The London School of Economics and Political Science

Three Papers on Asset Pricing

Petar Svilenov Sabtchevsky

Thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of

Doctor of Philosophy

April 2018

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Statement of conjoint work

I confirm that Chapter 3 is jointly co-authored with Philippe Mueller, Andrea Vedolin, and Paul Whelan. I contributed 25% of the work for Chapter 3.

I declare that my thesis consists of 36 814 words.

Acknowledgment

I would like to thank my academic supervisors Ian Martin and Georgy Chabakauri for their guidance and support. This thesis would not have been possible without the help I received from them. Both Ian and Georgy have been very patient, helpful, inspirational, and valuable advisers throughout the course of my PhD studies. Whereas Ian introduced me to academic research during my first year on the PhD programme, Georgy's lectures and research papers solidified my interest in continuous time general equilibrium models. I would also like to thank Peter Kondor. I immensely benefited from many discussions with him. He was always very approachable and willing to discuss early stage research ideas. Additionally, I benefited from discussions with Thummim Cho, Victor DeMiguel, Christian Julliard, Martin Oehmke, Andrea Tamoni, and Dimitri Vayanos.

I am particularly indebted to my co-authors Andrea Vedolin, Philippe Mueller, and Paul Whelan. I would not be where I am now without their support. Andrea, Philippe, and Paul introduced me to the fascinating world of empirical research. I learned tremendously from numerous lively and eye-opening discussions with them. I am particularly grateful to Philippe and Paul for their continuous guidance and support.

I would also like to thank the PhD programme directors, Daniel Ferreira, Christian Julliard, and Mike Burkart. I am also very grateful to Mary Comben, the PhD programme manager. She was always immensely kind, responsive, and helpful. The time at the LSE was very enjoyable thanks to my fellow PhD students from my cohort, Lukas Kremens, Dimitris Papadimitriou, Bernardo De Oliveira Guerra Ricca, Su Wang, and the track one students, Lorenzo Bretscher, Brandon Han, and Gosia Ryduchowska. The senior PhD students, Hoyong Choi, Sergei Glebkin, Jesus Gorrin, Olga Obizhaeva, Michael Punz, Una Savic, and Seyed Seyedan were always an invaluable source of advice and great role models. I have been lucky to also been surrounded by Fabrizio Core, James Guo, Alberto Pellicioli, Marco Pelosi, Karamfil

Todorov, and Yue Yuan.

My biggest gratitude goes to my family. I am indebted to my parents, Petia and Svilen, and my sister Joana for their unconditional love, belief in me, positivity, continuous support, and advice.

Abstract

This thesis consists of three papers on asset pricing.

In the first paper, I analyse the effects of volatility management (a trading strategy in which risky asset exposure is inversely proportional to the level of volatility) in a general equilibrium heterogeneous agent model. Two distinct types of agents populate the model economy, an unconstrained investor endowed with logarithmic utility over instantaneous consumption and a volatility-managed portfolio. My model goes a long way towards the rationalization of the behaviour of investment vehicles that follow investment management strategies that are isomorphic to the ones implied by the principles of volatility management. Whereas my theoretical approach offers a high degree of tractability, it is subject to some important caveats. Specifically, the model implies unrealistically high leverage for the unconstrained investor.

In the second paper, I propose a general equilibrium intermediary asset pricing model featuring a heterogeneous intermediary sector. Two distinct types of intermediaries populate the financial intermediary sector: equity-constrained intermediaries and shadow financial intermediaries. The main theoretical contribution of this paper is threefold. First, I show that over the region of the state space where the intermediation constraint binds, the risk premium on the intermediated risky asset is decreasing in the degree of intermediary sector heterogeneity. Second, intermediary sector heterogeneity allows for rich leverage dynamics within the intermediary sector and at the level of the aggregate intermediary sector. Third, the constrained region shrinks relative to the benchmark model in which the intermediary sector is homogeneous.

The third paper (co-authored with Philippe Mueller, Andrea Vedolin, and Paul Whelan), studies variance risk premia (VRP). We document a number of novel stylized facts related to the equity and the Treasury VRP (EVRP and TVRP) and show that (1) the short maturity TVRP predict excess returns on short maturity bonds; (2) long maturity TVRP and the EVRP predict excess returns on long

maturity bonds; and (3) whereas the EVRP predicts equity returns for horizons of up to 6 months, the long maturity TVRP contain robust information for equity returns at longer horizons. Finally, we present evidence that expected inflation is a powerful determinant of the dynamics of the EVRP, the TVRP, and their co-movement.

Contents

List of Tables

List of Figures

Chapter 1

Volatility-Managed Portfolios in General Equilibrium

1.1. Introduction

The volatility-managed trading strategy requires that the size of the risky asset position is *inversely* proportional to the level of volatility. Traders that follow volatility-managed trading strategies decrease their risky asset exposure in high-volatility states and increase risky asset exposure in low-volatility states. Given that volatility tends to be high in adverse states of the world, investors following a volatility-managed trading strategy would decrease their exposure at times of severe market dislocations and re-enter the market at times when market conditions are more favorable.

Volatility management constitutes a profitable trading strategy so long as asset volatility and the risk premium do not move in lockstep; that is, an increase in volatility is not immediately followed by a corresponding increase in the risk premium. If that is the case, an increase in asset volatility leads to a drop in the Sharpe ratio that renders the risky asset less attractive[1] and justifies portfolio rebalancing towards the risk-free asset.

The sign of the correlation between expected returns and conditional variance is the subject of a heated debate. Conventional wisdom suggests that it is prudent to increase risk taking, or at the very least keep it constant, during economic downturns when volatility tends to be high. Investors failing to follow this advice are warned that they risk missing out on a once-in-a-generation buying opportunity at their own peril. This type of advice is put forward not only by leading practitioners, such as Warren Buffet, CEO of Berkshire Hathaway, but also by leading financial economists such as John Cochrane of the University of Chicago. Cochrane goes so far as to argue that

> If you're less leveraged, less affected by recessions, and have a longer horizon than the average, it makes sense to buy. If you're more leveraged, more affected by recession or have a shorter horizon, it might be the time

[1] As is seen from the vantage point of an unconstrained, but myopic investor.

> to sell, even though you might be cashing out at the bottom. If you're about the same as everyone else, do nothing and relax. If you're wrong, at least you will have excellent company.

In a widely cited paper, Fama and French (1989) show, in line with the conventional view, that expected returns are high during recessions. On the theoretical front, the trading recommendation is broadly the same. Prominent off-the-shelf, single-agent equilibrium models, Bansal and Yaron (2004) and Campbell and Cochrane (1999), among others, imply a positive sign for the correlation between conditional volatility and expected returns. Clearly, the conventional view advocates for a portfolio trading strategy that is in many respects the opposite of a portfolio strategy devised in the spirit of volatility management.

In a recent paper, Moreira and Muir (2017) challenge this conventional view. They empirically show that conditional variance is only weakly related to future expected returns and the increase in expected returns is nowhere big enough to compensate investors for the increase in volatility. They also show that volatility-managed portfolios, scaled by the inverse of previous month's realized variance, earn large risk-adjusted alphas across a wide range of asset pricing factors, hinting at the possibility that the expected return conditional volatility trade-off weakens in high-volatility states of the world. In a companion paper, Moreira and Muir (2016) show that volatility timing is optimal for a very wide range of investors, both short-term and long-term investors, and volatility timing leads to substantial utility gains on the order of 35% of lifetime utility. These results are, however, subject to some important caveats. First, the results are sensitive to the volatility measurement horizon. The authors use a relatively short horizon of one month. Second, the authors use *realized* volatility as opposed to forward-looking *implied volatility*. As a result of this, they do not take into account the variance risk premium, the compensation that investors are willing to pay in order to hedge their exposure to volatility. In a recent paper, Martin (2017) uses implied—as opposed to realized—variance and shows that expected returns over the next year are high when *implied* volatility is high.

The recommendation for volatility timing implied by Moreira and Muir (2016) and Moreira and Muir (2017) poses a formidable puzzle, because it is not only at odds with the conventional view described above but it is also somewhat problematic from the perspective of standard general equilibrium asset pricing models. This is because standard equilibrium models imply countercyclical risk aversion, countercyclical risk premia, and countercyclical volatility.

The main objective of this chapter is to rationalize the volatility-managed trading strategy and, by doing so, reconcile the puzzle posed by Moreira and Muir (2016) and Moreira and Muir (2017). To this end, I develop a heterogeneous agent model cast in continuous time and study the model-implied pure exchange economy with heterogeneous agents. Unconstrained agents endowed with logarithmic utility over instantaneous consumption and investors following a volatility-managed investment strategy are the two different types of agents that populate the model economy. While the unconstrained investors admit the interpretation of sophisticated financial market participants, the volatility timers map well to fund managers with an investment mandate to time and manage volatility.

It is instructive to note that the modeling approach I take is very different from the modeling approach in Moreira and Muir (2017). In Moreira and Muir (2017), the authors take a partial equilibrium perspective, and I propose a general equilibrium model featuring heterogeneous agents. Whereas in the partial equilibrium setups of Moreira and Muir (2017) and Moreira and Muir (2016) volatility timers are price takers, the model I propose allows me to study the general equilibrium implications of volatility timing and also take into account equilibrium feedback effects. Moreira and Muir (2016) exogenously postulate the price and volatility dynamics. In my model, they are determined in equilibrium. I am particularly interested in analyzing how the volatility-managed portfolios affect the unconstrained investors, the centerpiece of my model. Additionally, the model allows me to study how results change when the relative size of the industry engaged in volatility timing changes.

The main theoretical contribution of this chapter is threefold. First, my model goes a long way towards the rationalization of the volatility-managed trading strategy (i.e., the risky asset position inversely proportional to the level of volatility). Namely, I show that the volatility-managed trading strategy constitutes a profitable trading strategy in my model economy, so long as some mild parameter restrictions are satisfied. This is because, in my model economy, the model-implied volatility and the risk premium are negatively correlated over the entire state space, and the risk-return trade-off, as measured by the Sharpe ratio, deteriorates in high-volatility states. Thus, an investor leaving the market in high-volatility states does not sacrifice the opportunity of earning a high risk premium. On the other hand, an investor that loads on the market in low-volatility states earns very high expected returns in risk-adjusted terms.

Second, I argue that the parametric restriction that delivers the above results is likely to hold in reality. For volatility timing to be profitable in my model economy, the fund management fee charged by volatility-managed portfolios should be lower than the subjective discount factor of the unconstrained investors. This is likely to hold in reality for the following reason. Volatility-managed portfolios effectively follow a passive investment strategy that is easy and cheap to implement. Fund managers following passive investment strategies are only able to charge very low fund management fees, and their fees are usually on the order of a few dozen basis points (i.e., very close to zero). Given that the subjective discount factor (i.e., the consumption-to-wealth ratio) of the unconstrained investor is a positive number, the above parameter restriction is likely to hold.

Third, I show that my model allows for rich equilibrium dynamics. A parameter constellation—although not an entirely realistic one—allows me to generate positively correlated conditional volatility and expected return dynamics. In this special case, the risk-return trade-off, as measured by the Sharpe ratio, improves in high-volatility states. This particular equilibrium outcome is consistent with the predictions of standard consumption-based homogeneous agent general equilibrium models that imply countercyclical risk premia.

Additionally, volatility timers exert positive externalities on the unconstrained agents, the centerpiece of my model. Being insensitive to the risk-return trade-off, volatility timers only use conditional volatility as an input to their portfolio construction process. This trading behavior, on the part of the volatility timers, creates distortions that the unconstrained investors can exploit to their advantage.

Finally, the model allows me to study how the volatility timers affect equilibrium quantities of interest for different relative sizes of the asset management industry explicitly engaged in volatility timing.

The theoretical results that I report are subject to some caveats. That is to say, the model-implied leverage of the unconstrained investor is counterfactually high over the entire state space.

Related Literature

This paper closely relates to three different strands of the literature. First, it is related to the literature on portfolio management and volatility management. Second, it is related to the literature studying the effects of portfolio insurance in general equilibrium. Third and finally, it also relates to the literature on equilibrium models with heterogeneous agents.

In a classic paper, Merton (1971) solves an intertemporal portfolio choice problem in continuous time. In his model, both the drift and diffusion coefficients of the risky asset are constant. Campbell and Viceira (1999) study optimal portfolio choice when returns are time varying. Notwithstanding that the literature that studies time variation in expected returns is well developed, the effect of time variation in second moments on equilibrium portfolio choice has to a large extent resisted formal theoretical treatment. In one of the very few academic studies on the topic, Chacko and Viceira (2005) show that the effect of time variation in volatility on hedging demand, and hence on equilibrium portfolio choice, can be sizeable.

