have

$$\begin{aligned}
&\left|\left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} - 1\right| \\
&\leq \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2} - I\right|\right|_{\mathcal{S}} \\
&= \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\left[\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right]\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \\
&\leq \left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}}\left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}} \qquad \text{(A.107)}
\end{aligned}$$

The first inequality is due to the triangular inequality for spectral norm. The second inequality is because for each unit vector $\mathbf{v}$ $in\mathbb{R}^d$,

$$\begin{aligned}
&\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\left[\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right]\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\mathbf{v} \\
&\leq \lambda_M\left(\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right)\left|\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|^2 \\
&= \left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v} \\
&\leq \left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}}\left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}.
\end{aligned}$$

Therefore, the results of this corollary follow directly from the inequality (A.107) and the results of Proposition 25. □

Proposition 26. *Under the assumptions in Subsection A.3.3, for any $\eta > 0$ there exists $\delta > 0$ such that*

$$\mathbb{P}_{0,n}\left\{\sup_{\theta\in\Omega(\theta_0,\delta)} s_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}s_{\mathbb{P},n}(\theta) < \eta\right\}^c = o(1).$$

Proof. Due to the continuity, we know that there exists $\delta_1 > 0$ such that for all $\theta \in \Omega(\theta_0, \delta_1)$,

$$\frac{1}{2} < \lambda_m \left(\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right) \leq \lambda_M \left(\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right) < 2.$$

It follows that for all $\theta \in \Omega(\theta_0, \delta_1)$

$$\begin{aligned} s_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}s_{\mathbb{P},n}(\theta) &\leq 2s_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}s_{\mathbb{P},n}(\theta) \\ &\leq 4s_{\mathbb{P},n}(\theta_0)^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}s_{\mathbb{P},n}(\theta_0) \\ &\quad + 4\left[s_{\mathbb{P},n}(\theta) - s_{\mathbb{P},n}(\theta_0)\right]^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\left[s_{\mathbb{P},n}(\theta) - s_{\mathbb{P},n}(\theta_0)\right]. \end{aligned}$$

By Taylor's expansion of the score function $s_{\mathbb{P},n}(\theta)$ around θ_0, we know that there exists $\tilde{\theta}$ between θ_0 and θ such that

$$\begin{aligned} &\left[s_{\mathbb{P},n}(\theta) - s_{\mathbb{P},n}(\theta_0)\right]^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\left[s_{\mathbb{P},n}(\theta) - s_{\mathbb{P},n}(\theta_0)\right] \\ &= (\theta - \theta_0)^T\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta - \theta_0) \\ &\leq \underline{\lambda}^{-1}(\theta - \theta_0)^T\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})^2(\theta - \theta_0). \end{aligned} \tag{A.108}$$

where the inequality above is due to the fact that $\underline{\lambda} \leq \underline{\lambda}(\theta_0)$. According to Proposition 25, there exists $\delta_2 \in (0, \delta_1)$ such that

$$\mathbb{P}_{0,n}\mathcal{J}_{1,n}(\delta_2, 1)^c = o\left(\frac{1}{n}\right),$$

where

$$\begin{aligned} &\mathcal{J}_{1,n}(\delta_2, 1) \\ &\quad \equiv \left\{\left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}} \leq \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1},\ \forall\, \theta \in \Omega(\theta_0, \delta_2) \text{ and } \tilde{\theta} \in \Omega(\theta, \delta_2)\right\}. \end{aligned}$$

Therefore, we only need to focus on the big probability set $\mathcal{J}_{1,n}(\delta_2, 1)$. Thus, by the triangular inequality, for any $\theta \in \Omega(\theta_0, \delta_2)$, we know that

$$\left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\right|\right|_{\mathcal{S}} \leq ||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathcal{S}} + \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1} \leq \overline{\lambda} + \underline{\lambda}.$$

Then, following the inequality (A.108), if we restrict on the big probability set $\mathcal{I}_{1,n}(\delta_2, 1)$, it follows that

$$\left[s_{\mathbb{P},n}(\theta) - s_{\mathbb{P},n}(\theta_0)\right]^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \left[s_{\mathbb{P},n}(\theta) - s_{\mathbb{P},n}(\theta_0)\right] \leq \underline{\lambda}^{-1}(\overline{\lambda} + \underline{\lambda})^2|\theta - \theta_0|^2.$$

Therefore, we choose

$$\delta = \min\left\{\delta_1, \delta_2, \sqrt{\frac{\eta}{8\underline{\lambda}^{-1}(\overline{\lambda} + \underline{\lambda})^2}}\right\},$$

and when $\theta \in \Omega(\theta_0, \delta)$ and $\mathbf{x^n} \in \mathcal{I}_{1,n}(\delta, 1)$,

$$s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) \leq 4 s_{\mathbb{P},n}(\theta_0)^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} s_{\mathbb{P},n}(\theta_0) + \frac{\eta}{2}.$$

By Markov's inequality, it is straightforward to see that

$$s_{\mathbb{P},n}(\theta_0)^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} s_{\mathbb{P},n}(\theta_0) \to 0 \text{ in } \mathbb{P}_{0,n}.$$

Therefore, we have shown that

$$\mathbb{P}_{0,n}\left\{\sup_{\theta \in \Omega(\theta_0,\delta)} s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) < \eta\right\}^c = o(1).$$

□

Proposition 27. *Let's define*

$$\mathcal{S}_n(\delta, \eta) \equiv \left\{\left|\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) d\theta - D_{\Theta}\right| < \eta\right\}.$$

Suppose that the assumptions in Subsection A.3.3 hold. For any $\eta > 0$, there exists $\delta > 0$ such *that*

$$\mathbb{P}_{0,n}\mathcal{S}_n(\delta,\eta)^c = o(1).$$

Proof. We first show that for any $\eta > 0$

$$\mathbb{P}_{0,n}\left\{\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})S_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}S_{\mathbb{P},n}(\theta)\mathrm{d}\theta > D_\Theta + \eta\right\} = o(1). \tag{A.109}$$

According to Corollary 8, we know that there exists $\delta_1 > 0$ such that

$$\mathbb{P}_{0,n}\mathcal{J}_{3,n}\left(\delta_1, \frac{\eta}{2D_\Theta}\right)^c = o\left(\frac{1}{n}\right),$$

where

$$\begin{aligned}
&\mathcal{J}_{3,n}\left(\delta_1, \frac{\eta}{2D_\Theta}\right) \\
&\equiv \left\{\left(1-\frac{\eta}{2D_\Theta}\right)^{1/2} \le \left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}} \le \left(1+\frac{\eta}{2D_\Theta}\right)^{1/2},\right. \\
&\qquad\qquad \left.\forall\, \theta \in \Omega(\theta_0,\delta_1) \text{ and } \tilde{\theta} \in \Omega(\theta,\delta_1)\right\}.
\end{aligned}$$

By Proposition 15, the set $\mathcal{A}_n(\delta_1) \equiv \left\{\widehat{\theta}^{\mathbb{P}} \in \Omega(\theta_0,\delta_1)\right\}$ has probability going to 1. Thus, on the big probability event $\mathcal{J}_{3,n}\left(\delta_1, \frac{\eta}{2D_\Theta}\right) \cap \mathcal{A}_n(\delta_1)$, by Taylor's expansion, we have

$$\begin{aligned}
&S_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}S_{\mathbb{P},n}(\theta) \\
&= n(\theta-\widehat{\theta}^{\mathbb{P}})^T\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}}) \\
&= n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta)^{1/2}\left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right]^2\mathbf{I}_{\mathbb{P}}(\theta)^{1/2}(\theta-\widehat{\theta}^{\mathbb{P}}) \\
&\le n\left(1+\frac{\eta}{2D_\Theta}\right)^{1/2}(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta)(\theta-\widehat{\theta}^{\mathbb{P}})
\end{aligned}$$

where $\tilde{\theta}$ is between θ and $\widehat{\theta}^{\mathbb{P}}$.

By the continuity, we know that there exists $\delta_2 > 0$ such that for all $\theta \in \Omega(\theta_0, \delta_2)$

$$\left|\left|\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\mathbf{I}_{\mathbb{P}}(\theta)\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\right|\right|_{\mathcal{S}} \leq \left(1 + \frac{\eta}{2D_\Theta}\right)^{1/2}.$$

Choose $\delta \equiv \min\{\delta_1, \delta_2\}$. Thus, when considering $\theta \in \Omega(\theta_0, \delta)$ and restricting on the event $\mathcal{I}_{3,n}\left(\delta, \frac{\eta}{2D_\Theta}\right) \cap \mathcal{A}_n(\delta)$, we have

$$S_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}^{-1}(\theta)S_{\mathbb{P},n}(\theta) \leq \left(1 + \frac{\eta}{2D_\Theta}\right) n(\theta - \widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta - \widehat{\theta}^{\mathbb{P}}).$$

Therefore, we have

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})S_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}^{-1}(\theta)S_{\mathbb{P},n}(\theta)\mathrm{d}\theta \\
&\leq \left(1 + \frac{\eta}{2D_\Theta}\right)\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})n(\theta - \widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta - \widehat{\theta}^{\mathbb{P}})\mathrm{d}\theta.
\end{aligned}$$

According to Theorem 1 and Proposition 3 of Chernozhukov and Hong (2003), we know that

$$\left(1 + \frac{\eta}{2D_\Theta}\right)\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})n(\theta - \widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta - \widehat{\theta}^{\mathbb{P}})\mathrm{d}\theta \to D_\Theta + \frac{\eta}{2} \text{ in } \mathbb{P}_{0,n}.$$

Therefore, the limit result in (A.109) holds.

The proof of the following limit result is quite similar, $\forall\ \eta > 0$

$$\mathbb{P}_{0,n}\left\{\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})S_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}S_{\mathbb{P},n}(\theta)\mathrm{d}\theta > D_\Theta - \eta\right\} = o(1). \tag{A.110}$$

So, we ignore the detailed proof. □

Proposition 28. *Let's define*

$$\mathcal{S}_{\mathbf{v},n}(\delta,\eta) \equiv \left\{\left|\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})\frac{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}S_{\mathbb{P},n}(\theta)S_{\mathbb{P},n}(\theta)^T\mathbf{I}_{\mathbb{P}}(\theta)\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}d\theta - 1\right| < \eta\right\}.$$

Suppose that the assumptions in Subsection A.3.3 hold. For any $\eta > 0$, there exists $\delta > 0$ such *that*

$$\mathbb{P}_{0,n}\mathcal{S}_{\mathbf{v},n}(\delta,\eta)^c = o(1).$$

Proof. The proof is similar to that of Proposition 27. □

Proposition 29. *Under the assumptions in Subsection A.3.3, for any open subset $\mathcal{N} \subset \Theta$ and open set $\mathcal{N}_0 \subset \mathcal{N}$ such that $d_L(\mathcal{N}^c, \mathcal{N}_0) > \delta$ for some $\delta > 0$, there exist positive constants C, ξ_1 and ξ_2 such that*

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \leq e^{\xi_1 n}\int_{\mathcal{N}^c}\pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)d\vartheta\right\} \leq Ce^{-\xi_2 n}.$$

Proof. According to Lemma 4 and Lemma 5 at the end of Section A.3.3, the assumptions in Subsection A.3.3 guarantee the existence of **strongly uniformly exponentially consistent (SUEC)** hypothesis tests.

In particular, for any open subset $\mathcal{N} \subset \Theta$ and $\mathcal{N}_0 \subset \mathcal{N}$ such that $d_L(\mathcal{N}, \mathcal{N}_0) > \delta$ for some $\delta > 0$, for each $\theta \in \mathcal{N}_0$, there exists a sequence of tests with acceptance region $\mathcal{A}_{\theta,n}$ for null hypothesis $\theta' = \theta$ versus $\theta' \in \mathcal{N}^c$ such that

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\mathcal{A}^c_{\theta,n} \leq Ce^{-\xi n} \text{ and } \sup_{\theta\in\mathcal{N}_0}\sup_{\theta'\in\mathcal{N}^c} \mathbb{P}_{\theta',n}\mathcal{A}_{\theta,n} < Ce^{-\xi n}, \text{ for some } \xi > 0.$$

Denote the mixture distribution of $\mathbb{P}_{\theta,n}$ with respect to the conditional prior distribution $\pi_{\mathbb{P}}(\cdot|\mathcal{N}^c)$ by $\mathbb{P}_{\mathcal{N}^c,n}$ with density $\pi_{\mathbb{P}}(\mathbf{x^n}|\mathcal{N}^c)$. More precisely, we define

$$\pi_{\mathbb{P}}(\mathbf{x^n}|\mathcal{N}^c) \equiv \int_{\mathcal{N}^c}\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\pi_{\mathbb{P}}(\theta|\mathcal{N}^c)\mathrm{d}\theta. \tag{A.111}$$

Following Lemma 6, we can show that there exist a real number $r > 0$ such that

$$||\mathbb{P}_{\theta,n} - \mathbb{P}_{\mathcal{N}^c,n}||_{TV} \geq 2(1-2e^{-rn}), \;\; \forall\;\; \theta \in \mathcal{N}_0.$$

For any positive sequence ϵ_n, by Markov's inequality, it follows that for each $\theta \in \mathcal{N}_0$

$$\mathbb{P}_{\theta,n}\left\{\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\mathcal{N}^\mathbf{c})}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)} > \epsilon_n\right\} \leq \frac{1}{\epsilon_n^{1/2}} \int \pi_{\mathbb{P}}(\mathbf{x^n}|\theta)^{1/2} \pi_{\mathbb{P}}(\mathbf{x^n}|\mathcal{N}^c)^{1/2} d\mathbf{x^n}$$
$$= \frac{1}{\epsilon_n^{1/2}} \alpha_H(\mathbb{P}_{\theta,n}, \mathbb{P}_{\mathcal{N}^c,n}) \leq \frac{1}{\epsilon_n^{1/2}} \sqrt{1 - \left(\frac{1}{2}||\mathbb{P}_{\theta,n} - \mathbb{P}_{\mathcal{N}^c,n}||_{TV}\right)^2}$$
$$\leq \frac{1}{\epsilon_n^{1/2}} \sqrt{1 - (1 - 2e^{-rn})^2} = \frac{2e^{-\frac{rn}{2}}}{\epsilon_n^{1/2}} \sqrt{1 - e^{-rn}} \leq \frac{2e^{-\frac{rn}{2}}}{\epsilon_n^{1/2}}.$$

If we choose $\epsilon_n = e^{-\frac{rn}{4}}$, then we have

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\mathcal{N}^c)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)} > e^{-\frac{rn}{4}}\right\} \leq 2e^{-\frac{rn}{4}}.$$

□

Proposition 30. *Under the assumptions in Subsection A.3.3, for any open subset $\mathcal{N} \subset \Theta$ and open set $\mathcal{N}_0 \subset \mathcal{N}$ such that $d_L(\mathcal{N}^c, \mathcal{N}_0) > \delta$ for some $\delta > 0$, then there exist positive constants C, ξ_1, ξ_2 such that*

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \leq e^{\xi_1 n} \int_{\mathcal{N}_{-1}(\theta_{(1)})^c} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)}) d\theta'_{(-1)}\right\}$$
$$\leq Ce^{-\xi_2 n}. \tag{A.112}$$

Proof. The proof is the same as that of Proposition 29. □

Proposition 31. *Under the assumptions in Subsection A.3.3, for any open neighborhood $\mathcal{N} \subset \Theta$ of θ_0 there exist positive constants C and ξ such that*

$$\mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0) \leq e^{\xi n} \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta) \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\vartheta) d\vartheta\right\} = o(1). \tag{A.113}$$

Proof. For any open neighborhood $\mathcal{N}$ of θ_0, from the identification assumption (i.e. **Assumption ID**) and the compactness of Θ, it follows that there exists $\epsilon > 0$ such that $\min_{\theta \in \mathcal{N}^c} H_{\mathbb{Q}}(\theta) \geq \epsilon$ where $H_{\mathbb{Q}}(\theta)$ is defined in (A.96). Consider the large probability set

$$\mathcal{A}_n \equiv \left\{ \sup_{\theta\in\Theta} |\widehat{H}_{\mathbb{Q},n}(\theta) - H_{\mathbb{Q}}(\theta)| < \epsilon/2 \right\},$$

where $\widehat{H}_{\mathbb{Q},n}(\theta)$ is defined in (A.97).

From Proposition 18, we know that $\mathbb{Q}_{0,n}\mathcal{A}_n \to 1$ as $n \to \infty$. Thus, we only need to focus on event $\mathcal{A}_n$. Then, we have

$$\begin{aligned}
&\mathbb{Q}_{0,n}\mathcal{A}_n \left\{ \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0) \le e^{n\epsilon/4} \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\theta')\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta')\mathrm{d}\theta' \right\} \\
&= \mathbb{Q}_{0,n}\mathcal{A}_n \left\{ e^{-n\widehat{H}_{\mathbb{Q},n}(\theta_0)} \le e^{n\epsilon/4} \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\theta')e^{-n\widehat{H}_{\mathbb{Q},n}(\theta)}\mathrm{d}\theta' \right\} \\
&\le \mathbb{Q}_{0,n} \left\{ e^{-n\widehat{H}_{\mathbb{Q},n}(\theta_0)} \le e^{n\epsilon/4} \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\theta')e^{-n\left[H_{\mathbb{Q}}(\theta')-\epsilon/2\right]}\mathrm{d}\theta' \right\} \\
&\le \mathbb{Q}_{0,n} \left\{ e^{-n\widehat{H}_{\mathbb{Q},n}(\theta_0)} \le e^{n\epsilon/4} \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\theta')e^{-n\epsilon/2}\mathrm{d}\theta' \right\} \\
&\le \mathbb{Q}_{0,n} \left\{ e^{-n\widehat{H}_{\mathbb{Q},n}(\theta_0)} \le e^{-n\epsilon/4} \right\} \le e^{-n\epsilon/16}\mathbb{E}_{\mathbb{Q}_0} e^{n\widehat{H}_{\mathbb{Q},n}(\theta_0)/4}
\end{aligned}$$

Because $n\widehat{H}_{\mathbb{Q},n}(\theta_0)$ converges to a chi-squire random variable with degree of freedom D_Θ in distribution, we know that $\mathbb{E}_{\mathbb{Q}_0} e^{n\widehat{H}_{\mathbb{Q},n}(\theta_0)/4} \to 2^{D_\Theta/2}$. Thus,

$$\mathbb{Q}_{0,n}\mathcal{A}_n \left\{ \pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta_0) \le e^{n\epsilon/4} \int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\theta')\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n}|\theta')\mathrm{d}\theta' \right\} \le Ce^{-n\epsilon/16}$$

for some constant $C > 0$. Therefore, if we take $\xi = \epsilon/4$, the proof is completed. □

We introduce Le Cam's theory on hypothesis testing (see, e.g. Le Cam and Yang, 2000, Chapter 8).

Proposition 32. *Under the regularity conditions in Subsection A.3.3, there are test functions $\mathcal{A}_n$ and positive coefficients C, ξ, ϵ and K such that $\mathbb{P}_{0,n}(1 - \mathcal{A}_n) \to 0$ and $\mathbb{P}_{\theta,n}\mathcal{A}_n \le Ce^{-n\xi|\theta-\theta_0|^2/2}$ for all θ such that $K/\sqrt{n} \le |\theta - \theta_0| \le \epsilon$.*

Proof. For all $z \in \mathbb{R}^{K(D_x+D_y)}$, define the rectangular $F_z \equiv (-\infty, z_1] \times (-\infty, z_2] \times \cdots \times (-\infty, z_{K(D_x+D_y)}]$. The empirical process is defined as $\widehat{\Pi}_n(z) \equiv \widehat{\mathbb{P}}_n F_z = n^{-1}\sum_{t=1}^n \mathbf{1}_{\{\mathbf{z_t}\in F_z\}}$. We define $\Pi_\theta(z) \equiv \mathbb{P}_\theta F_z$. According to Le Cam and Yang (2000, Page 250), there exists positive constants c and ϵ such that $\sup_x \left|\Pi_{\theta_0}(x) - \Pi_\theta(x)\right| > c|\theta - \theta_0|$ for $|\theta - \theta_0| \leq \epsilon$. Denote the expectation to be

$$\mu_n(\theta) \equiv \mathbb{E}_{\mathbb{P}_\theta} \sup_x \left|\Pi_{\theta_0}(x) - \Pi_\theta(x)\right|.$$

By the classical result of weak convergence for the Kolmogorov-Smirnov statistic

$$\sqrt{n} \sup_x \left|\Pi_{\theta_0}(x) - \Pi_\theta(x)\right|,$$

we know that there exists a large constant M such that $\mu_n(\theta) \leq \frac{M}{2c\sqrt{n}}$ for all $|\theta - \theta_0| \leq \epsilon$. We choose $K \equiv \frac{4M}{c}$. Consider the test functions $\mathcal{A}_n = \{\sup_z \left|\Pi_{\theta_0}(z) - \Pi_\theta(z)\right| < K/\sqrt{n}\}$. Using triangular inequality, we obtain

$$\mathbb{P}_{\theta,n}\mathcal{A}_n \leq \mathbb{P}_{\theta,n}\left\{\frac{c}{2}|\theta - \theta_0| \leq \sup_z \left|\widehat{\Pi}_n(z) - \Pi_\theta(z)\right| - \mu_n(\theta)\right\}$$

Using DvoretzkyŰ-KieferŰ-Wolfowitz type inequality for uniform mixing variables in Samson (2000, Theorem 3), we can show there exists $\xi > 0$ such that

$$\mathbb{P}_{\theta,n}\left\{\frac{c}{2}|\theta - \theta_0| \leq \sup_z \left|\widehat{\Pi}_n(z) - \Pi_\theta(z)\right| - \mu_n(\theta)\right\} \leq e^{-\xi|\theta-\theta_0|^2/2}$$

for all $K/\sqrt{n} \leq |\theta - \theta_0| \leq \epsilon$. Thus, $\mathbb{P}_{\theta,n}\mathcal{A}_n \leq e^{-\xi|\theta-\theta_0|^2/2}$ for all $K/\sqrt{n} \leq |\theta - \theta_0| \leq \epsilon$. By the same inequality, it is straightforward to get $\mathbb{P}_{0,n}(1 - \mathcal{A}_n) \to 0$. ☐

Proposition 33. *Under the assumptions in Subsection A.3.3, for any open subsets $\mathcal{N} \subset \Theta$ and any positive constant ξ, there exists a neighborhood $\mathcal{N}_0$ of θ_0 such that*

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \geq e^{\xi n} \int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)d\vartheta\right\} = o\left(\frac{1}{n}\right). \tag{A.114}$$

Proof. Let $r_n = 1/\sqrt{n}$ and it is sufficient to show that

$$\sup_{\theta \in \mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \geq e^{\xi n} \int_{\Omega(\theta,r_n)} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\} = o\left(\frac{1}{n}\right).$$

It is equivalent to show that

$$\sup_{\theta\in\mathcal{N}_0}\mathbb{P}_{\theta,n}\left\{\ln\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta,r_n))}\geq\xi_n n\right\}=o\left(\frac{1}{n}\right),$$

where

$$\xi_n\equiv\xi-\frac{1}{n}\ln\pi_{\mathbb{P}}(\Omega(\theta,r_n))$$

with

$$\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta,r_n))\equiv\int_{\Omega(\theta,r_n)}\frac{\pi_{\mathbb{P}}(\vartheta)}{\pi_{\mathbb{P}}(\Omega(\theta,r_n))}\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta.$$

In fact, we have

$$\pi_{\mathbb{P}}(\Omega(\theta,r_n))=\int_{\Omega(\theta,r_n)}\pi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta\geq m_\pi\Gamma_{D_\Theta}\left(\frac{1}{n}\right)^{D_\Theta/2},$$

where Γ_{D_Θ} is the volume of the unit ball in $\mathbb{R}^{D_\Theta}$. Thus,

$$\xi_n=\xi-O(n^{-1}\ln n).$$

Therefore, for all large n, $\xi_n\geq\xi/2$ and hence it suffices to show that

$$\sup_{\theta\in\mathcal{N}_0}\mathbb{P}_{\theta,n}\left\{\ln\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta,r_n))}\geq\xi n/2\right\}=o\left(\frac{1}{n}\right).$$

For each $\theta\in\Theta$, by Markov's inequality, we have

$$\mathbb{P}_{\theta,n}\left\{\ln\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta,r_n))}\geq\xi n/2\right\}\leq\frac{4}{n^2\xi^2}\mathbb{E}_{\mathbb{P}_\theta}\left[\ln\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta,r_n))}\right]^2\tag{A.115}$$

We consider the set

$$\mathcal{I}_{4,n}(\theta,\delta,1)\equiv\left\{\left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right|\right|_{\mathcal{S}}\leq 2,\ \forall\,\tilde{\theta}\in\Omega(\theta,\delta)\right\}.$$

According to Corollary 8, it follows that there exist positive constants δ and δ_0 such that

$$\sup_{\theta \in \Omega(\theta_0, \delta_0)} \mathbb{P}_{\theta,n} \mathcal{I}_{4,n}(\theta, \delta, 1)^c = o\left(\frac{1}{n}\right).$$

Therefore, we only need to focus on the big probability set $\mathcal{I}_{4,n}(\theta, \delta, 1)$ for each $\theta \in \Omega(\theta_0, \delta_0)$.

We choose $\mathcal{N}_0 = \Omega(\theta_0, \delta_0)$. We have for each $\theta \in \mathcal{N}_0$ the following equality holds

$$\begin{aligned}
\ln \frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)} &= \ln \int_{\Omega(\theta,r_n)} \pi_{\mathbb{P}}(\vartheta|\Omega(\theta, r_n)) \frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)} \mathrm{d}\vartheta \\
&= \ln \int_{\Omega(\theta,r_n)} \pi_{\mathbb{P}}(\vartheta|\Omega(\theta, r_n)) e^{\sqrt{n} S_{\mathbb{P},n}(\theta)^T(\vartheta-\theta) - \frac{1}{2} n(\vartheta-\theta)^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\vartheta-\theta)} \mathrm{d}\vartheta,
\end{aligned}$$

where $\tilde{\theta}$ is between θ and ϑ.

On the one hand, on the event $\mathcal{I}_{4,n}(\theta, \delta, 1)$, we have

$$\begin{aligned}
\ln \frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)} &\leq \ln \int_{\Omega(\theta,r_n)} \pi_{\mathbb{P}}(\vartheta|\Omega(\theta, r_n)) e^{|S_{\mathbb{P},n}(\theta)| + ||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathcal{S}}} \mathrm{d}\vartheta \\
&= |S_{\mathbb{P},n}(\theta)| + ||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathcal{S}}.
\end{aligned}$$

Thus, on the other hand, on the event $\mathcal{I}_{4,n}(\theta, \delta, 1)$, we have by Jensen's inequality

$$\begin{aligned}
&\ln \frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)} \\
&\geq \int_{\Omega(\theta,r_n)} \left[\sqrt{n} S_{\mathbb{P},n}(\theta)^T(\vartheta - \theta) - \frac{1}{2}(\vartheta - \theta)^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\vartheta - \theta)\right] \pi_{\mathbb{P}}(\vartheta|\Omega(\theta, r_n)) \mathrm{d}\vartheta \\
&\geq \int_{\Omega(\theta,r_n)} \left[\sqrt{n} S_{\mathbb{P},n}(\theta)^T(\vartheta - \theta) - (\vartheta - \theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\vartheta - \theta)\right] \pi_{\mathbb{P}}(\vartheta|\Omega(\theta, r_n)) \mathrm{d}\vartheta \\
&\geq -|S_{\mathbb{P},n}(\theta)| - ||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathcal{S}}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\left[\ln \frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta, r_n))}\right]^2 \\
&\quad \leq \left[|S_{\mathbb{P},n}(\theta)| + ||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathcal{S}}\right]^2 \leq 2|S_{\mathbb{P},n}(\theta)|^2 + 2\,||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathcal{S}}^2. \qquad \text{(A.116)}
\end{aligned}$$

Combining (A.115) and (A.116), we know that

$$\mathbb{P}_{\theta,n}\left\{\ln\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}|\Omega(\theta,r_n))}\geq \xi n/2\right\}\leq\frac{8}{n^2\xi^2}\left[\mathbb{E}_{\mathbb{P}_\theta}|S_{\mathbb{P},n}(\theta)|^2+\mathbb{E}_{\mathbb{P}_\theta}||\mathbf{I}_{\mathbb{P}}(\theta)||_{\mathbb{S}}^2\right]$$
$$\leq\frac{8}{n^2\xi^2}\left[\mathrm{tr}\left(\mathbf{I}_{\mathbb{P}}(\theta)\right)+\lambda_M\left(\mathbf{I}_{\mathbb{P}}(\theta)\right)^2\right]\leq\frac{8(D_\Theta\overline{\lambda}+\overline{\lambda}^2)}{n^2\xi^2}.$$

□

Proposition 34. *Under the assumptions in Subsection A.3.3, for any open subsets $\mathcal{N}\subset\Theta$ and any positive constant ξ, we have*

$$\mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)\geq e^{\xi n}\int_{\mathcal{N}}\pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)d\vartheta\right\}=o\left(\frac{1}{n}\right),\quad \textit{for every } \theta\in\mathcal{N}.$$

Proof. Let $r_n=1/\sqrt{n}$ and it is sufficient to show that

$$\mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)\geq e^{\xi n}\int_{\Omega(\theta,r_n)}\pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\}=o\left(\frac{1}{n}\right).$$

It is equivalent to show that

$$\mathbb{Q}_{0,n}\left\{\ln\frac{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)}{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\Omega(\theta,r_n))}\geq \xi_n n\right\}=o\left(\frac{1}{n}\right),$$

where

$$\xi_n\equiv\xi-\frac{1}{n}\ln\pi_{\mathbb{P}}(\Omega(\theta,r_n))$$

with

$$\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\Omega(\theta,r_n))\equiv\int_{\Omega(\theta,r_n)}\frac{\pi_{\mathbb{P}}(\vartheta)}{\pi_{\mathbb{P}}(\Omega(\theta,r_n))}\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta.$$

Thus, we have

$$\pi_{\mathbb{P}}(\Omega(\theta,r_n))=\int_{\Omega(\theta,r_n)}\pi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta\geq m_\pi\Gamma_{D_\Theta}\left(\frac{1}{n}\right)^{D_\Theta/2},$$

where Γ_{D_Θ} is the volume of the unit ball in $\mathbb{R}^{D_\Theta}$. Thus,

$$\xi_n = \xi - O(n^{-1}\ln n).$$

Therefore, for all large n, $\xi_n \geq \xi/2$ and hence it suffices to show that

$$\mathbb{Q}_{0,n}\left\{\ln\frac{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)}{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\Omega(\theta,r_n))} \geq \xi n/2\right\} = o\left(\frac{1}{n}\right).$$

By Markov's inequality, we have

$$\begin{aligned}\mathbb{Q}_{0,n}&\left\{\ln\frac{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)}{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\Omega(\theta,r_n))} \geq \xi n/2\right\}\\ &\leq \frac{4}{n^2\xi^2}\mathbb{E}_{\mathbb{Q}_0}\left[\ln\frac{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)}{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\Omega(\theta,r_n))}\right]^2 \qquad \text{(A.117)}\end{aligned}$$

We consider the set

$$\mathcal{I}_n \equiv \left\{\left|\left|\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{Q},n}(\tilde{\theta})\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1/2}\right|\right|_{\mathcal{S}} \leq 2,\ \ \forall\,\tilde{\theta}\in\Omega(\theta_0,r_n)\right\}.$$

According to Proposition 21, it follows that $\mathbb{Q}_{0,n}\mathcal{I}_n^c = o\,(1)$. Therefore, we only need to focus on the big probability set $\mathcal{I}_n$. It holds that

$$\begin{aligned}\ln\frac{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\Omega(\theta_0,r_n))}{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0)} &= \ln\int_{\Omega(\theta_0,r_n)}\pi_{\mathbb{P}}(\theta|\Omega(\theta_0,r_n))\frac{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta)}{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0)}\mathrm{d}\theta\\ &= \ln\int_{\Omega(\theta_0,r_n)}\pi_{\mathbb{P}}(\theta|\Omega(\theta_0,r_n))e^{n\widehat{H}_{\mathbb{Q},n}(\theta_0)-n\widehat{H}_{\mathbb{Q},n}(\theta)}\mathrm{d}\theta. \qquad \text{(A.118)}\end{aligned}$$

Now, because

$$\sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n}g(\mathbf{x^n},\mathbf{y^n},\theta)\right] = \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n}g(\mathbf{x_t},\mathbf{y_t},\theta_0)\right] + \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\tilde{\theta})}{\partial\theta}\right](\theta-\theta_0),$$

where $\tilde{\theta}$ is between θ_0 and θ. Thus,we have

$$\begin{aligned}
\left| n\widehat{H}_{\mathbb{Q},n}(\theta_0) - n\widehat{H}_{\mathbb{Q},n}(\theta) \right| &\le \frac{1}{2} n(\theta-\theta_0)^T \widehat{G}_n(\tilde{\theta})^T S_0^{-1} \widehat{G}_n(\tilde{\theta})(\theta-\theta_0) \\
&\quad + n \left\{ \left[\frac{1}{n} \sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta_0) \right]^T S_0^{-1} \left[\frac{1}{n} \sum_{t=1}^{n} g(\mathbf{x_t}, \mathbf{y_t}, \theta_0) \right] \right\}^{1/2} \\
&\quad \times \left\{ (\theta-\theta_0)^T \widehat{G}_n(\tilde{\theta})^T S_0^{-1} \widehat{G}_n(\tilde{\theta})(\theta-\theta_0) \right\}^{1/2} \\
&= \frac{1}{2} n(\theta-\theta_0)^T \widehat{\mathbf{I}}_{\mathbb{Q},n}(\theta_0)(\theta-\theta_0) + \left\{ n\widehat{H}_{\mathbb{Q},n}(\theta_0) \right\}^{1/2} \left\{ n(\theta-\theta_0)^T \widehat{\mathbf{I}}_{\mathbb{Q},n}(\theta_0)(\theta-\theta_0) \right\}^{1/2}.
\end{aligned}$$

Thus, on the event $\mathcal{I}_n$, we have

$$\begin{aligned}
&\left| n\widehat{H}_{\mathbb{Q},n}(\theta_0) - n\widehat{H}_{\mathbb{Q},n}(\theta) \right| \\
&\quad \le n(\theta-\theta_0)^T \mathbf{I}_{\mathbb{Q}}(\theta_0)(\theta-\theta_0) \\
&\qquad + \left\{ n\widehat{H}_{\mathbb{Q},n}(\theta_0) \right\}^{1/2} \left\{ n(\theta-\theta_0)^T \mathbf{I}_{\mathbb{Q}}(\theta_0)(\theta-\theta_0) \right\}^{1/2} \\
&\quad \le ||\mathbf{I}_{\mathbb{Q}}(\theta_0)||_{\mathcal{S}} + \left\{ n\widehat{H}_{\mathbb{Q},n}(\theta_0) \right\}^{1/2} ||\mathbf{I}_{\mathbb{Q}}(\theta_0)||_{\mathcal{S}}^{1/2}.
\end{aligned}$$

Combining (A.118), we know that on the event $\mathcal{I}_n$, it holds that

$$\left[\frac{\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n} | \Omega(\theta_0, r_n))}{\pi_{\mathbb{Q}}(\mathbf{x^n}, \mathbf{y^n} | \theta_0)} \right]^2 \le ||\mathbf{I}_{\mathbb{Q}}(\theta_0)||_{\mathcal{S}}^2 + ||\mathbf{I}_{\mathbb{Q}}(\theta_0)||_{\mathcal{S}} \left[n\widehat{H}_{\mathbb{Q},n}(\theta_0) \right].$$

In fact, it easy to see that

$$\mathbb{E}_{\mathbb{Q}_0} \left[n\widehat{H}_{\mathbb{Q},n}(\theta_0) \right] \to D_{\Theta}, \text{ as } n \to \infty.$$

Combining (A.117) and (A.118), we know that

$$\limsup_{n\to\infty} \mathbb{Q}_{0,n} \left\{ \ln \frac{\pi_{\mathbb{P}}(\mathbf{x^n}, \mathbf{y^n} | \theta)}{\pi_{\mathbb{P}}(\mathbf{x^n}, \mathbf{y^n} | \Omega(\theta, r_n))} \ge \xi n/2 \right\} \le \frac{8}{n^2 \xi^2} \left[D_{\Theta} ||\mathbf{I}_{\mathbb{Q}}(\theta_0)||_{\mathcal{S}} + ||\mathbf{I}_{\mathbb{Q}}(\theta_0)||_{\mathcal{S}}^2 \right].$$

□

Proposition 35. *Assume the regularity conditions in Subsection A.3.3 hold. For any open* *neighborhood $\mathcal{N}$ of θ_0, there is an $\xi > 0$ and a open neighborhood $\mathcal{N}_0$ of θ_0 such that*

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)d\vartheta \geq e^{\xi n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)d\vartheta\right\} = o(1).$$

Proof. According to Proposition 33, for any positive constant ξ_1, there exists a neighborhood $\mathcal{N}_0$ of θ_0 such that $d_L(\mathcal{N}^c, \mathcal{N}_0) > \delta$ for some $\delta > 0$ and

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\mathcal{A}_n(\theta,\xi_1) = o\left(\frac{1}{n}\right),$$

with

$$\mathcal{A}_n(\theta,\xi_1) \equiv \left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \geq e^{\xi_1 n}\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\}.$$

By Proposition 29, there exist positive constants C, ξ' and ξ_2 such that

$$\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \leq e^{\xi' n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\} \leq Ce^{-\xi_2 n}.$$

We take a constant $0 < \xi < \xi'$ and let $\xi'' = \xi' - \xi$. Then,

$$\begin{aligned}
&\sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta \leq e^{\xi n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\}\\
&\quad\leq \sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\mathcal{A}_n(\theta,\xi'')\left\{\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta \leq e^{\xi n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\}\\
&\qquad\qquad + \sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\mathcal{A}_n(\theta,\xi'')^c\\
&\quad\leq \sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \leq e^{n(\xi-\xi'')}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\} + o(1)\\
&\quad\leq \sup_{\theta\in\mathcal{N}_0} \mathbb{P}_{\theta,n}\left\{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \leq e^{n\xi'}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta\right\} + o(1)\\
&\quad\leq Ce^{-\xi_2 n} + o(1) = o(1).
\end{aligned}$$

□

Proposition 36. *Assume the regularity conditions in Subsection A.3.3 hold. For any open* *neighborhood $\mathcal{N}$ of θ_0, there is an $\xi > 0$ such that*

$$\mathbb{Q}_{0,n}\left\{\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)d\theta \geq e^{\xi n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)d\vartheta\right\} = o(1).$$

Proof. According to Proposition 34, for any positive constant ξ_1, there exists a neighborhood $\mathcal{N}_0$ of θ_0 such that $d_L(\mathcal{N}^c, \mathcal{N}_0) > \delta$ for some $\delta > 0$ and

$$\mathbb{Q}_{0,n}\mathcal{A}_n(\xi_1) = o\left(\frac{1}{n}\right),$$

with

$$\mathcal{A}_n(\xi_1) \equiv \left\{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0) \geq e^{\xi_1 n}\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\}.$$

By Proposition 31, there exist positive constants C and ξ' such that

$$\mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0) \leq e^{\xi' n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\} = o(1).$$

We take a constant $0 < \xi < \xi'$ and let $\xi'' = \xi' - \xi$. Then,

$$\begin{aligned}
&\mathbb{Q}_{0,n}\left\{\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta \leq e^{\xi n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\} \\
&\quad \leq \mathbb{Q}_{0,n}\mathcal{A}_n(\xi'')\left\{\int_{\mathcal{N}} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta \leq e^{\xi n}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\} \\
&\qquad + \mathbb{Q}_{0,n}\mathcal{A}_n(\xi'')^c \\
&\quad \leq \mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0) \leq e^{n(\xi-\xi'')}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\} + o(1) \\
&\quad \leq \mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\theta_0) \leq e^{n\xi'}\int_{\mathcal{N}^c} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{Q}}(\mathbf{x^n},\mathbf{y^n}|\vartheta)\mathrm{d}\vartheta\right\} + o(1) = o(1).
\end{aligned}$$

□

Proof of Theorem 1 in the Paper

Due to **Assumption FF**, together with the fact that the definition of our "dark matter" measure and other assumptions in Subsection A.3.3 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta) = \theta_1 \equiv (\theta_{(1)}, \cdots, \theta_{(D_f)})^T$ and hence $\partial f(\theta)/\partial\theta^T \equiv \mathbf{v} = \left[I_{D_f}, 0_{D_\Theta - D_f}\right]^T$. Therefore, under the restriction for optimization in constrained GMM, the free parameters are actually θ_2 which is the vector $\theta_2 \equiv (\theta_{D_f+1}, \cdots, \theta_{D_\Theta})$. Denote the inverse mapping to be $\theta = a(\theta_2)$ and it is trivial to see that $\partial a(\theta_2)/\partial\theta_2 \equiv \left[0_{D_\Theta - D_f}, I_{D_f}\right]^T$.

From Proposition 35 and $\sqrt{n}-$consistency of $\widehat{\theta}^{\mathbb{P}}$, we know that we only need to focus on the large probability event $\mathcal{A}_{1,n}$ on which $\theta_0 \in \Omega(\widehat{\theta}^{\mathbb{P}}, M)$ and

$$\int_{\Omega(\widehat{\theta}^{\mathbb{P}}, M)^c} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta < e^{-\xi n} \quad \text{for some } M > 0 \text{ and } \xi > 0.$$

Thus, based on the dominating assumptions on moment functions, we know that

$$\int_{\Omega(\widehat{\theta}^{\mathbb{P}}, M)^c} d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta \leq n \left| \frac{1}{n} \sum_{t=1}^{n} a_1(\mathbf{x_t}, \mathbf{y_t}) \right|^2 \lambda_m(S_0)^{-1} e^{-\xi n} \to 0, \;\; \text{in } \mathbb{P}_0.$$

Now, we define

$$\begin{aligned} h_{S_0}\{\mathbf{x^n}, \mathbf{y^n}\} &\equiv n \left[f(\widehat{\theta}^{\mathbb{Q}}) - f(\widehat{\theta}^{\mathbb{P}})\right]^T \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v}\right]^{-1} \left[f(\widehat{\theta}^{\mathbb{Q}}) - f(\widehat{\theta}^{\mathbb{P}})\right] \\ &= n(\widehat{\theta}^{\mathbb{Q}} - \widehat{\theta}^{\mathbb{P}})^T \mathbf{v} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v}\right]^{-1} \mathbf{v}^T (\widehat{\theta}^{\mathbb{Q}} - \widehat{\theta}^{\mathbb{P}}). \end{aligned}$$

According to Newey (1985), we know that $h_{S_0}\{\mathbf{x^n}, \mathbf{y^n}\}$ converges in distribution to h^∞ which has the following expression

$$\begin{aligned} X^T &\left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v} - \mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v}\right]^{1/2} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v}\right]^{-1} \\ &\times \left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v} - \mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v}\right]^{1/2} X \end{aligned}$$

where X is a D_Θ dimensional standard multivariate normal random variable. The limit distribution has mean

$$\mathbf{tr}\left\{\left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v}\right]^{-1} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v}\right]\right\} - D_f.$$

We denote the demeaned random variable to be

$$\varepsilon \equiv h^{\infty} - \left[\mathbf{tr} \left\{ \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v} \right]^{-1} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v} \right] \right\} - D_f \right].$$

Let's define

$$\Delta_{d,n} \equiv \int_{|\theta - \widehat{\theta}^{\mathbb{P}}| \le M} d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta - h_{S_0}\{\mathbf{x^n}, \mathbf{y^n}\}.$$

It is sufficient for us to show that for any $\epsilon > 0$ such that when n is large enough

$$\mathbb{Q}_{0,n} \left\{ \left| \Delta_{d,n} - \mathbf{tr} \left\{ \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v} \right]^{-1} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v} \right] \right\} \right| > \epsilon \right\} < \epsilon. \tag{A.119}$$

We know that, for any $K > 0$, the triangular inequality implies

$$\begin{aligned} &\left| \Delta_{d,n} - \mathbf{tr} \left\{ \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v} \right]^{-1} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v} \right] \right\} \right| \\ &\quad \le \left| \Delta_{d,n,K} - \mathbf{tr} \left\{ \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v} \right]^{-1} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v} \right] \right\} \right| && \text{(A.120)} \\ &\qquad + \int_{\frac{K}{\sqrt{n}} < |\theta - \widehat{\theta}^{\mathbb{P}}| \le M} d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta, && \text{(A.121)} \end{aligned}$$

where

$$\Delta_{d,n,K} \equiv \int_{|\theta - \widehat{\theta}^{\mathbb{P}}| \le \frac{K}{\sqrt{n}}} d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta - h_{S_0}\{\mathbf{x^n}, \mathbf{y^n}\}.$$

Based on (A.120)and (A.121), we complete our proof in the following two steps. First, we show that for any $\epsilon > 0$, there exists K_1 such that for each $K \ge K_1$, for large enough n

$$\mathbb{Q}_{0,n} \left\{ \int_{\frac{K}{\sqrt{n}} < |\theta - \widehat{\theta}^{\mathbb{P}}| \le M} d_{S_0}\{\mathbf{x^n}, \mathbf{y^n}, f(\theta)\} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \mathrm{d}\theta > \frac{\epsilon}{2} \right\} < \frac{\epsilon}{2}. \tag{A.122}$$

By definition, we know that $0 \le d_{S_0}(\mathbf{x^n}, \mathbf{y^n}, f(\theta)) \le d_{S_0}(\mathbf{x^n}, \mathbf{y^n}, \theta)$. Thus, it suffices to prove that for any $\epsilon > 0$, there exists K_1 such that for each $K \geq K_1$,

$$\mathbb{Q}_{0,n}\left\{\int_{\frac{K}{\sqrt{n}}<|\theta-\widehat{\theta}^{\mathbb{P}}|\leq M} d_{S_0}\{\mathbf{x^n},\mathbf{y^n},\theta\}\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta > \frac{\epsilon}{2}\right\} < \frac{\epsilon}{2} \text{ for large enough } n.$$

By triangular inequality, we have

$$\begin{aligned} d_{S_0}\{\mathbf{x^n},\mathbf{y^n},\theta\} \leq & \left|2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta) - 2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0)\right| \\ & + 2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0) - 2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\widehat{\theta}^{\mathbb{Q}}). \end{aligned}$$

Now, because $2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0) - 2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\widehat{\theta}^{\mathbb{Q}}) = O_p(1)$ and, by Proposition 19 or Chernozhukov and Hong (2003, Theorem 1 and Proposition 3),

$$\int_{\frac{K}{\sqrt{n}}<|\theta-\widehat{\theta}^{\mathbb{P}}|\leq M} \pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta \to \int_{|u|>K} \varphi_{\mathbb{P}}(u)\mathrm{d}u,$$

with $\varphi_{\mathbb{P}}(u) = \sqrt{\frac{\det\mathbf{I}_{\mathbb{P}}(\theta_0)}{(2\pi)^{D_\Theta}}}\exp\left[-\frac{1}{2}u^T\mathbf{I}_{\mathbb{P}}(\theta_0)u\right]$, we know that there exists $K_1' > 0$ such that for each $K \geq K_1'$, it follows that, for large enough n,

$$\mathbb{Q}_{0,n}\left\{\int_{\frac{K}{\sqrt{n}}<|\theta-\widehat{\theta}^{\mathbb{P}}|\leq M}\left[2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0) - 2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n},\widehat{\theta}^{\mathbb{Q}})\right]\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta > \frac{\epsilon}{4}\right\} < \frac{\epsilon}{4}.$$

Because[10]

$$\begin{aligned} & \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x^n},\mathbf{y^n},\theta)\right] \\ & = \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x_t},\mathbf{y_t},\widehat{\theta}^{\mathbb{P}})\right] + \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\tilde{\theta})}{\partial\theta}\right](\theta-\widehat{\theta}^{\mathbb{P}}), \end{aligned} \tag{A.123}$$

for some $\tilde{\theta}$ between θ and $\widehat{\theta}^{\mathbb{P}}$. We denote $\widehat{G}_n(\theta) \equiv \frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{x_t},\mathbf{y_t},\theta)}{\partial\theta}$. It holds that, by

[10]Rigorously speaking, the values of $\tilde{\theta}$ are usually different for different elements of the vector-valued function $\sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n} g(\mathbf{x^n},\mathbf{y^n},\theta)\right]$. Here, we follow the tradition in GMM literature and write the Mean Value Theorem in a less-rigorous yet simple way in hope of significantly simplifying the exposition of the algebras.

triangular inequality and Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left|2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta)-2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0)\right| \\
\leq &\left|2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\widehat{\theta^{\mathbb{P}}})-2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0)\right|+n(\theta-\widehat{\theta}^{\mathbb{P}})^T\widehat{G}_n(\tilde{\theta})^T S_0^{-1}\widehat{G}_n(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}}) \\
&+2n\left\{\left[\frac{1}{n}\sum_{t=1}^{n}g(\mathbf{x_t},\mathbf{y_t},\widehat{\theta}^{\mathbb{P}})\right]^T S_0^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}g(\mathbf{x_t},\mathbf{y_t},\widehat{\theta}^{\mathbb{P}})\right]\right\}^{1/2} \\
&\times\left\{(\theta-\widehat{\theta}^{\mathbb{P}})^T\widehat{G}_n(\tilde{\theta})^T S_0^{-1}\widehat{G}_n(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}})\right\}^{1/2} \\
\leq &\, O_p(1)\left[1+\sqrt{n}|\theta-\widehat{\theta}^{\mathbb{P}}|+n|\theta-\widehat{\theta}^{\mathbb{P}}|^2\right].
\end{aligned}
$$

Thus, by changing variable $\theta=\widehat{\theta}^{\mathbb{P}}+u/\sqrt{n}$, we have

$$
\begin{aligned}
&\int_{\frac{K}{\sqrt{n}}<|\theta-\widehat{\theta}^{\mathbb{P}}|\leq M}\left|2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta)-2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta_0)\right|\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta \\
\leq &O_p(1)\int_{|u|>K}\left[1+|u|+|u^2|\right]\pi_{\mathbb{P}}(\widehat{\theta}^{\mathbb{P}}+u/\sqrt{n}|\mathbf{x^n})\mathrm{d}u.
\end{aligned}
$$

According to Chernozhukov and Hong (2003, Theorem 1 and Proposition 3), we know that

$$
\begin{aligned}
&\int_{\frac{K}{\sqrt{n}}<|\theta-\widehat{\theta}^{\mathbb{P}}|\leq M}\left[1+\sqrt{n}|\theta-\widehat{\theta}^{\mathbb{P}}|+n|\theta-\widehat{\theta}^{\mathbb{P}}|^2\right]\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta \\
&\qquad\to\int_{|u|>K}\left[1+|u|+|u|^2\right]\varphi_{\mathbb{P}}(u)\mathrm{d}u,
\end{aligned}
\tag{A.124}
$$

with $\varphi_{\mathbb{P}}(u)=\sqrt{\frac{\det\mathbf{I}_{\mathbb{P}}(\theta_0)}{(2\pi)^{D_\Theta}}}\exp\left[-\frac{1}{2}u^T\mathbf{I}_{\mathbb{P}}(\theta_0)u\right]$. Therefore, it follows that there exists $K_1''>0$ such that for each $K\geq K_1''$, it follows that, for large enough n,

$$
\mathbb{Q}_{0,n}\left\{\int_{\frac{K}{\sqrt{n}}<|\theta-\widehat{\theta}^{\mathbb{P}}|\leq M}\left|2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n};\theta)-2J_{n,S_0}(\mathbf{x^n},\mathbf{y^n},\theta_0)\right|\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta>\frac{\epsilon}{4}\right\}<\frac{\epsilon}{4}.
$$

By taking $K_1\equiv\max(K_1',K_1'')$, the condition (A.122) holds.

Second, we show that for any ϵ, when K and n are large enough, it holds that

$$\mathbb{Q}_{0,n}\left\{\left|\Delta_{d,n,K}-\mathbf{tr}\left\{\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}\right]\right\}\right|>\frac{\epsilon}{4}\right\}<\frac{\epsilon}{4}.$$

We define the "Wald test statistic" to be

$$\begin{aligned}w_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\theta)\}&\equiv n\left[f(\widehat{\theta}^{\mathbb{Q}})-f(\theta)\right]^T\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\left[f(\widehat{\theta}^{\mathbb{Q}})-f(\theta)\right]\\&=n(\widehat{\theta}^{\mathbb{Q}}-\theta)^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\widehat{\theta}^{\mathbb{Q}}-\theta).\end{aligned}$$

According to Newey and West (1987b) and Lee (2005), under the regularity conditions in Subsection A.3.3 and using uniform law of large numbers, we know that

$$\alpha_n(K)\equiv\sup_{|\vartheta|\le K}\left|d_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\widehat{\theta}^{\mathbb{P}}+\vartheta/\sqrt{n})\}-w_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\widehat{\theta}^{\mathbb{P}}+\vartheta/\sqrt{n})\}\right|\to 0\ \text{ in }\mathbb{Q}_{0,n}.$$

Thus, for each $K>0$ we have, in $\mathbb{Q}_{0,n}$, the following stochastic integral converges to zero,

$$\begin{aligned}\int_{|\vartheta|\le K}&\left|d_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\widehat{\theta}^{\mathbb{P}}+\vartheta/\sqrt{n})\}-w_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\widehat{\theta}^{\mathbb{P}}+\vartheta/\sqrt{n})\}\right|\\&\times\pi_{\mathbb{P}}(\widehat{\theta}^{\mathbb{P}}+\vartheta/\sqrt{n}|\mathbf{x^n})\mathrm{d}\vartheta.\end{aligned}$$

Therefore, we only need to show that there exists a constant $K_2>0$ such that for every $K\ge K_2$ it holds that when n is large enough

$$\mathbb{Q}_{0,n}\left\{\left|\Delta_{w,n,K}-\mathbf{tr}\left\{\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}\right]\right\}\right|>\frac{\epsilon}{8}\right\}<\frac{\epsilon}{8}\qquad\text{(A.125)}$$

with

$$\Delta_{w,n,K}\equiv\int_{|\theta-\widehat{\theta}^{\mathbb{P}}|\le\frac{K}{\sqrt{n}}}w_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\theta)\}\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta-h_{S_0}\{\mathbf{x^n},\mathbf{y^n}\}.$$

We consider the decomposition

$$w_{S_0}\{\mathbf{x^n},\mathbf{y^n},f(\theta)\} = n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\theta-\widehat{\theta}^{\mathbb{P}}) \tag{A.126}$$

$$+2n(\widehat{\theta}^{\mathbb{P}}-\widehat{\theta}^{\mathbb{Q}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\theta-\widehat{\theta}^{\mathbb{P}}) \tag{A.127}$$

$$+n(\widehat{\theta}^{\mathbb{P}}-\widehat{\theta}^{\mathbb{Q}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\widehat{\theta}^{\mathbb{P}}-\widehat{\theta}^{\mathbb{Q}}). \tag{A.128}$$

For the term (A.173), according to Theorem 1 and Proposition 3 of Chernozhukov and Hong (2003), it follows that, in $\mathbb{P}_0$,

$$\begin{aligned}&\int_{|\theta-\widehat{\theta}^{\mathbb{P}}|\le\frac{K}{\sqrt{n}}} n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\theta-\widehat{\theta}^{\mathbb{P}})\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta\\ &\to \mathbf{tr}\left\{\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}\right]\right\}\\ &\qquad -\mathbf{tr}\left\{\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T\int_{|\vartheta|>K}\vartheta\vartheta^T\varphi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta\mathbf{v}\right\}.\end{aligned}$$

with $\varphi_{\mathbb{P}}(\vartheta)=\sqrt{\frac{\det\mathbf{I}_{\mathbb{P}}(\theta_0)}{(2\pi)^{D_\Theta}}}\exp\left[-\frac{1}{2}\vartheta^T\mathbf{I}_{\mathbb{P}}(\theta_0)\vartheta\right]$. We know that

$$\left|\left|\int_{|\vartheta|>K}\vartheta\vartheta^T\varphi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta\right|\right|_{\mathbb{S}}\to 0 \text{ as } K\to+\infty.$$

Thus, there exists $K_2'>0$ such that for each $K\ge K_2'$, it holds that for large enough n,

$$\begin{aligned}\mathbb{Q}_{0,n}\Bigg\{\Bigg|&\int_{|\theta-\widehat{\theta}^{\mathbb{P}}|\le\frac{K}{\sqrt{n}}} n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\theta-\widehat{\theta}^{\mathbb{P}})\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta\\ &-\mathbf{tr}\left\{\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}\right]\right\}\Bigg|>\frac{\epsilon}{24}\Bigg\}<\frac{\epsilon}{24}.\end{aligned} \tag{A.129}$$

For the term (A.174), according to Theorem 1 and Proposition 3 of Chernozhukov and

Hong (2003), it follows that

$$\int_{|\theta-\widehat{\theta}^{\mathbb{P}}|\leq K} n(\widehat{\theta}^{\mathbb{P}}-\widehat{\theta}^{\mathbb{Q}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\theta-\widehat{\theta}^{\mathbb{P}}) \\ \times \pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta \rightsquigarrow \tilde{X}^T\int_{|\vartheta|\leq K}\vartheta\varphi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta.$$

with $\varphi_{\mathbb{P}}(\vartheta) = \sqrt{\frac{\det\mathbf{I}_{\mathbb{P}}(\theta_0)}{(2\pi)^{D_\Theta}}}\exp\left[-\frac{1}{2}\vartheta^T\mathbf{I}_{\mathbb{P}}(\theta_0)\vartheta\right]$ and $\tilde{X}$ to be an multivariate normal random variable with zero mean and covariance matrix

$$\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}-\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\right]^{-1}\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T.$$

We know that

$$\int_{|\vartheta|\leq K}\vartheta\varphi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta \to 0, \quad \text{as } K\to+\infty.$$

Thus, there exists $K_2''>0$ such that for each $K\geq K_2''$, it holds that for large enough n, the following probability is strictly bounded by $\frac{\epsilon}{24}$,

$$\mathbb{Q}_{0,n}\left\{\left|\int_{|\theta-\widehat{\theta}^{\mathbb{P}}|\leq\frac{K}{\sqrt{n}}} n(\widehat{\theta}^{\mathbb{Q}}-\widehat{\theta}^{\mathbb{P}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\theta-\widehat{\theta}^{\mathbb{P}})\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta\right|>\frac{\epsilon}{24}\right\}.$$

For the term (A.175), according to Theorem 1 and Proposition 3 of Chernozhukov and Hong (2003), it follows that

$$n(\widehat{\theta}^{\mathbb{P}}-\widehat{\theta}^{\mathbb{Q}})^T\mathbf{v}\left[\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}\right]^{-1}\mathbf{v}^T(\widehat{\theta}^{\mathbb{P}}-\widehat{\theta}^{\mathbb{Q}}) \\ \times\left[\int_{|\theta-\widehat{\theta}^{\mathbb{P}}|\leq K}\pi_{\mathbb{P}}(\theta|\mathbf{x^n})-1\right] \rightsquigarrow -h^\infty\int_{|\vartheta|>K}\varphi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta.$$

We know that

$$\int_{|\vartheta|>K}\varphi_{\mathbb{P}}(\vartheta)\mathrm{d}\vartheta \to 0, \text{ as } K\to+\infty.$$

Thus, there exists $K_2''' > 0$ such that for each $K \geq K_2'''$, it holds that for large enough n,

$$\mathbb{Q}_{0,n} \left\{ \int_{|\theta - \widehat{\theta}^{\mathbb{P}}| \leq \frac{K}{\sqrt{n}}} n(\widehat{\theta}^{\mathbb{P}} - \widehat{\theta}^{\mathbb{Q}})^T \mathbf{v} \left[\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1} \mathbf{v} \right]^{-1} \mathbf{v}^T (\widehat{\theta}^{\mathbb{P}} - \widehat{\theta}^{\mathbb{Q}}) \right. \\ \left. \times \left| \int_{|\theta - \widehat{\theta}^{\mathbb{P}}| \leq K} \pi_{\mathbb{P}}(\theta | \mathbf{x^n}) - 1 \right| > \frac{\epsilon}{24} \right\} < \frac{\epsilon}{24}. \tag{A.130}$$

Combining the inequalities (A.129) - (A.130), by taking $K_2 = \max\{K_2', K_2'', K_2'''\}$, we know that the condition (A.125) is satisfied. Now, we have shown that

$$\varrho^f(\mathbf{x^n}, \mathbf{y^n}) \rightsquigarrow 2\varrho_a^{\mathbf{v}}(\theta_0) - D_f + \varepsilon, \quad \mathbb{E}_{\mathbb{Q}_0}[\varepsilon] = 0. \tag{A.131}$$

In the following, we shall use the Dominated Convergence Theorem (see, e.g. Serfling, 1980, Section 1.4) to show the convergence of expectations, i.e.

$$\mathbb{E}_{\mathbb{Q}_0} \left[\varrho^f(\mathbf{x^n}, \mathbf{y^n}) \right] \to 2\varrho_a^{\mathbf{v}}(\theta_0) - D_f, \quad \text{as } n \to \infty. \tag{A.132}$$

According to **Assumption D**, we can find a dominating random variable for $\varrho^f(\mathbf{x^n}, \mathbf{y^n})$:

$$|\varrho^f(\mathbf{x^n}, \mathbf{y^n})| \leq \frac{1}{\lambda_m(S_0)n} \sum_{t=1}^{n} a_1(\mathbf{x_t}, \mathbf{y_t}). \tag{A.133}$$

The dominating random variable is integrable, due to **Assumption D**. That is,

$$\mathbb{E}_{\mathbb{Q}_0} \left| \frac{1}{\lambda_m(S_0)n} \sum_{t=1}^{n} a_1(\mathbf{x_t}, \mathbf{y_t}) \right| = \frac{1}{\lambda_m(S_0)} \mathbb{E}_{\mathbb{Q}_0} a_1(\mathbf{x_1}, \mathbf{y_1}) < \infty. \tag{A.134}$$

Proof of Proposition 12 in the Paper

Because of Assumption **FF** and the assumptions in Subsection A.3.3 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta) = \theta_{(1)}$ and hence $\mathbf{v} = (1, 0, \cdots, 0)^T$. [11]. Let us denote

$$\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) = \frac{1}{\sqrt{2\pi\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}} \exp\left\{-\frac{1}{2\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2\right\}$$

$$\varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) = \frac{1}{\sqrt{2\pi\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}} \exp\left\{-\frac{1}{2\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}(\theta_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)})^2\right\}$$

where $\widehat{\theta}^{\mathbb{P}}$ and $\widehat{\theta}^{\mathbb{Q}}$ are MLE estimator and GMM estimator, respectively, and $\widehat{\theta}^{\mathbb{P}}_{(1)}$ and $\widehat{\theta}^{\mathbb{Q}}_{(1)}$ are the first elements of $\widehat{\theta}^{\mathbb{P}}$ and $\widehat{\theta}^{\mathbb{Q}}$, respectively.

Let's now focus on the decomposition of the relative entropy between constrained and unconstrained posterior distributions:

$$\mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})\right) = A_n + B_n + C_n \tag{A.135}$$

where

$$A_n = \int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})}{\varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})} d\theta_{(1)}, \tag{A.136}$$

$$B_n = \int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{\varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})}{\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})} d\theta_{(1)}, \tag{A.137}$$

$$C_n = \int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})}{\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})} d\theta_{(1)}. \tag{A.138}$$

We show that

$$A_n \to 0 \text{ in } \mathbb{Q}_{0,n}, \tag{A.139}$$

[11] A traditional result is Theorem 3 in Lin, Pittman, and Clarke (2007). The contributions of our results and proofs can be mainly illustrated from four aspects. First, our results extend the traditional results to the more general GMM framework (see Hansen, 1982) leveraging on the Bayesian GMM formulation (see e.g. Kim, 2002; Chernozhukov and Hong, 2003). Second, our results allow for general weak dependence among the observations, which makes our results suitable for studying time series data in finance and economics. Third, another major advantage of our results relative to the traditional results is that our results are adapted to "Bayesian learning" framework, because we allow the "prior" in each time period to be updated with extra data according to Bayesian rule. This extension poses nontrivial theoretical challenges since we have to establish some probabilistic inequalities that hold uniformly over a group of probabilistic models. At last, although some of our proofs are similar to the proof of Theorem 3 in Lin, Pittman, and Clarke (2007), their proof contains mistakes and non-rigorous arguments which are corrected and made rigorous in our proofs.

and

$$B_n - \frac{n}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} \left(\widehat{\theta}^{\mathbb{P}}_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)}\right)^2$$
$$\to \frac{1}{2} \ln \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} + \frac{1}{2} \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} - \frac{1}{2} \text{ in } \mathbb{Q}_{0,n}, \tag{A.140}$$

and

$$C_n \to 0 \text{ in } \mathbb{Q}_{0,n}. \tag{A.141}$$

Step 1: we prove the weak convergence of A_n in (A.139).

In fact, $A_n = \mathbf{D}_{KL}\left(\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) || \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})\right)$. And, according to Corollary 6, we know $A_n \to 0$ in $\mathbb{Q}_{0,n}$.

Step 2: we prove the weak convergence of B_n in (A.140).

We know that

$$B_n = \frac{1}{2} \int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \times \left[\ln \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} - n\frac{(\theta_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)})^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} + n\frac{(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\right] d\theta_{(1)} \tag{A.142}$$

Now let's define

$$\Delta_{B,n} \equiv B_n - \frac{n}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} \left(\widehat{\theta}^{\mathbb{P}}_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)}\right)^2 - \frac{1}{2} \ln \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} - \frac{1}{2} \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} + 1/2. \tag{A.143}$$

We consider the following decomposition for $\Delta_{B,n}$:

$$\Delta_{B,n} = \frac{1}{2}\int \left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] \mathrm{d}\theta_{(1)} \ln \frac{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} \tag{A.144}$$

$$-\frac{1}{2}\int \left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \tag{A.145}$$

$$+\frac{1}{2}\int \left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \tag{A.146}$$

The term in (A.144) is denoted as $B_{n,1}$, the term in (A.145) is denoted as $B_{n,2}$, and the term (A.146) is denoted as $B_{n,3}$.

Step 2.1: we show that $B_{n,1} \to 0$ in $\mathbb{Q}_{0,n}$.

We know that, in $\mathbb{Q}_{0,n}$,

$$|B_{n,1}| \leq \frac{1}{2}\ln\frac{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}\int \left|\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\right| \mathrm{d}\theta_{(1)} \to 0 \tag{A.147}$$

where the convergence result in (A.147) is due to the fact that the squared total variation distance[12] is upper bounded by the relative entropy (i.e. Kullback-Leibler distance) (see e.g. Kullback, 1967) and due to the result in Proposition 6.

Step 2.2: we show that $B_{n,2} \to 0$ in $\mathbb{Q}_{0,n}$. Equivalently, we show that

$$\int_{\Theta}\left[\pi_{\mathbb{Q}}(\theta|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)} - \widehat{\theta}^{\mathbb{Q}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta \to 0 \text{ in } \mathbb{Q}_{0,n}.$$

This is actually a direct implication from Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003).

Step 2.3: Similar argument can be used to prove that $B_{n,3} \to 0$ in $\mathbb{Q}_{0,n}$. More precisely, it is equivalent to show that

$$\frac{1}{2}\int \left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right] n\frac{(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} \to 0 \text{ in } \mathbb{Q}_{0,n},$$

[12] The total variation distance between the constrained posterior on $\theta_{(1)}$ and normal distribution is $\int \left|\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) - \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\right| \mathrm{d}\theta_{(1)}$.

and

$$\frac{1}{2}\int\left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})-\varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right]n\frac{2(\theta_{(1)}-\widehat{\theta}^{\mathbb{Q}}_{(1)})(\theta^{\mathbb{Q}}_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)}\to 0\ \text{ in } \mathbb{Q}_{0,n},$$

and

$$\frac{1}{2}\int\left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})-\varphi_{\mathbb{Q}}\left(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}\right)\right]n\frac{(\theta^{\mathbb{Q}}_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)}\to 0\ \text{ in } \mathbb{Q}_{0,n}.$$

According to Chernozhukov and Hong (2003, Theorem 1 and Proposition 1), all the three limiting conditions above are satisfied.

Step 3: we prove the weak convergence of C_n in (A.141).