In a series of recent papers, Moreira and Muir (2017) and Moreira and Muir (2016) empirically show that trading strategies devised in the spirit of volatility management deliver high Sharpe ratios and high risk-adjusted alphas. They further show that conditional volatility and expected returns are only weakly correlated, and the correlation weakens in high-volatility states. In a follow-up paper, Moreira and Muir (2016), the authors show that volatility management is optimal for a wide range of investors, both short term and long term.

Conceptually, the volatility-managed portfolios are the modern-day reincarnation of portfolio insurance, a dynamic hedging strategy developed by Hayne Leland, John O'Brien, and Mark Rubinstein in the 1970s. Portfolio insurance gained popularity in the 1980s as a way to protect investment portfolios against market downturns. Two prominent papers, Grossman and Zhou (1996) and Basak (1995), use martingale techniques to study portfolio insurance in a general equilibrium context. In their classic paper, Grossman and Zhou (1996) show that portfolio insurance increases price volatility, induces mean reversion in asset returns, and increases the Sharpe ratio and volatility in bad states of the world. There are similarities between these two papers on portfolio insurance and my paper. In the papers on portfolio insurance, portfolio insurers exit the market when the market drops. In my paper, volatility-managed portfolios exit the market when volatility rises. To the extent that market declines and volatility increases are perfectly correlated, portfolio insurance is similar to volatility timing.

My paper is also closely related to the literature on general equilibrium asset pricing models. In their paper, Danielsson, Shin, and Zigrand (2010) study endogenous risk in a general equilibrium model with heterogeneous agents. They show that risk-neutral investors who are subject to a value-at-risk (VaR) constraint effectively behave as if they are risk averse. This insight is instrumental in the construction of the volatility timers that populate my model economy. In contrast to my paper, the model economy in Danielsson, Shin, and Zigrand (2010) does not feature unconstrained investors. Instead, a group of passive investors faces a downwardsloping demand curve. Rytchkov (2014) studies the equilibrium effects of dynamic margin constraints in a dynamic heterogeneous agent economy cast in continuous time. In his paper, Rytchkov (2014) shows that a dynamic margin constraint, where the tightness of the constraint is proportional to the volatility of the total return process, results in a portfolio choice that is consistent with the principles of volatility management. The main difference between my paper and that of Rytchkov (2014) is that the volatility limit in my model binds over the entire state space. More importantly, the model proposed in Rytchkov (2014) is much less tractable compared to my model. Whereas Rytchkov (2014) resorts to numerical solution techniques, I derive all equilibrium quantities of interest in closed form. My paper is also related to Basak and Pavlova (2013), who study the effects of institutional investors on asset prices in equilibrium. Retail and institutional investors are the two types of agents that populate the model economy in Basak and Pavlova (2013). Whereas retail investors are endowed with logarithmic utility over terminal wealth, the marginal utility of institutional investors is increasing in the level of the stock market index. Interestingly, the optimal portfolio policy function of the institutional investors in Basak and Pavlova (2013) features a component that very much resembles a volatility-managed portfolio. This component is, however, small in magnitude, a fact that renders the model proposed by Basak and Pavlova (2013) impractical for the study of volatility management in equilibrium.

The remainder of the chapter is organized as follows. Section 1.2 discusses the economic setup. In particular, I describe the agents of the model, describe the assets available for trading, and solve for the optimal portfolio policies of the agents. I conclude the section by elaborating on the equilibrium conditions and previewing the model solution approach. In Section 1.3, I analytically solve the heterogeneous agent model. After solving the model, I discuss some of the main results and provide economic intuition. In Section 1.4, I analyze the equilibrium outcome. In Section 1.5, I report the equilibrium outcome for the counterfactual case in which my baseline parametric restriction is not satisfied. In Section 1.6, I put my results in a broader perspective and discuss how they relate to some prominent papers on volatility management. Section 1.7 concludes. Finally, in the mathematical Appendix, I provide detailed derivations of all results from the main body of the chapter.

1.2. Model

I consider an infinite-horizon heterogeneous agent pure exchange economy cast in continuous time. The economy is populated by two distinct types of agents, dynamic volatility timers and unconstrained investors optimizing over instantaneous consumption.

The dynamic volatility timers follow a volatility-managed trading strategy. Namely, they increase their risky asset exposure in low-volatility states and decrease their risky asset exposure in high-volatility states. As I show below, volatility timing constitutes an optimal portfolio strategy from the vantage point of a risk-neutral investor subject to a risk limit (volatility budget) that is reminiscent of a value-at-risk (VaR) constraint. Given that they are risk neutral, the volatility timers populating the model economy admit the interpretation of trading desks at a large financial intermediary operating under a value-at-risk (VaR) constraint. It is also possible to map the volatility timers of the model to investment management funds with an investment mandate stipulating the maintenance of a certain volatility level (volatility budgeting) or even outright volatility timing. Risk parity funds, for example, Bridgewater All Weather, AQR Risk Parity Fund, and Invesco Balanced Risk Allocation Fund, are prominent examples of the former, and volatility-managed portfolios, for example, Goldman Sachs Global Markets Navigator Fund, and AllianceBernstein Dynamic Asset Allocation Portfolio, are examples of the latter.

The unconstrained investors, the centerpiece of my model, are the second group of agents populating the model economy. In the model, the unconstrained investors are endowed with logarithmic utility over instantaneous consumption and are unconstrained in their portfolio choice. It is instructive to note that one can map the unconstrained investors to the group of sophisticated and unconstrained financial market participants with relatively long investment horizons or to asset/wealth managers with long-term discretionary investment mandates. It is one of the main objectives of this chapter to study how the volatility-managed portfolios affect unconstrained investors.

I solve the model in terms of the wealth share of the unconstrained investor, which is the state variable of the model. My model is unusually tractable for a model that belongs to the class of heterogeneous agent models. That all equilibrium quantities of interest are available in closed form allows me to present the theoretical results of the chapter in a very accessible and straightforward manner.

1.2.1. Assets

Here, I describe the financial assets on offer to the agents of the model. There is a risk-free asset with an instantaneous rate of return equal to $r(\cdot)$. The risk-free asset is in zero-net supply, and I solve for the endogenous risk-free interest rate. Below, I show that $r(\cdot)$ is a function of the state variable of the model, which is the wealth share of the unconstrained investor. There is also a single risky asset that is a claim to the dividend stream, $\{D_t\}$. The risky asset admits the interpretation of a dividend-paying common stock. Following convention in the literature, I assume that the risky asset is in a positive net supply of one unit. The stochastic differential equation

$$\frac{dD_t}{D_t} = \mu_D dt + \sigma_D dB_t \tag{1.1}$$

governs the dividend growth process, $\{D_t\}$. The drift, $\mu_D \in \mathbb{R}_{++}$, and the diffusion, $\sigma_D \in \mathbb{R}_{++}$, coefficients are positive and exogenous constants. The diffusion coefficient, σ_D, measures the amount of *fundamental* risk in the economy. Clearly, dividend growth is *i.i.d.* as implied by the above stochastic differential equation. $\{B_t\}$ is a standard one-dimensional Wiener process.[2]

[2]The Wiener process, $\{B_t\}$, is defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$. I denote by $\mathbf{F}$ the augmented filtration generated by the Wiener process. The filtered probability space

Let S_t be the time-t price of the risky asset. Then, the total return process, $\{R_t\}$, follows the drift-diffusion process

$$dR_t = \frac{D_t dt + dS_t}{S_t},$$

where $D_t/S_t dt$ admits the interpretation of income gain (dividend-income) per unit of time-dt, and dS_t/S_t is capital gain. In the following sections, I derive an expression for the endogenous price-dividend ratio, S_t/D_t, of the risky asset and show that it is a function of the state variable of the model.

In the model, the Brownian shocks driving the dividend growth process are the only source of randomness. Consequently, as I am going to show in the sequel, the dividend process, $\{D_t\}$, and the risky asset price process, $\{S_t\}$, are driven by the same Brownian motion. I conjecture, and will later on verify the conjecture, that the total return process follows a drift-diffusion,

$$dR_t = \mu_R(\cdot)dt + \sigma_R(\cdot)dB_t,$$

where $\mu_R(\cdot)$ and $\sigma_R(\cdot)$ are the equilibrium drift and diffusion coefficients, respectively. I allow all agents of the model to trade both the risk-free and the risky assets.

1.2.2. Volatility-Managed Portfolios

Here, I introduce volatility-managed portfolios that are isomorphic to the ones implied by the principles of volatility timing.

One possible way to generate a managed volatility portfolio is to solve a dynamic constrained portfolio optimization problem from the vantage point of a short-horizon myopic risk-neutral investor subject to a risk limit (volatility budget). I denote the time-t wealth of the risk-neutral investor by $\tilde{W}_t$, her instantaneous consumption by

satisfies the *usual conditions*; that is, the filtration is complete and right-continuous. A process in this chapter is by definition a stochastic process that is progressively measurable with respect to $\{\mathcal{F}_t\}$.

$\tilde{C}_t$, and I assume that the agent consumes a constant fraction, ρ_v, of her wealth,

$$\tilde{C}_t = \rho_v \tilde{W}_t.$$

It is sensible to assume that the instantaneous consumption of the fund manager equals a constant fraction of wealth under management, where the latter admits the interpretation of a fund-management fee levied on assets under management (exogenous compensation contract). Let $\tilde{\theta}_t$ be the portfolio policy of the fund manager expressed as a fraction of time-t wealth. Then, the stochastic differential equation

$$d\tilde{W}_t = (r_t \tilde{W}_t - \tilde{C}_t)dt + \tilde{\theta}_t \tilde{W}_t (dR_t - r_t dt)$$

fully characterizes the intertemporal wealth evolution of the volatility timer. The first term on right-hand side, $r_t \tilde{W}_t dt$, is the instantaneous risk-free return of the investor, and $\tilde{\theta}_t \tilde{W}_t$ is the dollar size of her risky asset position. Using the fact that the consumption rate is constant, the stochastic differential equation simplifies to

$$d\tilde{W}_t = (r_t - \rho_v)\tilde{W}_t dt + (\mu_R(\cdot) - r_t)\tilde{\theta}_t \tilde{W}_t dt + \sigma_R(\cdot)\tilde{\theta}_t \tilde{W}_t dB_t,$$

where $\mu_R(\cdot)$ and $\sigma_R(\cdot)$ are the conjectured drift and diffusion coefficients of the total return process, $\{R_t\}$. Clearly, the portfolio policy, $\tilde{\theta}_t$, is the only choice variable of the volatility timer. The myopic volatility timer maximizes next-period consumption, $\tilde{C}_{t+dt}$, subject to a volatility budget, and to a dynamic budget constraint,

$$\begin{aligned}
\max_{\{\tilde{\theta}_t\}} \quad & \mathbb{E}_t \left(\tilde{C}_{t+dt} \right), \\
\text{s.t.} \quad & \tilde{\theta}_t \geq 0, \\
& \beta \sqrt{\frac{\mathbb{V}\mathrm{ar}_t(d\tilde{W}_t)}{dt}} \leq \tilde{W}_t, \\
& d\tilde{W}_t = (r_t \tilde{W}_t - \tilde{C}_t)dt + \tilde{\theta}_t \tilde{W}_t (dR_t - r_t dt),
\end{aligned}$$

where the risk limit (volatility budget) is the second constraint, and the dynamic budget constraint is the third one. The long-only restriction, $\tilde{\theta}_t \geq 0$, squares well with the stylized fact that volatility-managed portfolios tend to be structured as long-only investment vehicles.