For a constant $r > 0$, we decompose the term C_n as follows

$$C_n = C_{n,1} + C_{n,2} - C_{n,3}, \tag{A.148}$$

where

$$C_{n,1} = \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\ln\frac{\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})}{\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})}\mathrm{d}\theta_{(1)} \tag{A.149}$$

$$C_{n,2} = \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\ln\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})\mathrm{d}\theta_{(1)} \tag{A.150}$$

$$C_{n,3} = \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\ln\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})\mathrm{d}\theta_{(1)}. \tag{A.151}$$

Step 3.1: we show that $C_{n,3}\to 0$ in $\mathbb{Q}_{0,n}$. Equivalently, we show that for any $\epsilon > 0$

$$\limsup_{n\to+\infty}\mathbb{Q}_{0,n}\left\{|C_{n,3}|>\epsilon\right\}<\epsilon. \tag{A.152}$$

Let

$$\mathcal{A}_n(\eta)\equiv\left\{\frac{\int_{\Omega(\theta_{0,(1)},r)^c}\int_{\Theta_{-1}(\theta_{(1)})}\pi_{\mathbb{P}}(\theta_{(1)},\theta_{(-1)})\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}}{\int_{\Omega(\theta_{0,(1)},r)}\int_{\Theta_{-1}(\theta_{(1)})}\pi_{\mathbb{P}}(\theta_{(1)},\theta_{(-1)})\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}}<\eta\right\}.$$

By Proposition 35, we know that for any $\eta > 0$

$$\mathbb{Q}_{0,n}\mathcal{A}_n(\eta)^c = o(1).$$

Define the set

$$\mathcal{B}_n(\eta) \equiv \left\{ \sup_{\theta\in\Theta} \left| \widehat{H}_{\mathbb{P},n}(\theta) - H_{\mathbb{P}}(\theta) \right| < \eta \right\},$$

where $H_{\mathbb{P}}(\theta)$ and $\widehat{H}_{\mathbb{P},n}(\theta)$ are defined in (A.96) and (A.98), respectively.

By Proposition 18, we know that for any $\eta > 0$

$$\mathbb{Q}_{0,n}\mathcal{B}_n(\eta)^c = o(1).$$

Define the set

$$\begin{aligned}
&\mathcal{I}_{1,n}(\delta,\eta) \\
&\quad \equiv \left\{ \left|\left| \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta) \right|\right|_{\mathcal{S}} \leq \eta \left|\left| \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \right|\right|_{\mathcal{S}}^{-1},\ \forall\, \theta \in \Omega(\theta_0,\delta) \text{ and } \tilde{\theta} \in \Omega(\theta,\delta) \right\},
\end{aligned}$$

By Proposition 25, we know that for any $\eta > 0$ there exists $\delta > 0$ such that

$$\mathbb{Q}_{0,n}\mathcal{I}_{1,n}(\delta,\eta)^c = o\left(\frac{1}{n}\right).$$

Define the set

$$\mathcal{E}_n(\delta) \equiv \left\{ \widehat{\theta}^{\mathbb{P}} \in \Omega(\theta_0,\delta) \right\}.$$

By Proposition 15, we know that

$$\mathbb{Q}_{0,n}\mathcal{E}_n(\delta)^c = o(1).$$

Let

$$\mathcal{K}_n(\delta,\eta) \equiv \left\{ \pi_{\mathbb{P}}(\mathbf{x^n}) \leq (1+\eta) \int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta \right\}.$$

According to Proposition 35, we know that for any $\delta > 0$ and $\eta > 0$ we have

$$\mathbb{Q}_{0,n}\mathcal{K}_n(\delta,\eta)^c \to 0.$$

We define

$$\mathcal{M}_n(\delta,\eta) \equiv \mathcal{A}_n(\eta) \cap \mathcal{B}_n(\eta) \cap \mathcal{I}_{1,n}(\delta,\eta) \cap \mathcal{E}_n(\delta) \cap \mathcal{K}_n(\delta,\eta),$$

then, we have

$$\mathbb{Q}_{0,n}\mathcal{M}_n(\delta,\eta)^c = o(1).$$

Let's consider the decomposition

$$\mathbb{Q}_{0,n}\left\{|C_{n,3}| > \epsilon\right\} \leq \mathbb{Q}_{0,n}\mathcal{M}_n(\delta,\eta) \cap \left\{|C_{n,3}| > \epsilon\right\} + \mathbb{Q}_{0,n}\mathcal{M}_n(\delta,\eta)^c.$$

Our strategy of proving the result (A.152) is to find random variables $\overline{C}_{n,3}$ and $\underline{C}_{n,3}$ such that

$$\underline{C}_{n,3} \leq C_{n,3} \leq \overline{C}_{n,3} \quad \text{on } \mathcal{A}_n(\eta) \text{ (of course on) } \mathcal{M}_n(\delta,\eta)$$

and

$$\underline{C}_{n,3} \to 0 \text{ in } \mathbb{Q}_{0,n} \text{ and } \overline{C}_{n,3} \to 0 \text{ in } \mathbb{Q}_{0,n}.$$

Thus, we have

$$\begin{aligned}
\limsup_{n\to+\infty} \mathbb{Q}_{0,n}\left\{|C_{n,3}| > \epsilon\right\} &\leq \limsup_{n\to+\infty} \mathbb{Q}_0^n\mathcal{M}_n(\delta,\eta) \cap \left\{|C_{n,3}| > \epsilon\right\} \\
&\leq \limsup_{n\to+\infty} \mathbb{Q}_{0,n}\mathcal{M}_n(\delta,\eta) \cap \left\{\max\{|\overline{C}_{n,3}|,\ |\underline{C}_{n,3}|\} > \epsilon\right\} \\
&\leq \limsup_{n\to+\infty} \mathbb{Q}_{0,n}\mathcal{M}_n(\delta,\eta) \cap \left\{|\overline{C}_{n,3}| > \epsilon\right\} \\
&\qquad + \limsup_{n\to+\infty} \mathbb{Q}_{0,n}\mathcal{M}_n(\delta,\eta) \cap \left\{|\underline{C}_{n,3}| > \epsilon\right\} = 0.
\end{aligned}$$

Now, let's figure out the limits of $\overline{C}_{n,3}$ and $\underline{C}_{n,3}$. On the event $\mathcal{M}_n(\delta, \eta)$, we have that

$$\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) = \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta_{(-1)}}{\int_{\Theta} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta} \tag{A.153}$$

$$= \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\theta|\mathbf{x^n})\mathrm{d}\theta_{(-1)}}{\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta \left(1 + \frac{\int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta}{\int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta}\right)} \tag{A.154}$$

$$\geq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{(1+\eta) \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta} \tag{A.155}$$

Because $\widehat{\theta}^{\mathbb{P}}$ is the MLE under $\mathbb{P}_\theta$, we have

$$\pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \leq \pi_{\mathbb{P}}(\mathbf{x^n}|\widehat{\theta}^{\mathbb{P}}) = \exp\left\{-\ln \frac{1}{\pi_{\mathbb{P}}(\mathbf{x^n}|\widehat{\theta}^{\mathbb{P}})}\right\} = \exp\left\{-n\widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}})\right\} \tag{A.156}$$

Thus, we have

$$\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \geq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{(1+\eta)\pi_{\mathbb{P}}(\Omega(\theta_{0,(1)},r))\exp\{-n\widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}})\}} \tag{A.157}$$

Plug (A.157) into the expression for $C_{n,3}$, we have

$$
\begin{aligned}
C_{n,3} &\geq \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \\
&\quad \times \ln \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{(1+\eta)\pi_{\mathbb{P}}(\Omega(\theta_{0,(1)},r))\exp\{-n\widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}})\}}\mathrm{d}\theta_{(1)} \\
&\geq \int_{\Omega(\theta_{0,(1)},r)^c}\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \\
&\qquad \times \ln\left[\pi_{\mathbb{P}}(\theta_{(1)})\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\right]\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \\
&\quad - \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \\
&\qquad \times \left[\ln(1+\eta) + \ln \pi_{\mathbb{P}}(\Omega(\theta_{0,(1)},r)) - nH_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}})\right]\mathrm{d}\theta_{(1)} \\
&= \int_{\Omega(\theta_{0,(1)},r)^c}\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \\
&\qquad \times \ln \pi_{\mathbb{P}}(\theta_{(1)})\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)} \\
&\quad - \pi_{\mathbb{P}}(\Omega(\theta_{0,(1)},r)^c|\mathbf{x^n},\mathbf{y^n}) \\
&\qquad \times \left[\ln(1+\eta) + \ln \pi_{\mathbb{P}}(\Omega(\theta_{0,(1)},r)) - n\widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}})\right]
\end{aligned} \tag{A.158}
$$

We define the term in (A.158) to be $\underline{C}_{n,3}$. Thus, we can further decompose $\underline{C}_{n,3}$ as follows,

$$\underline{C}_{n,3} = \underline{C}_{n,3,1} - \underline{C}_{n,3,2}$$

where

$$
\begin{aligned}
\underline{C}_{n,3,1} = &\int_{\Omega(\theta_{0,(1)},r)^c}\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \\
&\times \ln\left[\pi_{\mathbb{P}}(\theta_{(1)})\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\right]\mathrm{d}\theta_{(1)}\mathrm{d}\theta_{(-1)}
\end{aligned} \tag{A.159}
$$

and

$$\underline{C}_{n,3,2} = \pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)},r)^c|\mathbf{x^n},\mathbf{y^n})\left[\ln(1+\eta) + \ln \pi_{\mathbb{P}}(\Omega(\theta_{0,(1)},r)) - n\widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}})\right].$$

We have

$$
\begin{aligned}
|\underline{C}_{n,3,1}| \leq & \left| \int_{\Omega(\theta_{0,(1)},r)^c} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \ln \pi_{\mathbb{P}}(\theta_{(1)}) \mathrm{d}\theta_{(1)} \mathrm{d}\theta_{(-1)} \right| \\
& + \left| \int_{\Omega(\theta_{0,(1)},r)^c} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \right. \\
& \quad \left. \times \ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta) \mathrm{d}\theta_{(1)} \mathrm{d}\theta_{(-1)} \right| \\
& \leq \int_{\Omega(\theta_{0,(1)},r)^c} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \left| \ln \pi_{\mathbb{P}}(\theta_{(1)}) \right| \mathrm{d}\theta_{(1)} \mathrm{d}\theta_{(-1)} \quad \text{(A.160)} \\
& + \int_{\Omega(\theta_{0,(1)},r)^c} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \\
& \quad \times n|\widehat{H}_{\mathbb{P},n}(\theta) - H_{\mathbb{P}}(\theta)| \mathrm{d}\theta_{(1)} \mathrm{d}\theta_{(-1)} \quad \text{(A.161)} \\
& + \int_{\Omega(\theta_{0,(1)},r)^c} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) n |H_{\mathbb{P}}(\theta)| \mathrm{d}\theta_{(1)} \mathrm{d}\theta_{(-1)} \quad \text{(A.162)} \\
& + o_p(1).
\end{aligned}
$$

The term (A.160) can be bounded from above by

$$
\begin{aligned}
& M_1 \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \mathrm{d}\theta_{(1)} \\
& \qquad = M_1 \pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)},r)^c|\mathbf{x^n},\mathbf{y^n}) \to 0 \text{ in } \mathbb{Q}_{0,n},
\end{aligned}
$$

where the existence of such constant M_1 is due to the compactness of $\Theta \subset \mathbb{R}^d$ and the continuity of $\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)})|\ln \pi_{\mathbb{P}}(\theta_{(1)})|$ and we have

$$
\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)})|\ln \pi_{\mathbb{P}}(\theta_{(1)})| \mathrm{d}\theta_{(-1)} \leq M_1. \quad \text{(A.163)}
$$

The term (A.161), for large enough n, is bounded from above by

$$
n \int_{\Omega(\theta_{0,(1)},r)^c} \int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \mathrm{d}\theta_{(1)} \mathrm{d}\theta_{(-1)}, \quad \text{(A.164)}
$$

because

$$\sup_{\theta \in \Theta} |\widehat{H}_{\mathbb{P},n}(\theta) - H_{\mathbb{P}}(\theta)| < 1 \text{ for large enough } n. \tag{A.165}$$

The term (A.164) can be further bounded from above by

$$\begin{aligned} M_2 n \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \mathrm{d}\theta_{(1)} \\ = M_2 n \pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)}, r)^c | \mathbf{x^n}, \mathbf{y^n}) \to 0 \text{ in } \mathbb{Q}_{0,n}, \end{aligned} \tag{A.166}$$

where the existence of such constant M_2 is due to the compactness of the $\Theta \subset \mathbb{R}^d$ and the continuity of $\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) = \frac{\pi_{\mathbb{P}}(\theta_{(1)},\theta_{(-1)})}{\pi_{\mathbb{P}}(\theta_{(1)})}$,

$$\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)}) \mathrm{d}\theta_{(-1)} \leq M_2. \tag{A.167}$$

The term (A.162) is bounded from above by

$$\begin{aligned} M_3 n \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \\ = M_3 n \pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)}, r) | \mathbf{x^n}, \mathbf{y^n}) \to 0 \text{ in } \mathbb{Q}_{0,n}. \end{aligned} \tag{A.168}$$

Therefore, the term $\underline{C}_{n,3,1} \to 0$ in $\mathbb{Q}_{0,n}$. It is straightforward to see that $\underline{C}_{n,3,2}$ converges to zero in probability, because $n\pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)}, r)|\mathbf{x^n}, \mathbf{y^n}) \to 0$ in $\mathbb{Q}_{0,n}$ and $\widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}}) \to H_{\mathbb{P}}(\theta_0)$ in $\mathbb{Q}_{0,n}$. Now, let's construct $\overline{C}_{n,3}$ and show it indeed converges to zero in probability. By restricting the domain to $\Omega(\theta_0, \delta)$, we have

$$\pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \leq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{\int_{\Omega(\theta_0,\delta)} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta}. \tag{A.169}$$

By Taylor expansion of $\ln \pi_{\mathbb{P}}(\mathbf{x^n}|\theta)$ around $\widehat{\theta}^{\mathbb{P}}$, we have

$$\begin{aligned}
\pi_{\mathbb{P}}(\theta|\mathbf{x^n}) &\leq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}}{m_\pi \pi_{\mathbb{P}}(\mathbf{x^n}|\widehat{\theta}^{\mathbb{P}}) \int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}})\}\mathrm{d}\theta} \\
&\leq \frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\mathrm{d}\theta_{(-1)}}{m_\pi \int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}})\}\mathrm{d}\theta} \\
&\leq \frac{M_4}{\int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}})\}\mathrm{d}\theta}
\end{aligned}$$

where $\tilde{\theta}$ is on the segment between $\widehat{\theta}^{\mathbb{P}}$ and θ. And, the existence of the constant M_4 such that

$$\frac{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\mathrm{d}\theta_{(-1)}}{m_\pi} \leq M_4. \tag{A.170}$$

Thus, we have

$$C_{n,3} \leq \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \times \ln \frac{M_4 \mathrm{d}\theta_{(1)}}{\int_{\Omega(\theta_0,\delta)} \exp\{-\frac{n}{2}(\theta-\widehat{\theta}^{\mathbb{P}})^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}})\}\mathrm{d}\theta} \tag{A.171}$$

We define $\overline{C}_{n,3}$ to be the term on the right hand side of the inequality (A.171). On the event $\mathcal{M}_n(\delta,\eta)$, we have

$$(\theta-\widehat{\theta}^{\mathbb{P}})^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta-\widehat{\theta}^{\mathbb{P}}) \leq 2(\theta-\widehat{\theta}^{\mathbb{P}})^T \mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})$$

and

$$C_{n,3} \leq \int_{\Omega(\theta_{0,(1)},r)^c} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{M_4 \mathrm{d}\theta_{(1)}}{\int_{\Omega(\theta_0,\delta)} \exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T \mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})\}\mathrm{d}\theta}$$

By the normal distribution and the $\sqrt{n}-$ consistency of MLE $\widehat{\theta}^{\mathbb{P}}$, we know that for any $\nu > 0$, it follows that

$$\mathbb{Q}_{0,n}\left\{\frac{\int_{\Omega(\theta_0,\delta)}\exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})\}d\theta}{\int_{\mathbb{R}^{D_\Theta}}\exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})\}d\theta}<1-\nu\right\}\to 0. \tag{A.172}$$

Thus, we have

$$\begin{aligned}
&\limsup_{n\to+\infty}\mathbb{Q}_{0,n}\{|\overline{C}_{n,3}|>\epsilon\}\\
&\leq\limsup_{n\to+\infty}\mathbb{Q}_{0,n}\left\{\int_{\Omega(\theta_{0,(1)},r)^c}\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\right.\\
&\qquad\left.\times\ln\frac{(1-\nu)^{-1}M_4 d\theta_{(1)}}{\int_{\mathbb{R}^{D_\Theta}}\exp\{-n(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})\}d\theta}>\epsilon\right\}\\
&=\limsup_{n\to+\infty}\mathbb{Q}_{0,n}\left\{\int_{\Omega(\theta_{0,(1)},r)^c}\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\left|\ln\frac{M_4(2\pi)^{D_\Theta/2}}{(1-\nu)|2n\mathbf{I}_{\mathbb{P}}(\theta_0)|^{1/2}}\right|>\epsilon\right\}\\
&=\limsup_{n\to+\infty}\mathbb{Q}_{0,n}\left\{\pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)},r)^c|\mathbf{x^n},\mathbf{y^n})\left|\ln\frac{M_4(2\pi)^{D_\Theta/2}}{(1-\nu)|2n\mathbf{I}_{\mathbb{P}}(\theta_0)|^{1/2}}\right|>\epsilon\right\}=0,
\end{aligned}$$

where the last limiting result is a direct implication of Proposition 31. Thus, $\underline{C}_{n,3}\to 0$ in $\mathbb{Q}_{0,n}$. Therefore, we have $C_{n,3}\to 0$ in $\mathbb{Q}_{0,n}$.

Step 3.2: we show $C_{n,2}$ goes to zero in $\mathbb{Q}_{0,n}$. The expression (A.150) for $C_{n,2}$ can be rewritten as

$$\begin{aligned}
C_{n,2}=&\pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)},r)^c|\mathbf{x^n},\mathbf{y^n})\ln\frac{|n\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{v}|^{1/2}}{(2\pi)^{1/2}}\\
&\qquad-\int_{\Omega(\theta_{0,(1)},r)^c}\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}d\theta_{(1)}
\end{aligned}$$

Thus, the term $C_{n,2}$ can be decomposed as follows

$$C_{n,2} = \pi_{\mathbb{Q}}(\Omega(\theta_{0,(1)}, r)^c|\mathbf{x^n}, \mathbf{y^n}) \ln \frac{|n\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{v}|^{1/2}}{(2\pi)^{1/2}} \tag{A.173}$$

$$- \int_{\Omega(\theta_{0,(1)}, r)^c} \left[\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) - \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n})\right] \tag{A.174}$$

$$\frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{v}} \mathrm{d}\theta_{(1)}$$

$$- \int_{\Omega(\theta_{0,(1)}, r)^c} \varphi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} \tag{A.175}$$

It is easy to see that the first term (A.173) goes to zero in $\mathbb{Q}_{0,n}$. The second term (A.174) and the third term (A.175) go to zero in probability according to Theorem 1 and Proposition 1 in Chernozhukov and Hong (2003) and the fact that $n(\widehat{\theta}^{\mathbb{Q}} - \widehat{\theta}^{\mathbb{P}})^2 = O_p(1)$. Therefore, we have shown that $C_{n,2} \to 0$ in $\mathbb{Q}_{0,n}$.

Step 3.3: we need to prove $C_{n,1}$ goes to zero in $\mathbb{Q}_{0,n}$. According to Corollary 8, we know that for any $\eta > 0$ there exists $\delta_0 > 0$ such that

$$\mathbb{P}_{0,n}\mathcal{I}_n(\delta_0, \eta) \to 1, \text{ as } n \to \infty,$$

with

$$\mathcal{I}_n(\delta_0, \eta) \equiv \left\{1 - \eta \le \left|\left|\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1/2}\right|\right|_{\mathcal{S}} \le 1 + \eta, \text{for all } \tilde{\theta} \in \Omega(\theta_0, \delta_0)\right\}.$$

Also, by the continuity and positivity of the prior density $\pi_{\mathbb{P}}(\theta)$, we know that for any $\eta > 0$, there exists δ_1 small enough it holds that $1 - \eta \le \pi_{\mathbb{P}}(\theta)/\pi_{\mathbb{P}}(\theta') \le 1 + \eta$ for all $\theta, \theta' \in \Omega(\theta_0, \delta_1)$. According to the consistency of MLE $\widehat{\theta}^{\mathbb{P}}$, we shall only focus on the event $\mathcal{A}_n(\delta) \equiv \widehat{\theta}^{\mathbb{P}} \in \Omega(\theta_0, \delta)$ with $\delta = \min(\delta_0, \delta_1)$. On the joint large probability event

$\mathcal{I}_n(\delta,\eta) \cap \mathcal{A}_n(\eta)$, we have for each $\theta \in \Omega(\theta_0, r)$ with $r < \delta$, when n is large enough,

$$\begin{aligned}
\pi_{\mathbb{P}}(\theta|\mathbf{x^n}) &\leq \frac{\pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)}{\int_{\Omega(\theta_0,r)} \pi_{\mathbb{P}}(\vartheta)\pi_{\mathbb{P}}(\mathbf{x^n}|\vartheta)\mathrm{d}\vartheta} \\
&\leq (1+\eta)^2 \left[\frac{n(1+\eta)}{2\pi}\right]^{D_\Theta/2} [\det \mathbf{I}_{\mathbb{P}}(\theta_0)]^{1/2}\, e^{-\frac{1}{2}(1-\eta)n(\theta-\widehat{\theta}^{\mathbb{P}})^T \mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})}
\end{aligned}$$

On the event $\mathcal{I}_n(\delta,\eta) \cap \mathcal{A}_n(\eta)$ we have

$$\begin{aligned}
C_{n,1} \geq & \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \frac{1}{\sqrt{2\pi n^{-1}|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}|}} \mathrm{d}\theta_{(1)} \\
& - \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} \\
& - \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \ln \Bigg[\int_{\Theta_{-1}(\theta_{(1)})} e^{-\frac{n}{2}(1-\eta)(\theta-\widehat{\theta}^{\mathbb{P}})^T \mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})} \times \\
& \qquad \frac{[(1-\eta)n]^{D_\Theta/2} [\det \mathbf{I}_{\mathbb{P}}(\theta_0)]^{1/2}}{(2\pi)^{D_\Theta/2}} \mathrm{d}\theta_{(-1)}\Bigg] \mathrm{d}\theta_{(1)} \\
& - \left[\frac{D_\Theta}{2} + 2\right] \ln(1+\eta) + \frac{D_\Theta}{2} \ln(1-\eta).
\end{aligned}$$

There exists open square centered at θ_0 which is denoted as

$$\mathcal{O} = \Omega(\theta_{0,(1)}, r) \otimes \Omega(\theta_{0,(-1)}, \delta).$$

First, we have

$$\begin{aligned}
\int_{\Theta_{-1}(\theta_{(1)})} e^{-\frac{n}{2}(1-\eta)(\theta-\widehat{\theta}^{\mathbb{P}})^T \mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})} \frac{[(1-\eta)n]^{D_\Theta/2} [\det \mathbf{I}_{\mathbb{P}}(\theta_0)]^{1/2}}{(2\pi)^{D_\Theta/2}} \mathrm{d}\theta_{(-1)} \\
\leq \frac{e^{-(1-\eta)\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}}}{\sqrt{2\pi n^{-1}\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}/(1-\eta)}}.
\end{aligned}$$

Thus, on the large probability event $\mathcal{I}_n(\delta,\eta) \cap \mathcal{A}_n(\eta)$ we have, when n is large

$$C_{n,1} \geq -\left[\frac{D_\Theta}{2}+2\right]\ln(1+\eta)+\left[\frac{D_\Theta-1}{2}\right]\ln(1-\eta)$$
$$+\eta\int_{\Omega(\theta_{0,(1)},r)}\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)}.$$

According to Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003), we know that

$$\int_{\Omega(\theta_{0,(1)},r)}\pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n})\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}\mathrm{d}\theta_{(1)} = O_p(1).$$

Therefore, it follows that

$$C_{n,1} \geq -\left[\frac{D_\Theta}{2}+2\right]\ln(1+\eta)+\left[\frac{D_\Theta-1}{2}\right]\ln(1-\eta)+\eta O_p(1). \tag{A.176}$$

On the other hand, by Proposition 35, we know that for any $\eta'>0$ and $\delta'>0$

$$\mathbb{P}_{0,n}\mathcal{X}_n(\delta',\eta') \to 1, \quad \text{as } n\to\infty,$$

with

$$\mathcal{X}_n(\delta',\eta') \equiv \left\{\pi_{\mathbb{P}}(\mathbf{x^n}) \leq (1+\eta')\int_{\Omega(\theta_0,\delta')}\pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta\right\}.$$

Then, on the large probability event $\mathcal{I}_n(\delta,\eta)\cap\mathcal{A}_n(\eta)\cap\mathcal{X}_n(\delta,\eta)$ and taking $r<\delta$, it holds that

$$
\begin{aligned}
C_{n,1} &= \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \left[\ln \frac{\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})\pi_{\mathbb{P}}(\mathbf{x^n})}{\int_{\Theta_{-1}(\theta_{(1)})} \pi_{\mathbb{P}}(\theta)\pi_{\mathbb{P}}(\mathbf{x^n}|\theta)\mathrm{d}\theta_{(-1)}} \right] \mathrm{d}\theta_{(1)} \\
&\leq \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \\
&\qquad \times \ln \left[\frac{(2\pi)^{-1/2}|n^{-1}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}|^{-1/2} e^{-\frac{n(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}}}{\int_{\Theta_{-1}(\theta_{(1)})}(1-\eta)e^{-(1+\eta)\frac{n}{2}(\theta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\theta-\widehat{\theta}^{\mathbb{P}})}\mathrm{d}\theta_{(-1)}} \right] \mathrm{d}\theta_{(1)} \\
&\quad + \int_{\Omega(\theta_{0,(1)},r)} \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \\
&\qquad \times \ln \left[\int_{\Omega(\theta_0,\delta)} (1+\eta) e^{-(1-\eta)\frac{n}{2}(\vartheta-\widehat{\theta}^{\mathbb{P}})^T\mathbf{I}_{\mathbb{P}}(\theta_0)(\vartheta-\widehat{\theta}^{\mathbb{P}})}\mathrm{d}\vartheta \right] \mathrm{d}\theta_{(1)}.
\end{aligned}
$$

Calculating the integrations, we have

$$
\begin{aligned}
C_{n,1} \leq \frac{D_\Theta+1}{2}\ln(1+\eta) &- \frac{D_\Theta+2}{2}\ln(1-\eta) \\
&- \inf_{\theta_{(1)}\in\Omega(\theta_{0,(1)},r)} \ln \Phi_n(\Theta_{-1}(\theta_{(1)})|\theta_{(1)}) \\
&+ \eta \int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n},\mathbf{y^n}) \frac{n(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{v}}\mathrm{d}\theta.
\end{aligned}
$$

where $\Phi_n(\cdot|\theta_{(1)})$ is the multivariate normal probability measure on $\mathbb{R}^{D_\Theta-1}$ and effectively it is the conditional distribution of $\theta_{(-1)}$ given $\theta_{(1)}$ where $\theta = (\theta_{(1)},\theta_{(-1)}) \sim N(\widehat{\theta}^{\mathbb{P}},(1+\eta)^{-1}n^{-1}\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1})$. In fact, $\Phi_n(\cdot|\theta_{(1)})$ is multivariate normal with distribution

$$
N\left(\widehat{\theta}^{\mathbb{P}}_{(-1)} + \Sigma_{21}\Sigma_{11}^{-1}(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)}), (1+\eta)^{-1}n^{-1}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\right)
$$

with

$$
\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.
$$

We choose δ and r small enough such that $\widehat{\theta}^{\mathbb{P}}_{(-1)} + \Sigma_{21}\Sigma_{11}^{-1}(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)})$ is in the interior of $\Theta_{-1}(\theta_{(1)})$ and there exists $\tau_0 > 0$ such that $d_L(\widehat{\theta}^{\mathbb{P}}_{(-1)} + \Sigma_{21}\Sigma_{11}^{-1}(\theta_{(1)}-\widehat{\theta}^{\mathbb{P}}_{(1)}), \partial\Theta_{-1}(\theta_{(1)})) >$

τ_0 for all $\theta_{(1)} \in \Omega(\theta_{0,(1)})$. Thus,

$$\inf_{\theta_{(1)} \in \Omega(\theta_{0,(1)}, r)} \Phi_n(\Theta_{-1}(\theta_{(1)})|\theta_{(1)}) \to 1.$$

Therefore, when n is large enough, we have

$$\inf_{\theta_{(1)} \in \Omega(\theta_{0,(1)}, r)} \Phi_n(\Theta_{-1}(\theta_{(1)})|\theta_{(1)}) > \frac{1}{2}\ln(1-\eta).$$

And hence, we have

$$C_{n,1} \leq \frac{D_\Theta + 1}{2}\ln(1+\eta) - \frac{D_\Theta + 3}{2}\ln(1-\eta) + \eta \int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{v}} \mathrm{d}\theta.$$

According to Theorem 1 and Proposition 1 of Chernozhukov and Hong (2003), we know that $\int \pi_{\mathbb{Q}}(\theta_{(1)}|\mathbf{x^n}, \mathbf{y^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)\mathbf{v}} \mathrm{d}\theta = O_p(1)$. Therefore, we can get

$$C_{n,1} \leq \frac{D_\Theta + 1}{2}\ln(1+\eta) - \frac{D_\Theta + 3}{2}\ln(1-\eta) + \eta O_p(1). \tag{A.177}$$

Combine the bounds in (A.176) and (A.177) where the constant η can be arbitrarily small, we know that $C_{n,1} \to 0$ in $\mathbb{Q}_{0,n}$.

Proposition 37. *Consider the feature function $f : \mathbb{R}^{D_\Theta} \to \mathbb{R}$ with $\mathbf{v} = \partial f(\theta_0)/\partial\theta$. Under the assumptions in Subsection A.3.3, if we define*

$$\lambda \equiv \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}},$$

then it follows that

$$\frac{n}{\mathbf{v}^T \mathbf{I}_{\mathbb{Q}}(\theta_0)^{-1}\mathbf{v}} (f(\widehat{\theta}^{\mathbb{P}}) - f(\widehat{\theta}^{\mathbb{Q}}))^2 \rightsquigarrow (1-\lambda^{-1})\chi^2_1. \tag{A.178}$$

Moreover,

$$\mathbf{D}_{KL}(\pi_{\mathbb{Q}}(f(\theta)|\mathbf{x^n}, \mathbf{y^n})||\pi_{\mathbb{P}}(f(\theta)|\mathbf{x^n})) \rightsquigarrow \frac{1-\lambda^{-1}}{2}(\chi_1^2 - 1) + \frac{1}{2}\ln(\lambda), \tag{A.179}$$

Proof. The asymptotic distribution result in (A.178) is directly from Proposition 24. The asymptotic result in (A.179) is based on Proposition 12, limit result (A.178), and the Slutsky Theorem. □

Proof of Proposition 13 in the Paper

Because of Assumption **FF** and the assumptions in Subsection A.3.3 are invariant under invertible and second-order smooth transformations, without loss of generality, we assume that $f(\theta) = \theta_{(1)}$ and hence $\mathbf{v} = (1, 0, \cdots, 0)^T$.

We want to show when m and n go to infinity and $m/n \to \varrho$, we have for any $\epsilon > 0$ that

$$\limsup_{n\to+\infty} \mathbb{P}_{0,n}\left\{\left|\int \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int \pi(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x^n})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)})} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta - \frac{1}{2}\ln\frac{n}{m+n}\right| > \epsilon\right\} < \epsilon.$$

Denote

$$R_n \equiv \int \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int \pi(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\mathbf{x^n})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)})} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta - \frac{1}{2}\ln\frac{n}{m+n}.$$

Then, we are going to show for any $\epsilon > 0$ it holds that

$$\limsup_{n\to+\infty} \mathbb{P}_{0,n}\left\{|R_n| > \epsilon\right\} < \epsilon \tag{A.180}$$

We can further decompose R_n as follows:

$$R_n = \int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \frac{\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})}{\pi_{\mathbb{P}}(\theta_{(1)})} \mathrm{d}\theta_{(1)} \tag{A.181}$$

$$+ \int \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \ln \frac{\int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta')\pi_{\mathbb{P}}(\theta')\mathrm{d}\theta'}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \tag{A.182}$$

$$- \int \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \tag{A.183}$$

$$\times \ln \frac{\int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta$$

$$- \frac{1}{2} \ln \frac{n}{m+n}.$$

We denote the term in (A.181) as $R_{n,1}$, denote the term in (A.182) as $R_{n,2}$, and denote the term in (A.183) as $R_{n,3}$.

For the term $R_{n,1}$, we can further decompose it as follows

$$R_{n,1} = \int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \frac{\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})}{\varphi(\theta_{(1)}|\mathbf{x^n})} \mathrm{d}\theta_{(1)} + \frac{1}{2}\ln(n) - \frac{1}{2}\ln\left(2\pi\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}\right)$$
$$- \int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} - \int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \pi_{\mathbb{P}}(\theta_{(1)})\mathrm{d}\theta_{(1)},$$

where

$$\varphi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) = \frac{1}{\sqrt{2\pi\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}} \exp\left\{-\frac{1}{2\frac{1}{n}\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}}(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2\right\}.$$

According to Corollary 6, we know that

$$\int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \frac{\pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n})}{\varphi(\theta_{(1)}|\mathbf{x^n})} \mathrm{d}\theta_{(1)} \to 0 \text{ in } \mathbb{P}_{0,n}. \tag{A.184}$$

The same argument as in proving the weak convergence of the term involving B_n in (A.142) can be used to show that

$$\int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \frac{n(\theta_{(1)} - \widehat{\theta}^{\mathbb{P}}_{(1)})^2}{2\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} \mathrm{d}\theta_{(1)} \to \frac{1}{2} \text{ in } \mathbb{P}_{0,n}. \tag{A.185}$$

Note that the prior $\pi_{\mathbb{P}}(\theta_{(1)})$ is assumed to be continuous on the compact domain. Again, because of Corollary 6 and the fact that total variation distance is bounded by relative entropy, we know that

$$\int \pi_{\mathbb{P}}(\theta_{(1)}|\mathbf{x^n}) \ln \pi_{\mathbb{P}}(\theta_{(1)}) \mathrm{d}\theta_{(1)} \to \ln \pi_{\mathbb{P}}(\theta_{0,(1)}) \text{ in } \mathbb{P}_{0,n}. \tag{A.186}$$

Thus, following (A.184) - (A.186), it holds that

$$\begin{aligned} R_{n,1} - \frac{1}{2}\ln(n) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}\right) \\ - \frac{1}{2} - \ln \pi_{\mathbb{P}}(\theta_{0,(1)}) \to 0 \text{ in } \mathbb{P}_{0,n}. \end{aligned} \tag{A.187}$$

We shall also show that, in $\mathbb{P}_{0,n}$,

$$\begin{aligned} R_{n,2} + \frac{D_\Theta}{2}\ln(n+m) - \frac{D_\Theta}{2}\ln(2\pi) - \frac{1}{2}\ln\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right] \\ - \ln \pi_{\mathbb{P}}(\theta_0) - \frac{D_\Theta}{2} \to 0. \end{aligned} \tag{A.188}$$

and

$$\begin{aligned} R_{n,3} + \frac{D_\Theta - 1}{2}\ln(n+m) - \frac{D_\Theta - 1}{2}\ln(2\pi) \\ - \frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} - \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \frac{D_\Theta - 1}{2} \to 0 \text{ in } \mathbb{P}_{0,n}. \end{aligned} \tag{A.189}$$

Combining the weak convergence results (A.187) - (A.189), we can achieve the weak convergence of R_n, or equivalently the result in (A.180).

The proofs of the result (A.188) and the result (A.189) are quite similar, though the proof of the result (A.189) is a little bit more involving. Without tedious repeating the same proofs, we shall only provide the proof for the result (A.189). We further define the left-hand side of (A.189) as $R^*_{n,3}$, that is,

$$R^*_{n,3} \equiv R_{n,3} + \frac{D_\Theta - 1}{2}\ln(n+m) - \frac{D_\Theta - 1}{2}\ln(2\pi) \\ - \frac{1}{2}\ln\frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}} - \ln\pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \frac{D_\Theta - 1}{2}. \tag{A.190}$$

In the rest of the proof, we shall show that for any $\epsilon > 0$, it holds that

$$\limsup_{n\to+\infty}\mathbb{P}_{0,n}\left\{|R^*_{n,3}| > \epsilon\right\} < \epsilon \tag{A.191}$$

Step 1: We first define the big probability events which we shall focus on in order to show (A.191). We define the big probability event

$$\mathcal{A}_n(\theta_0,\delta,\xi) \equiv \left\{\pi_{\mathbb{P}}(\Omega(\theta_0,\delta)^c|\mathbf{x^n}) \le e^{-\xi n}\right\}. \tag{A.192}$$

According to Theorem 1 and Proposition 3 in Chernozhukov and Hong (2003), we know that it suffices to show that

$$\limsup_{n\to+\infty}\mathbb{P}_{0,n}\mathcal{A}_n(\theta_0,\delta,\xi)\left\{|R^*_{n,3}| > \epsilon\right\} < \epsilon. \tag{A.193}$$

Define the sets

$$\mathcal{J}_{1,n}(\delta,\eta) \equiv \left\{\left|\left|\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}} \le \eta\left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1},\ \forall\,\theta\in\Omega(\theta_0,\delta)\text{ and }\tilde{\theta}\in\Omega(\theta,\delta)\right\},$$

$$\mathcal{J}_{2,m}(\theta,\delta,\eta) \equiv \left\{\left|\left|\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta}) - \mathbf{I}_{\mathbb{P}}(\theta)\right|\right|_{\mathcal{S}} \le \eta\left|\left|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right|\right|_{\mathcal{S}}^{-1},\ \forall\,\tilde{\theta}\in\Omega(\theta,\delta)\right\}.$$

Appealing to Propositions 25, we know that for any $\eta > 0$ there exists small enough positive constants δ_1 and δ such that

$$\mathbb{P}_{0,n}\mathcal{J}_{1,n}(\delta,\eta)^c = o\left(\frac{1}{n}\right)\quad\text{and}\quad\sup_{\theta\in\Omega(\theta_0,\delta_1)}\mathbb{P}_{\theta,m}\mathcal{J}_{2,m}(\theta,\delta,\eta)^c = o\left(\frac{1}{m}\right).$$

Define

$$\mathcal{H}_{1,n}(\delta) \equiv \left\{ \left| \widehat{H}_{\mathbb{P},n}(\widehat{\theta}^{\mathbb{P}}) - H(\theta_0) \right| < \delta \right\},$$
$$\text{and } \ \mathcal{H}_{2,n}(\delta) \equiv \left\{ \sup_{\theta\in\Theta} \left| \widehat{H}_{\mathbb{P},n}(\theta) - H(\theta) \right| < \delta \right\}.$$

According to Proposition 18, we know that

$$\mathbb{P}_{0,n}\mathcal{H}_{1,n}(\delta)^c = o(1) \quad \text{and} \quad \mathbb{P}_{0,n}\mathcal{H}_{2,n}(\delta)^c = o(1).$$

Define

$$\mathcal{B}_m(\theta,\delta,\xi)$$
$$\equiv \left\{ \pi_{\mathbb{P}}(\mathbf{x^m}|\theta) > e^{\xi m} \int_{\Omega_{(-1)}(\theta,\delta)^c} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\pi_{\mathbb{P}}(\mathbf{x^m}|\theta_{(1)},\theta'_{(-1)})\mathrm{d}\theta'_{(-1)} \right\}. \quad \text{(A.194)}$$

According to Proposition 29, for any $\delta_0 \in (0,\delta)$, there exists $\xi > 0$ such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{B}_m(\theta,\delta,\xi)^c = O(e^{-\xi m}).$$

Define

$$\mathcal{C}_m(\theta,\delta,\xi)$$
$$\equiv \left\{ \pi_{\mathbb{P}}(\mathbf{x^m}|\theta) < e^{\xi m} \int_{\Omega(\theta,\delta)} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\pi_{\mathbb{P}}(\mathbf{x^m}|\theta_{(1)},\theta'_{(-1)})\mathrm{d}\theta'_{(-1)} \right\}. \quad \text{(A.195)}$$

According to Proposition 33, for any $\delta_0 \in (0,\delta)$, there exists $\xi > 0$ such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{C}_m(\theta,\delta_0,\xi)^c = o\left(\frac{1}{n}\right).$$

We define

$$\mathcal{L}_n(\delta,\eta) \equiv \left\{ s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) < \eta, \ \forall\, \theta \in \Omega(\theta_0,\delta) \right\}. \tag{A.196}$$

According to Proposition 26, we know that for any $\eta > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}_{0,n}\mathcal{L}_n(\delta,\eta)^c = o(1).$$

<u>**Step 2:**</u> We capture the asymptotically essential component in $R_{n,3}$.

For any $\eta \in (0, 1/2)$, according to the discussion in Step 1, we know that there exists $\delta, \delta_1 \in (0,\eta)$ such that

$$\mathbb{P}_{0,n}\mathcal{I}_{1,n}(\delta,\eta)^c = o\left(\frac{1}{n}\right) \text{ and } \sup_{\theta\in\Omega(\theta,\delta_1)} \mathbb{P}_{\theta,m}\mathcal{I}_{2,m}(\theta,\delta,\eta)^c = o\left(\frac{1}{m}\right).$$

For the given δ above, we know that there exist positive constants $\delta_2 < \delta$, ξ_1 and ξ_2 such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_2)} \mathbb{P}_{\theta,m}\mathcal{B}_m(\theta,\delta,\xi_1/\varrho)^c \le e^{-\xi_2 m} \text{ and } \sup_{\theta\in\Omega(\theta_0,\delta_2)} \mathbb{P}_{\theta,m}\mathcal{C}_m(\theta,\delta,\xi_1/8)^c = o\left(\frac{1}{m}\right)$$

where $\mathcal{B}_m(\theta,\delta,\xi_1)$ and $\mathcal{C}_m(\theta,\delta,\xi_1)$ are defined in (A.194) and (A.195), respectively.

Because $H_{\mathbb{P}}(\theta)$ is continuous in θ, then there exists $\delta_3 > 0$ such that

$$\sup_{\theta\in\Omega(\theta_0,\delta_3)} |H_{\mathbb{P}}(\theta) - H_{\mathbb{P}}(\theta_0)| < \xi_1/8.$$

For the given δ, according to Proposition 26, we know that there exists $\delta_4 > 0$ such that

$$\mathbb{P}_{0,n}\mathcal{L}_n(\delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta)^c = o(1).$$

We choose $\delta_0 \equiv \min\{\delta_1,\delta_2,\delta_3,\delta_4\}$. According to Proposition 29, there exist $\xi_0 > 0$ such that $\mathbb{P}_{0,n}\mathcal{A}_n(\delta_0,\xi_0)^c \le e^{-\xi_0 n}$.

By restricting on the event

$$\mathcal{M}_n \equiv \mathcal{A}_n(\theta_0, \delta_0, \xi_0) \cap \mathcal{H}_{1,n}(\xi_1/8) \cap \mathcal{H}_{2,n}(\xi_1/8) \cap \mathcal{I}_{1,n}(\delta, \eta) \cap \mathcal{L}_n(\delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta)$$

and then focusing on the event, for a given $\theta \in \Omega(\theta_0, \delta_0)$,

$$\mathcal{N}_m(\theta) \equiv \mathcal{B}_m(\theta, \delta, \xi_1) \cap \mathcal{I}_{2,m}(\theta, \delta, \eta) \cap \mathcal{C}_m(\theta, \delta, \xi_1).$$

We show that the following term, denoted as $R^e_{n,3}$, is the asymptotically essential term of $R_{n,3}$

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \times \ln \frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)}) \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta. \tag{A.197}$$

That is, there exists a function $a(\eta)$ with $\lim_{\eta\to 0} a(\eta) = 0$ such that on the event $\mathcal{M}_n$

$$|R_{n,3} - R^e_{n,3}| \leq a(\eta) + o_p(1).$$

We consider the decomposition

$$\begin{aligned}
&\ln \frac{\int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)}) \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \\
&= \ln \left[\frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)}) \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \right. \\
&\qquad \left. + \frac{\int_{\Omega_{(-1)}(\theta,\delta)^c} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)}) \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta)} \right]
\end{aligned} \tag{A.198}$$

On the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$, we know that the second term in the log term of

(A.198) can be upper bounded by

$$\begin{aligned}\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})} &= e^{-n\left[\widehat{H}_{\mathbb{P},n}(\theta_{(1)},\theta'_{(-1)})-\widehat{H}_{\mathbb{P},n}(\theta_{(1)},\theta_{(-1)})\right]}\\ &\le e^{-n\left[H_{\mathbb{P}}(\theta_{(1)},\theta'_{(-1)})-H_{\mathbb{P}}(\theta_{(1)},\theta_{(-1)})\right]+n\xi_1/4}\\ &\le e^{-n\left[H_{\mathbb{P}}(\theta_{(1)},\theta'_{(-1)})-H_{\mathbb{P}}(\theta_{0,(1)},\theta_{0,(-1)})\right]+n\xi_1 3/8} \le e^{n\xi_1 3/8}.\end{aligned}$$

Thus, we have on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$

$$\begin{aligned}&\frac{\int_{\Omega_{(-1)}(\theta,\delta)^c}\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta_{(1)},\theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta)}\\ &\le e^{n\xi_1 3/8}\frac{\int_{\Omega_{(-1)}(\theta,\delta)^c}\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)},\theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)},\theta_{(-1)})} \le e^{-n\xi_1 5/8}.\end{aligned}$$

On the other hand, we have on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$

$$\begin{aligned}\frac{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathbb{P}}(\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})} &= e^{-n\left[\widehat{H}_{\mathbb{P},n}(\theta_{(1)},\theta'_{(-1)})-\widehat{H}_{\mathbb{P},n}(\theta_{(1)},\theta_{(-1)})\right]}\\ &\ge e^{-n\left[H_{\mathbb{P}}(\theta_{(1)},\theta'_{(-1)})-H_{\mathbb{P}}(\theta_{(1)},\theta_{(-1)})\right]-n\xi_1/4} \ge e^{-n\xi_1 3/8}.\end{aligned}$$

Thus, we have on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$

$$\begin{aligned}&\frac{\int_{\Omega_{(-1)}(\theta,\delta)}\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta_{(1)},\theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}},\mathbf{x^n}|\theta)}\\ &\ge e^{-n\xi_1 3/8}\frac{\int_{\Omega_{(-1)}(\theta,\delta)}\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)},\theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta_{(1)},\theta_{(-1)})} \ge e^{-n\xi_1/2}.\end{aligned}$$

Therefore, on the event $\mathcal{M}_n$ and $\mathcal{N}_m(\theta)$, for the positive constant $\eta > 0$, we know that when n is large enough, it holds that

$$\frac{\int \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)} \\ \leq (1+\eta)\frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)}$$

Because $H_{\mathbb{P}}(\theta)$ is continuous on Θ, we can define

$$M_H \equiv \sup_{\theta\in\Theta} H_{\mathbb{P}}(\theta) - \inf_{\theta\in\Theta} H_{\mathbb{P}}(\theta). \tag{A.199}$$

Then, on the event $\mathcal{M}_n$, we have $|R_{n,3} - R^e_{n,3}|$ is upper bounded by

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)^c} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \times \\ \left| \ln \frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta)} \right| \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}\mathrm{d}\theta \\ + \ln(1+\eta) + \pi_{\mathbb{P}}(\Omega(\theta_0,\delta_0)^c|\mathbf{x}^{\mathbf{n}})(m+n)M_H.$$

The first term in the long expression above is upper bounded by

$$(m+n)M_H \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{N}_m(\theta)^c \\ \leq (m+n)M_H \Bigg[\sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{B}_m(\theta,\delta,\xi_1/\varrho)^c \\ + \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{\theta,m}\mathcal{I}_{2,m}(\theta,\delta,\eta)^c + \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_{0,m}\mathcal{C}_m(\theta,\delta,\xi_1/8)^c \Bigg] \\ = o(1).$$

On the set $\mathcal{M}_n$, we know that

$$\pi_{\mathbb{P}}(\Omega(\theta_0,\delta_0)^c|\mathbf{x^n})(m+n)M_H = O\left(ne^{-\xi_0 n}\right). \quad \text{(A.200)}$$

Therefore, by the arbitrariness of positive constant η, we know that the asymptotically essential component of $R_{n,3}$ is $R^e_{n,3}$ with expression

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\mathbf{\tilde{x}^m}|\theta) \times \ln \frac{\int_{\Omega_{(-1)}(\theta,\delta)} \pi_{\mathbb{P}}(\mathbf{\tilde{x}^m},\mathbf{x^n}|\theta_{(1)},\theta'_{(-1)})\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)}}{\pi_{\mathbb{P}}(\mathbf{\tilde{x}^m},\mathbf{x^n}|\theta)} \mathrm{d}\mathbf{\tilde{x}^m}\mathrm{d}\theta$$

Step 3: We first show that the term $R^*_{n,3}$ is upper-bounded, asymptotically, by zero.
Step 3.1: We find the upper bound for the log term in the expression of $R^e_{n,3}$ when $\mathbf{x^n} \in \mathcal{M}_n$, $\theta \in \Omega(\theta_0,\delta_0)$ and $\mathbf{\tilde{x}^m} \in \mathcal{N}_m(\theta)$. By Taylor's expansion, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\mathbf{\tilde{x}^m},\mathbf{x^n}|\theta_{(1)},\theta'_{(-1)})}{\pi_{\mathbb{P}}(\mathbf{\tilde{x}^m},\mathbf{x^n}|\theta_{(1)},\theta_{(-1)})}\pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)} \\
&= \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})](\theta'-\theta)} \\
&\quad \times \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)})\mathrm{d}\theta'_{(-1)},
\end{aligned}$$

where $\tilde{\theta}$ is between θ' and θ, and

$$\theta' \equiv \begin{pmatrix} \theta_{(1)} \\ \theta'_{(-1)} \end{pmatrix}, \quad \text{(A.201)}$$

and

$$s_{\mathbb{P},n}(\theta) \equiv \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta), \text{ and } s_{\mathbb{P},m}(\theta) \equiv \frac{1}{m}\sum_{t=1}^{m} \frac{\partial}{\partial\theta} \ln \pi_{\mathbb{P}}(\mathbf{\tilde{x}_t};\theta),$$

and

$$\widehat{\mathbf{I}}_{\mathbb{P},n}(\theta) \equiv -\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \ln \pi_{\mathbb{P}}(\mathbf{x_t};\theta), \text{ and } \widehat{\mathbf{I}}_{\mathbb{P},m}(\theta) \equiv -\frac{1}{m}\sum_{t=1}^{m} \frac{\partial^2}{\partial\theta\partial\theta^T} \ln \pi_{\mathbb{P}}(\mathbf{\tilde{x}_t};\theta).$$

Let's define

$$\rho(\delta) \equiv \sup_{\theta\in\Omega(\theta_0,2\delta)} \left| \ln \frac{\pi_{\mathbb{P}}(\theta_{(-1)}|\theta_{(1)})}{\pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)})} \right|. \tag{A.202}$$

Because $\theta \in \Omega(\theta_0, \delta_0)$ and $\theta'_{(-1)} \in \Omega_{(-1)}(\theta, \delta)$ imply that $\theta' \in \Omega(\theta_0, 2\delta)$, we know that

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\leq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} \\
&\qquad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\delta).
\end{aligned}$$

It is obvious that $\rho(\cdot)$ is increasing a univariate increasing function. Then, it holds that $\rho(\delta) \leq \rho(\eta)$ since $\delta < \eta$. Thus, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\leq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} \\
&\qquad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\eta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (A.202).

On the event $\mathcal{I}_{1,n}(\delta, \eta)$, we have for all $\theta \in \Omega(\theta_0, \delta_0)$

$$(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\theta' - \theta) \geq (1 - \eta)(\theta' - \theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta' - \theta).$$

On the event $\mathcal{I}_{2,m}(\theta, \delta, \eta)$, we have

$$(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathbb{P},m}(\theta' - \theta) \geq (1 - \eta)(\theta' - \theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta' - \theta).$$

Thus, it follows that

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\leq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(1-\eta)(\theta'-\theta)^T[n\mathbf{I}_{\mathbb{P}}(\theta)+m\mathbf{I}_{\mathbb{P}}(\theta)](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} \\
&\qquad\qquad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\eta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (A.202).

Denote

$$\alpha = \frac{n}{n+m}.$$

Let's consider the following identities

$$\begin{aligned}
&(\theta' - \theta)^T [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)] - \frac{1}{2}(1-\eta)(\theta'-\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-\theta) \\
&= -\frac{1-\eta}{2}(\theta'-u)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta-u) \\
&\qquad + \frac{1}{2(1-\eta)}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]
\end{aligned}$$

where

$$u \equiv \theta + \frac{1}{1-\eta} \mathbf{I}_{\mathbb{P}}(\theta)^{-1} [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]. \tag{A.203}$$

Therefore, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\eta) \\
&\leq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-u)} \mathrm{d}\theta'_{(-1)} \\
&\qquad + \frac{m+n}{2(1-\eta)}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)],
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (A.202).

Further, we have

$$\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-u)} \mathrm{d}\theta'_{(-1)} \leq \int_{\mathbb{R}^{d-1}} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-u)} \mathrm{d}\theta'_{(-1)}$$
$$= \frac{(2\pi)^{D_\Theta/2} |(1-\eta)^{-1}\mathbf{I}_{\mathbb{P}}(\theta)^{-1}|^{1/2}}{(m+n)^{D_\Theta/2}} \times$$
$$\int_{\mathbb{R}^{D_\Theta-1}} \frac{(m+n)^{D_\Theta/2}}{(2\pi)^{D_\Theta/2}|(1-\eta)^{-1}\mathbf{I}_{\mathbb{P}}(\theta)^{-1}|^{1/2}} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-u)} \mathrm{d}\theta'_{(-1)}$$

Thus, we can obtain

$$\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1-\eta)(\theta'-u)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-u)} \mathrm{d}\theta'_{(-1)}$$
$$\leq \frac{(2\pi)^{D_\Theta/2}\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{(1-\eta)^{D_\Theta/2}(m+n)^{D_\Theta/2}} \frac{(m+n)^{1/2}}{(2\pi)^{1/2}|(1-\eta)^{-1}\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}}$$
$$= \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} (1-\eta)^{-\frac{D_\Theta-1}{2}}.$$

Therefore,

$$\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)}$$
$$\leq \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) + \frac{1}{2}\ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|} - \frac{m+n}{2}(1-\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}$$
$$+ \frac{m+n}{2(1-\eta)}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha)s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha)s_{\mathbb{P},m}(\theta)]$$
$$+ \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\eta) - \frac{D_\Theta - 1}{2}\ln(1-\eta), \tag{A.204}$$

where $\rho(\eta)$ is the function defined in (A.202).

Step 3.2: We find the upper bound for the asymptotically essential term of $R^e_{n,3}$.

We take integrations over θ and $\tilde{\mathbf{x}}^{\mathbf{m}}$ in (A.197) over each term on the right hand side of the inequality (A.204).

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$\leq \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right)$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$\leq \frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}|} + \eta$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$= \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$- \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)^c} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$= \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta + o(1),$$

and recall the definition of θ' in (A.201) which implies that $\theta'_{(1)} \equiv \theta_{(1)}$ and remember the definition of u in (A.203), we have

$$\begin{aligned}
&\int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \\
&= -\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \left\{ \int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)\right] \right. \\
&\qquad \left. \times \left[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)\right]^T \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \right\} \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v} \\
&\qquad \times \left[2(m+n)^{-1}(1-\eta)^2 \mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v} \right]^{-1} \\
&= -\frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \left\{ \alpha^2 s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T + (1-\alpha)^2 \mathbf{I}_{\mathbb{P}}(\theta)/m \right\} \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}{2(m+n)^{-1}(1-\eta)^2 \mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} \\
&= -\frac{\alpha}{(1-\eta)^2} \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} - \frac{1-\alpha}{2(1-\eta)^2},
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}(1-\eta) \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&= -\frac{\alpha}{2(1-\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} \mathrm{d}\theta \\
&\quad - \frac{1-\alpha}{2(1-\eta)} + o(1) \\
&\leq -\frac{\alpha}{2} - \frac{1-\alpha}{2(1-\eta)} + o(1) \leq -\frac{1}{2} + o(1),
\end{aligned}$$

and

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{m+n}{2(1-\eta)} \\
&\quad \times [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&= \frac{\alpha}{2(1-\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) \mathrm{d}\theta + \frac{1-\alpha}{2(1-\eta)} D_\Theta + o(1) \\
&\leq \frac{\alpha}{2(1-\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) \mathrm{d}\theta \\
&\quad + \frac{1-\alpha}{2} D_\Theta + D_\Theta \eta + o(1) \\
&\leq \frac{\alpha(D_\Theta + \eta)}{2(1-\eta)} + \frac{1-\alpha}{2} D_\Theta + D_\Theta \eta + o(1) \\
&\leq \frac{D_\Theta}{2} + (\alpha + D_\Theta + \alpha D_\Theta)\eta + o(1).
\end{aligned}$$

and

$$\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \\
&\quad \times \left[\ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\eta) - \frac{D_\Theta - 1}{2} \ln(1-\eta) \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&= \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) + \rho(\eta) - \frac{D_\Theta - 1}{2} \ln(1-\eta) + o(1),
\end{aligned}$$

where $\rho(\eta)$ is defined in (A.202).

Therefore, we know that the asymptotically essential component (A.197) of the term $R_{n,3}$ is upper bounded by

$$\begin{aligned}
&\frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) + \frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v}|} + \frac{D_\Theta - 1}{2} + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) \\
&+ \rho(\eta) + (1 + \alpha + D_\Theta + \alpha D_\Theta)\eta - \frac{D_\Theta - 1}{2} \ln(1-\eta) + o(1),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (A.202).

Because of the fact that $\lim_{x \to 0} \rho(x) = 0$ and the arbitrariness of η, we know that $R^*_{n,3}$ defined in (A.190) is asymptotically upper bounded by zero.

Step 4: We then show that the difference $R^*_{n,3}$ is lower-bounded, asymptotically, by zero.

Step 4.1: We find the lower bound for the log term in the expression of $R^e_{n,3}$ when $\mathbf{x^n} \in \mathcal{M}_n(\theta_0)$, $\theta \in \Omega(\theta_0, \delta_0)$ and $\tilde{\mathbf{x}}^{\mathbf{m}} \in \mathcal{N}_m(\theta)$.

By Taylor's expansion, we have

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) d\theta'_{(-1)} \\
&= \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})](\theta'-\theta)} \\
&\qquad \times \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) d\theta'_{(-1)},
\end{aligned}$$

where $\tilde{\theta}$ is between θ' and θ, and

$$\theta' \equiv \begin{pmatrix} \theta_{(1)} \\ \theta'_{(-1)} \end{pmatrix}.$$

Because $\theta \in \Omega(\theta_0, \delta_0)$ and $\theta'_{(-1)} \in \Omega_{(-1)}(\theta, \delta)$ imply that $\theta' \in \Omega(\theta_0, 2\delta)$, we know that

$$\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x^n}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) d\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})](\theta'-\theta)} d\theta'_{(-1)} \\
&\qquad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\delta),
\end{aligned}$$

where the function $\rho(\eta)$ is defined in (A.202).

It is obvious that $\rho(\cdot)$ is increasing a univariate increasing function. Then,

$$\rho(\delta) \leq \rho(\eta) \quad \text{since} \quad \delta < \eta.$$

Thus, we have

$$
\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(\theta'-\theta)^T[n\widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})+m\widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} \\
&\qquad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta),
\end{aligned}
$$

where the function $\rho(\eta)$ is defined in (A.202).

On the event $\mathcal{I}_{1,n}(\delta, \eta)$, we have for all $\theta \in \Omega(\theta_0, \delta_0)$

$$
(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathbb{P},n}(\tilde{\theta})(\theta' - \theta) \leq (1 + \eta)(\theta' - \theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta' - \theta).
$$

On the event $\mathcal{I}_{2,m}(\theta, \delta, \eta)$, we have

$$
(\theta' - \theta)^T \widehat{\mathbf{I}}_{\mathbb{P},m}(\tilde{\theta})(\theta' - \theta) \leq (1 + \eta)(\theta' - \theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta' - \theta).
$$

Thus, it follows that

$$
\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}}|\theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)}|\theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{(\theta'-\theta)^T[ns_{\mathbb{P},n}(\theta)+ms_{\mathbb{P},m}(\theta)]-\frac{1}{2}(1+\eta)(\theta'-\theta)^T[n\mathbf{I}_{\mathbb{P}}(\theta)+m\mathbf{I}_{\mathbb{P}}(\theta)](\theta'-\theta)} \mathrm{d}\theta'_{(-1)} \\
&\qquad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta),
\end{aligned}
$$

where the function $\rho(\eta)$ is defined in (A.202).

Denote

$$
\alpha = \frac{n}{n+m}.
$$

Let's consider the following identities

$$
\begin{aligned}
&(\theta' - \theta)^T [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)] - \frac{1}{2}(1+\eta)(\theta'-\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-\theta) \\
&= -\frac{1+\eta}{2}(\theta' - v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta - v) + \frac{1}{2(1+\eta)}[\alpha s_{\mathbb{P},n}(\theta) \\
&\qquad + (1-\alpha) s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]
\end{aligned}
$$

where

$$
v \equiv \theta + \frac{1}{1+\eta} \mathbf{I}_{\mathbb{P}}(\theta)^{-1}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)].
$$

Therefore, we have

$$
\begin{aligned}
&\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}} | \theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}} | \theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)} | \theta_{(1)}) \mathrm{d}\theta'_{(-1)} \\
&\geq \ln \int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&\quad + \frac{m+n}{2(1+\eta)}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)] \\
&\quad + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)} | \theta_{0,(1)}) - \rho(\eta),
\end{aligned}
$$

where the function $\rho(\eta)$ is defined in (A.202). Further, we have

$$
\begin{aligned}
&\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&\quad = \int_{\mathbb{R}^{D_\Theta - 1}} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&\qquad - \int_{\Omega_{(-1)}(\theta,\delta)^c} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_{\mathbb{R}^{D_\Theta - 1}} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&= \frac{(2\pi)^{D_\Theta/2}\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{(1+\eta)^{D_\Theta/2}(m+n)^{D_\Theta/2}} \times \\
&\quad \int_{\mathbb{R}^{D_\Theta - 1}} \frac{(m+n)^{D_\Theta/2}(1+\eta)^{D_\Theta/2}}{(2\pi)^{D_\Theta/2}\left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
&= \frac{(2\pi)^{D_\Theta/2}\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{(1+\eta)^{D_\Theta/2}(m+n)^{D_\Theta/2}} \frac{(m+n)^{1/2}}{(2\pi)^{1/2}|(1+\eta)^{-1}\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} \\
&\quad \times e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} \\
&= \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta - 1}{2}} \frac{\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} \\
&\quad \times e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta - 1}{2}}.
\end{aligned}
$$

Using the definition of v, we know that for any θ' such that $\theta'_{(-1)} \in \Omega_{(-1)}(\theta, \delta)^c$

$$
\begin{aligned}
&(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v) \\
&= \left[\theta' - \theta - \frac{1}{1+\eta}\mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta)\right]^T \mathbf{I}_{\mathbb{P}}(\theta) \left[\theta' - \theta - \frac{1}{1+\eta}\mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta)\right] \\
&\geq \frac{1}{2}(\theta'-\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-\theta) - \frac{1}{(1+\eta)^2} s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) \\
&\geq \frac{\underline{\lambda}\delta^2}{2(1+\eta)^2} - \frac{\underline{\lambda}\delta^2\eta}{2(1+\eta)^2} \text{ because of } \mathcal{L}_n(\theta_0, \delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta) \\
&= \frac{\underline{\lambda}\delta^2(1-\eta)}{2(1+\eta)^2}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& -\int_{\Omega_{(-1)}(\theta,\delta)^c} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
& \geq -\int_{\Omega_{(-1)}(\theta,\delta)^c} e^{-\frac{(m+n)(1-\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} e^{-\frac{m+n}{4}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
& \geq -e^{-\frac{(m+n)(1-\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} \int_{\mathbb{R}^{D_\Theta - 1}} e^{-\frac{m+n}{4}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)} \\
& = -e^{-\frac{(m+n)(1-\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} 2^{\frac{D_\Theta - 1}{2}} \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} \\
& \qquad \times e^{-\frac{m+n}{4}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta-1}{2}} \\
& \geq -e^{-\frac{(m+n)(1-2\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} 2^{\frac{D_\Theta - 1}{2}} \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} \\
& \qquad \times e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta-1}{2}}
\end{aligned}
$$

where the last inequality is due to the fact that

$$
\frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} = \frac{1}{(1+\eta)^2} \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}.
$$

and

$$
\begin{aligned}
& \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}} \\
& \leq \sup_{|\mathbf{u}|=1} \frac{\mathbf{u}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{u}}{\mathbf{u}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{u}} \\
& = \lambda_m \left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2} s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right] \\
& \leq \operatorname{tr}\left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1/2} s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1/2}\right] \\
& = s_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} s_{\mathbb{P},n}(\theta) < \frac{\underline{\lambda}\delta^2\eta}{2(1+\eta)^2} \text{ because of } \mathcal{L}_n(\theta_0, \delta_0, \frac{1}{2}\underline{\lambda}\delta^2\eta).
\end{aligned}
$$

and hence

$$\int_{\Omega_{(-1)}(\theta,\delta)} e^{-\frac{m+n}{2}(1+\eta)(\theta'-v)^T \mathbf{I}_{\mathbb{P}}(\theta)(\theta'-v)} \mathrm{d}\theta'_{(-1)}$$

$$\geq \left[1 - e^{-\frac{(m+n)(1-2\eta)}{8(1+\eta)}\underline{\lambda}\delta^2} 2^{\frac{D_\Theta - 1}{2}}\right] \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}}$$

$$\times e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta-1}{2}}$$

$$\geq \left(\frac{2\pi}{m+n}\right)^{\frac{D_\Theta-1}{2}} \frac{\left[\det \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]^{1/2}}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|^{1/2}} e^{-\frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}} (1+\eta)^{-\frac{D_\Theta}{2}},$$

for large n, m.

Therefore,

$$\ln \int_{\Omega_{(-1)}(\theta,\delta)} \frac{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}} | \theta_{(1)}, \theta'_{(-1)})}{\pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}, \mathbf{x}^{\mathbf{n}} | \theta_{(1)}, \theta_{(-1)})} \pi_{\mathbb{P}}(\theta'_{(-1)} | \theta_{(1)}) \mathrm{d}\theta'_{(-1)}$$

$$\geq \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) + \frac{1}{2} \ln \frac{|\mathbf{I}_{\mathbb{P}}(\theta)^{-1}|}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|}$$

$$- \frac{m+n}{2}(1+\eta)\frac{\left(\theta'_{(1)}-u_{(1)}\right)^2}{\mathbf{v}^T\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}$$

$$+ \frac{m+n}{2(1+\eta)}[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha)s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha)s_{\mathbb{P},m}(\theta)]$$

$$+ \ln \pi_{\mathbb{P}}(\theta_{0,(-1)} | \theta_{0,(1)}) - \rho(\eta) - \frac{D_\Theta}{2}\ln(1+\eta), \tag{A.205}$$

where the function $\rho(\eta)$ is defined in (A.202).

<u>**Step 4.2:**</u> We find the lower bound for the asymptotically essential term $R^e_{n,3}$. We take integrations over θ and $\tilde{\mathbf{x}}^{\mathbf{m}}$ over each term on the right hand side of the inequality

(A.205).

$$
\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&\quad \geq \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) \pi_{\mathbb{P}}\left(\Omega(\theta_0,\delta_0)|\mathbf{x^n}\right) \left[1 - \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_\theta^m \mathcal{N}_m(\theta)\right] \\
&\quad = \frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) + o(1), \text{ due to } \mathcal{A}_n(\theta_0,\delta_0,\xi_0).
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}|} \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&\quad \geq \left[\frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}|} - \eta\right] \pi_{\mathbb{P}}\left(\Omega(\theta_0,\delta_0)|\mathbf{x^n}\right) \left[1 - \sup_{\theta\in\Omega(\theta_0,\delta_0)} \mathbb{P}_\theta^m \mathcal{N}_m(\theta)\right] \\
&\quad = \frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\mathbf{v}|} - \eta + o(1), \text{ due to } \mathcal{A}_n(\theta_0,\delta_0,\xi_0).
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&\quad = \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&\qquad - \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)^c} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta \\
&\quad = \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1}\mathbf{v}}\right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta + o(1),
\end{aligned}
$$

and

$$\int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2} \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}}$$

$$= -\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \left\{ \int_{\mathcal{X}^m} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)\right] \right.$$

$$\left. \times \left[\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)\right]^T \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \right\} \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}$$

$$\times \left[2(m+n)^{-1}(1+\eta)^2 \mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v} \right]^{-1}$$

$$= -\frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \left\{ \alpha^2 s_{\mathbb{P},n}(\theta) s_{\mathbb{P},n}(\theta)^T + (1-\alpha)^2 \mathbf{I}_{\mathbb{P}}(\theta)/m \right\} \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}}{2(m+n)^{-1}(1+\eta)^2 \mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}}$$

$$= -\frac{\alpha}{(1+\eta)^2} \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}}{2\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}} - \frac{1-\alpha}{2(1+\eta)^2},$$

Thus,

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \left[-\frac{m+n}{2}(1-\eta) \frac{\left(\theta'_{(1)} - u_{(1)}\right)^2}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}} \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$

$$= -\frac{\alpha}{2(1+\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x}^{\mathbf{n}}) \frac{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}}{\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} \mathbf{v}} \mathrm{d}\theta$$

$$- \frac{1-\alpha}{2(1+\eta)} + o(1)$$

$$\geq -\frac{\alpha}{2} - \frac{1-\alpha}{2(1+\eta)} + o(1) \geq -\frac{1}{2} + o(1).$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta) \frac{m+n}{2(1+\eta)}$$
$$\times [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)]^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} [\alpha s_{\mathbb{P},n}(\theta) + (1-\alpha) s_{\mathbb{P},m}(\theta)] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$= \frac{\alpha}{2(1+\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) \mathrm{d}\theta + \frac{1-\alpha}{2(1+\eta)} D_\Theta + o(1)$$
$$\geq \frac{\alpha}{2(1+\eta)} \int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) S_{\mathbb{P},n}(\theta)^T \mathbf{I}_{\mathbb{P}}(\theta)^{-1} S_{\mathbb{P},n}(\theta) \mathrm{d}\theta + \frac{1-\alpha}{2} D_\Theta - D_\Theta \eta + o(1)$$
$$\geq \frac{\alpha(D_\Theta - \eta)}{2(1+\eta)} + \frac{1-\alpha}{2} D_\Theta - D_\Theta \eta + o(1) \geq \frac{D_\Theta}{2} - (D_\Theta + \alpha D_\Theta)\eta + o(1).$$

and

$$\int_{\Omega(\theta_0,\delta_0)} \pi_{\mathbb{P}}(\theta|\mathbf{x^n}) \int_{\mathcal{N}_m(\theta)} \pi_{\mathbb{P}}(\tilde{\mathbf{x}}^{\mathbf{m}}|\theta)$$
$$\times \left[\ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta) - \frac{D_\Theta}{2} \ln(1+\eta) \right] \mathrm{d}\tilde{\mathbf{x}}^{\mathbf{m}} \mathrm{d}\theta$$
$$= \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)}) - \rho(\eta) - \frac{D_\Theta}{2} \ln(1+\eta) + o(1),$$

where the function $\rho(\eta)$ is defined in (A.202).