The risk limit (volatility budget) very much resembles a value-at-risk (VaR) constraint. Namely, β-times the risk of the portfolio, as measured by the forward-looking standard deviation of portfolio returns, admits the interpretation of value-at-risk. Parameter $\beta \in \mathbb{R}_{++}$ is an exogenous parameter controlling the tightness of the constraint. Clearly, the tightness of the constraint is increasing in parameter β. The exogeneity of this parameter is without loss of generality. It is usually imposed on financial intermediaries by regulatory bodies. In the context of asset management, the investment mandate of the investment vehicle establishes what values β should take. The identity

$$\mathbb{V}\text{ar}_t(d\tilde{W}_t) = d\langle \tilde{W}, \tilde{W} \rangle_t$$

allows me to calculate the instantaneous variance of wealth from the quadratic variation of $\{W_t\}$. Following convention, I use the notation $\langle \cdot, \cdot \rangle$ to denote the square-bracket process (quadratic variation). Given the facts that $\tilde{W}_t$ depends on past portfolio choices, $\tilde{C}_t = \rho_v \tilde{W}_t$, and ρ_v is a constant, $\tilde{C}_t$ is pre-determined as of time-t. Consequently,

$$\text{argmax}(\mathbb{E}_t(\tilde{C}_{t+dt})) = \text{argmax}(\mathbb{E}_t(d\tilde{C}_t)) = \text{argmax}(\mathbb{E}_t(d\tilde{W}_t)).$$

Therefore, it is enough for an agent with the objective of maximizing next-period consumption to maximize the increase in the value of assets under management, dW_t. As I disallow for fund flows in the model, capital appreciation is the only way towards the realization of this objective. These considerations allow me to rewrite the portfolio choice problem in a form reminiscent of the model specification in

Danielsson, Shin, and Zigrand (2010),

$$
\begin{aligned}
\max_{\{\tilde{\theta}_t\}} \quad & \mathbb{E}_t\left(\frac{d\tilde{W}_t}{dt}\right), \\
\text{s.t.} \quad & \tilde{\theta}_t \geq 0, \\
& \beta\sqrt{\frac{\mathbb{V}\mathrm{ar}_t(d\tilde{W}_t)}{dt}} \leq \tilde{W}_t, \\
& d\tilde{W}_t = (r_t\tilde{W}_t - \tilde{C}_t)dt + \tilde{\theta}_t\tilde{W}_t(dR_t - r_t dt).
\end{aligned}
$$

The constraints are the same as above. The only difference between the two optimization problems is in the objective function. Here, I maximize expected wealth growth as opposed to next-period consumption.

Absent the risk limit (volatility budget) and under the assumption that the expected total return on the risky asset is positive, the risk-neutral investor sets $\tilde{\theta}_t = \infty$ and, by doing so, prices the unconstrained investor out of the market. In the presence of a risk limit (volatility budget), however, this strategy is no longer feasible, and the agent behaves as if she is risk averse. Namely, the volatility timer sets the size of her risky asset position in such a way so as not to violate her volatility budget. Additionally, the tightness of the constraint is inversely proportional to the endogenous wealth level, $\tilde{W}_t$, of the agent. Negative shocks to wealth (assets under management), $\tilde{W}_t$, erode the capital position of the agent and inhibit her ability to take large risky asset positions.

Expected wealth growth and its conditional variance immediately follow from the stochastic differential equation governing the intertemporal wealth evolution of the wealth of the volatility timer,

$$
\begin{aligned}
\mathbb{E}_t(d\tilde{W}_t) =& (r_t - \rho_v)\tilde{W}_t dt + (\mu_R(\cdot) - r_t)\tilde{\theta}_t\tilde{W}_t dt, \\
\mathbb{V}\mathrm{ar}_t(d\tilde{W}_t) =& (\sigma_R(\cdot)\tilde{\theta}_t\tilde{W}_t)^2 dt.
\end{aligned}
$$

Below, I show that in the model the risk premium on the risky asset is positive over the entire state space and the risky asset earns a *positive* risk premium. This fact immediately implies that the constrained portfolio optimization problem of the risk-neutral investor admits the solution:

$$\tilde{\theta}_t = \frac{1}{\beta}\frac{1}{\sigma_R(\cdot)}.$$

This is because $\tilde{\theta}_t$ only takes non-negative values, and the constraint associated with the risk limit(volatility budget),

$$\beta\sqrt{\frac{\mathbb{V}\mathrm{ar}_t(d\tilde{W}_t)}{dt}} \leq \tilde{W}_t,$$

binds with equality. It is convenient to define $\bar{\sigma} = 1/\beta$. The newly defined parameter, $\bar{\sigma}$, admits the interpretation of a risk limit. Clearly, the tightness of the risk limit is decreasing in $\bar{\sigma}$. For high values of $\bar{\sigma}$, the agent can take high leverage. Using the newly defined parameter, I can rewrite $\tilde{\theta}_t$,

$$\tilde{\theta}_t = \frac{\bar{\sigma}}{\sigma_R(\cdot)}.$$

I assume that the principal of the constrained risk-neutral investor sets the exogenous risk limit in such a way so as to satisfy the inequality $\bar{\sigma} < \sigma_D$. This parameter restriction inhibits the ability of the volatility timer to maintain a high leverage ratio. It is, however, without loss of generality and broadly in line with prevailing practice in the fund management industry. Funds with investment mandates to time volatility tend to maintain low leverage ratios, usually below one.

It is important to keep in mind that $\tilde{\theta}_t$ expresses the portfolio choice of the agent as a fraction of total wealth. Consequently, the dollar size of the risky asset position of the agent is

$$\frac{\bar{\sigma}}{\sigma_R(\cdot)}\tilde{W}_t.$$

The portfolio choice of the risk-neutral agent facing a risk limit (volatility budget) is a valid volatility-managed portfolio. This is because the size of the risky asset position is inversely proportional to the level of volatility of the risky asset. More importantly, the expected return on the risky asset, $\mu_R(\cdot)$, appears nowhere in this expression. This is the defining characteristic of the volatility-managed portfolio. At the same time, it is also the main point of difference between a myopic mean-variance portfolio and a volatility-managed portfolio. Additionally, the size of the risky asset position is inversely proportional to the tightness of the constraint. Finally, the volatility-managed portfolio depends on the wealth level, $\tilde{W}_t$, of the agent. In this respect, the volatility-managed portfolio resembles the optimal portfolio of an investor endowed with logarithmic utility, or the portfolio of a constant relative risk aversion (CRRA) investor, if we were to abstract from the hedging demand term. The agent is nominally risk neutral, but behaves as if risk averse, because of the constraint. The effective risk aversion is inversely proportional to the wealth level.

1.2.3. Unconstrained Investors

Here, I introduce the group of unconstrained logarithmic utility investors. They admit the interpretation of sophisticated financial players or asset managers with flexible (discretionary) long-term investment mandates. The unconstrained investors are the centerpiece of my model. One of my main objectives is to study how the unconstrained and fully rational investors are affected by the volatility-managed portfolios that they co-exist with in the model economy. In equilibrium, prices have to adjust so that the unconstrained investors are happy to take the other side of the trade. The unconstrained investors are identical and form a continuum of measure one. The representative unconstrained investor solves the intertemporal consumption-portfolio choice problem,

$$\max_{\{C_t,\theta_t\}} \quad \mathbb{E}\left(\int_0^\infty e^{-\rho t} u(C_t) dt\right) \quad \rho > 0,$$

subject to the dynamic constraint

$$dW_t = r_t W_t dt - C_t dt + \theta_t W_t (dR_t - r_t dt), \tag{1.2}$$

where the stochastic differential equation governs the intertemporal wealth evolution (between time t and time $t + dt$) of the representative unconstrained agent and

$$u(C_t) = \ln(C_t), \quad C_t > 0.$$

The first term on the right-hand side of (1.2) is the instantaneous return on the investment in the risk-free asset. The second and the third terms are instantaneous consumption and risky asset excess return, respectively. Additionally, $\rho \in \mathbb{R}_{++}$ is the subjective discount factor of the agent, W_t is wealth, and $dR_t - r_t dt$ is the total *excess* return on the risky asset. Finally, θ_t is the portfolio choice expressed as a fraction of total wealth, W_t, and $\theta_t W_t$ is the dollar size of the risky asset position.

In the main version of the model, I assume that $\rho > \rho_v$. Below, I am going to argue that this inequality is likely to hold in reality. While ρ is the consumption-to-wealth ratio of the unconstrained investor, ρ_v admits the interpretation of the fund management fee that the volatility-managed portfolios charge. Given that in its simplicity the volatility-managed strategy resembles a passive trading strategy, it is not unrealistic to assume that volatility-managed portfolios are only able to charge very low fund management fees on the order of a few dozen basis points. So long as volatility management portfolios charge a low fund management fee, ρ_v will be positive but will be very close to zero, and $\rho > \rho_v$ is likely to hold.

The unconstrained investor chooses her portfolio at time t and rebalances it at time $t + dt$ should there be a need to do so. For convenience purposes, instead of directly solving for the optimal portfolio policy, θ_t, of the unconstrained investor I derive it from the market clearing condition. The optimal portfolio choice admits the representation

$$\theta_t = \frac{1}{x_t} - \frac{\bar{\sigma}}{\sigma_R(\cdot)} \frac{1 - x_t}{x_t},$$

where

$$x_t \doteq \frac{W_t}{S_t}$$

is the state variable of the model, which is the wealth share of the unconstrained investor. Clearly, in a pure exchange economy $x_t \in [0, 1]$. In the following sections, I explicitly solve for $\sigma_R(\cdot)$ and show that it is a function of the state variable, x_t.

1.2.4. Equilibrium Conditions

In this subsection, I formally define the equilibrium concept used to solve the model. After enlisting all equilibrium conditions, I outline the model solution strategy that I follow in the following sections.

Definition 1. *An equilibrium is a set of price processes and investment policies* $\{\theta(\cdot), \tilde{\theta}(\cdot)\}$ *such that the investment policies solve the dynamic portfolio optimization problems of the volatility timer and of the unconstrained investor, respectively.*

1. *Given the price process, the unconstrained investor and the volatility timer solve their respective portfolio optimization problems.*

2. *The unconstrained investor is unconstrained in its portfolio choice and the risk-neutral volatility timer faces a risk limit (volatility budget).*

3. *The goods market clears,*

$$C_t + \tilde{C}_t = D_t.$$

4. The market for the risky asset clears,

$$\theta_t W_t + \tilde{\theta}_t \tilde{W}_t = S_t.$$

5. The market for the risk-free asset clears by Walras' law.

Given that I model a pure exchange Lucas (1978) economy, it should be the case that in equilibrium total wealth equals the price of the risky asset,

$$W_t + \tilde{W}_t = S_t.$$

Equipped with the above equilibrium conditions, I solve for the equilibrium outcome in terms of the state variable of the model, $x_t \in [0, 1]$, which is the wealth share of the unconstrained investor. It is instructive to note that the model I propose delivers more tractability than what is typical for the class of equilibrium models with heterogeneous agents.

1.3. Characterization of Equilibrium

In this section, I solve for the main version of the model. The logarithmic utility solution is very tractable and allows me to present some of the main results of the chapter in a very accessible way. After fully characterizing the equilibrium, I thoroughly analyze the equilibrium outcome and provide economic intuition.

1.3.1. Main Results

Given that the unconstrained investor is endowed with logarithmic utility over instantaneous consumption, it is optimal for her to consume a constant fraction, ρ, of her wealth. Therefore,

$$C_t = \rho W_t.$$

In the logarithmic utility case, the consumption-to-wealth ratio, ρ, is also equal to the subjective discount factor of the agent.

The interplay between the diffusion coefficient of the total return process, $\sigma_R(\cdot)$, and the equilibrium risk premium sways the performance of any trading strategy devised in the spirit of volatility timing. Consequently, the analysis concerned with the study of volatility timing in the context of general equilibrium necessitates the derivation of these two equilibrium quantities. The following proposition provides an equilibrium expression for the diffusion coefficient.

Proposition 1. (Model-Implied Total Return Process Volatility)
The volatility of the total return process, $\sigma_R(\cdot)$, is given by

$$\sigma_R(x_t) = \frac{1}{1-A}(\sigma_D - \bar{\sigma}A) + \frac{A}{1-A}(\bar{\sigma} - \sigma_D)x_t, \tag{1.3}$$

where x_t is the state variable of the model, σ_D and $\bar{\sigma}$ are exogenous constants, and

$$A \doteq 1 - \frac{\rho}{\rho_v}.$$

Please refer to the mathematical Appendix for a detailed proof of the proposition. The total return process volatility depends on the quantity of fundamental risk in the economy, σ_D, on the risk limit imposed on the risk-neutral investor, $\bar{\sigma}$, on the ratio of the consumption-to-wealth ratios, ρ_v/ρ, through A, and on the state variable of the model, which is the wealth share of the unconstrained investor, $x_t = W_t/S_t$.

In the following proposition, I summarize some of the main properties of the volatility of the total return process.

Proposition 2. (Properties of Total Return Process Volatility)

- *The volatility of the total return process, $\sigma_R(x_t)$, is increasing in the wealth share of the unconstrained investor.*

- *The sensitivity of $\sigma_R(x_t)$ to x_t is increasing in the wedge between $\bar{\sigma}$ and σ_D.*

Please refer to the mathematical Appendix for a detailed proof of the proposition. The expression for volatility admits a very intuitive decomposition: a term that only depends on exogenous parameters and a second term that is a function of the state variable of the model, x_t. Given that $\bar{\sigma} - \sigma_D$ is negative by construction, the sign of the loading on x_t depends on the sign of A, which in turn depends on ρ and ρ_v, the consumption-to-wealth ratios of the unconstrained investor, and the volatility-managed portfolio, respectively. For $\rho > \rho_v$ (the case I consider in this section), the volatility of the risky asset is increasing in the wealth share of the unconstrained investor. In other words, low-volatility states are states in which the unconstrained investor is undercapitalized, and the agent who is engaged in volatility timing owns most of the wealth in the economy. On the other hand, states in which the unconstrained investor owns most of the wealth in the economy are characterized by high levels of volatility and admit the interpretation of adverse states of the world. The economic intuition is as follows. When $\rho > \rho_v$, an increase in x_t leads to an increase in the share of impatient agents in the economy, and this increases volatility.

Interestingly, the volatility of the total return process, $\sigma_R(\cdot)$, is linear in the state variable of the model, which is the wealth share of the unconstrained investor. More importantly, the sensitivity of $\sigma_R(\cdot)$ with respect to the state variable depends on the wedge between fundamental risk in the economy, σ_D, and the risk limit imposed on the volatility timer, $\bar{\sigma}$. The bigger the wedge, the higher the sensitivity. When the discrepancy between the two is large, the volatility-managed portfolios can take large positions in the risky asset. Consequently, even small changes in the wealth share lead to large portfolio rebalancing, and this heightens volatility.

The unconstrained investor does not face any portfolio constraints and is always marginal in the market for the risky asset. This allows me to directly derive the risk premium on the risky asset from the Euler equation of the unconstrained investor,

$$-\rho dt - \mathbb{E}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{V}\mathrm{ar}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{E}_t(dR_t) = \mathbb{C}\mathrm{ov}_t\left(\frac{dC_t}{C_t}, dR_t\right).$$

The Euler equation holds for any tradable asset. Consequently, the corresponding Euler equation for the risk-free asset admits the following representation:

$$r_t dt = \rho dt + \mathbb{E}_t\left(\frac{dC_t}{C_t}\right) - \mathbb{V}\mathrm{ar}_t\left(\frac{dC_t}{C_t}\right).$$

Subtracting the expression for $r_t dt$ from the expression for $\mathbb{E}(dR_t)$, I obtain an expression for the risk premium on the risky asset. It is a very fundamental result that under logarithmic utility consumption growth equals wealth growth, $dC_t/C_t = dW_t/W_t$. Consequently, the risk premium on the risky asset takes the simple form

$$\begin{aligned}
\mathbb{E}_t(dR_t) - r_t dt =& \mathbb{C}\mathrm{ov}_t\left(\frac{dW_t}{W_t}, dR_t\right), \\
\mathbb{E}_t(dR_t) - r_t dt =& \theta_t \mathbb{V}\mathrm{ar}_t(dR_t),
\end{aligned}$$

where $\mathbb{V}\mathrm{ar}_t(dR_t) = \sigma_R^2(x_t)dt$, and the equilibrium expression for $\sigma_R(x_t)$ follows from (1.3) above. The following proposition reports the risk premium on the risky asset.

Proposition 3. (Risk Premium on the Risky Asset)
The risk premium on the risky asset is given by

$$\mathbb{E}_t(dR_t) - r_t dt = \left(\frac{1}{x_t} - \frac{\bar{\sigma}}{\sigma_R(x_t)}\frac{1-x_t}{x_t}\right)\sigma_R^2(x_t)dt.$$

Under the parameter restriction $\bar{\sigma} < \sigma_D$, the risk premium takes positive values over the entire state space. The risk premium is decreasing in x_t to the left of $\hat{x}$ and increasing in x_t to the right of $\hat{x}$, where

$$\hat{x} \doteq \sqrt{\frac{\bar{\sigma}A - \sigma_D}{A(\bar{\sigma} - A\sigma_D)}}.$$

Please see the mathematical Appendix for a detailed proof of the proposition. The risk premium on the risky asset takes positive values over the entire state space and exhibits non-linear behavior with respect to the volatility of the total return process, $\sigma_R(x_t)$. The shape of the risk premium depends on the tightness of the risk limit, $\bar{\sigma}$. For very low values of $\bar{\sigma}$ (very tight volatility budget), the second term in brackets goes to zero and the risk premium is proportional to the variance of the risky asset. So long as the increase in the denominator of the coefficient $1/x_t$ is not big enough to offset the increase in variance, the risk premium is increasing in variance.

The analysis of the joint dynamics of the model-implied volatility of the total return process and the risk premium is one of the main objectives of this paper. This analysis necessitates the derivation of the quadratic co-variation between the risk premium and the diffusion coefficient of the total return process. The following proposition reports the quadratic co-variation implied by the model.

Proposition 4. (Quadratic Co-variation)
The quadratic co-variation between the risk premium on the risky asset and its conditional volatility is given by

$$d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t = \left(\tilde{A}_1(x_t)\frac{A}{1-A}\frac{1}{x_t^2} + \tilde{A}_2(x_t)\frac{1}{x_t}\right) d\langle x, x\rangle_t,$$

where the coefficients $\tilde{A}_1(x_t)$, $\tilde{A}_2(x_t)$, and the quadratic variation, $d\langle x, x\rangle_t$, are given by

$$\begin{aligned}
\tilde{A}_1(x_t) &\doteq \sigma_R(x_t)(\bar{\sigma}-\sigma_D)(\bar{\sigma}-\sigma_R(x_t)),\\
\tilde{A}_2(x_t) &\doteq (2\sigma_R(x_t) - \bar{\sigma}(1-x_t))\left(\frac{(\bar{\sigma}-\sigma_D)A}{1-A}\right)^2,\\
d\langle x, x\rangle_t &= ((\theta_t - 1)\sigma_R(x_t)x_t)^2 dt.
\end{aligned}$$

Please refer to the mathematical Appendix for a detailed proof of the proposition. Intuitively, the quadratic co-variation between volatility and the risk premium admits the interpretation of instantaneous covariance. Thus, the sign thereof is of particular interest for the analysis of volatility timing. A positive sign for the quadratic co-variation violates the most fundamental assumption put forward by the proponents of volatility timing and renders their arguments, related to the profitability of volatility timing as a trading strategy, untenable. Below, I check the sign of the quadratic co-variation over the entire state space.

Proposition 5. (Properties of Quadratic Co-variation)
The sign of the quadratic co-variation depends on the sign of

$$\operatorname{sgn}(d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t) = \operatorname{sgn}\left(\tilde{A}_1(x_t)\frac{A}{1-A}\frac{1}{x_t^2} + \tilde{A}_2(x_t)\frac{1}{x_t}\right).$$

- *The sign of the quadratic co-variation is negative to the left of $\hat{x}$ and positive to the right of $\hat{x}$, where*

$$\hat{x} \doteq \sqrt{\frac{\bar{\sigma}A - \sigma_D}{A(\bar{\sigma} - \sigma_D A)}}.$$

- *Under the parameter restriction,*

$$\left(1 - \frac{\rho}{\rho_v}\right)^2 \leq 1,$$

  $\hat{x} \geq 1$ *and the quadratic co-variation is negative over the entire state space.*

Please refer to the mathematical Appendix for a detailed proof of the proposition. Given that $\hat{x}$ is always positive, there are two distinct regions of the state space. For $x \in [0, \hat{x}]$, the quadratic co-variation is negative. For $x \in (\hat{x}, 1]$, conditional volatility and the risk premium are positively correlated. The case in which $\hat{x} \geq 1$ and the quadratic co-variation is negative over the entire state space is of particular interest. This inequality is satisfied when $A^2 \leq 1$, that is, when ρ is not too far away from ρ_v.

The sensitivities that I report in the proposition below are useful in analyzing the equilibrium.

Proposition 6. (Portfolio Sensitivities)
The sensitivities of the portfolio of the volatility timer with respect to the state variable, x_t, and with respect to $\sigma_R(x_t)$ are as follows:

- *Delta with respect to x_t*

$$\Delta_x \doteq \frac{\partial \tilde{\theta}_t}{\partial x_t} = -\frac{\bar{\sigma}}{\sigma_R^2(x_t)} \frac{A}{(1-A)}(\bar{\sigma} - \sigma_D).$$

- *Gamma with respect to x_t*

$$\Gamma_x \doteq \frac{\partial^2 \tilde{\theta}_t}{\partial x_t^2} = \frac{2\bar{\sigma}(A(\bar{\sigma} - \sigma_D))^2}{\sigma_R^3(x_t)(1-A)^2}.$$

- *Vega with respect to $\sigma_R(x_t)$*

$$\mathcal{V} \doteq \frac{\partial \tilde{\theta}_t}{\partial \sigma_R(x_t)} = -\frac{\bar{\sigma}}{\sigma_R^2(x_t)}.$$

- *Volga (Volatility Gamma) with respect to $\sigma_R(x_t)$*

$$\mathcal{V}^2 \doteq \frac{\partial^2 \tilde{\theta}_t}{\partial \sigma_R^2(x_t)} = \frac{2\bar{\sigma}}{\sigma_R^3(x_t)}.$$

By construction, the risky asset position of the volatility timer, $\tilde{\theta}_t$, is inversely proportional to the level of volatility. This implies a negative vega (calculated with respect to $\sigma_R(x_t)$) over the entire state space. More importantly, the absolute value of the vega is inversely proportional to the instantaneous conditional *variance* of the risky asset. On the other hand, the volga of $\tilde{\theta}_t$ (calculated with respect to $\sigma_R(x_t)$) is always positive and decreasing in the volatility of the total return process. The sign of delta (calculated with respect to x_t) is negative. This is because volatility is increasing in x_t over the entire state space.

1.4. Analysis of Equilibrium (Main Case, $\rho > \rho_v$)

In the model, the volatility on the total rerun process is *increasing* in the wealth share of the unconstrained investor, x_t, over the entire state space. States in which the volatility-managed portfolios own a small fraction of the total wealth in the economy are high-volatility states. On the other hand, the dynamics of the risk premium are much more nuanced. Going back to the expression for the risk premium,

$$\mathbb{E}_t(dR_t) - r_t dt = \frac{1}{x_t}\left(1 - \frac{\bar{\sigma}}{\sigma_R(x_t)}(1 - x_t)\right)\sigma_R^2(x_t)dt,$$

it is immediate to see that $1/x_t$ is decreasing in x_t, and the remaining terms are increasing in x_t. In the vicinity of the lower boundary of the state space, $1/x_t$ dominates, and the risk premium is decreasing in the wealth share of the unconstrained investor. In the region of the state space, where the unconstrained investor owns most of the wealth in the economy, the sign of the derivative of the risk premium with respect to x_t very much depends on the sensitivity of $\sigma_R(x_t)$ to x_t. In Proposition 3, I show that the risk premium is decreasing in x_t to the left of the cutoff

$$\hat{x} = \sqrt{\frac{\bar{\sigma}A - \sigma_D}{A(\bar{\sigma} - A\sigma_D)}}$$

and is increasing in x_t to the right of $\hat{x}$. In the calibration of the model that I will consider below, the quadratic co-variation between conditional volatility and expected returns is negative over the entire state space. This implies that the risk premium is decreasing in x_t over the entire state space.

[Insert Figure 1.1]

[Insert Figure 1.2]

Figures 1.1 and 1.2 offer a convenient graphical representation of my results. Whereas Figure 1.2 plots the equilibrium quantities of interest over the entire state space, Figure 1.1 excludes the region that is in the immediate vicinity of the lower boundary of the state space. Therefore, Figure 1.1 is a zoomed-in version of Figure 1.2. Notwithstanding that all equilibrium quantities are well defined over the entire state space, they take very extreme values in the vicinity of the lower boundary of the state space, where $x_t \to 0$. For this reason, Figure 1.1 is particularly useful in analyzing my results. It plots the volatility of the risky asset, $\sigma_R(\cdot)$, (top-left panel of the figure), the risk premium, $(\mathbb{E}_t(dR_t) - r_t dt)/dt$, (top-right panel), the portfolio choice of the unconstrained investor, θ_t, (bottom-left panel), and the portfolio choice of the volatility timer, $\tilde{\theta}_t$, (bottom-right panel) as functions of the state variable of the model, which is the wealth share of the unconstrained investor. The vertical blue line in the top-right panel of the figure passes through the value of x_t that minimizes the risk premium. Given that in my calibration the risk premium is decreasing in x_t over the entire state space, the blue vertical line passes through $x_t = 1$.