Therefore, we know that the asymptotically essential component (A.197) of the term $R_{n,3}$ is lower bounded by

$$\frac{D_\Theta - 1}{2} \ln\left(\frac{2\pi}{m+n}\right) + \frac{1}{2} \ln \frac{\det\left[\mathbf{I}_{\mathbb{P}}(\theta_0)^{-1}\right]}{|\mathbf{v}^T \mathbf{I}_{\mathbb{P}}(\theta_0)^{-1} \mathbf{v}|} + \frac{D_\Theta - 1}{2} + \ln \pi_{\mathbb{P}}(\theta_{0,(-1)}|\theta_{0,(1)})$$
$$- \rho(\eta) - (D_\Theta + \alpha D_\Theta)\eta - \frac{D_\Theta}{2} \ln(1+\eta) + o(1),$$

where the function $\rho(\eta)$ is defined in (A.202).

Because of the fact that $\lim_{x \to 0} \rho(x) = 0$ and the arbitrariness of η, we know that $R^*_{n,3}$ defined in (A.190) is asymptotically lower bounded by zero.

Proof of Theorem 2 in the Paper

It follows immediately from the results in Proposition 37 implied by Proposition 12 and Proposition 13.

Appendix B

Appendix: The Volatility of International Capital Flows and Foreign Assets

This appendix presents additional theoretical results on the expectation correspondence in Section B.1, the proof of Proposition 4 in Section B.2, the proof of Theorem 3 in Section B.3, two simple cases with analytical solutions in Section B.4, a description of the numerical algorithm in Section B.5, and some additional empirical results on the data in Section **??** and the model in Section B.6.

B.1 Equations of Expectation Correspondence Φ

The optimization problem of each country $i \in \{1,2\}$ can be re-formulated using Lagrangian multipliers, for each $s \in \mathcal{S}$ and $e \in \overline{\mathbb{R}}_+$,

$$U_i(W_i) = \min_{\mu_i^j \geq 0, \mu_i^b \geq 0, \mu_{i,\tilde{s}}^b \geq 0} \max_{c_i^j, \vartheta_i^j, b_i^j} \frac{C_i^{1-\psi_i^{-1}}}{1-\psi_i^{-1}} + \beta \mathbb{E}\left[U_i(\tilde{W}_i)^{\theta_i} | s, e\right]^{1/\theta_i}$$
$$+ \sum_{j=1}^{2} \mu_i^j \vartheta_i^j - \mu_i^b b_i^i + \sum_{\tilde{s} \in \mathcal{S}} \mu_{i,\tilde{s}}^b \mathbb{C}_i(s^t, \tilde{s})$$

subject to the intra-termporal “budget constraint”

$$C_i = G_i(c_i^1, c_i^2),$$

the inter-temporal budget constraint

$$W_i = \sum_{j=1}^{2} p_j c_i^j + \sum_{j=1}^{2} \vartheta_i^j q_j + \sum_{j=1}^{2} b_i^j q_j^b,$$

$$\tilde{W}_i = \tilde{p}_i \tilde{\omega}_i + \sum_{j=1}^{2} \vartheta_i^j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right) + \sum_{j=1}^{2} b_i^j \tilde{p}_j.$$

Using textbook arguments, we can show that each value function $U_i(W_i)$ is concave, continuous, and increasing. Then, the standard variational argument leads to the envelop condition:

$$U_{i,W}(W_i) = C_i^{-\psi_i^{-1}} G_{i,c^2}\left(c_i^1, c_i^2\right), \tag{B.1}$$

where $U_{i,W}$ is the partial derivative of U_i w.r.t. W and G_{i,c^2} is the partial derivative of G_i w.r.t. c^2.

For notational simplicity, we denote $\mathbb{E}[\cdot|s,e] \equiv \mathbb{E}_{s,e}[\cdot]$. The first-order condition for ϑ_i^j, with $i, j \in \{1,2\}$, gives

$$\begin{aligned} & q_j C_1^{-\psi_1^{-1}} G_{1,c^2}\left(c_1^1, c_1^2\right) \\ & = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_1\right)^{\theta_1}\right]^{1/\theta_1 - 1} \mathbb{E}_{s,e}\left[\tilde{U}_{1,W}\left(\tilde{U}_1\right)^{\theta_1 - 1}\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right)\right] \\ & \quad + \mu_1^j + \sum_{\tilde{s} \in \mathcal{S}} \mu_{1,\tilde{s}}^b \eta_j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right), \end{aligned} \tag{B.2}$$

and

$$
\begin{aligned}
& q_j C_2^{-\psi_2^{-1}} G_{2,c^2}\left(c_2^1, c_2^2\right) \\
& = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_2\right)^{\theta_2}\right]^{1/\theta_2 - 1} \mathbb{E}_{s,e}\left[\tilde{U}_{2,W}\left(\tilde{U}_2\right)^{\theta_2 - 1}\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right)\right] \\
& \quad + \mu_2^j + \sum_{\tilde{s} \in \mathcal{S}} \mu_{2,\tilde{s}}^b\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right).
\end{aligned} \tag{B.3}
$$

If we plug the Envelop condition (B.1) into the FOC in (B.3), we can get

$$
\begin{aligned}
& q_j C_i^{-\psi_i^{-1}} G_{i,c^2}\left(c_i^1, c_i^2\right) \\
& = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_i\right)^{\theta_i}\right]^{1/\theta_i - 1} \mathbb{E}_{s,e}\left[\left(\tilde{C}_i\right)^{-\psi_i^{-1}} G_{i,c^2}\left(\tilde{c}_i^1, \tilde{c}_i^2\right)\left(\tilde{U}_i\right)^{\theta_i - 1}\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right)\right] \\
& + \mu_i^j + \sum_{\tilde{s} \in \mathcal{S}} \mu_{i,\tilde{s}}^b\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right).
\end{aligned} \tag{B.4}
$$

Note that when $\theta_i = 1$, the condition above is simplified as

$$
\begin{aligned}
& q_j C_i^{-\psi_i^{-1}} G_{i,c^2}\left(c_{i,1}, c_{i,2}\right) \\
& = \beta \mathbb{E}_{s,e}\left[\left(\tilde{C}_i\right)^{-\psi_i^{-1}} G_{i,c^2}\left(\tilde{c}_i^1, \tilde{c}_i^2\right)\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right)\right] \\
& + \mu_i^j + \sum_{\tilde{s} \in \mathcal{S}} \mu_{i,\tilde{s}}^b\left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right).
\end{aligned} \tag{B.5}
$$

This is simply the first-order condition for CRRA utility functions. Similar equation is derived in Stepanchuk and Tsyrennikov (2015).

Similarly, the first-order condition for bond holdings of agent 1 (i.e. b_1^j) and Envelop condition together lead to

$$
\begin{aligned}
& q_1^b C_1^{-\psi_1^{-1}} G_{1,c^2}\left(c_1^1, c_1^2\right) \\
& = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_1\right)^{\theta_1}\right]^{1/\theta_1 - 1} \mathbb{E}_{s,e}\left[\left(\tilde{C}_1\right)^{-\psi_1^{-1}} G_{1,c^2}\left(\tilde{c}_1^1, \tilde{c}_1^2\right)\left(\tilde{U}_1\right)^{\theta_1 - 1} \tilde{p}_1\right] \\
& - \mu_1^b + \sum_{\tilde{s} \in \mathcal{S}} \mu_{1,\tilde{s}}^b \tilde{p}_1.
\end{aligned} \tag{B.6}
$$

and

$$
\begin{aligned}
& q_2^b C_1^{-\psi_1^{-1}} G_{1,c^2}\left(c_1^1, c_1^2\right) \\
& = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_1\right)^{\theta_1}\right]^{1/\theta_1 - 1} \mathbb{E}_{s,e}\left[\left(\tilde{C}_1\right)^{-\psi_1^{-1}} G_{1,c^2}\left(\tilde{c}_1^1, \tilde{c}_1^2\right)\left(\tilde{U}_1\right)^{\theta_1 - 1} \tilde{p}_2\right].
\end{aligned} \tag{B.7}
$$

Also, the first-order condition for bond holdings of agent 2 (i.e. b_2^j) and Envelop condition together lead to, for $j = 1, 2$,

$$
\begin{aligned}
& q_1^b C_2^{-\psi_i^{-1}} G_{2,c^2}\left(c_2^1, c_2^2\right) \\
& \quad = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_i\right)^{\theta_2}\right]^{1/\theta_2 - 1} \mathbb{E}_{s,e}\left[\left(\tilde{C}_2\right)^{-\psi_2^{-1}} G_{2,c^2}\left(\tilde{c}_2^1, \tilde{c}_2^2\right)\left(\tilde{U}_2\right)^{\theta_2 - 1} \tilde{p}_1\right] \\
& \quad + \sum_{\tilde{s} \in \mathcal{S}} \mu_{2,\tilde{s}}^b \tilde{p}_1.
\end{aligned} \tag{B.8}
$$

and

$$
\begin{aligned}
& q_2^b C_2^{-\psi_i^{-1}} G_{2,c^2}\left(c_2^1, c_2^2\right) \\
& = \beta \mathbb{E}_{s,e}\left[\left(\tilde{U}_i\right)^{\theta_2}\right]^{1/\theta_2 - 1} \mathbb{E}_{s,e}\left[\left(\tilde{C}_2\right)^{-\psi_2^{-1}} G_{2,c^2}\left(\tilde{c}_2^1, \tilde{c}_2^2\right)\left(\tilde{U}_2\right)^{\theta_2 - 1} \tilde{p}_2\right] \\
& \quad + \sum_{\tilde{s} \in \mathcal{S}} \mu_{2,\tilde{s}}^b \tilde{p}_2 \\
& \quad - \mu_2^b.
\end{aligned} \tag{B.9}
$$

The intra-temporal Euler conditions for country $i \in \{1, 2\}$ is

$$
p_1 G_{i,c^2}\left(c_i^1, c_i^2\right) = p_2 G_{i,c^1}\left(c_i^1, c_i^2\right). \tag{B.10}
$$

Therefore, the expectation correspondence $\widehat{\Phi}$ consists of the following five groups of conditions for all $i, j \in \{1, 2\}$:

(1) The intra-temporal Euler equations in (B.10);

(2) The inter-temporal Euler equations about equity holdings in (B.2) and (B.3);

(3) The inter-temporal Euler equations about bond holdings in (B.6), (B.7), (B.8) and (B.9) and the feasibility conditions, for all $\tilde{s} \in \mathcal{S}$,

$$\sum_{j=1}^{2} \eta_j \vartheta_1^j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right) + b_1^1 \tilde{p}_1 \geq 0$$

$$\text{and} \quad \tilde{p}_2 \tilde{\omega}_2 + \sum_{j=1}^{2} \vartheta_2^j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right) + \sum_{j=1}^{2} b_2^j \tilde{p}_j \geq 0,$$

and slackness conditions, for all $\tilde{s} \in \mathcal{S}$,

$$\mu_{1,\tilde{s}}^b \left[\sum_{j=1}^{2} \vartheta_1^j \eta_j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right) + b_1^1 \tilde{p}_1\right] = 0$$

$$\text{and} \quad \mu_{2,\tilde{s}}^b \left[\tilde{p}_2 \tilde{\omega}_2 + \sum_{j=1}^{2} \vartheta_2^j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right) + \sum_{j=1}^{2} b_2^j \tilde{p}_j\right] = 0;$$

(4) The inter-temporal budget constraints, for all $\tilde{s} \in \mathcal{S}$,

$$\tilde{w} \left(\sum_{i=1}^{2} \tilde{p}_i \tilde{e}_i + \sum_{j=1}^{2} \tilde{q}_j\right) = \tilde{p}_1 \tilde{\omega}_1 + \sum_{j=1}^{2} \vartheta_1^j \left(\tilde{q}_j + \tilde{p}_j \tilde{d}_j\right) + \sum_{j=1}^{2} b_1^j \tilde{p}_j \tag{B.11}$$

and

$$\tilde{w} \left(\sum_{i=1}^{2} \tilde{p}_i \tilde{e}_i + \sum_{j=1}^{2} \tilde{q}_j\right) = \sum_{j=1}^{2} \tilde{p}_j \tilde{c}_1^j + \sum_{j=1}^{2} \tilde{\vartheta}_1^j \tilde{q}_j + \sum_{j=1}^{2} \tilde{b}_1^j \tilde{q}_j^b; \tag{B.12}$$

(5) The commodity market clearing conditions, for all $\tilde{s} \in \mathcal{S}$,

$$\tilde{c}_1^j + \tilde{c}_2^j = (c_1^j + c_2^j)\zeta(\tilde{s}). \tag{B.13}$$

B.2 Proof of Proposition 4

Suppose that when $e = 1$, a wealth-recursive equilibrium exists and has the policy functions with the following form

$$\Pi(w,s,1) \equiv \left\{ c_i^j(w,s), \vartheta_i^j(w,s), b_i^j(w,s), p_i(w,s), q_i(w,s), q_i^b(w,s), \mu_i^j(w,s), \mu_{i,\tilde{s}}^b(w,s) \right\}.$$

More precisely, the policy functions in $\Pi(w,s)$, the transition map $\Omega(w,s)$ and the value function $U_i(w,s)$ satisfy the following conditions, for all $(w,s) \in [0,1] \times \mathcal{S}$:

(0) The vectors of endogenous variables lie in $\mathcal{Y}$ defined in (4.10) - (4.11), i.e.

$$\Pi(w,s,1), \Pi(\tilde{w},\tilde{s},\zeta(\tilde{s})) \in \mathcal{Y}.$$

(1) The intra-temporal Euler equations are held:

$$p_1(w,s) G_{i,c^2}\left(c_i^1(w,s), c_i^2(w,s)\right) = p_2(w,s) G_{i,c^1}\left(c_i^1(w,s), c_i^2(w,s)\right). \tag{B.14}$$

(2) The inter-temporal Euler equations about equity positions are held:

$$\begin{aligned} & q_j(w,s) C_i(w,s)^{-\psi_i^{-1}} G_{i,c^2}\left(c_i^1(w,s), c_i^2(w,s)\right) \\ & = \beta \mathbb{E}_{w,s}\left[U_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))^{\theta_i}\right]^{1/\theta_i - 1} \\ & \times \mathbb{E}_{w,s}\Big[C_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))^{-\psi_i^{-1}} G_{i,c^2}\left(c_i^1(\tilde{w},\tilde{s},\zeta(\tilde{s})), c_i^2(\tilde{w},\tilde{s},\zeta(\tilde{s}))\right) \\ & \times U_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))^{\theta_i - 1}(q_j(\tilde{w},\tilde{s},\zeta(\tilde{s})) + p_j(\tilde{w},\tilde{s},\zeta(\tilde{s})) d_j(\tilde{s}) \zeta(\tilde{s}))\Big] \\ & + \mu_i^j(w,s) + \sum_{\tilde{s} \in \mathcal{S}} \mu_{i,\tilde{s}}^b(w,s) \left(q_j(\tilde{w},\tilde{s},\zeta(\tilde{s})) + p_j(w,\tilde{s},\zeta(\tilde{s})) d_j(\tilde{s}) \zeta(\tilde{s})\right). \end{aligned} \tag{B.15}$$

(3) The inter-temporal Euler equations about bond are held

$$
\begin{aligned}
q_j^b(w,s)C_i(w,s)^{-\psi_i^{-1}}G_{i,c^2}\left(c_i^1(w,s),c_i^2(w,s)\right) &= \beta\mathbb{E}_{w,s}\left[U_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))^{\theta_i}\right]^{1/\theta_i-1} \\
&\times \mathbb{E}_{w,s}\Big[\tilde{C}_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))^{-\psi_i^{-1}}G_{i,c^2}\left(c_i^1(\tilde{w},\tilde{s},\zeta(\tilde{s})),c_i^2(\tilde{w},\tilde{s},\zeta(\tilde{s}))\right) \\
&\times \left(U_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))\right)^{\theta_i-1}p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))\Big] \\
&+ \sum_{\tilde{s}\in\mathcal{S}}\mu_{i,\tilde{s}}^b p_j(\tilde{w},\tilde{s},\zeta(\tilde{s})).
\end{aligned}
$$

and the feasibility conditions, for all $\tilde{s}\in\mathcal{S}$,

$$
\begin{aligned}
p_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))\omega_i(\tilde{s})\zeta(\tilde{s}) &+ \sum_{j=1}^{2}\vartheta_i^j(w,s)\left[q_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))+p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))d_j(\tilde{s})\zeta(\tilde{s})\right] \\
&+ \sum_{j=1}^{2}b_i^j(w,s)p_j(\tilde{w},\tilde{s},\zeta(\tilde{s})) \geq 0,
\end{aligned}
$$

and slackness conditions, for all $\tilde{s}\in\mathcal{S}$,

$$
\begin{aligned}
\mu_{i,\tilde{s}}^j(w,s)\Bigg[p_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))\omega_i(\tilde{s})\zeta(\tilde{s}) &+ \sum_{j=1}^{2}\vartheta_i^j(w,s)\left[q_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))+p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))d_j(\tilde{s})\zeta(\tilde{s})\right] \\
&+ \sum_{j=1}^{2}b_i^j(w,s)p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))\Bigg] = 0,
\end{aligned}
$$

(4) The inter-temporal budget constraints, for all $\tilde{s}\in\mathcal{S}$, the wealth share corresponding to exogenous shock in the period $\tilde{s}$ is $\tilde{w}=\Omega(w,s;\tilde{s})$. More precisely,

$$
\begin{aligned}
\tilde{w}\left(\sum_{i=1}^{2}p_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))e_i(\tilde{s})\zeta(\tilde{s})+\sum_{j=1}^{2}q_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))\right) \qquad \text{(B.16)} \\
= p_1(\tilde{w},\tilde{s},\zeta(\tilde{s}))\omega_1(\tilde{s})\zeta(\tilde{s}) \\
+ \sum_{j=1}^{2}\vartheta_1^j(w,s)\left[q_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))+p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))d_j(\tilde{s})\zeta(\tilde{s})\right] \\
+ \sum_{j=1}^{2}b_1^j(w,s)p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))
\end{aligned}
$$

and

$$\tilde{w}\left(\sum_{i=1}^{2} p_i(\tilde{w},\tilde{s},\zeta(\tilde{s}))e_i(\tilde{s})\zeta(\tilde{s}) + \sum_{j=1}^{2} q_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))\right) \tag{B.17}$$
$$= \sum_{j=1}^{2} p_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))c_1^j(\tilde{w},\tilde{s},\zeta(\tilde{s})) + \sum_{j=1}^{2} \vartheta_1^j(\tilde{w},\tilde{s},\zeta(\tilde{s}))q_j(\tilde{w},\tilde{s},\zeta(\tilde{s}))$$
$$+ \sum_{j=1}^{2} b_1^j(\tilde{w},\tilde{s},\zeta(\tilde{s}))q_j^b(\tilde{w},\tilde{s},\zeta(\tilde{s}));$$

(5) The commodity market clearing conditions, for all $\tilde{s} \in \mathcal{S}$,

$$c_1^j(\tilde{w},\tilde{s},\zeta(\tilde{s})) + c_2^j(\tilde{w},\tilde{s},\zeta(\tilde{s})) = (c_1^j(w,s) + c_2^j(w,s))\zeta(\tilde{s}). \tag{B.18}$$

For general value of e, we plug the expressions of (4.11) - (4.14) into the Bellman equation and conditions (0) - (5) above, and then we can see that the current size e is perfectly canceled out. By assumption for the case of $e = 1$, we know that they are policy functions, transition map, and value functions for wealth-recursive Markov equilibrium for any current size e.

B.3 Proof of Theorem 3

We prove the existence by construction which combines important ideas of the proofs in Duffie, Geanakoplos, Mas-Colell, and McLennan (1994), Kubler and Schmedders (2003) and Geanakoplos and Zame (2013). The existence results of equilibria are standard for finite-horizon economy even with incomplete market, while for the infinite-horizon economy the proofs are much more involving. The key idea of the proofs in the literature[1] is basically backward induction and is based on the existence of competitive equilibria on all finitely-truncated economy whose equilibrium variables are

[1]Examples for existence of competitive equilibria in infinite-horizon incomplete market economy with heterogeneous agents include Levine and Zame (1996), Magill and Quinzii (1996), and Hernandez and Santos (1996), among others. Examples for existence of recursive Markov equilibria include Duffie, Geanakoplos, Mas-Colell, and McLennan (1994) and Kubler and Schmedders (2003), among others.

uniformly bounded. We extend the proof in Kubler and Schmedders (2003) to allow for Epstein-Zin preferences including those which are not bounded below (i.e. EIS parameter is bigger than one).

The T-truncated economy is defined to be a finite-horizon economy built on an event tree, denoted by $\mathbb{S}^T$, which consists of all the nodes and edges along the path $s^T = (s_0, s_1, \cdots, s_T)$ in the original event tree $\mathbb{S}$. The endowments and asset payoffs at the nodes of the truncated tree, as well as agents' preferences and portfolio constraints at these nodes, are the identical to the original infinite-horizon economy. The sequential budget constraint of agent i in the T-truncated economy is $\mathbb{B}_{\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T})$ which is a collection of consumption plans $\mathcal{C}_i^{\mathbb{S}^T} = \{c_i^1(s^t), c_i^2(s^t)\}_{s^t \in \mathbb{S}^T}$ and portfolio choice plans $\mathcal{A}_i^{\mathbb{S}^T} = \{\vartheta_i^1(s^t), \vartheta_i^2(s^t), b_i^1(s^t), b_i^2(s^t)\}_{s^t \in \mathbb{S}^T}$ such that, at each node of the event tree $s^t \in \mathbb{S}^T$, the portfolio positions satisfy the short-selling constraint (4.5) and borrowing constraint (4.6) and at each node s^t on the event tree $\mathbb{S}^T$,

$$\begin{aligned} \sum_{j=1}^{2} p_j(s^t) c_i^j(s^t) + \sum_{j=1}^{2} q_j(s^t) \vartheta_i^j(s^t) + \sum_{j=1}^{2} q_j^b(s^t) b_i^j(s^t) \\ = p_i(s^t)\omega_i(s^t) + \sum_{j=1}^{2} \left[q_j(s^t) + p_j(s^t) d_j(s^t) \right] \vartheta_i^j(s^{t-1}) \\ + \sum_{j=1}^{2} p_j(s^t) b_j(s^{t-1}) \end{aligned} \tag{B.19}$$

where s^{t-1} is the ancestor node of the node s^t on the event tree and $s^0 = s_0$ is the initial node.

Inspired by the result in Proposition 4, we take off the scaling effect of the economy by assuming the world tree always has size one, i.e. $e(s^t) \equiv 1$ for all $s^t \in \mathbb{S}$. We first show that the competitive equilibria exist and the equilibrium variables are uniformly bounded over $T \geq 1$. We first formally introduce the following lemma and leave its proof to Appendix B.3.1.

Lemma 7. *For all $T \geq 1$, there exists a competitive equilibrium for the T-truncated economy in which all equilibrium variables, including consumptions, portfolio holdings and prices, all* *lie in a compact set* $\mathcal{Y}^* \subset \mathcal{Y}$.

For any compact set $\mathcal{K} \subset \mathcal{Y}$, and a policy correspondence $\mathrm{Y} : \mathcal{S} \times [0,1] \rightrightarrows \mathcal{K}$, we define an operator $O_{\mathcal{K}}$, that maps the policy correspondence $\mathrm{Y} : \mathcal{S} \times \Delta \rightrightarrows \mathcal{K}$ to another policy correspondence $O_{\mathcal{K}}(\mathrm{Y})$ such that for all $s \in \mathcal{S}$ and $w \in [0,1]$

$$\begin{aligned} & O_{\mathcal{K}}(\mathrm{Y})(s,w) \\ & = \left\{ y \in \mathcal{K} \ : \ \exists \ \left((\tilde{w}_1, \tilde{y}_1), \cdots, (\tilde{w}_{|\mathcal{S}|}, \tilde{y}_{|\mathcal{S}|}) \right) \in \Phi(w,y,s,1) \text{ s.t. } \tilde{y}_{\tilde{s}} \in \mathrm{Y}(\tilde{s}, \tilde{w}_{\tilde{s}}), \ \forall \ \tilde{s} \in \mathcal{S} \right\}. \end{aligned}$$

The correspondence $O_{\mathcal{K}}(\mathrm{Y})(w,s)$ is basically computing the endogenous variables $y \in \mathcal{Y}$ given the state variables (w,s) in the current period and the next period's equilibrium endogenous variables $(w_1, y_1), \cdots, (w_{|\mathcal{S}|}, y_{|\mathcal{S}|})$.

Define constant correspondence Y^0 by $\mathrm{Y}^0(y,w) \equiv \mathcal{Y}^*$ for all $w \in [0,1]$ and all $y \in \mathcal{Y}$. Given a correspondence Y^n, we define recursively $\mathrm{Y}^{n+1} = O_{\mathcal{Y}^*}(\mathrm{Y}^n)$. First, for each n, the set Y^n is nonempty. This is because of Lemma 7, which implies that for all n there exists a n-horizon competitive equilibrium whose endogenous variables lie in the compact set $\mathcal{Y}^*$. Second, we show that Y^n is closed for each n. We prove it by induction. It is obvious that $\mathrm{Y}^0 \equiv \mathcal{Y}^*$ is closed. Suppose Y^n is closed, then $\mathrm{Y}^{n+1} = O_{\mathcal{Y}^*}(\mathrm{Y}^n)$ is also closed because the graph of Φ is closed and the graph of Y^n is closed. Third, for each n, $\mathrm{Y}^{n+1} \subset \mathrm{Y}^n$. By definition, it is obvious that $\mathrm{Y}^1 \subset \mathrm{Y}^0 \equiv \mathcal{Y}^*$. Suppose that $\mathrm{Y}^n \subset \mathrm{Y}^{n+1}$, then we have $\mathrm{Y}^{n+1} \subset \mathrm{Y}^{n+2}$. This is because $O_{\mathcal{Y}^*}(\mathrm{Y}^n) \subset O_{\mathcal{Y}^*}\left(\mathrm{Y}^{n+1}\right)$ by definition.

We define a correspondence Y^* such that for all $(w,s) \in [0,1] \times \mathcal{S}$

$$\mathrm{Y}^*(w,s) \equiv \cap_{n=0}^{\infty} \mathrm{Y}^n(w,s). \tag{B.20}$$

Because for each $(w,s) \in [0,1] \times \mathcal{S}$, the sequence of sets $\{\mathrm{Y}^n(w,s)\}$ are compact, nested, and nonempty, thus $\mathrm{Y}^*(w,s)$ is a closed and nonempty set. $\mathrm{Y}^*(w,s)$ is policy correspondence in recursive Markov equilibria and the definition of operator $O_{\mathcal{Y}^*}$ implies the existence of a transition for the recursive Markov equilibrium.

B.3.1 Proof of Lemma 7

We first show that the policy functions in equilibria are uniformly bounded for all $T \geq 1$ if equilibria exist. The agent i's budget constraint $\mathbb{B}_{i,\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T})$ contains the portfolio constraints including:

$$\vartheta_i^j(s^t) \geq 0, \ -\overline{D}e_i(s^t) \leq b_i^j(s^t) \leq \overline{D}e_i(s^t)\ , \ \text{and}$$
$$p_i(s^t)\omega_i(s^t) + \sum_{j=1}^{2} \vartheta_i^j(s^{t-1})\left[q_j(s^t) + p_j(s^t)d_j(s^t)\right] + \sum_{j=1}^{2} p_j(s^t)b_i^j(s^{t-1}) \geq 0,$$

for $s^{t-1}, s^t \in \mathbb{S}^T$.

For all $T \geq 1$, in equilibria, we know that the consumptions lie in the interval $[0, \overline{e}]$ where $\overline{e} \equiv \max_{s \in \mathbb{S}} \{e_1(s) + e_2(s)\}$ and we know that by nonnegativity and commodity market clearing

$$0 \leq c_i^j(s^t) \leq e_j(s^t) \leq \overline{e}, \quad \text{for } s^t \in \mathbb{S}^T, \tag{B.21}$$

and also by short-selling constraint and equity market clearing

$$0 \leq \vartheta_j(s^t) \leq 1, \quad \text{for } s^t \in \mathbb{S}^T. \tag{B.22}$$

By the debt ceiling and bond market clearing, we know that

$$-\overline{D}\overline{e} \leq b_i^j(s^t) \leq \overline{D}\overline{e} \ \ \text{for } s^t \in \mathbb{S}^T. \tag{B.23}$$

The intra-temporal Euler equations must hold in equilibria,

$$\frac{c_1^1(s^t)}{c_1^2(s^t)} = \left[p_1(s^t)\frac{1-s}{s}\right]^{\frac{1}{\rho-1}}, \tag{B.24}$$

and

$$\frac{e_1(s^t) - c_1^1(s^t)}{e_2(s^t) - c_1^2(s^t)} = \left[p_1(s^t)\frac{s}{1-s}\right]^{\frac{1}{\rho-1}}. \tag{B.25}$$

Thus, combining (B.24) and (B.25), we have

$$e_1(s^t) - e_2(s^t)\left[p_1(s^t)\frac{s}{1-s}\right]^{\frac{1}{\rho-1}}$$
$$= c_1^2(s^t)p_1(s^t)^{\frac{1}{\rho-1}}\left\{\left(\frac{1-s}{s}\right)^{\frac{1}{\rho-1}} - \left(\frac{s}{1-s}\right)^{\frac{1}{\rho-1}}\right\}.$$

On the one hand, because $c_1^1(s^t) \geq 0$ and $\rho \leq 1$, then

$$p_1(s^t) \geq \frac{1-s}{s}\left[\frac{e_1(s^t)}{e_2(s^t)}\right]^{\frac{1}{1-\rho}} \geq \frac{1-s}{s}\left(\frac{1}{\kappa}\right)^{\frac{1}{1-\rho}} \equiv \underline{P}_1, \tag{B.26}$$

with $\kappa \equiv \max_{s\in\mathbb{S},i_1,i_2=1,2}\frac{e_{i_1}(s)}{e_{i_2}(s)} > 1$. And, on the other hand, because $c_1^1(s^t) \leq e_1(s^t)$, then

$$p_1(s^t) \leq \frac{s}{1-s}\kappa^{\frac{1}{\rho-1}} \equiv \overline{P}_1. \tag{B.27}$$

Now, we consider the prices of bonds. According to Santos and Woodford (1997), if aggregate endowment is bounded away from zero, then any stationary and recursive preference ordering does satisfy the form of impatience that for each agent $i \in I$, there exist $K > 0$ and $0 \leq \delta < 1$ such that for every $s^t \in \mathbb{S}$,

$$\Big((c_i^1(s^t), c_i^2(s^t) + Ke_1(s^t)), (\delta c_{i,+}^1(s^t), \delta c_{i,+}^2(s^t))\Big)$$
$$\succ^i \Big((c_i^1(s^t), c_i^2(s^t)), (c_{i,+}^1(s^t), c_{i,+}^2(s^t))\Big)$$

for all consumption plans satisfying $c_i^j(s^t) \leq e_j(s^t)$ for all $s^t \in \mathbb{S}$ and $c_{i,+}^j(s^t)$ represents the consumption of agent i for goods j over all remaining nodes; i.e. $s^r \in \mathbb{S}$ such that $s^r \succcurlyeq s^t$ and $r > t$. It is obvious that in our economy for a given price system $\mathcal{P}^T$, if a consumption plan c can be supported by an initial wealth W, then the consumption plan δc can be supported by the initial wealth δW for any constant $\delta \in (0,1)$. Thus, we know that for each agent $i \in I$, there exist a $K > 0$ and $0 \leq \delta < 1$ such that for every

$s^t \in \mathbb{S}$,

$$\Big((c_i^1(s^t), c_i^2(s^t) + Ke_1(s^t)), (\delta W_i(s^{t+1}) \ : \ s^{t+1} \succcurlyeq s^t) \Big) \\ \succ^i \Big((c_i^1(s^t), c_i^2(s^t)), (W_i(s^{t+1}) \ : \ s^{t+1} \succcurlyeq s^t) \Big)$$

for all current consumption satisfying $c_i^j(s^t) \leq e_j(s^t)$ for all $s^t \in \mathbb{S}$ and wealth in the beginning of the next period satisfying $W_i(s^{t+1}) \leq \sum_{i=1}^{2} p_i(s^{t+1}) e_i(s^{t+1}) + q_i(s^{t+1})$ for all $s^t \in \mathbb{S}$. Let $Q_2^b \equiv \frac{K\bar{e}}{w_m(1-\delta)}$, and are going to show that the bond price $q_2^b(s^t)$ cannot be higher than Q_2^b by contradiction. Suppose in an equilibrium, there is a note s^t such that $q_2^b(s^t) > Q_2^b$ in an equilibrium of the T-truncated economy. At this node s^t, there must be one agent who is not borrowing in net position, and hence her wealth in the state $s^{t+1} \succcurlyeq s^t$ in the next period is at least w_m. Let's just assume she is agent 1, without loss of any generosity. Suppose her current consumption and next period's wealth plan is $\big((c_i^1(s^t), c_i^2(s^t)), (W_i(s^{t+1}) \ : \ s^{t+1} \succcurlyeq s^t) \big)$. If the agent 1 sells $w_m(1-\delta)$ unit of bond 2 at the node s^t (i.e. borrow $w_m(1-\delta)$ unit more bond 2), she could gain at least K amount of proceeds and then use all of the proceeds to buy at least $Ke_2(s^t)$ units of commodity 2 which are consumed at s^t. However, this selling of bond 2 makes her wealth plan in the next period is not lower than $\big(\delta W_i(s^{t+1}) \ : \ s^{t+1} \succcurlyeq s^t \big)$. Therefore, the new plan strictly preferred relative to the original plan and at the same time the new plan is in the budget constraint given the price system, which contradicts with the agent optimization condition for general equilibrium. Similarly, we can show that there is a large constant $Q_1^b < +\infty$ such that the equilibrium price of bond 1 satisfies $q_1^b(s^t) \leq Q_1^b$ for all $s^t \in \mathbb{S}^T$ in the T-truncated economy and all $T \geq 1$.