In the figures, the volatility is increasing in x_t over the entire state space (top-left panels). Whereas states in which the unconstrained investor owns most of the wealth in the economy are high-volatility states, those in which the volatility timer owns most of the wealth in the economy are low-volatility states. This result is intuitive. In states in which the volatility timer owns most of the wealth in the economy, volatility should be low in order to induce her to take a large risky asset position and the market for the risky asset to clear. Given that the volatility timer maintains a leverage ratio below one over the entire state space (bottom-right panel), the unconstrained investor is indispensable for market clearing. For the undercapitalized unconstrained investor to be willing to invest in the risky asset and for the market to clear, the risk premium has to be very high in the vicinity of the lower boundary of the state space (top-right panel). In the other extreme, where x_t is close to the upper boundary of the state space, the unconstrained investor occupies the driving seat and the effect of the volatility timer in the price formation process is very limited. This is because for high values of x_t, the wealth of the volatility timer is small, and this limits the size of the risky asset position she can take. The effective risk aversion of the unconstrained investor decreases as her wealth increases. Consequently, in this region of the state space, the agent is content with holding the risky asset even if the risk premium is relatively low.

One of the main objectives of the chapter is to study the joint dynamics of conditional volatility and the risk premium. To this end, in Figure 1.5, I plot the quadratic covariation between volatility and the risk premium.

[Insert Figure 1.5]

In this particular calibration of the model, the parameter restriction $A^2 \leq 1$ is satisfied and the quadratic co-variation takes negative values over the entire state space (please see the top-left and bottom-left panels of the figure). In other words, conditional volatility and expected returns move in the opposite direction. When volatility rises, the risk premium decreases and vice versa. Consequently, any trading strategy that is consistent with the principles of volatility timing constitutes a profitable trading strategy. This is because increases in volatility are not followed by corresponding increases in the risk premium and the risk-return trade-off, as measured by the Sharpe ratio (top-right panel of Figure 1.5), deteriorates in high-volatility states. This is particularly true, when the share of the industry that follows a volatility-managed strategy, $\tilde{W}_t/S_t$, is large relative to the total size of the economy, S_t.

As discussed above, x_t admits the interpretation of the size of the asset management industry with an investment mandate that stipulates volatility timing. In the model, x_t is endogenous. One can, however, conduct a thought experiment. By assigning different values to x_t, one can analyze how the equilibrium changes as a function of the relative size of the sector explicitly engaged in volatility management. In the limit case, where $1-x_t = 0$, the wealth share of the volatility-managed industry is equal to zero, and the heterogeneous agent economy reduces to a homogeneous agent economy solely populated by the unconstrained logarithmic utility investors. By setting $1-x_t$ close to one, one can analyze the equilibrium outcome for the case in which passive investment in general, and volatility timing in particular, dominate the market. It is instructive to note that the volatility timers exert positive externalities on the unconstrained investors. This is because the Sharpe ratio on the risky asset is increasing in the wealth share of the volatility timers. In states in which the volatility-managed industry is large, the few surviving unconstrained investors have the opportunity to earn very high Sharpe ratios.

1.4.1. Comparative Statics

Here, I conduct a comparative statics exercise. In particular, I examine how the equilibrium quantities of interest react to changes in the exogenous parameters of the model.

[Insert Figure 1.10]

[Insert Figure 1.11]

Figure 1.11 and the zoomed-in version of that figure in Figure 1.10 plot the sensitivities of conditional volatility, the risk premium, and the portfolio choice to the tightness of the risk limit, $\bar{\sigma}$. A decrease in $\bar{\sigma}$ tightens the volatility budget of the volatility-managed portfolio. As a result of this, both the conditional volatility and the leverage ratio of the volatility-managed portfolio decrease over the entire state space. Given that for low values of $\bar{\sigma}$ the volatility timer can take smaller risky asset positions (compared to the case in which $\bar{\sigma}$ is high), volatility should decrease more in the region of the state space, where the unconstrained investor is undercapitalized, so that the volatility timer can take a large enough risky asset position and the market for the risky asset to clear.

The curvature of the risk premium increases in the tightness of the risk limit (top-right panel). The volatility timer maintains a leverage ratio that takes values below one over the entire state space (for all values of $\bar{\sigma}$). Consequently, the unconstrained investor has to take a long position in order for the market for the risky asset to clear. The more constrained the volatility timers are, the larger position the unconstrained investor has to take. The unconstrained investor invests in the risky asset only if the risk premium is sufficiently high. For this reason, the curvature of the risk premium is increasing in the tightness of the volatility limit of the volatility-managed portfolio.

1.5. Analysis of Equilibrium for $\rho < \rho_v$

The model that I develop here allows for rich equilibrium dynamics. For the case in which the consumption-to-wealth ratio of the unconstrained investor is lower than the fund management fee that the volatility-managed portfolio charges, $\rho < \rho_v$, the model delivers equilibrium dynamics that are consistent with the predictions of standard heterogeneous agent models. Namely, under this parameter restriction, the conditional volatility implied by my model is positively correlated with expected returns over the entire state space.

The consumption-to-wealth ratio is a good proxy for the degree of patience. An agent with a high consumption-to-wealth ratio admits the interpretation of an impatient agent. On the other hand, an agent with a low consumption-to-wealth ratio admits the interpretation of a patient agent. Using this terminology, the parametric restriction $\rho < \rho_v$ is likely to hold in a market environment, where the unconstrained investors are more patient than are the volatility-managed portfolios.

It is instructive to compare the equilibrium implied by my heterogeneous agent model featuring volatility timers to the equilibrium implied by a heterogeneous agent economy devoid of agents engaged in volatility timing. Below, I refer to the latter economy as the baseline economy (model). This comparison is useful, because the baseline economy very much resembles the economic setup in Longstaff and Wang (2012), a widely cited and standard heterogeneous agent model.

To fill the void resulting from the removal of the volatility timers, I augment the baseline model by adding a new type of agent. In the interest of simplicity, I implicitly characterize the new agent through her portfolio choice, $\theta_{CM,t}$,

$$\theta_{CM,t} = \frac{1}{\gamma_{CM}} \frac{\mu_R(x_t) - r_t}{\sigma_R^2(x_t)},$$

where the abbreviation in the subscript, CM, stands for model of comparison and γ_{CM} is the constant risk aversion coefficient of the new agent. I set $\theta_{CM,t}$ so that it is as comparable as possible to the volatility-managed portfolio,

$$\tilde{\theta}_t = \frac{1}{\tilde{\lambda}_t}\frac{\mu_R(x_t) - r_t}{\sigma_R^2(x_t)} = \frac{\bar{\sigma}}{\sigma_R(x_t)},$$

where $\tilde{\lambda}_t \doteq \lambda_t/\bar{\sigma}^2$, and λ_t is the Lagrange multiplier associated with the volatility budget. Please see the mathematical Appendix for a closed form expression for λ_t. Whereas the volatility-managed portfolio, $\tilde{\theta}_t$, is insensitive to the risk premium, the newly defined portfolio, $\theta_{CM,t}$, is a function thereof. In other words, the agent that augments the baseline model internalizes the risk-return trade-off at the portfolio construction stage. This is the main point of difference between the new agent and the volatility timer. Notably, $\theta_{CM,t}$ resembles a mean-variance portfolio. I opt for this particular parametrization in order to enhance the comparability between the volatility-managed portfolio and $\theta_{CM,t}$. In the proposition that follows, I summarize the main equilibrium quantities of interest that the baseline model implies. Please refer to the mathematical Appendix for a detailed proof of the proposition.

Proposition 7. (Equilibrium, Baseline Model)

The portfolio choice of the unconstrained investor, the volatility of the risky asset, the risk premium, and the quadratic co-variation between the risk premium and volatility in the baseline model are given by

- *Portfolio choice of the unconstrained investor*

$$\theta_t = \frac{\gamma_{CM}}{1 + x_t(\gamma_{CM} - 1)},$$

- *Volatility of the total return process*

$$\sigma_R(x_t) = \sigma_D(1 - x_t\breve{A})\frac{1 + x_t(\gamma_{CM} - 1)}{1 + x_t(\gamma_{CM} - 1) - x_t\gamma_{CM}\breve{A}},$$

  where

$$\breve{A} = 1 - \frac{\rho}{\rho_{CM}}.$$

- *Risk premium*

$$\mathbb{E}_t(dR_t - r_t dt) = \frac{\gamma_{CM}}{1 + x_t(\gamma_{CM} - 1)}\sigma_R^2(x_t)dt,$$

- *Quadratic co-variation between the volatility of the risky asset and the risk premium*

$$d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t = \Psi_1(x_t)\Psi_2(x_t)d\langle x, x\rangle_t,$$

  where

$$\begin{aligned}
\Psi_1(x_t) \doteq & -\frac{\gamma_{CM}(\gamma_{CM}-1)}{(1+x_t(\gamma_{CM}-1))^2}\sigma_R^2(x_t) + \frac{2\sigma_R(x_t)\gamma_{CM}}{1+x_t(\gamma_{CM}-1)}\Psi_2(x_t),\\
\Psi_2(x_t) \doteq & -\sigma_D\breve{A}\tilde{B} + \frac{\gamma_{CM}\breve{A}\tilde{A}}{(1+x_t(\gamma_{CM}-1) - x_t\gamma_{CM}\breve{A})^2},\\
\tilde{A} \doteq & \sigma_D(1 - x_t\breve{A}),\\
\tilde{B} \doteq & \frac{1+x_t(\gamma_{CM}-1)}{1+x_t(\gamma_{CM}-1) - x_t\gamma_{CM}\breve{A}},\\
d\langle x, x\rangle_t = & ((\theta_t - 1)\sigma_R(x_t)x_t)^2 dt.
\end{aligned}$$

In Figure 1.3, I plot the risk premium, volatility, and portfolio policies as functions of the state variable, x_t. The shape of the risk premium resembles the shape of the risk premium for the case in which $\rho > \rho_v$. The dynamics of volatility are, however, markedly different. For $\rho < \rho_v$, the volatility of the risky asset is decreasing in the wealth share of the unconstrained investor. In other words, high-volatility states are those in which the unconstrained investor is undercapitalized and the agent engaged in volatility timing owns most of the wealth in the economy. On the other hand, states in which the unconstrained investor owns most of the wealth in the economy are characterized by low levels of volatility and admit the interpretation of expansion states. The economic intuition behind this result is fairly simple. An increase in x_t increases the wealth share of the patient investor, and this dampens volatility.

[Insert Figure 1.3]

[Insert Figure 1.4]

In all panels, the equilibrium quantities implied by the heterogeneous agent model featuring volatility timers are in red. The equilibrium quantities implied by the baseline model described in Proposition 7 are in black. For ease of interpretation, in Figure 1.3, I plot the results for $x \in [0.15, 1]$. In Figure 1.4, I plot the results over the entire state space.

It is instructive to compare the equilibrium quantities implied by the model featuring volatility timers to the equilibrium quantities implied by the baseline model. The baseline model implies a hump-shaped volatility. When x_t is close to the lower and upper boundaries of the state space one type of agent dominates, the leverage in economy is low, and volatility is subdued. In the baseline model, the portfolio policies of the agents of the model are positively correlated, and the logarithmic investor takes lower leverage in the vicinity of the lower boundary of the state space, compared to the model featuring volatility timers. Finally, the risk premium implied by the baseline model is not very sensitive to the state variable of the model. This is because the two types of agents that populate the baseline economy are very similar. Given that the newly added agent in the baseline economy is slightly more risk averse than the unconstrained investor, the risk premium in the baseline model increases for low values of x_t.

A brief description of the underlying economic intuition (for the heterogeneous agent model featuring volatility-managed portfolios) is in order. In the vicinity of the lower boundary of the state space, the asset demand of the unconstrained investor is very elastic with respect to her wealth, W_t. Consequently, for low values of x_t, even marginal changes in W_t result in large adjustments in the size of the risky asset position of the unconstrained investor. This trading behavior inevitably elevates volatility. On the other hand, for high values of x_t the unconstrained investor owns most of the wealth in the economy. Her risky asset demand becomes less sensitive to changes in wealth, and this dampens volatility. Similarly, the risk premium is high for low values of x_t in order to induce the unconstrained investor to increase her leverage and for the market for the risky asset to clear. For $\rho < \rho_v$, high-volatility states are also high risk premium states. An investor willing to support market prices in high-volatility states of the world has the opportunity to earn a high risk premium. The sign of the quadratic co-variation (see Figure 1.6) supports this result.

[Insert Figure 1.6]

1.6. Broader Perspective and Discussion

In this section, I revisit some of the most important theoretical results from the previous sections and put them in a broader context. In particular, I will argue that, in some special cases, some of my theoretical results are broadly consistent with what standard homogeneous agent general equilibrium models predict. I then compare my results against the main implications of Moreira and Muir (2017) and Moreira and Muir (2016).

1.6.1. Implications of Homogeneous Agent General Equilibrium Models

A wide range of consumption-based general equilibrium asset pricing models, such as the homogeneous agent economies in Campbell and Cochrane (1999) and Bansal and Yaron (2004) and the intermediary asset pricing framework of He and Krishnamurthy (2013) featuring heterogeneous agents, imply a positive relation between asset volatility and risk premia. For example, in Campbell and Cochrane (1999) and He and Krishnamurthy (2013), negative shocks to consumption increase the volatility of consumption, leading to an endogenous increase in the conditional volatility of the risky asset. Consequently, in all these standard off-the-shelf models, risky asset volatility endogenously goes up at times when the market price of risk is high. As a result of this, in the above class of models, volatility timing does not constitute a sensible approximation to the optimal trading strategy. In the counter-factual case, where I set $\rho < \rho_v$, the above results carry over to my heterogeneous agent economy.

1.6.2. Relation to Moreira and Muir (2017)

In a recent paper, Moreira and Muir (2017) empirically show that trading strategies that scale standard asset pricing factors, which are widely used in the empirical cross-sectional asset pricing literature, by realized variance over a one-month horizon deliver high alphas. The authors further claim that in their particular dataset the correlation between future expected returns and conditional volatility is very week. In other words, increases in volatility do not lead to increases in expected returns, or if expected returns go up, the increase is marginal relative to the increase in volatility.

The authors use these results to argue that volatility management delivers high returns in risk-adjusted terms. In a companion paper, the authors argue that volatility timing is optimal for a very wide range of different types of investors, both short-term and long-term investors (Moreira and Muir (2016)). The optimality of volatility timing for long-term investors is particularly puzzling, especially if returns are mean reverting and/or volatility shocks are transitory as opposed to persistent.

The approach of Moreira and Muir (2016) and Moreira and Muir (2017) is very different from my approach along several dimensions. First, in the theory sections of the above two papers, the authors resort to partial equilibrium analysis, and I develop a general equilibrium model featuring heterogeneous agents. The general equilibrium framework allows me to take into account equilibrium feedback effects. More importantly, in the above two papers, the authors exogenously postulate the dynamics that drive the price process and its volatility. In my paper, I endogenously solve for the equilibrium price and volatility dynamics. Second, in their framework, the investors engaged in volatility management are price takers and do not affect prices. In my model, I show that volatility timers are instrumental in the price formation process and have a big effect on the Sharpe ratio of the risky asset. As a result of this, the unconstrained investors, the centerpiece of my model, are materially affected by the volatility-managed portfolios.

The implications of my model for the realistic case in which $\rho > \rho_v$ are consistent with the stylized empirical facts reported by Moreira and Muir (2016).

1.6.3. Relation to Martin (2017)

In a recent paper, Martin (2017) derives a lower bound on the equity premium in terms of SVIX, an implied volatility index calculated from the prices of index options. He shows that the SVIX forecast is positively correlated with subsequent returns. He further shows that a contrarian market timing strategy, using SVIX as a signal, delivers a Sharpe ratio considerably higher than the Sharpe ratio of the market. In essence, the trading strategy in Martin (2017) is the opposite of volatility timing as defined here. A trader following the trading strategy in Martin (2017) would increase his risky asset exposure when implied volatility is high and decrease risky asset exposure when volatility is low. It is important to note, however, that the results in Martin (2017) and Moreira and Muir (2017) are not directly comparable. This is because the two papers use different measures of volatility. Whereas Martin (2017) uses forward-looking *implied* volatility, which is extracted from option prices, Moreira and Muir (2017) use past *realized* volatility.

Furthermore, Martin (2017) shows that the equity risk premium perceived by an unconstrained rational investor with logarithmic utility who is fully invested in the market is proportional to the risk-neutral variance. The risk-premium implied by my model is

$$\mathbb{E}_t(dR_t) - r_t dt = \theta_t \mathbb{V}\text{ar}_t(dR_t).$$

Given that there are no jumps, realized variance equals implied variance, $\mathbb{V}\text{ar}_t(dR_t) =$

$\mathbb{V}\text{ar}_t^{\mathbb{Q}}(dR_t)$. For the case in which $\theta_t = 1$,

$$\mathbb{E}_t(dR_t) - r_t dt = \mathbb{V}\text{ar}_t^{\mathbb{Q}}(dR_t).$$

In Figures 1.8 and 1.9, I plot the risk premium (black line) and the variance of the total return process (red line) over the state space.

[Insert Figure 1.8]

[Insert Figure 1.9]

In the limit, where the unconstrained investor owns all the wealth in the economy, $x_t = 1$, $\theta_t = 1$, and the risk premium is equal to the variance. Over the remaining region of the state space, $x_t < 1$, the variance of the risky asset is a lower bound on the risk premium.

1.7. Conclusion

In this chapter, I have studied the effects of volatility timing in a general equilibrium heterogeneous agent model. Two distinct types of agents populate the model economy, an unconstrained investor endowed with logarithmic utility over instantaneous consumption and a risk-neutral agent subject to a volatility budget. My model goes a long way towards the rationalization of the behavior of investment vehicles that follow investment management strategies that are isomorphic to the ones implied by the principles of volatility management. Whereas my theoretical approach offers a high degree of tractability, it is subject to some important caveats. Specifically, the model implies unrealistically high leverage for the unconstrained investor.

1.8. Mathematical Appendix

1.8.1. Portfolio Choice of the Risk-neutral Investor

In this section, I derive the optimal portfolio choice of the volatility-managed portfolio. Here, the main objective is to derive a closed form expression for the Lagrange multiplier, associated with the volatility budget, and express the portfolio choice of the volatility-managed portfolio as a function thereof.

The risk-neutral volatility timer solves the constrained portfolio optimization problem

$$
\begin{aligned}
\max_{\{\tilde{\theta}_t\}} \quad & \mathbb{E}_t\left(\frac{d\tilde{W}_t}{dt}\right), \\
\text{s.t.} \quad & \tilde{\theta}_t \geq 0, \\
& \beta\sqrt{\frac{\mathbb{V}\text{ar}_t(d\tilde{W}_t)}{dt}} \leq \tilde{W}_t, \\
& d\tilde{W}_t = (r_t\tilde{W}_t - \tilde{C}_t)dt + \tilde{\theta}_t\tilde{W}_t(dR_t - r_t dt).
\end{aligned}
$$

Given that $\tilde{C}_t = \rho_v\tilde{W}_t$, the SDE for the dynamic budget constraint of the agent takes the form

$$
d\tilde{W}_t = (r_t - \rho_v)\tilde{W}_t dt + \tilde{\theta}_t\tilde{W}_t(\mu_R(\cdot) - r_t)dt + \tilde{\theta}_t\tilde{W}_t\sigma_R(\cdot)dB_t.
$$

It is then immediate to see that

$$
\begin{aligned}
\mathbb{E}_t(d\tilde{W}_t) =& (r_t - \rho_v)\tilde{W}_t dt + \tilde{\theta}_t\tilde{W}_t(\mu_R(\cdot) - r_t)dt, \\
d\langle\tilde{W}, \tilde{W}\rangle_t =& (\tilde{\theta}_t\tilde{W}_t\sigma_R(\cdot))^2 dt.
\end{aligned}
$$

Consequently, $\mathbb{V}\text{ar}(dW_t) = (\tilde{\theta}_t\tilde{W}_t\sigma_R(\cdot))^2 dt$. The corresponding Lagrangian is

$$
\mathcal{L} = (r_t - \rho_v)\tilde{W}_t + (\mu_R(\cdot) - r_t)\tilde{\theta}_t\tilde{W}_t + \lambda_t(\tilde{W}_t - \beta\sqrt{(\tilde{\theta}_t\tilde{W}_t\sigma_R(\cdot))^2}),
$$

where λ_t is the Lagrangian multiplier associated with the risk limit (volatility budget). The first order condition with respect to $\tilde{\theta}_t$ gives

$$\tilde{\theta}_t = \frac{\mu_R(\cdot)\tilde{W}_t - r\tilde{W}_t}{\lambda_t \beta(\tilde{\theta}_t \tilde{W}_t \sigma_R(\cdot))^{-1}}(\tilde{W}_t \sigma_R(\cdot))^{-2}.$$

In Proposition 3, I prove that the risk premium on the risky asset is positive over the entire state space. Consequently, it is optimal for the risk-neutral agent to increase the size of her risky asset position up to the volatility limit. Therefore, the dynamic constraint holds with equality, i.e.,

$$\beta\sqrt{\frac{\mathbb{V}\mathrm{ar}_t(d\tilde{W}_t)}{dt}} = \tilde{W}_t.$$

This allows me to derive an expression for $\tilde{W}_t$,

$$\tilde{W}_t = \beta\frac{\mu_R(\cdot)\tilde{W}_t - r_t\tilde{W}_t}{\lambda_t \beta(\tilde{\theta}_t \tilde{W}_t \sigma_R(\cdot))^{-1}}(\sigma_R(\cdot)\tilde{W}_t)^{-1}.$$

Additionally, it is useful to note that

$$(\tilde{\theta}_t \tilde{W}_t \sigma_R(\cdot))^{-1} = \frac{\beta}{\tilde{W}_t}.$$

The next step is to solve for the Lagrange multiplier,

$$\lambda_t = \frac{\mu_R(\cdot) - r_t}{\beta\sigma_R(\cdot)}.$$

I substitute the expression for the Lagrange multiplier into the first order condition, with respect to $\tilde{\theta}_t$, and after a straightforward simplification I obtain the optimal portfolio policy,

$$\tilde{\theta}_t = \frac{1}{\beta\sigma_R(\cdot)} = \frac{\bar{\sigma}}{\sigma_R(\cdot)}.$$

■

1.8.2. Portfolio Choice of the Unconstrained Investor

In this sub-section of the mathematical Appendix, I derive the optimal portfolio policy of the unconstrained investor, θ_t, who is endowed with logarithmic utility over instantaneous consumption. I start from the market clearing condition,

$$\theta_t \frac{W_t}{S_t} + \tilde{\theta}_t \frac{\tilde{W}_t}{S_t} = 1,$$

where the constant on the right-hand side is the total supply of the risky asset. Given that I model a pure exchange economy, $\tilde{W}_t = S_t - W_t$. Consequently,

$$\begin{aligned} \theta_t \frac{W_t}{S_t} + \tilde{\theta}_t \frac{S_t - W_t}{S_t} &= 1, \\ \theta_t x_t + \tilde{\theta}_t (1 - x_t) &= 1. \end{aligned}$$

The last step is to solve the above expression for θ_t. The optimal portfolio policy of the unconstrained investor,

$$\theta_t = \frac{1}{x_t} - \frac{1 - x_t}{x_t} \tilde{\theta}_t = \frac{1}{x_t} - \frac{1 - x_t}{x_t} \frac{\bar{\sigma}}{\sigma_R(\cdot)},$$

follows immediately.

■

1.8.3. Price-Dividend Ratio

In this sub-section of the mathematical Appendix, I derive an expression for the price-dividend ratio on the risky asset. I impose market clearing in the consumption goods market,

$$C_t + \tilde{C}_t = D_t,$$

and note that $C_t = \rho W_t$, and $\tilde{C}_t = \rho_v \tilde{W}_t$. Using the fact that $\tilde{W}_t = S_t - W_t$, I obtain

$$\rho W_t + \rho_v (S_t - W_t) = D_t.$$

I then solve the above expression for S_t and obtain

$$S_t = \frac{1}{\rho_v} D_t + \left(1 - \frac{\rho}{\rho_v}\right) W_t.$$

To obtain an expression for the price-dividend ratio divide both sides by D_t,

$$\begin{aligned} \frac{S_t}{D_t} &= \frac{1}{\rho_v} + \left(1 - \frac{\rho}{\rho_v}\right) \frac{W_t}{D_t} \frac{S_t}{S_t} \\ &= \frac{1}{\rho_v} + \left(1 - \frac{\rho}{\rho_v}\right) \frac{S_t}{D_t} x_t. \end{aligned}$$

Solving for S_t/D_t, I obtain an expression for the price-dividend ratio,

$$\frac{S_t}{D_t} = \frac{1}{\rho_v - x_t(\rho_v - \rho)}.$$

This completes the derivation of the price-dividend ratio on the risky asset.