Now, let's consider the equity prices. For all $T \geq 1$, the agent i's value function at each node $s^t \in \mathbb{S}^T$ is upper bounded by $\frac{\bar{e}^{1-\psi_i^{-1}}}{1-\psi_i^{-1}} + U_i(\bar{e}, \bar{e}, \cdots)$, where $U_i(\bar{e}, \bar{e}, \cdots)$ denotes the value function for the consumption plan of consuming constant $\bar{e}$ of both commodities over the infinite-horizon tree $\mathbb{S}$. It is easy to get $U_i(\bar{e}, \bar{e}, \cdots) = \frac{1}{1-\beta} \frac{\bar{e}^{1-\psi_i^{-1}}}{1-\psi_i^{-1}}$.

Therefore, for any consumption plan

$$\left((c_i^1(s^t), c_i^2(s^t)), (c_{i,+}^1(s^t), c_{i,+}^2(s^t))\right)$$

such that $c_i^j(s^r) \leq e_j(s^r)$ for all $i, j = 1, 2$ and $s^r \succcurlyeq s^t$, we have

$$U_i\left((c_i^1(s^t), c_i^2(s^t)), (c_{i,+}^1(s^t), c_{i,+}^2(s^t))\right) \leq \frac{2}{1-\beta}\frac{\overline{e}^{1-\psi_i^{-1}}}{1-\psi^{-1}}, \text{ for } i = 1, 2.$$

Due to the assumption that $\psi_i \geq 1$, there exists large constant K such that for $i = 1, 2$,

$$\frac{K^{1-\psi_i^{-1}}}{1-\psi_i^{-1}} \geq \frac{2}{1-\beta}\frac{\overline{e}^{1-\psi_i^{-1}}}{1-\psi^{-1}}.$$

Thus, we know that for each agent $i \in I$, there exist a $K > 0$ such that for every $s^t \in \mathbb{S}$,

$$\left((c_i^1(s^t) + K, c_i^2(s^t) + K), (\underbrace{0, 0, \cdots, 0}_{|\mathbb{S}|})\right) \succ^i \left((c_i^1(s^t), c_i^2(s^t)), (W_i(s^{t+1}) \ : \ s^{t+1} \succcurlyeq s^t)\right)$$

for all current consumption satisfying $c_i^j(s^t) \leq e_j(s^t)$ for all $s^t \in \mathbb{S}$ and wealth in the beginning of the next period satisfying $W_i(s^{t+1}) \leq \sum_{i=1}^2 p_i(s^{t+1})e_i(s^{t+1}) + q_i(s^{t+1})$ for all $s^t \in \mathbb{S}$. We define a constant $Q_2 \equiv 4\max\left\{K(1+\overline{P}_1), (Q_1^b + Q_2^b)\overline{D}\overline{e}\right\}$ and show that the equity prices are uniformly bounded from above by this large constant by contradiction. Suppose that there exists a node $s^t \in \mathbb{S}^T$ such that $q_2(s^t) > Q_2$ in a T-truncated equilibrium. There must one agent whose position on equity 2 is no less than $1/2$ in a equilibrium. Without loss of generosity, we assume that the agent 1 holds no less than $1/2$ of equity 2. If agent 1 sells $1/4$ shares of equity 2 and consumes the proceeds for K units of goods 1 and K units of goods 2, then the new plan strictly preferred relative to the original plan and at the same time the new plan is in the budget constraint given the price system, which contradicts with the agent optimization condition for general equilibrium. Similarly, we can show that there is a large constant $Q_1 < +\infty$ such that the equilibrium price of equity 1 satisfies $q_1(s^t) \leq Q_1$ for all $s^t \in \mathbb{S}^T$ in an equilibrium of the T-truncated economy and all $T \geq 1$.

Therefore, we have shown that in equilibria of all T-truncated economies with $T \geq 1$ uniformly lie within a bounded rectangular area, denoted as $\mathcal{Y}^*$.

Now, we show that competitive equilibria exist for all $T \geq 1$. For the purpose of showing equilibrium existence, we change the price normalization following Kubler and Schmedders (2003). That is, instead of setting the price of consumption commodity 2 at every node $s^t \in \mathbb{S}^T$ to be one, we assume that the prices $\mathcal{P}^{\mathbb{S}^T}(s^t) := \{p_i(s^t), q_i(s^t), q_i^b(s^t)\}_{i=1,2}$ at each node s^t to lie in the unit simplex Δ, i.e. $\sum_{i=1}^2 p_i(s^t) + \sum_{i=1}^2 q_i(s^t) + \sum_{i=1}^2 q_i^b(s^t) = 1$ and every price is nonnegative. We define the truncated budget constraint to imposing the uniform bounds for the equilibria if they exist. We construct truncated budget sets in this economy by adding extra bounds on the allocations and holdings, where the truncation will not affect the equilibria under portfolio constraints. More precisely, we define the truncated budget set by, for $i = 1, 2$,

$$\overline{\mathbb{B}}_{i,\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T}) = \mathbb{A}_{i,\mathbb{S}^T} \cap \mathbb{B}_{i,\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T}),$$

where $\mathbb{A}_{i,\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T})$ imposes the uniform bounds on allocation and portfolio defined as

$$\begin{aligned}\mathbb{A}_{i,\mathbb{S}^T} \equiv \Big\{ & 0 \leq \vartheta_i^j(s^t) \leq 1, \ -\overline{D}\overline{e} \leq b_i^j(s^t) \leq \overline{D}\overline{e}\,, \\ & 0 \leq c_i^j(s^t) \leq \overline{e}, \quad \text{for} \quad s^{t-1}, s^t \in \mathbb{S}^T \ \text{ and } \ j = 1, 2 \Big\}.\end{aligned}$$

Based on the truncated budget constraint, we define the truncated demand correspondences which would be enough for our analysis. More precisely, we denote

$$\sigma_{i,T}(\mathcal{P}^{\mathbb{S}^T}) \equiv \underset{(\mathcal{C}_i^{\mathbb{S}^T}, \mathcal{A}_i^{\mathbb{S}^T}) \in \overline{\mathbb{B}}_{i,\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T})}{\operatorname{Arg\,max}} U_i(\mathcal{C}_i^{\mathbb{S}^T}). \tag{B.28}$$

Note that truncated demand exists at every price system $\mathcal{P}^{\mathbb{S}^T}$ because the equity holdings are lower bounded and the bond holdings are bounded. Absent such bounds, demand correspondence could be empty at some prices.

Denote the demand correspondence component at the node s^t to be $\sigma_{i,T}(\mathcal{P}^{\mathbb{S}^T}; s^t)$ and the aggregate excess demand at the node s^t is

$$\Sigma_T(\mathcal{P}^{\mathbb{S}^T}) \equiv \sum_{i=1}^{2} \sigma_{i,T}(\mathcal{P}^{\mathbb{S}^T}; s^t) - \left(e_1(s^t), e_2(s^t), 1, 1, 0, 0\right), \tag{B.29}$$

and define the excess demand of the T-truncated economy to be

$$\Sigma_T(\mathcal{P}^{\mathbb{S}^T}) \equiv \Pi_{s^t \in \mathbb{S}^T} \Sigma_T(\mathcal{P}^{\mathbb{S}^T}; s^t). \tag{B.30}$$

It's easy to check that $\Sigma_T(\mathcal{P}^{\mathbb{S}^T})$ is nonempty (because $\sigma_{i,T}(\mathcal{P}^{\mathbb{S}^T})$ is nonempty), compact-valued (because U_i is continuous), convex-valued (because U_i is quasi-concave) and upper hemi-continuous. Also, it is obvious that $\Sigma_T(\mathcal{P}^{\mathbb{S}^T})$ is uniformly bounded, because consumptions and asset holdings are all uniformly bounded in the truncated budget sets $\overline{\mathbb{B}}_{i,\mathbb{S}^T}(\mathcal{P}^{\mathbb{S}^T})$. That is, there exists $R > 0$ such that for all $\mathcal{P}^{\mathbb{S}^T} \in \Delta^{|\mathbb{S}^T|}$ it holds that

$$\Sigma_T(\mathcal{P}^{\mathbb{S}^T}) \subset [-R, R]^{|\mathbb{S}^T| \times I(J+E+B)}. \tag{B.31}$$

We further define the truncated space of endogenous variables

$$\left\{ \left(c_i^1, c_i^2\right)_{i=1,2}, \left(\vartheta_i^1, \vartheta_i^2, b_i^1, b_i^2\right)_{i=1,2} \right\}$$

as follows

$$\mathcal{Y}(s^t) \equiv \left\{ y \in \mathbb{R}^{I(J+E+B)} \ : \ ||y|| \leq R \right\}. \tag{B.32}$$

We first define the correspondence

$$P_T(\cdot; s^t) \ : \ \mathcal{Y}(s^t) \rightrightarrows \Delta \tag{B.33}$$

such that

$$P_T(y; s^t) \equiv \operatorname*{Arg\,max}_{\mathcal{P} \in \Delta} \mathcal{P} \cdot y. \tag{B.34}$$

It's obvious that $P_T(\cdot; s^t)$ is nonempty, compact-valued, convex-valued, and upper hemi-continuous correspondence. Now, we define the correspondence

$$F_T(\cdot,\cdot;s^t) \;:\; \Delta \times \mathcal{Y}(s^t) \rightrightarrows \Delta \times \mathcal{Y}(s^t) \tag{B.35}$$

such that

$$F_T(\mathcal{P}, y; s^t) = P_T(y; s^t) \times \Sigma_T(\mathcal{P}; s^t) \tag{B.36}$$

The product correspondence $F_T \;:\; \Delta^{|\mathbb{S}^T|} \times \Pi_{s^t \in \mathbb{S}^T} \mathcal{Y}(s^t)$ is defined as

$$F_T(\mathcal{P}^{\mathbb{S}^T}, y) = \Pi_{s^t \in \mathbb{S}^T} F_T\left(\mathcal{P}^{\mathbb{S}^T}(s^t), y(s^t); s^t\right). \tag{B.37}$$

It is obvious that F_T is nonempty, compact-valued, convex-valued, and upper hemi-continuous correspondence. Therefore, by Kakutani Theorem, we know that F_T has fixed point. We denote the collection of fixed points to be G_T.

We shall show that every fixed point $(\mathcal{P}^{\mathbb{S}^T}, y^{\mathbb{S}^T}) \in G_T$ constitutes an equilibrium for the T-truncated economy. Equivalently, we shall show that $\forall\ (\mathcal{P}^{\mathbb{S}^T}, y^{\mathbb{S}^T}) \in G_T$,

$$y^{\mathbb{S}^T} \equiv 0 \quad \text{, and} \quad \mathcal{P}^{\mathbb{S}^T} >> 0^{2}. \tag{B.38}$$

Our plan is to prove $y^{\mathbb{S}^T}(s^t) \equiv 0$ for all $s^t \in \mathbb{S}^T$ by induction first, and then show the positiveness of prices. At the initial node s^0, because of local non-satiation, we know that agent i's budget equation at node s^t is then

$$\begin{aligned} &\sum_{j=1}^{2} p_j(s^0)c_i^j(s^0) + \sum_{j=1}^{2} q_j(s^0)\vartheta_i^j(s^0) + \sum_{j=1}^{2} q_j^b(s^0)b_i^j(s^0) \\ &\qquad - p_i(s^0)w_i(s^0) - \sum_{j=1}^{2} \vartheta_i^j(s^{-1})(q_j(s^0) + p_j(s^0)d_j(s^0)) \\ &\qquad - \sum_{j=1}^{2} b_i^j(s^{-1})p_j(s^0) = 0, \end{aligned}$$

which is due to the assumption that $\sum_{i=1}^{2} \vartheta_i^j(s^{-1}) = 1$ and $\sum_{i=1}^{2} b_i^j(s^{-1}) = 0$ for $j = 1, 2$.

[2] This means that every element of $\mathcal{P}^{\mathbb{S}^T}$ is positive.

We sum over all agents and get

$$\sum_{j=1}^{2} p_j(s^0)\left[\sum_{i=1}^{2} c_i^j(s^0) - e_j(s^0)\right] + \sum_{j=1}^{2} q_j(s^0)\left[\sum_{i=1}^{2} \vartheta_i^j(s^0) - 1\right] + \sum_{j=1}^{2} q_j^b(s^0)\left[\sum_{i=1}^{2} b_i^j(s^0)\right] = 0.$$

This implies that

$$0 = \max_{\mathcal{P}\in\Delta} \mathcal{P}\cdot y^{\mathbb{S}^T}(s^0). \tag{B.39}$$

Suppose that there is positive excess demand in some market at the node s^0. Without loss of generality, we assume the largest excess demand is in the market of commodity 1. Then, the optimal solution for maximization problem (B.39) at the node s^0 would be to set $p_1(s^0) = 1$ and $p_2(s^0) = q_1(s^0) = q_2(s^0) = q_1^b(s^0) = q_2^b(s^0) = 0$. However, this leads to a positive value which contradicts with (B.39). On the other hand, suppose that there is negative excess demand in some market. Without loss of generality, we assume that the most negative excess demand is in the market of commodity 1. In this case, the price of commodity 1 must be zero, i.e. $p_1(s^0) = 0$, in order to make $\mathcal{P}^{\mathbb{S}^T}(s^0)$ to be the solution to (B.39). With zero price of commodity 1, the excess demand of commodity 1 should be positive for two agents because of monotonicity of preference, which is contradictory.

Suppose that $y^{\mathbb{S}^T}(s^t) \equiv 0$. For any node $s^{t+1} \succcurlyeq s^t$, because of local non-satiation, we know that agent i's budget equation at node s^{t+1} is then

$$\sum_{j=1}^{2} p_j(s^{t+1})c_i^j(s^{t+1}) + \sum_{j=1}^{2} q_j(s^{t+1})\vartheta_i^j(s^{t+1}) + \sum_{j=1}^{2} q_j^b(s^{t+1})b_i^j(s^{t+1}) - p_i(s^{t+1})w_i(s^{t+1}) - \sum_{j=1}^{2} \vartheta_i^j(s^t)(q_j(s^{t+1}) + p_j(s^{t+1})d_j(s^{t+1})) - \sum_{j=1}^{2} b_i^j(s^t)p_j(s^{t+1}) = 0,$$

which is due to the assumption that $\sum_{i=1}^{2} \vartheta_i^j(s^t) = 1$ and $\sum_{i=1}^{2} b_i^j(s^t) = 0$ for $j = 1, 2$. We sum over all agents and get

$$\sum_{j=1}^{2} p_j(s^{t+1}) \left[\sum_{i=1}^{2} c_i^j(s^{t+1}) - e_j(s^{t+1}) \right] + \sum_{j=1}^{2} q_j(s^{t+1}) \left[\sum_{i=1}^{2} \vartheta_i^j(s^{t+1}) - 1 \right]$$
$$+ \sum_{j=1}^{2} q_j^b(s^{t+1}) \left[\sum_{i=1}^{2} b_i^j(s^{t+1}) \right] = 0.$$

This implies that

$$0 = \max_{\mathcal{P} \in \Delta} \mathcal{P} \cdot y^{\mathbb{S}^T}(s^{t+1}). \tag{B.40}$$

Suppose that there is positive excess demand in some market at the node s^{t+1}. Without loss of generality, we assume the largest excess demand is in the market of commodity 1. Then, the optimal solution for maximization problem (B.40) at the node s^{t+1} would be to set $p_1(s^{t+1}) = 1$ and $p_2(s^{t+1}) = q_1(s^{t+1}) = q_2(s^{t+1}) = q_1^b(s^{t+1}) = q_2^b(s^{t+1}) = 0$. However, this leads to a positive value which contradicts with (B.39). On the other hand, suppose that there is negative excess demand in some market. Without loss of generality, we assume that the most negative excess demand is in the market of commodity 1. In this case, the price of commodity 1 must be zero, i.e. $p_1(s^{t+1}) = 0$, in order to make $\mathcal{P}^{\mathbb{S}^T}(s^{t+1})$ to be the solution to (B.39). With zero price of commodity 1, the excess demand of commodity 1 should be positive for two agents because of monotonicity of preference, which is contradictory. Therefore, we complete the induction step in the proof and hence we have shown that $y^{\mathbb{S}^T}(s^t) \equiv 0$ for all node $s^t \in \mathbb{S}^T$.

Because utility functions U_i are monotone, by Debreu (1959), we have the standard boundary condition which means the demand blows up when $\mathcal{P}^{\mathbb{S}^T} \to \partial \left(\Delta^{|\mathbb{S}^T|} \right)$. Thus, if there is an element of $\mathcal{P}^{\mathbb{S}^T}$ is zero, there must exist an element of $y^{\mathbb{S}^T}$ is nonzero. This is contradictory with the result we just proved above.

B.4 Two Simple Cases with Analytical Solutions

In the two simple examples, we consider the case where (1) agents have log utilities (i.e. $\gamma = \psi = 1$) and (2) agents have no portfolio constraints. The simple examples allow us to derive the analytical solution and hence exactly check our algorithm and numerical solution. The first-best consumption plan or the complete-market allocation can be characterized by the following risk-sharing problem with intratemporal budget constraints of both agents and the budget constraints of both agents being binding. That is, at each node of the tree $s^t \in \mathbb{S}$,

$$\max_{c^1_{1,t}, c^2_{1,t}, c^1_{2,t}, c^2_{2,t}} \lambda \log \left(s(c^1_{1,t})^\rho + (1-s)(c^2_{1,t})^\rho \right) + (1-\lambda) \log \left((1-s)(c^1_{2,t})^\rho + s(c^2_{2,t})^\rho \right) \tag{B.41}$$

such that

$$c^i_{1,t} + c^i_{2,t} = e_{i,t}, \quad \text{for } i = 1, 2. \tag{B.42}$$

Thus, the perfect risk sharing rule gives that the state price density (SPD) is

$$\pi_t = \frac{\lambda \beta^t}{p_{1,t}} \frac{s\rho (c^1_{1,t})^{\rho-1}}{s(c^1_{1,t})^\rho + (1-s)(c^2_{1,t})^\rho} = \frac{(1-\lambda)\beta^t}{p_{1,t}} \frac{(1-s)\rho(c^1_{2,t})^{\rho-1}}{(1-s)(c^1_{2,t})^\rho + s(c^2_{2,t})^\rho} \tag{B.43}$$

together with the Intratemporal Euler Equations

$$p_{1,t} = \frac{s}{1-s} \left(\frac{c^1_{1,t}}{c^2_{1,t}} \right)^{\rho-1} = \frac{1-s}{s} \left(\frac{c^1_{2,t}}{c^2_{2,t}} \right)^{\rho-1}, \quad \text{with } p_{2,t} \equiv 1. \tag{B.44}$$

Plug market clearing conditions into the equation above, we have

$$\left(\frac{s}{1-s} \right)^{\frac{2}{\rho-1}} = \frac{e_{1,t}/c^1_{1,t} - 1}{e_{2,t}/c^2_{1,t} - 1}. \tag{B.45}$$

Under the complete market assumption, the agents' sequential budget constraints can be summarized into static budget constraints of consumption claims.

$$\pi_t W^*_{1,t} = \mathbb{E}_t \left[\sum_{\tau \geq t} \pi_\tau \left(p_{1,\tau} c^1_{1,\tau} + c^2_{1,\tau} \right) \right] \tag{B.46}$$

and

$$\pi_t W^*_{2,t} = \mathbb{E}_t \left[\sum_{\tau \geq t} \pi_\tau \left(p_{1,\tau} c^1_{2,\tau} + c^2_{2,\tau} \right) \right] \tag{B.47}$$

From (B.43), (B.44) and (B.46), we have

$$\begin{aligned} \pi_t W^*_{1,t} &= \lambda \mathbb{E}_t \left[\sum_{\tau \geq t} \beta^\tau \frac{1}{p_{1,t}} \frac{s\rho (c^1_{1,\tau})^{\rho-1}}{s(c^1_{1,\tau})^\rho + (1-s)(c^2_{1,\tau})^\rho} (p_{1,\tau} c^1_{1,\tau} + c^2_{1,\tau}) \right] \\ &= \frac{\lambda \rho}{1-\beta}. \end{aligned}$$

And, similarly, we have

$$\begin{aligned} \pi_t W^*_{2,t} &= (1-\lambda) \mathbb{E}_t \left[\sum_{\tau \geq t} \beta^\tau \frac{1}{p_{1,t}} \frac{(1-s)\rho (c^1_{2,\tau})^{\rho-1}}{(1-s)(c^1_{2,\tau})^\rho + s(c^2_{2,\tau})^\rho} (p_{1,\tau} c^1_{2,\tau} + c^2_{2,\tau}) \right] \\ &= \frac{(1-\lambda)\rho}{1-\beta}. \end{aligned}$$

Thus, the wealth ratio is equal to the Pareto weight

$$\lambda = \frac{W^*_{1,t}}{W^*_{1,t} + W^*_{2,t}}. \tag{B.48}$$

The coincides above has a strong implication that the total wealth share λ is constant over time in the equilibrium. Also, we have

$$W^*_{1,t} = \frac{1}{1-\beta} \left(p_{1,t} c^1_{1,t} + c^2_{1,t} \right), \tag{B.49}$$

and

$$W^*_{2,t} = \frac{1}{1-\beta} \left(p_{1,t} c^1_{2,t} + c^2_{2,t} \right). \tag{B.50}$$

So far, we have only assumed log utility and complete market. The consumption policies are characterized by the consumption shares $\nu_{i,t}$ with $i = 1, 2$ where $c^i_1 = \nu_{i,t} e_i$

$$\left(\frac{s}{1-s} \right)^{\frac{2}{\rho - 1}} = \frac{1/\nu_{1,t} - 1}{1/\nu_{2,t} - 1} \tag{B.51}$$

and

$$\lambda = \frac{\nu_{1,t} e_{1,t} \frac{s}{1-s} \left(\frac{\nu_{1,t}}{\nu_{2,t}}\right)^{\rho-1} \left(\frac{e_{1,t}}{e_{2,t}}\right)^{\rho-1} + \nu_{2,t} e_{2,t}}{e_{1,t} \frac{s}{1-s} \left(\frac{\nu_{1,t}}{\nu_{2,t}}\right)^{\rho-1} \left(\frac{e_{1,t}}{e_{2,t}}\right)^{\rho-1} + e_{2,t}} = \frac{\nu_{1,t} \frac{s}{1-s} \left(\frac{\nu_{1,t}}{\nu_{2,t}}\right)^{\rho-1} \left(\frac{e_{1,t}}{e_{2,t}}\right)^{\rho} + \nu_{2,t}}{\frac{s}{1-s} \left(\frac{\nu_{1,t}}{\nu_{2,t}}\right)^{\rho-1} \left(\frac{e_{1,t}}{e_{2,t}}\right)^{\rho} + 1}. \tag{B.52}$$

Thus, in general, the first-best consumption plans (i.e. $\nu_{i,t}$) depend on the parameters ρ and s, as well as the total wealth share λ and the output ratio $e_{2,t}/e_{1,t}$. We consider two special cases where the consumption shares $\nu_{i,t}$ are constant over time and hence facilitates analytical solutions. One example is the well-known Cole-Obstfeld Economy[3], and the other is the Symmetric Economy.

B.4.1 Cole and Obstfeld Economy

Based on the two assumptions in the beginning of Appendix B.4, we further assume that the aggregator is Cobb-Douglas (i.e. $\rho = 0$) and the equity leverage ratio coefficient is zero (i.e. $\varrho = 0$) in our model. This economy is effectively the Cole-Obstfeld economy. The solution to equations (B.51) and (B.52) are simply

$$\nu_{1,t} \equiv \nu_1 \equiv \frac{1}{1 + \frac{1-s}{s}\frac{1-\lambda}{\lambda}} \quad \text{and} \quad \nu_{2,t} \equiv \nu_2 \equiv \frac{1}{1 + \frac{s}{1-s}\frac{1-\lambda}{\lambda}}. \tag{B.53}$$

[3]In their classic analysis of the irrelevance of asset markets for international risk sharing, Cole and Obstfeld (1991) show that in an open economy with two differentiated goods, agents with logarithmic preferences and Cobb-Douglas aggregator, and no trade costs, the central-plannerŠs allocation can be achieved even without trade in asset markets. This occurs because the endogenous response of the Term of Trade to supply shocks to the two goods is sufficient to implement the international wealth transfers that support the central plannerŠs consumption allocation. As is well known, the Cole and Obstfeld equilibrium features: perfectly correlated Home and Foreign stock markets, symmetric aggregate stock market portfolio holdings, zero holdings of risk-free bonds, equal consumption state by state, zero NX, and indeterminate NFA and CA. The exchange rate is either constant ($s = 0.5$) or positively related to the Term of Trade ($s > 0.5$).

Thus, the optimal consumptions are

$$c^1_{1,t} = \nu_1 e_{1,t},\ c^2_{1,t} = \nu_2 e_{2,t}, \tag{B.54}$$

$$c^1_{2,t} = (1-\nu_1)e_{1,t},\ c^2_{2,t} = (1-\nu_2)e_{2,t}. \tag{B.55}$$

The Term of Trade is

$$p_{1,t} = A\frac{e_{2,t}}{e_{1,t}}, \quad \text{with} \quad A \equiv \frac{s}{1-s}\frac{\nu_2}{\nu_1}, \tag{B.56}$$

and the real exchange rate is

$$Q_t \equiv p_{1,t}^{2s-1}. \tag{B.57}$$

Because the wealth share is constant, based on the Proposition 4, the Euler equation for equity prices are

$$\begin{aligned}
&e_2(s^t)^{-1}\left[q_1(s_t)e(s^t)\right] \\
&= \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) e_2(s^{t+1})^{-1}\left[q_1(s_{t+1})e(s^{t+1}) + p_1(s_{t+1})\bar{d}e_1(s^{t+1})\right] \\
&= \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) e_2(s^{t+1})^{-1}\left[q_1(s_{t+1})e(s^{t+1}) + A\bar{d}e_2(s^{t+1})\right]
\end{aligned} \tag{B.58}$$

and

$$\begin{aligned}
&e_2(s^t)^{-1}\left[q_2(s_t)e(s^t)\right] \\
&= \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) e_2(s^{t+1})^{-1}\left[q_2(s_{t+1})e(s^{t+1}) + \bar{d}e_2(s^{t+1})\right]
\end{aligned} \tag{B.59}$$

Because two equities have perfectly correlated dividend flows which are $\{A\bar{d}e_{2,t}\}_{t\geq 0}$ and $\{\bar{d}e_{2,t}\}_{t\geq 0}$, then it is straightforward to know that $q_1(s_t) \equiv Aq_2(s_t)$. Therefore, the US and ROW equities are perfectly correlated, and hence the equity holdings are indeterminate. Thus, one possible set of portfolio holdings are

$$\vartheta^1_{1,t} \equiv \frac{\nu_1 - (1-\bar{d})}{\bar{d}}, \quad \vartheta^2_{1,t} \equiv \frac{\nu_2}{\bar{d}}, \quad b^1_{1,t} \equiv b^2_{1,t} \equiv 0, \tag{B.60}$$

$$\vartheta_{2,t}^1 \equiv \frac{1-\nu_1}{\bar{d}}, \quad \vartheta_{2,t}^2 \equiv \frac{1-\nu_2-(1-\bar{d})}{\bar{d}}, \quad b_{2,t}^1 \equiv b_{2,t}^2 \equiv 0. \tag{B.61}$$

The bonds' prices are, for $i = 1, 2$,

$$e_2(s^t)^{-1}\left[q_i^b(s_t)e(s^t)\right] = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) e_2(s^{t+1})^{-1}\left[p_i(s_{t+1})e(s^{t+1})\right]. \tag{B.62}$$

Based on the structures of endowment processes specified in Section 4.2.1, the Euler equations for asset prices in (B.58), (B.59) and (B.62) can re-written as

$$x_{2,t}^{-1} q_1(s_t) = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1})\left[x_{2,t+1}^{-1} q_1(s_{t+1}) + A\bar{d}\right] \tag{B.63}$$

$$x_{2,t}^{-1} q_2(s_t) = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1})\left[x_{2,t+1}^{-1} q_2(s_{t+1}) + \bar{d}\right] \tag{B.64}$$

and

$$x_{2,t}^{-1} q_i^b(s_t) = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) x_{2,t+1}^{-1} p_i(s_{t+1}). \tag{B.65}$$

The equity prices are solution of the S by S linear equations, while the bond prices can be directly calculated. We can see that the normalized asset prices are independent of global component and the disaster probability.

B.4.2 Symmetric Economy

Based on the two assumptions in the beginning of Appendix B.4, we further assume that the consumption share coefficient (i.e. $s = 0.5$) and the equity leverage ratio coefficient is zero (i.e. $\varrho = 0$) in our model. The solution to equations (B.51) and (B.52) are simply

$$\nu_{1,t} \equiv \nu_{2,t} \equiv \lambda. \tag{B.66}$$

Thus, the optimal consumptions are

$$c_{1,t}^1 = \lambda e_{1,t}, \ c_{1,t}^2 = \lambda e_{2,t}, \tag{B.67}$$

$$c^1_{2,t} = (1-\lambda)e_{1,t},\ c^2_{2,t} = (1-\lambda)e_{2,t}. \tag{B.68}$$

The Term of Trade is

$$p_{1,t} = \left(\frac{e_{1,t}}{e_{2,t}}\right)^{\rho-1}, \tag{B.69}$$

and the real exchange rate is

$$Q_t \equiv 1. \tag{B.70}$$

Because the wealth share is constant, the Euler equation for equity prices are

$$\begin{aligned}
&\frac{e_2(s^t)^{\rho-1}}{e_1(s^t)^\rho + e_2(s^t)^\rho}\left[q_1(s_t)e(s^t)\right] \\
&= \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{e_2(s^{t+1})^{\rho-1}}{e_1(s^{t+1})^\rho + e_2(s^{t+1})^\rho}\left\{\left[q_1(s_{t+1})e(s^{t+1})\right] + p_1(s_{t+1})\bar{d}e_1(s^{t+1})\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{e_2(s^t)^{\rho-1}}{e_1(s^t)^\rho + e_2(s^t)^\rho}\left[q_2(s_t)e(s^t)\right] \\
&= \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{e_2(s^{t+1})^{\rho-1}}{e_1(s^{t+1})^\rho + e^2(s^{t+1})^\rho}\left\{\left[q_2(s_{t+1})e(s^{t+1})\right] + \bar{d}e_2(s^{t+1})\right\}
\end{aligned}$$

The Inter-temporal Euler equations above can be re-written as

$$\begin{aligned}
&\frac{q_1(s_t)e(s^t)}{p_1(s_t)e_1(s^t) + e_2(s^t)} \\
&= \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{q_1(s_{t+1})e(s^{t+1}) + p_1(s_{t+1})\bar{d}e_1(s^{t+1})}{p_1(s_{t+1})e_1(s^{t+1}) + e_2(s^{t+1})}
\end{aligned} \tag{B.71}$$

and

$$\frac{q_2(s_t)e(s^t)}{p_1(s_t)e_1(s^t) + e_2(s^t)} = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{q_2(s_{t+1})e(s^{t+1}) + \bar{d}e_2(s^{t+1})}{p_1(s_{t+1})e_1(s^{t+1}) + e_2(s^{t+1})}. \tag{B.72}$$

The equilibrium portfolio holdings are

$$\vartheta^1_{1,t} \equiv \frac{\lambda - (1-\bar{d})}{\bar{d}}, \quad \vartheta^2_{1,t} \equiv \frac{\lambda}{\bar{d}}, \quad b^1_{1,t} \equiv b^2_{1,t} \equiv 0, \tag{B.73}$$

$$\vartheta^1_{2,t} \equiv \frac{1-\lambda}{\bar{d}}, \quad \vartheta^2_{2,t} \equiv \frac{1-\lambda-(1-\bar{d})}{\bar{d}}, \quad b^1_{2,t} \equiv b^2_{2,t} \equiv 0. \tag{B.74}$$

The bonds' prices Euler equations are, for $i = 1, 2$,

$$\frac{q^b_i(s_t)e(s^t)}{p_1(s_t)e_1(s^t) + e_2(s^t)} = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{p_i(s_{t+1})e(s^t)}{p_1(s_{t+1})e_1(s^{t+1}) + e_2(s^{t+1})}. \tag{B.75}$$

Based on the structures of endowment processes specified in Section 4.2.1, the Euler equations for asset prices in (B.71), (B.72) and (B.75) can re-written as

$$\frac{q_1(s_t)}{p_1(s_t)x_{1,t} + x_{2,t}} = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{q_1(s_{t+1}) + p_1(s_{t+1})\bar{d}x_{1,t+1}}{p_1(s_{t+1})x_{1,t+1} + x_{2,t+1}}, \tag{B.76}$$

$$\frac{q_2(s_t)}{p_1(s_t)x_{1,t} + x_{2,t}} = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{q_2(s_{t+1}) + p_2(s_{t+1})\bar{d}x_{2,t+1}}{p_1(s_{t+1})x_{1,t+1} + x_{2,t+1}}, \tag{B.77}$$

and for $i = 1, 2$

$$\frac{q^b_i(s_t)}{p_1(s_t)x_{1,t} + x_{2,t}} = \beta \sum_{s_{t+1}\in\mathcal{S}} P(s_t, s_{t+1}) \frac{p_i(s_{t+1})\zeta(s_{t+1})^{-1}}{p_1(s_{t+1})x_{1,t+1} + x_{2,t+1}}, \tag{B.78}$$

where

$$p_1(s_t) = \left(\frac{x_{1,t}}{x_{2,t}}\right)^{\rho-1}. \tag{B.79}$$

So, the equity prices can be solved from the S by S linear system, and the bond prices can be directly calculated. We can see that the normalized asset prices are independent of global component and the disaster probability.

How about the value function $U_i(\lambda, s_t, e_t)$? From the definition, we know that

$$\begin{aligned} U_1(\lambda, s_t, e_t) &= \mathbb{E}_t \left\{ \sum_{\tau \geq t} \beta^{\tau-t} \frac{1}{\rho} \log \left[s(c^1_1(\lambda, s_\tau, e_\tau))^\rho + (1-s)(c^2_1(\lambda, s_\tau, e_\tau))^\rho \right] \right\} \\ &= \frac{1}{1-\beta}\log(\lambda) + \frac{1}{1-\beta}\log(e_t) + F_1(s_t), \end{aligned}$$

where

$$F_1(s_t) = \sum_{\tau > t} \beta^{\tau-t} \mathbb{E}_t \left[\log \left(\frac{e_\tau}{e_t} \right) \right] + \frac{1}{\rho} \sum_{\tau \geq t} \beta^{\tau-t} \mathbb{E}_t \left[\log(s x_{1,t}^\rho + (1-s) x_{2,t}^\rho) \right].$$

Plugging back into the recursive formulation of the value function, it follows that, for $s_t, s_{t+1} \in \mathcal{S}$,

$$\begin{aligned} F_1(s_t) &= \frac{1}{\rho} \log(s x_{1,t}^\rho + (1-s) x_{2,t}^\rho) \\ &\quad + \beta \sum_{s_{t+1} \in \mathcal{S}} P(s_t, s_{t+1}) \left[F_1(s_{t+1}) + \frac{1}{1-\beta} \log(\zeta_{t+1}) \right]. \end{aligned} \tag{B.80}$$

Thus, the function $F_1(s)$ can be solved out from the S by S linear system.

$$\begin{aligned} U_2(\lambda, s_t, e_t) &= \mathbb{E}_t \left\{ \sum_{\tau \geq t} \beta^{\tau-t} \frac{1}{\rho} \log \left[(1-s)(c_2^1(\lambda, s_\tau, e_\tau))^\rho + s(c_2^2(\lambda, s_\tau, e_\tau))^\rho \right] \right\} \\ &= \frac{1}{1-\beta} \log(1-\lambda) + \frac{1}{1-\beta} \log(e_t) + F_2(s_t), \end{aligned}$$

where

$$\begin{aligned} F_2(s_t) &= \sum_{\tau > t} \beta^{\tau-t} \mathbb{E}_t \left[\log \left(\frac{e_\tau}{e_t} \right) \right] \\ &\quad + \frac{1}{\rho} \sum_{\tau \geq t} \beta^{\tau-t} \mathbb{E}_t \left[\log((1-s) x_{1,t}^\rho + s x_{2,t}^\rho) \right]. \end{aligned}$$

Plugging back into the recursive formulation of the value function, it follows that, for $s_t, s_{t+1} \in \mathcal{S}$,

$$\begin{aligned} F_2(s_t) &= \frac{1}{\rho} \log((1-s) x_{1,t}^\rho + s x_{2,t}^\rho) \\ &\quad + \beta \sum_{s_{t+1} \in \mathcal{S}} P(s_t, s_{t+1}) \left[F_2(s_{t+1}) + \frac{1}{1-\beta} \log(\zeta_{t+1}) \right]. \end{aligned} \tag{B.81}$$

Thus, the function $F_2(s)$ can be solved out from the S by S linear system.

B.4.3 Financial Wealth Share as Endogenous State Variable

The total wealth (including the present value of labor income) distribution serves as a natural state variable in complete market to characterize the equilibrium, as we have shown above where the first-best allocation can be achieved even in an incomplete market and more generally described in standard textbooks of complete market, such as Magill and Quinzii (2002).

However, when the market is incomplete, a natural endogenous state variable characterizing the equilibrium would be the financial wealth (excluding the present value of non-tradable cash flows) distribution, instead of the total wealth distribution. In our simple examples above, the financial wealth share can be expressed in terms of total wealth share, asset prices and endowments.

B.5 Numerical Solution

The algorithm is a time iteration algorithm.

Step 0: Select an error tolerance ϵ for the stopping criterion and a grids discretizing the financial wealth ratio endogenous state variable w on $[0, 1]$. Denote the grids as $0 < \overline{w}_1 < \cdots < \overline{w}_N < 1$. We also choose the initial guess for Π and Ω as $\hat{\Pi}_0$ and $\hat{\Omega}_0$.

Step 1: For $k = 1, \cdots, K$, given piecewise-linear (or more general interpolation methods) functions $\hat{\Pi}_{k-1}$ and $\hat{\Omega}_{k-1}$, we solve out the policy functions and transition map $\hat{\Pi}_k$ and $\hat{\Omega}_k$. This is key part of time iteration algorithm.

(1) Solve out $\hat{\Pi}_k$ given $\hat{\Pi}_{k-1}$ and $\hat{\Omega}_{k-1}$:

(a) Given a grid point $\overline{w}_i$, $s \in \mathcal{S}$ and $\tilde{s} \in \mathcal{S}$, calculate $\tilde{w}$ based on $\hat{\Omega}_{k-1}$:

$$\tilde{w} = \hat{\Omega}_{k-1}(\overline{w}_i, s, \tilde{s})$$

(b) Interpolate/Extroplate the policy function $\hat{\Pi}_{k-1}$ at $\tilde{w}$ or you can also say evaluating the interpolated function $\hat{\Pi}_{k-1}$ at $\tilde{w}$.

(c) Solve out $\hat{\Pi}_k$ as current period's policies based on taking the next period's policies as those interpolated above.

(2) Update $\hat{\Omega}_k$ given $\hat{\Pi}_k$ and $\hat{\Omega}_{k-1}$:

(a) Given a grid point $\overline{w}_i$, $s \in \mathcal{S}$ and $\tilde{s} \in \mathcal{S}$, calculate $\tilde{w}$ based on $\hat{\Omega}_{k-1}$:

$$\tilde{w} = \hat{\Omega}_{k-1}(\overline{w}_i, s, \tilde{s})$$

(b) Calculate $\hat{\Omega}_k(\overline{w}_i, s, \tilde{s})$:

$$\hat{\Omega}_k(\overline{w}_i, s, \tilde{s}) = \frac{p_1(\tilde{w}, \tilde{s})w_1(\tilde{s}) + \sum_{j=1}^{2} \vartheta_1^j(\overline{w}_i, s)(q_j(\tilde{w}, \tilde{s}) + p_j(\tilde{w}, \tilde{s})d_j(\tilde{w}, \tilde{s})) + \sum_{j=1}^{2} b_1^j(\overline{w}_i, s)p_j(\tilde{w}, \tilde{s})}{\sum_{j=1}^{2} p_j(\tilde{w}, \tilde{s})e_j(\tilde{s}) + q_j(\tilde{w}, \tilde{s})}$$

Step 2: Check stropping criterion. If

$$\max_{\overline{w} \in \overline{\mathcal{W}}, z, z' \in \mathcal{Z}} \left\{ |\hat{\rho}_k(w, z) - \hat{\rho}_{k-1}(w, z)|, |\hat{\Omega}_k(w, z, z') - \hat{\Omega}_{k-1}(w, z, z')| \right\} < \epsilon$$

then go to Step 3. Otherwise, $k = k + 1$ and go to Step 1.

Step 3: The algorithm terminates. Set

$$\hat{\rho} = \hat{\rho}_k \text{ and } \hat{\Omega} = \hat{\Omega}_k.$$

B.6 Additional Simulation Results

We simulate an economy for 100 years at a quarterly frequency. This simulation is then repeated 30,000 times. In each simulation the model is hit with a disaster probability shock in the first quarter of year 96. The disaster probability jumps to a large value of $p_t = p_H \equiv 6\%$ and then decays back to average level according to its average convergence speed. The half life of the temporary risk shock is 10 quarters according to our calibration. All other shocks are randomly drawn. This experiment generates the average impact of a risk shock, where the average is taken over the distribution of aggregate and country-specific shocks.

Panel A of figure B-1 reports the path of the disaster probability p_t. The shock generates a sharp spike in the average disaster probability p_t across the 30,000 simulations, which dies out with a half-life of about 6 months.

Figure B-2 reports the impulse response functions of domestic and foreign asset pricing moments to a large buy temporary shock in the disaster probability.

The capital stock, scaled by GDP, is defined as:

$$CS_{i,t}^{j} \equiv \frac{\vartheta_{i,t}^{j} q_{j,t}}{Y_{i,t}}.$$

Panel B of Figure B-3 reports the response of the U.S. holdings of U.S. and ROW equity. The increase in the disaster probability leads to a sharp decreases the price-dividend and consumption-wealth ratios, and a sharp increase in equity risk premia. The U.S. capital stock of U.S. equity decreases sharply while the U.S. capital stock of the ROW equity increases slightly.

The change in capital stocks is due to a change in the value of the existing holdings and a change in the holdings, reflected in the international capital flows. We defined the capital flows, scaled by GDP, as equal to:

$$CF_{i,t+1}^{j} \equiv \frac{\vartheta_{i,t+1}^{j} q_{j,t+1} - \vartheta_{i,t}^{j} q_{j,t+1}}{Y_{i,t+1}}.$$

Panel C of Figure B-3 reports the gross and net equity flows from the perspective of the U.S. Inflows increase and outflows decrease immediately in response to the disaster probability shock. The net inflows are positive but small in comparison with the gross flows. The inflows and outflows immediately reverse after the shock, but their subsequent size are an order of magnitude smaller than their initial responses. Panel

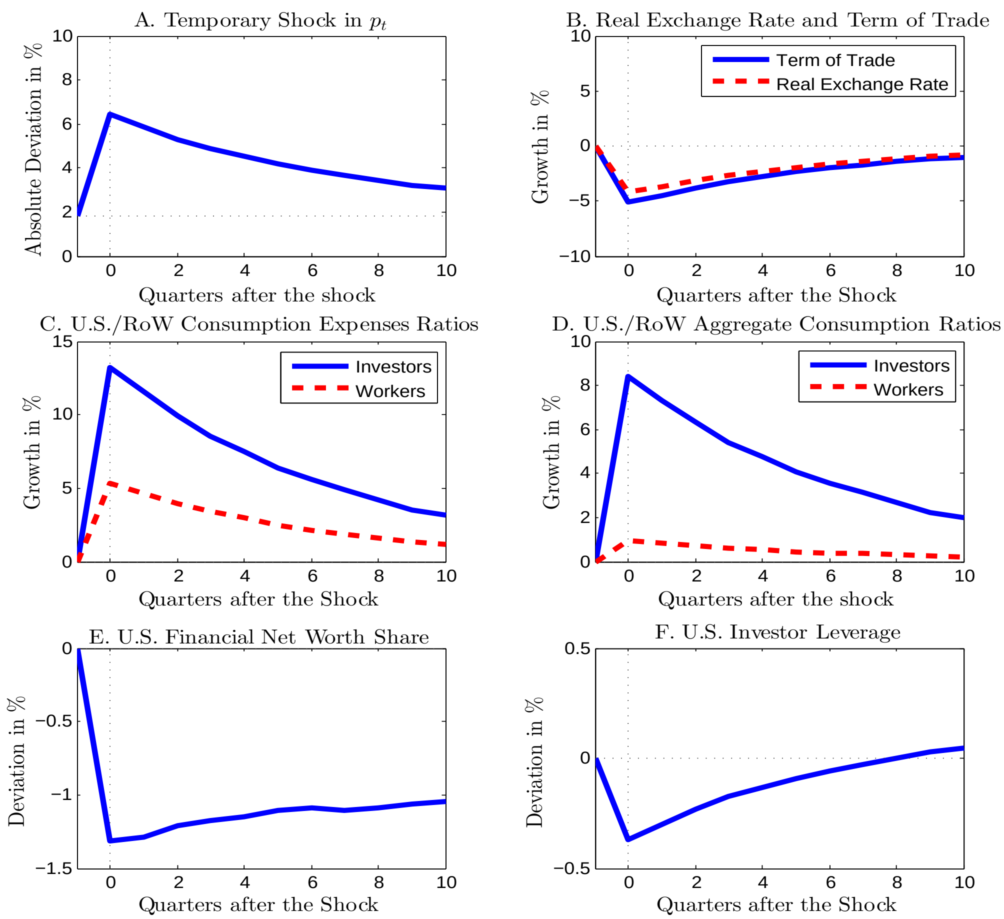

Figure B-1: Impulse-Response Functions of Basic Quantities to a Disaster Probability Shock in the Model.

D of Figure B-1 shows that the net debt flow is almost immune to the large risk shocks.

The change in capital stocks can be decomposed into a capital flow and a valuation component:

$$
\begin{aligned}
\Delta CS^j_{i,t+1} &= \frac{\vartheta^j_{i,t+1} q_{j,t+1}}{Y_{i,t+1}} - \frac{\vartheta^j_{i,t} q_{j,t}}{Y_{i,t}}, \\
\Delta CS^j_{i,t+1} &= \underbrace{CS^j_{i,t}\left(\frac{R^j_{i,t+1}}{R^Y_{i,t+1}} - 1\right)}_{\text{Valuation effects}} + \underbrace{CS^j_{i,t+1} - CS^j_{i,t}\frac{R^j_{i,t+1}}{R^Y_{i,t+1}}}_{\text{Flow effects}},
\end{aligned}
$$

where $R^j_{i,t+1} \equiv \frac{q_{j,t+1}}{q_{j,t}}$ and $R^Y_{i,t+1} \equiv \frac{Y_{i,t+1}}{Y_{i,t}}$. The first term above captures the impact of asset prices on the change of capital stocks. The second term corresponds to the capital flows, as defined previously:

$$
CS^j_{i,t+1} - CS^j_{i,t}\frac{R^j_{i,t+1}}{R^Y_{i,t+1}} = \frac{\vartheta^j_{i,t+1} q_{j,t+1} - \vartheta^j_{i,t} q_{j,t+1}}{Y_{i,t+1}} = CF^j_{i,t+1}.
$$

Figure B-3 reports the impulse response functions of capital stocks to a large shock in the disaster probability. The figure reports the dynamics of capital stocks levels, capital stock changes and capital flows.

Figure B-4 reports the dynamics of U.S. domestic and foreign holdings in response to a shock on the disaster probability. Panel B corresponds to the U.S. holdings of U.S. equity. Panel C corresponds to the U.S. holdings of ROW equity. And, Panel D corresponds to the U.S. holdings of international bond. In each panel, the stock changes are decomposed into their valuation and flow components.

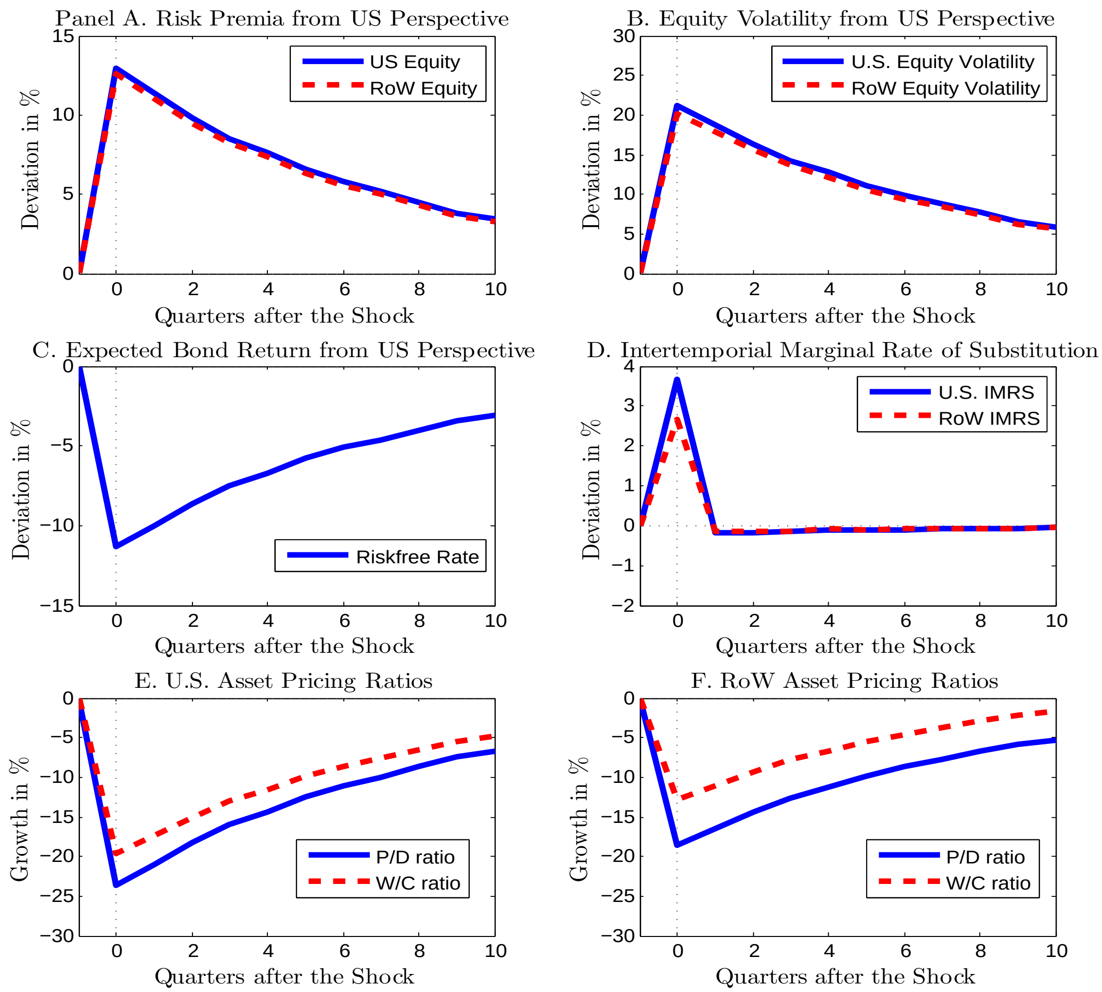

Figure B-2: The simulation of pricing moments with a large temporary risk shock.

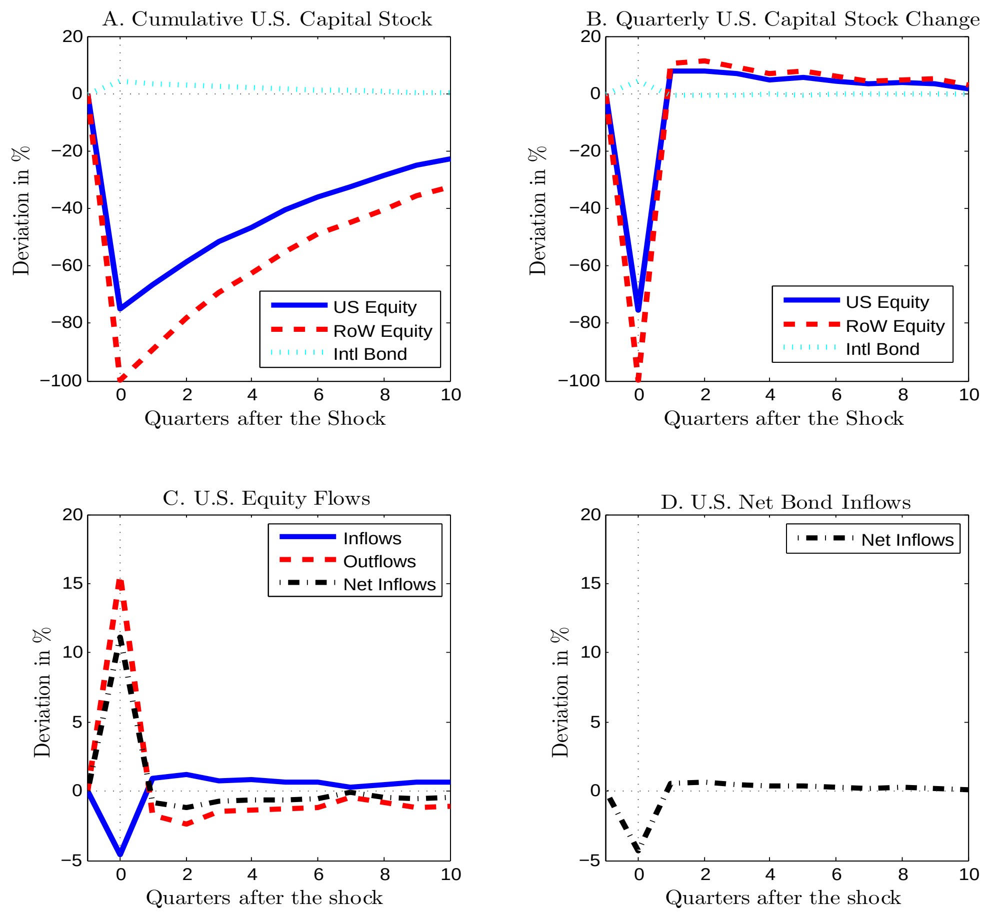

Figure B-3: The simulation of capital stocks and on capital stocks with a large risk shock.

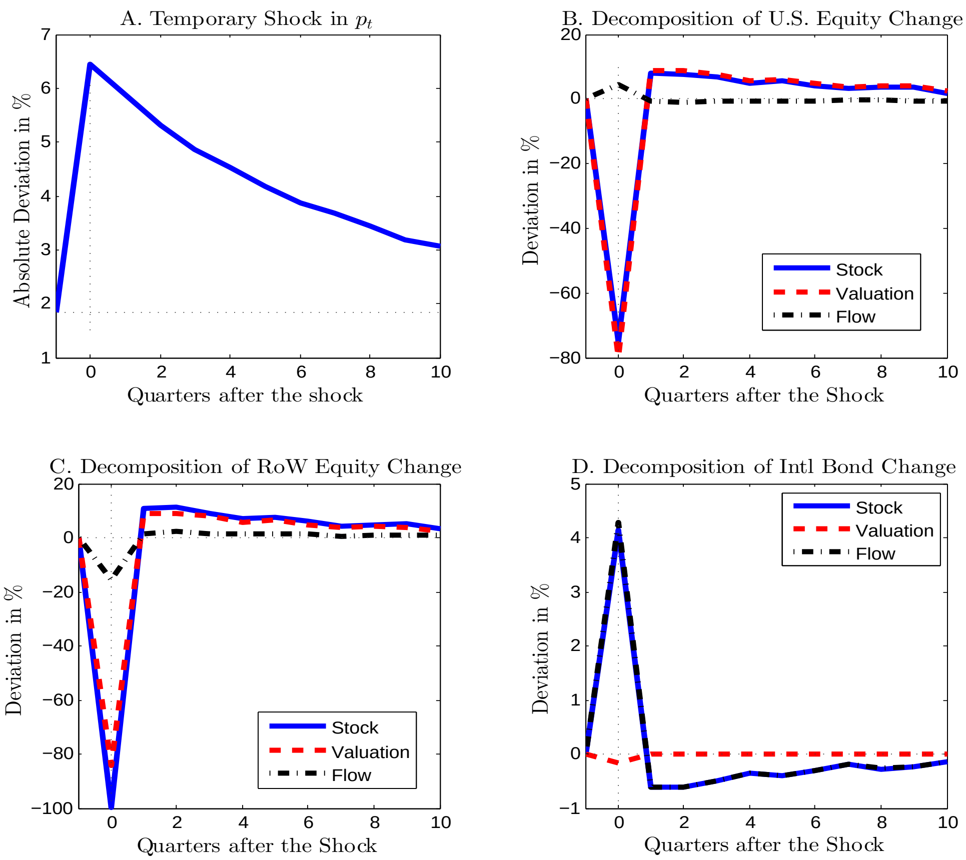

Figure B-4: Impulse-Response Functions of U.S. Capital Stocks Decompositions to a Large but Temporary Disaster Probability Shock.

Bibliography

Abel, A. B., 1983, "Optimal Investment under Uncertainty," *American Economic Review*, 73, 228–33.

Abel, A. B., A. Dixit, J. C. Eberly, and R. S. Pindyck, 1996, "Options, the Value of Capital, and Investment," *The Quarterly Journal of Economics*, 111, 753–77.

Adrian, T., and N. Boyarchenko, 2012, "Intermediary leverage cycles and financial stability," Discussion paper.

Ahmed, S., S. E. Curcuru, F. E. Warnock, and A. Zlate, 2015, "The Two Components of International Portfolio Flows," Working paper, Darden Graduate School of Business, University of Virginia.

Ai, H., M. M. Croce, and K. Li, 2012, "Toward a Quantitative General Equilibrium Asset Pricing Model with Intangible Capital," *Review of Financial Studies*, 26, 491–530.

Ai, H., and D. Kiku, 2013, "Growth to value: Option exercise and the cross section of equity returns," *Journal of Financial Economics*, 107, 325–349.

Ai, H., and D. Kiku, 2015, "Volatility Risks and Growth Options," *Management Sciences*.

Akaike, H., 1973, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory (Tsahkadsor, 1971)* . pp. 267–281, Akadémiai Kiadó, Budapest.

Allen, F., and D. Gale, 1994, "Limited Market Participation and Volatility of Asset Prices," *American Economic Review*, 84, 933–55.

Alvarez, F., A. Atkeson, and P. J. Kehoe, 2002, "Money, Interest Rates and Exchange Rates with Endogenously Segmented Markets," *Journal of Political Economy*, 110(1), 73–112.

Alvarez, F., A. Atkeson, and P. J. Kehoe, 2009, "Time-Varying Risk, Interest Rates and Exchange Rates in General Equilibrium," *Review of Economic Studies*, 76, 851 – 878.

Alvarez, F., and U. J. Jermann, 2000, "Efficiency, Equilibrium, and Asset Pricing with Risk of Default," *Econometrica*, 68, 775–798.

Alvarez, F., and U. J. Jermann, 2001, "Quantitative Asset Pricing Implications of Endogenous Solvency Constraints," *Review of Financial Studies*, 14, 1117–51.

Alvarez, F., and U. J. Jermann, 2005, "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73, 1977–2016.

Anderson, E. W., 2005, "The Dynamics of Risk-Sensitive Allocations," *Journal of Economic Theory*, 125, 93–150.

Anderson, E. W., L. P. Hansen, and T. J. Sargent, 2003, "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1, 68–123.

Ando, T., 2007, "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," *Biometrika*, 94, 443–458.

Andrews, D. W. K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–58.

Andrews, D. W. K., and J. C. Monahan, 1992, "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–66.

Ang, A., and G. Bekaert, 2002, "International Asset Allocation With Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Angeletos, G.-M., 2007, "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, 10, 1–30.

Arellano, C., Y. Bai, and P. Kehoe, 2011, "Financial Markets and Fluctuations in Uncertainty," Discussion paper.

Arellano, M., and S. Bond, 1991, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–97.

Bacchetta, P., and E. V. Wincoop, 2006, "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?," *American Economic Review*, 96, 552–576.

Bachmann, R., and C. Bayer, 2014, "Investment Dispersion and the Business Cycle," *American Economic Review*, 104, 1392–1416.

Bachmann, R., R. J. Caballero, and E. Engel, 2006, "Lumpy Investment in Dynamic General Equilibrium," Cowles foundation discussion papers, Cowles Foundation for Research in Economics, Yale University.

Bachmann, R., R. J. Caballero, and E. M. R. A. Engel, 2013, "Aggregate Implications of Lumpy Investment: New Evidence and a DSGE Model," *American Economic Journal: Macroeconomics*, 5, 29–67.

Backus, D. K., P. J. Kehoe, and F. E. Kydland, 1992, "International Real Business Cycles," *Journal of Political Economy*, 100, 745–775.

Backus, D. K., and G. Smith, 1993, "Consumption and Real Exchange Rates in Dynamic Economies with Non-Traded Goods," *Journal of International Economics*, 35, 297–316.

Bansal, R., D. Kiku, I. Shaliastovich, and A. Yaron, 2014, "Volatility, the Macroeconomy and Asset Prices," *Journal of Finance*, Forthcoming.

Bansal, R., D. Kiku, and A. Yaron, 2012a, "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 1, 183–221.

Bansal, R., D. Kiku, and A. Yaron, 2012b, "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 1, 183–221.

Bansal, R., and A. Yaron, 2004a, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Bansal, R., and A. Yaron, 2004b, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Bar-Ilan, A., and W. C. Strange, 1996, "Investment Lags," *American Economic Review*, 86, 610–22.

Barro, R., 2006a, "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121, 823–866.

Barro, R. J., 2006b, "Rare disasters and asset markets in the twentieth century," *The Quarter Journal of Economics*, 121, 823–866.

Barro, R. J., and J. F. Ursúa, 2008, "Macroeconomic Crises since 1870," *Brookings Papers on Economic Activity*, 39, 255–350.

Barron, A. R., 1989, "Uniformly powerful goodness of fit tests," *Ann. Statist.*, 17, 107–124.

Basak, S., and D. Cuoco, 1998, "An Equilibrium Model with Restricted Stock Market Participation," *Review of Financial Studies*, 11, 309–41.

Baxter, M., and U. Jermann, 1997, "The International Diversification Puzzle Is Worse Than You Think," *The American Economic Review*, 87(1), 179–180.

Benigno, P., and S. Nisticò, 2012, "International Portfolio Allocation under Model Uncertainty," *American Economic Journal: Macroeconomics*, 4, 144–89.

Bernanke, B., and M. Gertler, 1989, "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 79, 14–31.

Bernanke, B. S., 1983, "Irreversibility, Uncertainty, and Cyclical Investment," *The Quarterly Journal of Economics*, 98, 85–106.

Bernanke, B. S., M. Gertler, and S. Gilchrist, 1999, "The financial accelerator in a quantitative business cycle framework," in J. B. Taylor, and M. Woodford (ed.), *Handbook of Macroeconomics*, vol. 1 of *Handbook of Macroeconomics*, . chap. 21, pp. 1341–1393, Elsevier.

Bhamra, H. S., and R. Uppal, 2006, "The role of risk aversion and intertemporal substitution in dynamic consumption-portfolio choice with recursive utility," *Journal of Economic Dynamics and Control*, 30, 967 – 991.

Biais, B., T. Mariotti, G. Plantin, and J.-C. Rochet, 2007, "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications," *The Review of Economic Studies*, 74, 345–390.

Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve, 2010, "Large Risks, Limited Liability, and Dynamic Moral Hazard," *Econometrica*, 78, 73–118.

Black, F., and M. S. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–54.

Bloom, N., 2009, "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. Terry, 2013, "Really Uncertain Business Cycles," CEP Discussion Papers 1195, Centre for Economic Performance, LSE.

Bluedorn, J., R. Duttagupta, J. Guajardo, and P. Topalova, 2013, "Capital Flows are Fickle: Anytime, Anywhere," Working Paper, I.M.F.

Blum, M. G. B., 2010, "Approximate Bayesian computation: a nonparametric perspective," *J. Amer. Statist. Assoc.*, 105, 1178–1187, With supplementary material available online.

Bolton, P., H. Chen, and N. Wang, 2011, "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management," *Journal of Finance*, 66, 1545–1578.

Bolton, P., H. Chen, and N. Wang, 2013, "Market timing, investment, and risk management," *Journal of Financial Economics*, 109, 40–62.

Bolton, P., N. Wang, and J. Yang, 2013, "Liquidity Hoarding and Investment under Uncertainty," Working papers.

Box, G. E. P., 1976, "Science and statistics," *J. Amer. Statist. Assoc.*, 71, 791–799.

Bradley, R. C., 2005, "Basic properties of strong mixing conditions. A survey and some open questions," *Probab. Surv.*, 2, 107–144, Update of, and a supplement to, the 1986 original.

Brave, S., and R. A. Butters, 2011, "Monitoring financial stability: a financial conditions index approach," *Economic Perspectives*, pp. 22–43.

Brumm, J., M. Grill, F. Kubler, and K. H. Schmedders, 2013, "Margin Regulation and Volatility," Swiss finance institute research paper, Swiss Finance Institute.

Brunnermeier, M. K., and Y. Sannikov, 2014, "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 104, 379–421.

Bruno, V., and H. S. Shin, 2014, "Cross-Border Banking and Global Liquidity," *The Review of Economic Studies*, p. rdu042.

Bundick, B., and S. Basu, 2014, "Uncertainty shocks in a model of effective demand," Research working paper, Federal Reserve Bank of Kansas City.

Caballero, R., 1991, "On the Sign of the Investment-Uncertainty Relationship," *American Economic Review*, 81, 279–88.

Caballero, R. J., E. Farhi, and P.-O. Gourinchas, 2008, "An Equilibrium Model of Global Imbalances and Low Interest Rates," *American Economic Review*, 98, 358–93.

Cadenillas, A., J. Cvitanic, and F. Zapatero, 2007, "Optimal risk-sharing with effort and project choice," *Journal of Economic Theory*, 133, 403–440.

Campbell, J. Y., 1993, "Intertemporal Asset Pricing without Consumption Data," *American Economic Review*, 83, 487–512.

Campbell, J. Y., S. Giglio, and C. Polk, 2013, "Hard Times," *Review of Asset Pricing Studies*.

Campbell, J. Y., S. W. Giglio, C. Polk, and R. Turley, 2015, "An Intertemporal CAPM with Stochastic Volatility," CEPR Discussion Papers 10681, C.E.P.R. Discussion Papers.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu, 2001, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 56, 1–43.

Campbell, J. Y., and R. J. Shiller, 1988a, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43, 661–76.

Campbell, J. Y., and R. J. Shiller, 1988b, "Stock Prices, Earnings and Expected Dividends," *Journal of Finance*, 43, 661–76.

Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan, 2002, "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?," *The Review of Economic Studies*, 69, 533–564.

Chen, H., W. W. Dou, and L. Kogan, 2015, "Measuring the "Dark Matter" in Asset Pricing Models," Mit sloan working papers.

Chen, H., J. Miao, and N. Wang, 2010, "Entrepreneurial Finance and Nondiversifiable Risk," *Review of Financial Studies*, 23, 4348–4388.

Chernozhukov, V., and H. Hong, 2003, "An MCMC approach to classical estimation," *J. Econometrics*, 115, 293–346.

Chien, Y., H. Cole, and H. Lustig, 2012, "Is the Volatility of the Market Price of Risk Due to Intermittent Portfolio Rebalancing?," *American Economic Review*, 102, 2859–96.

Chien, Y., and H. Lustig, 2010, "The Market Price of Aggregate Risk and the Wealth Distribution," *Review of Financial Studies*, 23, 1596–1650.

Chien, Y., H. Lustig, and K. Naknoi, 2015, "Why Are Exchange Rates So Smooth? A Segmented Asset Markets Explanation," Working Paper UCLA Anderson School of Management.

Christiano, L., R. Motto, and M. Rostagno, 2010, "Financial factors in economic fluctuations," Working Paper Series 1192, European Central Bank.

Christiano, L. J., and J. D. M. Fisher, 2003, "Stock Market and Investment Goods Prices: Implications for Macroeconomics," Nber working papers, National Bureau of Economic Research, Inc.

Christiano, L. J., R. Motto, and M. Rostagno, 2014, "Risk Shocks," *American Economic Review*, 104, 27–65.

Clarke, B. S., 1999, "Asymptotic normality of the posterior in relative entropy," *IEEE Trans. Inform. Theory*, 45, 165–176.