■

1.8.4. Proof of Proposition 1 (Model-Implied Total Return Process Volatility)

I start from the definition of the total return process,

$$dR_t = \frac{D_t dt + dS_t}{S_t}.$$

It is then immediate to see that

$$\mu_R(\cdot)dt + \sigma_R(\cdot)dB_t = \frac{D_t dt}{S_t} + \frac{dS_t}{S_t}.$$

Clearly, in order to compute the diffusion coefficient of the total return process, $\sigma_R(\cdot)$, it is enough to find the diffusion coefficient of dS_t/S_t. To this end, I derive a stochastic differential equation for $\{S_t\}$. I note that ρ_v and ρ are constants and do Itô on

$$S_t = \frac{1}{\rho_v} D_t + \left(1 - \frac{\rho}{\rho_v}\right) W_t.$$

The stochastic differential equation

$$dS_t = \frac{1}{\rho_v} dD_t + \left(1 - \frac{\rho}{\rho_v}\right) dW_t$$

follows immediately. I then use the facts that the exogenous dividend process follows a geometric Brownian motion and the stochastic differential equation

$$dW_t = -C_t dt + r_t W_t dt + \theta_t W_t (dR_t - r_t dt)$$

governs the intertemporal wealth evolution of the unconstrained investor. The above SDE simplifies to

$$\begin{aligned} dS_t =& \frac{1}{\rho_v}(\mu_D D_t dt + \sigma_D D_t dB_t) + \left(1 - \frac{\rho}{\rho_v}\right)(r W_t dt - C_t dt + \theta_t W_t (dR_t - r_t dt)) \\ =& (\cdot) dt + \left(\frac{1}{\rho_v}\sigma_D D_t + \left(1 - \frac{\rho}{\rho_v}\right)\sigma_R(\cdot)\theta_t W_t\right) dB_t. \end{aligned}$$

I then match the coefficients and obtain

$$\sigma_R(\cdot) = \frac{1}{S_t}\left(\frac{1}{\rho_v} D_t\right)\sigma_D + \left(1 - \frac{\rho}{\rho_v}\right)\theta_t \sigma_R(\cdot)\frac{W_t}{S_t}.$$

Using the fact that

$$S_t = \frac{1}{\rho_v} D_t + \left(1 - \frac{\rho}{\rho_v}\right) W_t,$$

I obtain

$$\frac{1}{\rho_v} D_t = S_t - \left(1 - \frac{\rho}{\rho_v}\right) W_t.$$

I then substitute the latter in the expression for $\sigma_R(\cdot)$, set $x_t = W_t/S_t$, and obtain

$$\sigma_R(\cdot) = \sigma_D - \sigma_D\left(1 - \frac{\rho}{\rho_v}\right)x_t + \left(1 - \frac{\rho}{\rho_v}\right)\theta_t\sigma_R(\cdot)x_t.$$

To simplify the notation, I define

$$A \doteq 1 - \frac{\rho}{\rho_v}.$$

Using this newly introduced notation, the expression for $\sigma_R(\cdot)$ simplifies to

$$\sigma_R(\cdot) = \sigma_D - \sigma_D A x_t + A\theta_t\sigma_R(\cdot)x_t.$$

I then substitute the expression for the equilibrium portfolio weight,

$$\theta_t = \frac{1}{x_t} - \frac{1-x_t}{x_t}\tilde{\theta}_t = \frac{1}{x_t} - \frac{1-x_t}{x_t}\frac{\bar{\sigma}}{\sigma_R(\cdot)},$$

into the expression for $\sigma_R(x_t)$ and, after a straightforward simplification, I obtain

$$\begin{aligned}\sigma_R(x_t)(1-A) =& \sigma_D - x_t\sigma_D A - (1-x_t)\bar{\sigma}A,\\ \sigma_R(x_t) =& \frac{1}{1-A}(\sigma_D - \bar{\sigma}A) + \frac{A}{1-A}(\bar{\sigma} - \sigma_D)x_t.\end{aligned}$$

This completes the proof of the proposition.

■

1.8.5. Proof of Proposition 2 (Properties of Total Return Process Volatility)

In the main version of the model, $\rho > \rho_v$. Under this parametric restriction, $A < 0$. Additionally, throughout the paper I assume that $\bar{\sigma} < \sigma_D$. Therefore, the loading on x_t is positive,

$$\frac{A}{(1-A)}(\bar{\sigma} - \sigma_D) > 0.$$

Consequently, the total return process volatility,

$$\sigma_R(x_t) = \frac{1}{1-A}(\sigma_D - \bar{\sigma}A) + \frac{A}{1-A}(\bar{\sigma} - \sigma_D)x_t,$$

is increasing in the wealth share of the unconstrained investor, x_t. It attains a minimum at $x_t = 0$. For $x_t = 0$, $\sigma_R(x_t) > 0$ as σ_D and $\bar{\sigma}$ are positive constants. Consequently, the model-implied volatility is positive over the entire state space. The claim that the sensitivity of $\sigma_R(x_t)$ with respect to x_t is increasing in the wedge between $\bar{\sigma}$ and σ_D follows immediately. This competes the proof of the proposition.

■

In Section 1.5 of the paper, I analyze the equilibrium outcome for the case in which $\rho < \rho_v$. Below, I show that under this parametric restriction the total return process volatility is decreasing in x_t. Given that $\rho < \rho_v$, $A \in (0, 1)$. Throughout the paper, I assume that $\sigma_D > \bar{\sigma}$. Consequently, the loading on x_t is negative,

$$\frac{A}{1-A}(\bar{\sigma} - \sigma_D) < 0,$$

and the total return process volatility is decreasing in x_t over the entire state space. Volatility is linear in x_t and attains a minimum at $x_t = 1$. Consequently, if $\sigma_R(1) \geq 0$, then the volatility is positive over the entire state space. For $x_t = 1$, volatility is equal to

$$\sigma_R(1) = \frac{1}{1-A}(\sigma_D - \bar{\sigma}A) + \frac{A}{1-A}(\bar{\sigma} - \sigma_D) = \sigma_D > 0.$$

■

1.8.6. Proof of Proposition 3 (Risk Premium)

In the logarithmic utility case, the SDF is proportional to marginal utility,

$$\Lambda_t \propto e^{-\rho t} \frac{\partial u(C_t)}{\partial C_t}.$$

I use the fact that $d\Lambda_t R_t$ follows a martingale under the physical measure and note that

$$\frac{d\Lambda_t}{\Lambda_t} = -\rho dt + \frac{e^{-\rho t}}{\Lambda_t}\frac{\partial^2 u(C_t)}{\partial C_t^2}dC_t + \frac{1}{2}\frac{e^{-\rho t}}{\Lambda_t}\frac{\partial^3 u(C_t)}{\partial C_t^3}d\langle C, C\rangle_t.$$

The Euler equation of the unconstrained investor,

$$-\rho dt - \mathbb{E}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{V}\mathrm{ar}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{E}_t(dR_t) = \mathbb{C}\mathrm{ov}_t\left(\frac{dC_t}{C_t}, dR_t\right),$$

follows immediately. This equation holds for any tradable asset. By setting $dR_t = r_t dt$, and noting that

$$\mathbb{C}\mathrm{ov}_t\left(\frac{dC_t}{C_t}, r_t dt\right) = 0,$$

I obtain an expression for the risk-free interest rate, r_t. Finally, the risk premium, $\mathbb{E}_t(dR_t - r_t dt)$, is given by

$$\mathbb{E}_t(dR_t - r_t dt) = \mathbb{C}\mathrm{ov}_t\left(\frac{dC_t}{C_t}, dR_t\right).$$

In the model, the wealth and consumption of the unconstrained investor grow at the same rate, $dC_t/C_t = dW_t/W_t$, and the risk premium simplifies to

$$\begin{aligned}\mathbb{E}_t(dR_t) - r_t dt =& \mathbb{C}\mathrm{ov}_t\left(\frac{dW_t}{W_t}, dR_t\right),\\ \mathbb{E}_t(dR_t) - r_t dt =& \theta_t \mathbb{V}\mathrm{ar}_t(dR_t),\end{aligned}$$

where $\mathbb{V}\mathrm{ar}_t(dR_t) = \sigma_R^2(\cdot)dt$. Using the fact that

$$\theta_t = \frac{1}{x_t} - \frac{1-x_t}{x_t}\frac{\bar{\sigma}}{\sigma_R(\cdot)},$$

I obtain the expression for the risk premium on the risky asset

$$\mathbb{E}_t(dR_t) - r_t dt = \left(\frac{1}{x_t} - \frac{\bar{\sigma}}{\sigma_R(x_t)}\frac{1-x_t}{x_t}\right)\sigma_R^2(x_t)dt.$$

The next step is to prove the claim that the risk premium is positive over the entire state space. Given that $\sigma_R(x_t) \geq 0, \forall x_t$,

$$\operatorname{sgn}(\mathbb{E}_t(dR_t - r_t dt)) = \operatorname{sgn}(\theta_t).$$

A straightforward simplification of the equilibrium expression for θ_t yields

$$\theta_t = \frac{1}{x_t} - \tilde{\theta}_t\left(\frac{1-x_t}{x_t}\right) = \frac{1}{x_t}\left(1 - \tilde{\theta}_t\,(1-x_t)\right) \propto \left(1 - \tilde{\theta}_t\,(1-x_t)\right).$$

Consequently,

$$\operatorname{sgn}(\theta_t) = \operatorname{sgn}\left(1 - \tilde{\theta}_t\,(1-x_t)\right) = \operatorname{sgn}\left(\sigma_R(x_t) - \bar{\sigma}(1-x_t)\right).$$

I substitute out the expression for $\sigma_R(x_t)$ and simplify,

$$\operatorname{sgn}(\theta_t) = \operatorname{sgn}\left(\sigma_D - \bar{\sigma} + (\bar{\sigma} - A\sigma_D)x_t\right).$$

In the main version of the model, $\rho > \rho_v$. This implies that $A < 0$ and $\mathbb{E}_t(dR_t - r_t dt) > 0$, over the entire state space. The last step is to show that the risk premium is decreasing in x_t to the left of $\hat{x}$ and increasing in x_t to the right of $\hat{x}$, where

$$\hat{x} \doteq \sqrt{\frac{\bar{\sigma}A - \sigma_D}{A(\bar{\sigma} - \sigma_D A)}}.$$

Please see the proof of Proposition 5 for a derivation of $\hat{x}$. In Proposition 5, I prove that the quadratic co-variation between the risk premium and volatility is negative to the left of $\hat{x}$ and positive to the right of $\hat{x}$. Given that volatility is increasing in x_t over the entire state space, the risk premium is decreasing in x_t to the left of $\hat{x}$

and increasing in x_t to the right of $\hat{x}$. This completes the proof of the proposition. ∎

In Section 1.5 of the paper, I analyze the equilibrium outcome for the case in which $\rho < \rho_v$. In this case, the risk premium on the risky asset is positive over the entire state space. To show that, I note that when $\rho < \rho_v$, $A \in (0,1)$. As above, the sign of the risk premium depends on the sign of θ_t which in terms depends on

$$\text{sgn}(\theta_t) = \text{sgn}\left(\sigma_D - \bar{\sigma} + (\bar{\sigma} - A\sigma_D)x_t\right).$$

There are two possible cases:

- For $\bar{\sigma} - A\sigma_D \geq 0$, $\mathbb{E}_t(dR_t - r_t dt) > 0, \forall x \in [0,1]$.
- For $\bar{\sigma} - A\sigma_D < 0$, $\sigma_D - \bar{\sigma} + (\bar{\sigma} - A\sigma_D)x_t$ is decreasing in x_t. For $x = 1$, $\sigma_D - \bar{\sigma} + (\bar{\sigma} - A\sigma_D) > 0$. Therefore, $\mathbb{E}_t(dR_t - r_t dt) > 0, \forall x \in [0,1]$.