Clarke, B. S., and A. R. Barron, 1990, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inform. Theory*, 36, 453–471.

Clarke, B. S., and A. R. Barron, 1994, "Jeffreys' prior is asymptotically least favorable under entropy risk," *J. Statist. Plann. Inference*, 41, 37–60.

Cochrane, J. H., F. A. Longstaff, and P. Santa-Clara, 2008, "Two Trees," *Review of Financial Studies*, 21, 347–385.

Coeurdacier, N., 2009, "Do Trade Costs in Goods Market Lead to Home Bias in Equities?," *Journal of International Economics*, 77, 86–100.

Coeurdacier, N., and P.-O. Gourinchas, 2011, "When Bonds Matter: Home Bias in Goods and Assets," Nber working papers, National Bureau of Economic Research, Inc.

Coeurdacier, N., and H. Rey, 2013, "Home Bias in Open Economy Financial Macroeconomics," *Journal of Economic Literature*, 51, 63–115.

Colacito, R., M. M. Croce, S. W. Ho, and P. Howard, 2014, "BKK the EZ Way: International Long-Run Growth News and Capital Flows," Working Paper, University of North Carolina.

Cole, H. L., and M. Obstfeld, 1991, "Commodity Trade and International Risk Sharing : How Much Do Financial Markets Matter?," *Journal of Monetary Economics*, 28, 3–24.

Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer, 2013, "Parameter Learning in General Equilibrium: The Asset Pricing Implications," Working Paper.

Constantinides, G. M., and D. Duffie, 1996, "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104, 219–40.

Corsetti, G., L. Dedola, and S. Leduc, 2008, "International Risk Sharing and the Transmission of Productivity Shocks," *Review of Economic Studies*, 75, 443 – 473.

Cover, T. M., and J. A. Thomas, 1991, *Elements of information theory* . Wiley Series in Telecommunications, John Wiley & Sons Inc., New York, A Wiley-Interscience Publication.

Csiszár, I., 1975, "*I*-divergence geometry of probability distributions and minimization problems," *Ann. Probability*, 3, 146–158.

Cummins, J. G., and G. L. Violante, 2002, "Investment-Specific Technical Change in the US (1947-2000): Measurement and Macroeconomic Consequences," *Review of Economic Dynamics*, 5, 243–284.

Curcuru, S. E., T. Dvorak, and F. E. Warnock, 2010, "Decomposing the US External Returns Differential," *Journal of International Economics*, 80, 22–32.

Dai, Q., K. J. Singleton, and W. Yang, 2007, "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields," *Review of Financial Studies*, 20, 1669–1706.

Danthine, J.-P., and J. B. Donaldson, 2002, "Labour Relations and Asset Returns," *Review of Economic Studies*, 69, 41–64.

Debreu, G., 1959, *Theory of Value*, Yale University Press, New Haven.

DeMarzo, P., and M. J. Fishman, 2007, "Optimal Long-Term Financial Contracting," *Review of Financial Studies*, 20, 2079–2128.

DeMarzo, P. M., M. J. Fishman, Z. He, and N. Wang, 2012, "Dynamic Agency and the q Theory of Investment," *Journal of Finance*, 67, 2295–2340.

DeMarzo, P. M., and Y. Sannikov, 2006, "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *Journal of Finance*, 61, 2681–2724.

Devereux, M. B., and A. Sutherland, 2010, "Valuation Effects and the Dynamics of Net External Assets," *Journal of International Economics*, 80, 129 – 143.

Devereux, M. B., and A. Sutherland, 2011, "Country Portfolios In Open Economy Macro-Models," *Journal of the European Economic Association*, 9, 337–369.

Di Tella, S., 2014, "Uncertainty Shocks and Balance Sheet Recessions," Working paper.

DIXIT, A. K., AND R. S. PINDYCK (eds.), 1994, *Investment Under Uncertainty*, Princeton University Press.

Doob, J. L., 1950, "Recent Publications: On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation," *Amer. Math. Monthly*, 57, 497–498.

Dou, W. W., 2016, "Embrace or Fear Uncertainty: Growth Options, Limited Risk Sharing, and Asset Prices," Working papers.

Dou, W. W., and Y. Ji, 2015, "External Financing and Customer Capital: A Financial Theory of Markups," Working papers.

Dou, W. W., and A. Verdelhan, 2015, "The Volatility of International Capital Flows and Foreign Assets," Mit sloan working papers.

Drechsler, I., A. Savov, and P. Schnabl, 2014, "A Model of Monetary Policy and Risk Premia," Nber working papers, National Bureau of Economic Research, Inc.

Drechsler, I., and A. Yaron, 2011, "What's Vol Got to Do with It," *Review of Financial Studies*, 24, 1–45.

Duffie, D., and L. G. Epstein, 1992a, "Asset Pricing with Stochastic Differential Utility," *Review of Financial Studies*, 5, 411–36.

Duffie, D., and L. G. Epstein, 1992b, "Stochastic Differential Utility," *Econometrica*, 60, 353–94.

Duffie, D., J. Geanakoplos, A. Mas-Colell, and A. McLennan, 1994, "Stationary Markov Equilibria," *Econometrica*, 62, 745–81.

Duffie, D., and C. Skiadas, 1994, "Continuous-time security pricing : A utility gradient approach," *Journal of Mathematical Economics*, 23, 107–131.

Dumas, B., 1989, "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2, 157–188.

Dumas, B., K. K. Lewis, and E. Osambela, 2014, "Differences of opinion and international equity markets," National Bureau of Economic Research.

Dumas, B., and A. Lyasoff, 2012, "Incomplete-Market Equilibria Solved Recursively on an Event Tree," *Journal of Finance*, 67, 1897–1941.

Edmans, A., and X. Gabaix, 2011, "Tractability in Incentive Contracting," *Review of Financial Studies*, 24, 2865–2894.

Edmans, A., X. Gabaix, T. Sadzik, and Y. Sannikov, 2012, "Dynamic CEO Compensation," *The Journal of Finance*, 67, 1603–1647.

Edmans, A., and G. Manso, 2011, "Governance Through Trading and Intervention: A Theory of Multiple Blockholders," *Review of Financial Studies*, 24, 2395–2428.

Efron, B., 2004, "The estimation of prediction error: covariance penalties and cross-validation," *J. Amer. Statist. Assoc.*, 99, 619–642, With comments and a rejoinder by the author.

Eisfeldt, A. L., and D. Papanikolaou, 2013, "Organization Capital and the Cross-Section of Expected Returns," *Journal of Finance*, 68, 1365–1406.

Epstein, L., and S. Zin, 1989a, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns I: A Theoretical Framework," *Econometrica*, 57, 937–969.

Epstein, L. G., and M. Schneider, 2003, "Recursive multiple-priors," *Journal of Economic Theory*, 113, 1–31.

Epstein, L. G., and M. Schneider, 2010, "Ambiguity and Asset Markets," *Annual Review of Financial Economics*, 2, 315–346.

Epstein, L. G., and S. Zin, 1989b, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Epstein, L. G., and S. E. Zin, 1989c, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–69.

Evans, M. D. D., and V. Hnatkovska, 2005, "Solving General Equilibrium Models with Incomplete Markets and Many Assets," Nber technical working papers, National Bureau of Economic Research, Inc.

Fajgelbaum, P., E. Schaal, and M. Taschereau-Dumouchel, 2013, "Uncertainty Traps," 2013 Meeting Papers 677, Society for Economic Dynamics.

Fama, E. F., 1984, "Forward and Spot Exchange Rates," *Journal of Monetary Economics*, 14, 319–338.

Fama, E. F., and K. R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Favilukis, J., L. Garlappi, and S. Neamati, 2015, "The Carry Trade and Uncovered Interest Parity when Markets are Incomplete," Working Paper UBC.

Favilukis, J., and X. Lin, 2013, "Long run productivity risk and aggregate investment," *Journal of Monetary Economics*, 60, 737–751.

Fazzari, S. M., R. G. Hubbard, and B. C. Petersen, 1988, "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 19, 141–206.

Fearnhead, P., and D. Prangle, 2012, "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation," *Journal of the Royal Statistical Society Series B*, 74, 419–474.

Fernandez-Villaverde, J., P. Guerron-Quintana, J. F. Rubio-Ramirez, and M. Uribe, 2011, "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, 101, 2530–61.

Fisher, J. D. M., 2006, "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks," *Journal of Political Economy*, 114, 413–451.

Fisher, R. A., 1922, "On the mathematical foundations of theoretical statistics," *Phil. Trans. R. Soc. Lond. A*, 222, 309–68.

Foster, D. P., and E. I. George, 1994, "The risk inflation criterion for multiple regression," *Ann. Statist.*, 22, 1947–1975.

Fudenberg, D., B. Holmstrom, and P. Milgrom, 1990, "Short-term contracts and long-term agency relationships," *Journal of Economic Theory*, 51, 1–31.

Gabaix, X., 2012, "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *The Quarterly Journal of Economics*, 127, 645–700.

Garlappi, L., and Z. Song, 2014, "Can Investment Shocks Explain Value Premium and Momentum Profits?," Working paper, Sauder School of Business.

Gârleanu, N., L. Kogan, and S. Panageas, 2012, "Displacement risk and asset returns," *Journal of Financial Economics*, 105, 491–510.

Gârleanu, N., and S. Panageas, 2015, "Young, Old, Conservative, and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing," *Journal of Political Economy*.

Gârleanu, N., S. Panageas, and J. Yu, 2012, "Technological Growth and Asset Pricing," *Journal of Finance*, 67, 1265–1292.

Geanakoplos, J., and W. R. Zame, 2013, “Collateral Equilibrium: A Basic Framework,” Cowles foundation discussion papers, Cowles Foundation for Research in Economics, Yale University.

Gelfand, A. E., and S. K. Ghosh, 1998, “Model choice: a minimum posterior predictive loss approach,” *Biometrika*, 85, 1–11.

Gelman, A., J. Hwang, and A. Vehtari, 2013, “Understanding predictive information criteria for Bayesian models,” *Statsitics and Computing*.

Gentzkow, M., and J. M. Shapiro, 2013, “Measuring the Sensitivity of Parameter Estimates to Sample Statistics,” Chicago booth working papers.

Gertler, M., and P. Karadi, 2011, “A model of unconventional monetary policy,” *Journal of Monetary Economics*, 58, 17–34.

Gertler, M., and N. Kiyotaki, 2010, “Financial Intermediation and Credit Policy in Business Cycle Analysis,” in Benjamin M. Friedman, and Michael Woodford (ed.), *Handbook of Monetary Economics* . , vol. 3, Elsevier.

Ghosh, A., and G. Constantinides, 2015, “Asset Pricing with Countercyclical Household Consumption Risk,” Discussion paper.

Gilboa, I., and D. Schmeidler, 1989, “Maxmin expected utility with non-unique prior,” *Journal of Mathematical Economics*, 18, 141–153.

Gilchrist, S., J. W. Sim, and E. Zakrajsek, 2014, “Uncertainty, Financial Frictions, and Investment Dynamics,” Finance and economics discussion series, Board of Governors of the Federal Reserve System (U.S.).

Gilchrist, S., and E. Zakrajsek, 2012, “Credit Spreads and Business Cycle Fluctuations,” *American Economic Review*, 102, 1692–1720.

Glover, B., and O. Levine, 2015, “Uncertainty, investment, and managerial incentives,” *Journal of Monetary Economics*, 69, 121–137.

Gomes, J., L. Kogan, and L. Zhang, 2003, “Equilibrium Cross Section of Returns,” *Journal of Political Economy*, 111, 693–732.

Gordon, R. J., 1990, *The Measurement of Durable Goods Prices* . No. gord90-1 in NBER Books, National Bureau of Economic Research, Inc.

Gourinchas, P.-O., and H. Rey, 2007, “International Financial Adjustment,” *Journal of Political Economy*, 115, 665–703.

Gourinchas, P.-O., H. Rey, and N. Govillot, 2010, “Exorbitant Privilege and Exorbitant Duty,” Imes discussion paper series, Institute for Monetary and Economic Studies, Bank of Japan.

Gourio, F., 2012, "Disaster Risk and Business Cycles," *American Economic Review*, 102(6), 2734–66.

Gourio, F., M. Siemer, and A. Verdelhan, 2015, "Uncertainty and International Capital Flows," Discussion paper.

Gray, R. M., 1988, *Probability, random processes, and ergodic properties*, Springer-Verlag, New York.

Greenwood, J., Z. Hercowitz, and G. W. Huffman, 1988, "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78, 402–17.

Greenwood, J., Z. Hercowitz, and P. Krusell, 1997, "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 87, 342–62.

Greenwood, J., Z. Hercowitz, and P. Krusell, 2000, "The role of investment-specific technological change in the business cycle," *European Economic Review*, 44, 91–115.

Guvenen, F., 2009, "A Parsimonious Macroeconomic Model for Asset Pricing," *Econometrica*, 77, 1711–1750.

Haddad, V., 2014, "Concentrated Ownership and Equilibrium Asset Prices," Discussion paper.

Hahn, J., W. Newey, and R. Smith, 2011, "Tests for neglected heterogeneity in moment condition models," CeMMAP working papers CWP26/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–84.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press.

Hansen, L., and T. Sargent, 2008a, *Robustness*, Princeton University Press.

Hansen, L. P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–54.

Hansen, L. P., 2007, "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97, 1–30.

Hansen, L. P., 2014, "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122, 945 – 987.

Hansen, L. P., and T. J. Sargent, 1980, "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, 2, 7–46.

Hansen, L. P., and T. J. Sargent, 1991, *Rational Expectations Econometrics*, Westview Press, Boulder, Colorado.

Hansen, L. P., and T. J. Sargent, 2001, "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66.

Hansen, L. P., and T. J. Sargent, 2008b, *Robustness*, Princeton University Press, Princeton, New Jersey.

Hansen, L. P., and J. A. Scheinkman, 2014, "Stochastic Compounding and Uncertain Valuation," Working paper, University of Chicago.

Hartman, R., 1972, "The effects of price and cost uncertainty on investment," *Journal of Economic Theory*, 5, 258–266.

Havranek, T., R. Horvath, Z. Irsova, and M. Rusnak, 2013, "Cross-Country Heterogeneity in Intertemporal Substitution," Working papers, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Hayashi, F., 1982, "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50, 213–24.

He, Z., and A. Krishnamurthy, 2011, "A Model of Capital and Crises," *The Review of Economic Studies*.

He, Z., and A. Krishnamurthy, 2013, "Intermediary Asset Pricing," *American Economic Review*, 103, 732–70.

Heathcote, J., and F. Perri, 2013, "The International Diversification Puzzle Is Not as Bad as You Think," *Journal of Political Economy*, 121, 1108 – 1159.

Heaton, J., and D. Lucas, 1997, "Market Frictions, Savings Behavior, And Portfolio Choice," *Macroeconomic Dynamics*, 1, 76–101.

Heaton, J., and D. Lucas, 2000a, "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk," *Journal of Finance*, 55, 1163–1198.

Heaton, J., and D. Lucas, 2000b, "Portfolio Choice in the Presence of Background Risk," *Economic Journal*, 110, 1–26.

Heaton, J., and D. J. Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 104, 443–87.

Hernandez, A., and M. S. Santos, 1996, "Competitive equilibria for infinite-horizon economies with incomplete markets," *Journal of Economic Theory*, 71, 102–130.

Herskovic, B., B. T. Kelly, H. Lustig, and S. V. Nieuwerburgh, 2014, "The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications," Nber working papers, National Bureau of Economic Research, Inc.

Hodrick, R., and E. Prescott, 1997, "Post-War U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, 29, 1–16.

Hoffmann, F., and S. Pfeil, 2010, "Reward for Luck in a Dynamic Agency Model," *Review of Financial Studies*, 23, 3329–3345.

Holderness, C. G., 2003, "A survey of blockholders and corporate control," *Economic Policy Review*.

Holderness, C. G., 2009, "The Myth of Diffuse Ownership in the United States," *Review of Financial Studies*, 22, 1377–1408.

Holmstrom, B., and J. Tirole, 1997, "Financial Intermediation, Loanable Funds, and The Real Sector," *The Quarterly Journal of Economics*, 112, 663–691.

Holtz-Eakin, D., W. Newey, and H. S. Rosen, 1988, "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–95.

Ibragimov, I. A., 1962, "Some limit theorems for stationary processes," *Teor. Verojatnost. i Primenen.*, 7, 361–392.

Israelsen, R. D., 2010, "Investment Based Valuation and Managerial Expectations," Working paper, Kelley School of Business at Indiana University.

Ivashina, V., and D. Scharfstein, 2010, "Loan Syndication and Credit Cycles," *American Economic Review*, 100, 57–61.

Jaynes, E., 1982, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, 70, 939 – 952.

Jo, S., 2012, "The Effects of Oil Price Uncertainty on the Macroeconomy," Discussion paper.

Jones, L., 1989, "ApproximationâĂŞTheoretic Derivation of Logarithmic Entropy Principles for Inverse Problems and Unique Extension of the Maximum Entropy Method to Incorporate Prior Knowledge," *Siam Applied Math*.

Jovanovic, B., 2009, "Investment options and the business cycle," *Journal of Economic Theory*, 144, 2247–2265.

Judd, K. L., F. Kubler, and K. Schmedders, 2003, "Asset Trading Volume with Dynamically Complete Markets and Heterogeneous Agents," *Journal of Finance*, 58, 2203–2218.

Julliard, C., and A. Ghosh, 2012, "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25, 3037–3076.

Justiniano, A., G. Primiceri, and A. Tambalotti, 2011, "Investment Shocks and the Relative Price of Investment," *Review of Economic Dynamics*, 14, 101–121.

Kaplan, S. N., and L. Zingales, 1997, "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints," *The Quarterly Journal of Economics*, 112, 169–215.

Khan, A., and J. K. Thomas, 2008, "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics," *Econometrica*, 76, 395–436.

Khan, R., R. Dharwadkar, and P. Brandes, 2005, "Institutional ownership and CEO compensation: a longitudinal examination," *Journal of Business Research*, 58, 1078–1088.

Kim, J.-Y., 2002, "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107, 175–193.

King, R. G., C. I. Plosser, and S. T. Rebelo, 1988, "Production, growth and business cycles : I. The basic neoclassical model," *Journal of Monetary Economics*, 21, 195–232.

King, R. G., C. I. Plosser, and S. T. Rebelo, 2002, "Production, Growth and Business Cycles: Technical Appendix," *Computational Economics*, 20, 87–116.

Kiyotaki, N., and J. Moore, 1997, "Credit Cycles," *Journal of Political Economy*, 105, 211–48.

Klibanoff, P., M. Marinacci, and S. Mukerji, 2005, "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73, 1849–1892.

Knight, F. H., 1921, *Risk, Uncertainty and Profit*, Houghton Mifflin Co.

Kogan, L., and D. Papanikolaou, 2013, "Firm Characteristics and Stock Returns: The Role of Investment-Specific Shocks," *Review of Financial Studies*, 26, 2718–2759.

Kogan, L., and D. Papanikolaou, 2014, "Growth Opportunities, Technology Shocks, and Asset Prices," *Journal of Finance*, 69, 675–718.

Kogan, L., D. Papanikolaou, and N. Stoffman, 2015, "Winners and Losers: Creative Destruction and the Stock Market," Nber working papers, National Bureau of Economic Research, Inc.

Kollmann, R., 1995, "Consumption, Real Exchange Rates and the Structure of International Asset Markets," *Journal of International Money and Finance*, 14, 191–211.

Kraay, A., and J. Ventura, 2000, "Current accounts in debtor and creditor countries," *Quarterly Journal of Economics*, pp. 1137–1166.

Krantz, S. G., and H. R. Parks, 2013, *The implicit function theorem* . Modern Birkhäuser Classics, Birkhäuser/Springer, New York, History, theory, and applications, Reprint of the 2003 edition.

Kreps, D., and E. L. Porteus, 1978a, "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Kreps, D. M., and E. L. Porteus, 1978b, "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Krishnamurthy, A., and T. Muir, 2015, "Credit Spreads and the Severity of Financial Crises," Working paper, Stanford GSB.

Kubler, F., and H. Polemarchakis, 2004, "Stationary Markov Equilibria for Overlapping Generations," *Economic Theory*, 24, 623–643.

Kubler, F., and K. Schmedders, 2002, "Recursive Equilibria In Economies With Incomplete Markets," *Macroeconomic Dynamics*, 6, 284–306.

Kubler, F., and K. Schmedders, 2003, "Stationary Equilibria in Asset-Pricing Models with Incomplete Markets and Collateral," *Econometrica*, 71, 1767–1793.

Kullback, S., 1967, "A lower bound for discrimination information in terms of variation," *IEEE Transactions on Information Theory*, 13, 126–127.

Lambrecht, B. M., and S. C. Myers, 2007, "A Theory of Takeovers and Disinvestment," *Journal of Finance*, 62, 809–845.

Lambrecht, B. M., and S. C. Myers, 2008, "Debt and managerial rents in a real-options model of the firm," *Journal of Financial Economics*, 89, 209–231.

Lambrecht, B. M., and S. C. Myers, 2012, "A Lintner Model of Payout and Managerial Rents," *Journal of Finance*, 67, 1761–1810.

Lane, P. R., and G. M. Milesi-Ferretti, 2007, "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004," *Journal of international Economics*, 73, 223–250.

Le Cam, L., and G. L. Yang, 2000, *Asymptotics in statistics* . Springer Series in Statistics, Springer-Verlag, New York, second edn., Some basic concepts.

Lee, L.-f., 2005, "Classical Inference with ML and GMM Estimates with Various Rates of Convergence," Ohio state unviersity working papers.

Lehmann, E. L., and G. Casella, 1998, *Theory of point estimation* . Springer Texts in Statistics, Springer-Verlag, New York, second edn.

Lettau, M., and S. Ludvigson, 2001, "Consumption, Aggregate Wealth and Expected Stock Returns," *The Journal of Finance*, 56, 815–849.

Levine, D. K., and W. R. Zame, 1996, "Debt Constraints and Equilibrium in Infinite Horizon Economies with Incomplete Markets," *Journal of Mathematical Economics*, 26, 103–131.

Lewis, K. K., 1999, "Trying to Explain Home Bias in Equities and Consumption," *Journal of Economic Literature*, 37, 571–608.

Lin, X., J. Pittman, and B. Clarke, 2007, "Information conversion, effective samples, and parameter size," *IEEE Trans. Inform. Theory*, 53, 4438–4456.

Longstaff, F. A., and M. Piazzesi, 2004a, "Corporate earnings and the equity premium," *Journal of Financial Economics*, 74, 401–421.

Longstaff, F. A., and M. Piazzesi, 2004b, "Corporate Earnings and the Equity Premium," *Journal of financial Economics*, 74, 401–421.

Longstaff, F. A., and J. Wang, 2012, "Asset Pricing and the Credit Market," *Review of Financial Studies*, 25, 3169–3215.

Lorenzoni, G., 2008, "Inefficient Credit Booms," *Review of Economic Studies*, 75, 809–833.

Lucas, R. E., and T. J. Sargent, 1981, *Rational Expectations and Econometric Practice*, University of Minnesota Press, Minneapolis.

Lucas, R. J., and N. L. Stokey, 1984, "Optimal Growth with Many Consumers," *Journal of Economic Theory*, 32, 139–171.

Lustig, H., and S. V. Nieuwerburgh, 2010, "How Much Does Household Collateral Constrain Regional Risk Sharing?," *Review of Economic Dynamics*, 13, 265–294.

Lustig, H., A. Stathopoulos, and A. Verdelhan, 2015, "Nominal Exchange Rate Stationarity and Long-Term Bond Returns," Working Paper, MIT.

Lustig, H., S. Van Nieuwerburgh, and A. Verdelhan, 2013, "The Wealth-Consumption Ratio," *Review of Asset Pricing Studies*, p. rat002.

Lustig, H., and A. Verdelhan, 2007, "The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk," *American Economic Review*, 97, 89–117.

Maggiori, M., 2015, "Financial Intermediation, International Risk Sharing, and Reserve Currencies," Harvard Working Paper.

Magill, M., and M. Quinzii, 1996, "Incomplete Markets Over an Infinite Horizon: Long-Lived Securities and Speculative Bubbles," *Journal of Mathematical Economics*, 26, 133–170.

Magill, M., and M. Quinzii, 2002, *Theory of Incomplete Markets, Vol. 1*, The MIT Press, Cambridge.

Mankiw, N., and S. P. Zeldes, 1991, "The consumption of stockholders and nonstockholders," *Journal of Financial Economics*, 29, 97 – 112.

Mankiw, N. G., 1986, "The equity premium and the concentration of aggregate shocks," *Journal of Financial Economics*, 17, 211–219.

Martin, I., 2012, "Consumption-Based Asset Pricing with Higher Cumulants," Forthcoming, Review of Economic Studies.

Martin, I., 2013, "The Lucas Orchard," *Econometrica*, 81, 55–111.

Mendoza, E. G., V. Quadrini, and J.-V. Rios-Rull, 2009, "Financial Integration, Financial Development, and Global Imbalances," *Journal of Political Economy*, 117, 371–416.

Merton, R., 1973a, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–87.

Merton, R. C., 1969, "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *The Review of Economics and Statistics*, 51, 247–57.

Merton, R. C., 1973b, "Theory of Rational Option Pricing," *Bell Journal of Economics*, 4, 141–183.

Miao, J., and N. Wang, 2007, "Investment, consumption, and hedging under incomplete markets," *Journal of Financial Economics*, 86, 608–642.

Miller, M. H., and F. Modigliani, 1961, "Dividend Policy, Growth, and the Valuation of Shares," *Journal of Business*, 34, 411.

Modigliani, F., and M. H. Miller, 1958, "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, 48, 261–297.

Myers, S. C., 2000, "Outside Equity," *Journal of Finance*, 55, 1005–1037.

Nakamura, E., D. Sergeyev, and J. Steinsson, 2014, "Growth-Rate and Uncertainty Shocks in Consumption: Cross-Country Evidence," Nber working papers, National Bureau of Economic Research, Inc.

Newey, W. K., 1985, "Maximum likelihood specification testing and conditional moment tests," *Econometrica*, 53, 1047–1070.

Newey, W. K., and K. D. West, 1987a, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–08.

Newey, W. K., and K. D. West, 1987b, "Hypothesis testing with efficient method of moments estimation," *Internat. Econom. Rev.*, 28, 777–787.

Newey, W. K., and K. D. West, 1994, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–53.

Nieuwerburgh, S. V., and L. Veldkamp, 2009, "Information Immobility and the Home Bias Puzzle," *Journal of Finance*, 64, 1187–1215.

Oi, W., 1961, "The desirability of price instabilty under perfect competition," *Econometrica*, 29, 58–64.

Opp, C., 2014, "Venture Capital Cycles," Working paper, Wharton School of the Unversity Pennsylvania.

Panousi, V., and D. Papanikolaou, 2012, "Investment, Idiosyncratic Risk, and Ownership," *The Journal of Finance*, 67, 1113–1148.

Papanikolaou, D., 2011, "Investment Shocks and Asset Prices," *Journal of Political Economy, University of Chicago Press*, 119, 639 – 685.

Pástor, u., and P. Veronesi, 2006, "Was there a Nasdaq bubble in the late 1990s?," *Journal of Financial Economics*, 81, 61–100.

Pástor, u., and P. Veronesi, 2009, "Technological Revolutions and Stock Prices," *American Economic Review*, 99, 1451–83.

Patton, A., D. Politis, and H. White, 2009, "Correction to ŞAutomatic Block-Length Selection for the Dependent BootstrapŤ by D. Politis and H. White," *Econometric Reviews*, 28, 372–375.

Pavlova, A., 2002, "Adjustment Costs, Learning-by-Doing, and Technology Adoption Under Uncertainty," Working papers 4369-01, Massachusetts Institute of Technology (MIT), Sloan School of Management.

Pavlova, A., and R. Rigobon, 2010, "An Asset-Pricing View of External Adjustment," *Journal of International Economics*, 80, 144–156.

Pavlova, A., and R. Rigobon, 2012, "Equilibrium Portfolios and External Adjustment Under Incomplete Markets," AFA 2009 San Francisco Meetings Paper.

Perez, C., 2002, *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*, Edward Elgar.

Peter Ferderer, J., 1996, "Oil price volatility and the macroeconomy," *Journal of Macroeconomics*, 18, 1–26.

Petrosky-Nadeau, N., and L. Zhang, 2013, "Unemployment Crises," Nber working papers, National Bureau of Economic Research, Inc.

Pindyck, R. S., 1993, "Investments of uncertain cost," *Journal of Financial Economics*, 34, 53–76.

Politis, D., and H. White, 2004, "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*, 23, 53–70.

Politis, D. N., and J. P. Romano, 1994a, “Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap,” *Statist. Sinica*, 4, 461–476.

Politis, D. N., and J. P. Romano, 1994b, “The stationary bootstrap,” *J. Amer. Statist. Assoc.*, 89, 1303–1313.

Polson, N. G., 1992, “On the expected amount of information from a nonlinear model,” *J. Roy. Statist. Soc. Ser. B*, 54, 889–895.

Purnanandam, A., and U. Rajan, 2014, “Growth Option Exercise and Capital Structure,” Working papers, University of Michigan.

Rabitsch, K., S. Stepanchuk, and V. Tsyrennikov, 2015, “International Portfolios: A Comparison of Solution Methods,” WU Department of Economics Working Paper Series.

Ramey, V. A., and N. Francis, 2009, “A Century of Work and Leisure,” *American Economic Journal: Macroeconomics*, 1, 189–224.

Reinhart, C. M., and K. S. Rogoff, 2009, *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press, Princeton, New Jersey.

Rietz, T. A., 1988, “The Equity Premium: A Solution,” *Journal of Monetary Economics*, 22, 117–131.

Roberts, K., and M. L. Weitzman, 1981, “Funding Criteria for Research, Development, and Exploration Projects,” *Econometrica*, 49, 1261–88.

Rockafellar, R., 1970, *Convex Analysis* . Princeton mathematical series, Princeton University Press.

Rouhenworst, G., 1995, “Asset Pricing Implications of Real Business Cycle Models,” *Frontiers of business cycle research*.

Roussas, G. G., 1996, “Exponential probability inequalities with some applications,” in *Statistics, probability and game theory*, vol. 30 of *IMS Lecture Notes Monogr. Ser.*, . pp. 303–319, Inst. Math. Statist., Hayward, CA.

Samson, P.-M., 2000, “Concentration of measure inequalities for Markov chains and Φ-mixing processes,” *Ann. Probab.*, 28, 416–461.

Samuelson, P. A., 1969, “Lifetime Portfolio Selection by Dynamic Stochastic Programming,” *The Review of Economics and Statistics*, 51, 239–46.

Santos, M. S., and M. Woodford, 1997, “Rational Asset Pricing Bubbles,” *Econometrica*, 65, 19–58.

Saracoglu, R., and T. J. Sargent, 1978, "Seasonality and portfolio balance under rational expectations," *Journal of Monetary Economics*, 4, 435–458.

Schwartz, L., 1965, "On Bayes procedures," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 4, 10–26.

Schwarz, G., 1978, "Estimating the dimension of a model," *Ann. Statist.*, 6, 461–464.

Segal, G., I. Shaliasovich, and A. Yaron, 2015, "Good Uncertainty and Bad Uncertainty: Macroeconomic and Financial Market Implications," *Journal of Financial Economics*, 117, 369–397.

Serfling, R. J., 1980, *Approximation theorems of mathematical statistics*, John Wiley & Sons Inc., New York, Wiley Series in Probability and Mathematical Statistics.

Shaliastovich, I., 2015, "Learning, confidence, and option prices," *Journal of Econometrics*, 187, 18–42.

Shen, X., and J. Ye, 2002, "Adaptive model selection," *J. Amer. Statist. Assoc.*, 97, 210–221.

Smets, F., and R. Wouters, 2007, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde, 2002, "Bayesian measures of model complexity and fit," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64, 583–639.

Stathopoulos, A., 2012, "Portfolio Home Bias and External Habit Formation," Working paper, University of Southern California.

Stepanchuk, S., and V. Tsyrennikov, 2015, "Portfolio and Welfare Consequences of Debt Market Dominance," *Journal of Monetary Economics, forthcoming*.

Storesletten, K., C. Telmer, and A. Yaron, 2007, "Asset Pricing with Idiosyncratic Risk and Overlapping Generations," *Review of Economic Dynamics*, 10, 519–548.

Sung, J., 1997, "Corporate Insurance and Managerial Incentives," *Journal of Economic Theory*, 74, 297 – 332.

Taniguchi, M., and Y. Kakizawa, 2000, *Asymptotic theory of statistical inference for time series* . Springer Series in Statistics, Springer-Verlag, New York.

Tibshirani, R., and K. Knight, 1999, "The Covariance Inflation Criterion for Adaptive Model Selection," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 529–546.

Tille, C., and E. van Wincoop, 2010, "International Capital Flows," *Journal of International Economics*, 80, 157–175.

Timmermann, A., 2000, "Moments of Markov switching models," *Journal of Econometrics*, 96, 75–111.

Tryon, R., 1979, "Testing for Rational Expectations in Foreign Exchange Markets," International Finance Discussion Papers.

van der Vaart, A. W., 1998, *Asymptotic statistics* . , vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge.

Van Nieuwerburgh, S., and L. Veldkamp, 2006, "Learning asymmetries in real business cycles," *Journal of Monetary Economics*, 53, 753–772.

Ventura, J., 2001, "A portfolio view of the US current account deficit," *Brookings Papers on Economic Activity*, 2001, 241–253.

Vissing-Jorgensen, A., 2002a, "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110, 825–853.

Vissing-Jorgensen, A., 2002b, "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110, 825–853.

Wachter, J. A., 2013, "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *Journal of Finance*, 68, 987–1035.

Wang, J., 1996, "The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors," *Journal of Financial Economics*, 41, 75–110.

Weil, P., 1989, "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Weil, P., 1990, "Nonexpected Utility in Macroeconomics," *The Quarterly Journal of Economics*, 105, 29–42.

Welch, I., 2004, "Capital Structure and Stock Returns," *Journal of Political Economy*, 112, 106–131.

White, H., and I. Domowitz, 1984, "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143–61.

Ye, J., 1998, "On measuring and correcting the effects of data mining and model selection," *J. Amer. Statist. Assoc.*, 93, 120–131.

Yin, G., 2009, "Bayesian generalized method of moments," *Bayesian Analysis*, 4, 191–207.

Zellner, A., 1996, "Bayesian method of moments (BMOM) analysis of mean and regression models," in *Modelling and prediction (Hsinchu, 1994)* . pp. 61–74, Springer, New York.

Zellner, A., 1998, "The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches," *Journal of Econometrics*, 83, 185–212.

Zin, S. E., 2002, "Are behavioral asset-pricing models structural?," *Journal of Monetary Economics*, 49, 215–228.