1.8.7. Proof of Proposition 4 (Quadratic Co-variation)

I complete the proof of the proposition in steps. The first step is to derive the stochastic differential equation for the volatility of the total return process. To this end, I do Itô on

$$\sigma_R(x_t) = \frac{1}{1-A}(\sigma_D - \bar{\sigma}A) + \frac{A}{1-A}(\bar{\sigma} - \sigma_D)x_t.$$

Consequently, $\{\sigma_R(x_t)\}$ follows

$$d\sigma_R(x_t) = \frac{A}{1-A}(\bar{\sigma} - \sigma_D)dx_t.$$

Similarly, I do Itô on the risk premium,

$$\begin{aligned}
\text{RP}(x_t) =& \theta_t \sigma_R^2(x_t), \\
\text{RP}(x_t) =& \sigma_R^2(x_t)\frac{1}{x_t} - \bar{\sigma}\left(\frac{1-x_t}{x_t}\right)\sigma_R(x_t),
\end{aligned}$$

and obtain

$$
\begin{aligned}
d(\mathrm{RP}(x_t)) =& -\frac{1}{x_t^2}\sigma_R^2(x_t)dx + \frac{1}{x_t}d\sigma_R^2(x_t) + (\cdot)dt \\
& - \left(\bar{\sigma}\frac{\partial}{\partial x_t}\left(\frac{1-x_t}{x_t}\right)\sigma_R(x_t)dx + \bar{\sigma}\left(\frac{1-x_t}{x_t}\right)d\sigma_R(x_t) + (\cdot)dt\right) \\
=& \left(-\frac{1}{x_t^2}\sigma_R^2(x_t) + \frac{\bar{\sigma}}{x_t^2}\sigma_R(x_t)\right)dx_t + (\cdot)dt \\
& + \left(2\left(\frac{1}{x_t}\right)\sigma_R(x_t) - \bar{\sigma}\left(\frac{1-x_t}{x_t}\right)\right)d\sigma_R(x_t).
\end{aligned}
$$

The last step is to calculate the quadratic co-variation, $d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t$,

$$
\begin{aligned}
d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t =& \frac{1}{x_t^2}\sigma_R(x_t)(\bar{\sigma} - \sigma_R(x_t))d\langle x, \sigma_R(x)\rangle_t \\
& + \frac{1}{x_t}(2\sigma_R(x_t) - \bar{\sigma}(1-x_t))d\langle \sigma_R(x), \sigma_R(x)\rangle_t.
\end{aligned}
$$

In order to further simplify this expression, it is useful to note that

$$
\begin{aligned}
d\sigma_R(x_t) =& \frac{A}{1-A}(\bar{\sigma} - \sigma_D)dx_t, \\
d\langle \sigma_R(x), \sigma_R(x)\rangle_t =& \left(\frac{A}{1-A}(\bar{\sigma} - \sigma_D)\right)^2 d\langle x, x\rangle_t, \\
d\langle \sigma_R(x), x\rangle_t =& \frac{A}{1-A}(\bar{\sigma} - \sigma_D)d\langle x, x\rangle_t.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t =& \left(\sigma_R(x_t)(\bar{\sigma} - \sigma_D)(\bar{\sigma} - \sigma_R(x_t))\frac{A}{1-A}\frac{1}{x_t^2}\right)d\langle x, x\rangle_t \\
& + (2\sigma_R(x_t) - \bar{\sigma}(1-x_t))\left(\frac{(\bar{\sigma} - \sigma_D)A}{1-A}\right)^2\frac{1}{x_t}d\langle x, x\rangle_t.
\end{aligned}
$$

In order to further simplify the expression for the quadratic co-variation, I define

$$
\begin{aligned}
\tilde{A}_1(x_t) \doteq& \sigma_R(x_t)(\bar{\sigma} - \sigma_D)(\bar{\sigma} - \sigma_R(x_t)), \\
\tilde{A}_2(x_t) \doteq& (2\sigma_R(x_t) - \bar{\sigma}(1-x_t))\left(\frac{(\bar{\sigma} - \sigma_D)A}{1-A}\right)^2.
\end{aligned}
$$

The expression for the quadratic co-variation takes the form

$$d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t = \left(\tilde{A}_1(x_t)\frac{A}{1-A}\frac{1}{x_t^2} + \tilde{A}_2(x_t)\frac{1}{x_t}\right) d\langle x, x\rangle_t.$$

The last step in the proof of the proposition is to derive an expression for the quadratic variation of the state variable.

$$\begin{aligned} dx_t =& d\left(\frac{W_t}{S_t}\right) = \frac{1}{S_t}dW_t + W_t d\left(\frac{1}{S_t}\right) + (\cdot)dt \\ =& \frac{1}{S_t}\theta_t W_t \sigma_R(x_t) dB_t - \frac{W_t}{S_t}\frac{dS_t}{S_t} + (\cdot)dt \\ =& x_t\sigma_R(x_t)(\theta_t - 1)dB_t + (\cdot)dt. \end{aligned}$$

Therefore,

$$d\langle x, x\rangle_t = (x_t\sigma_R(x_t)(\theta_t - 1))^2 dt.$$

This completes the proof of the proposition.

■

1.8.8. Proof of Proposition 5 (Properties of Quadratic Co-variation)

Given that $d\langle x, x\rangle_t$ is non-negative, the sign of the quadratic co-variation,

$$d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t = \left(\tilde{A}_1(x_t)\frac{A}{1-A}\frac{1}{x_t^2} + \tilde{A}_2(x_t)\frac{1}{x_t}\right) d\langle x, x\rangle_t,$$

depends on the sign of

$$\tilde{A}_1(x_t)\frac{A}{1-A}\frac{1}{x_t^2} + \tilde{A}_2(x_t)\frac{1}{x_t}.$$

I substitute out the expressions for $\tilde{A}_1(x_t)$ and $\tilde{A}_2(x_t)$ and simplify to obtain

$$\frac{1}{x_t}\frac{A}{1-A}(\bar{\sigma} - \sigma_D)\left(\frac{1}{x_t}\sigma_R(x_t)(\bar{\sigma} - \sigma_R(x_t)) + (2\sigma_R(x_t) - \bar{\sigma}(1 - x_t))\frac{A}{1-A}(\bar{\sigma} - \sigma_D)\right).$$

Given that $\bar{\sigma} < \sigma_D$ and $A < 0$,

$$\frac{1}{x_t}\frac{A}{1-A}(\bar{\sigma}-\sigma_D) > 0.$$

Consequently, the sign of $d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t$ depends on the sign of $\Psi(x_t)$, where I define

$$\Psi(x_t) \doteq \frac{1}{x_t}\sigma_R(x_t)(\bar{\sigma}-\sigma_R(x_t)) + (2\sigma_R(x_t) - \bar{\sigma}(1-x_t))\frac{A}{1-A}(\bar{\sigma}-\sigma_D).$$

I then substitute $\sigma_R(x_t)$ out and simplify. After a straightforward simplification, I obtain

$$\Psi(x_t) = A(\bar{\sigma}-\sigma_D)(\bar{\sigma}-\sigma_D A)x_t^2 + (\sigma_D - \bar{\sigma}A)(\bar{\sigma}-\sigma_D).$$

The two roots of $\Psi(x_t)$ are

$$x_1 = -\sqrt{\frac{\bar{\sigma}A-\sigma_D}{A(\bar{\sigma}-\sigma_D A)}}, \quad \hat{x} = \sqrt{\frac{\bar{\sigma}A-\sigma_D}{A(\bar{\sigma}-\sigma_D A)}}.$$

While $x_1 < 0$, $\hat{x} > 0$. There are two possible cases. When $\hat{x} < 1$, the quadratic covariation is negative to the left of $\hat{x}$ and positive to the right of $\hat{x}$. When $\hat{x} \geq 1$, the quadratic co-variation is negative over the entire state space. The latter inequality holds when

$$\frac{\bar{\sigma}A-\sigma_D}{A(\bar{\sigma}-\sigma_D A)} \geq 1,$$

or

$$\left(1-\frac{\rho}{\rho_v}\right)^2 = A^2 \leq 1.$$

This completes the proof of the proposition.

■

In Section 1.5 of the paper, I analyze the equilibrium outcome for the case in which $\rho < \rho_v$. In this case, the sign of the quadratic co-variation is positive over the entire state space. This is because $A > 0$, $\tilde{A}_1(x_t) > 0$, and $\tilde{A}_2(x_t) > 0$ over the entire state space.

1.8.9. Proof of Proposition 6 (Portfolio Sensitivities)

Substitute the expression for volatility into the portfolio policy function of the volatility timer to obtain

$$\tilde{\theta}_t = \frac{\bar{\sigma}(1-A)}{(\sigma_D - \bar{\sigma}A) + Ax_t(\bar{\sigma} - \sigma_D)}.$$

The results follow immediately.

■

1.8.10. Proof of Proposition 7 (Equilibrium, Baseline Model)

I start from the market clearing condition,

$$\theta_t = \frac{1}{x_t} - \frac{1-x_t}{x_t}\theta_{CM,t}.$$

I then substitute

$$\theta_{CM,t} = \frac{\mu_R(\cdot) - r_t}{\gamma_{CM}\sigma_R^2(x_t)}$$

into the above expression and after a straightforward simplification I obtain

$$\theta_t = \frac{\gamma_{CM}}{1 + x_t(\gamma_{CM} - 1)}.$$

To solve for the volatility on the total return process, I substitute θ_t into

$$\sigma_R(\cdot) = \sigma_D - x_t\sigma_D\breve{A} + x_t\sigma_R(x_t)\theta_t\breve{A}.$$

and obtain

$$\sigma_R(x_t) = \sigma_D(1 - x_t\breve{A})\frac{1 + x_t(\gamma_{CM} - 1)}{1 + x_t(\gamma_{CM} - 1) - x_t\gamma_{CM}\breve{A}}.$$

The risk premium on the risky asset follows immediately. In order to derive an expression for the quadratic co-variation between return volatility and the risk premium, I first do Itô on the volatility of the total return process,

$$\sigma_R(x_t) = \sigma_D(1 - x_t\breve{A})\frac{1 + x_t(\gamma_{CM} - 1)}{1 + x_t(\gamma_{CM} - 1) - x_t\gamma_{CM}\breve{A}}$$

and obtain

$$d\sigma_R(x_t) = \left(-\sigma_D\breve{A}\tilde{B} + \frac{\gamma_{CM}\breve{A}\tilde{A}}{(1 + x_t(\gamma_{CM} - 1) - x_t\gamma_{CM}\breve{A})^2}\right)dx_t + (\cdot)dt,$$

where

$$\begin{aligned}\tilde{A} &\doteq \sigma_D(1 - x_t\breve{A}),\\ \tilde{B} &\doteq \frac{1 + x_t(\gamma_{CM} - 1)}{1 + x_t(\gamma_{CM} - 1) - x_t\gamma_{CM}\breve{A}}.\end{aligned}$$

I then do Itô on the risk premium and after a straightforward simplification obtain

$$d(\mathrm{RP}(x_t)) = -\frac{\gamma_{CM}(\gamma_{CM} - 1)}{(1 + x_t(\gamma_{CM} - 1))^2}\sigma_R^2(x_t)dx_t + 2\sigma_R(x_t)\frac{\gamma_{CM}}{1 + x_t(\gamma_{CM} - 1)}d\sigma_R(x_t) + (\cdot)dt.$$

The last step is to calculate the product of $d(\mathrm{RP}(x_t))$ and $d\sigma_R(x_t)$,

$$d\langle \mathrm{RP}(x), \sigma_R(x)\rangle_t = \Psi_1(x_t)\Psi_2(x_t)d\langle x, x\rangle_t,$$

where I define

$$\Psi_1(x_t) \doteq -\frac{\gamma_{CM}(\gamma_{CM} - 1)}{(1 + x_t(\gamma_{CM} - 1))^2}\sigma_R^2(x_t) + \frac{2\sigma_R(x_t)\gamma_{CM}}{1 + x_t(\gamma_{CM} - 1)}\Psi_2(x_t),$$

$$\Psi_2(x_t) \doteq -\sigma_D \breve{A}\tilde{B} + \frac{\gamma_{CM}\breve{A}\tilde{A}}{(1 + x_t(\gamma_{CM} - 1) - x_t\gamma_{CM}\breve{A})^2}.$$

The expression for $d\langle x, x\rangle_t$ is identical to the one derived in the previous proposition. This completes the proof of the proposition.

■

1.8.11. Derivation of the Equilibrium Risk-free Interest Rate

The market clearing condition in the goods market implies

$$\begin{aligned} \rho dW_t + \rho_v d\tilde{W}_t =& dD_t, \\ \mathbb{E}_t(\rho dW_t) + \mathbb{E}_t(\rho_v d\tilde{W}_t) =& \mathbb{E}_t(dD_t). \end{aligned}$$

I then note that

$$\begin{aligned} dW_t =& r_t W_t dt - C_t dt + \theta_t W_t (dR_t - r_t dt), \\ d\tilde{W}_t =& r_t \tilde{W}_t dt - \tilde{C}_t dt + \tilde{\theta}_t \tilde{W}_t (dR_t - r_t dt). \end{aligned}$$

Consequently,

$$\begin{aligned} &\mathbb{E}_t \rho[r_t W_t - C_t + \theta_t W_t(\mu_R(x_t) - r_t)]dt + \\ &\mathbb{E}_t \rho_v[r_t \tilde{W}_t - \tilde{C}_t + \tilde{\theta}_t \tilde{W}_t(\mu_R(x_t) - r_t]dt = \mathbb{E}_t(dD_t). \end{aligned}$$

I then use the fact that $\mu_R(x_t)dt - r_t dt = \text{RP}dt = \theta_t \sigma_R^2(x_t)dt$ and match the coefficients. The expression for the risk-free interest rate simplifies to

$$\begin{aligned} r_t =& \mu_D + \frac{1}{\rho x_t + \rho_v(1 - x_t)} \left(\rho^2 x_t + \rho_v^2(1 - x_t) - \theta_t^2 \sigma_R^2(x_t)\rho x_t - \theta_t \sigma_R^2(x_t)\rho_v \frac{\bar{\sigma}}{\sigma_R(x_t)}(1 - x_t) \right) \\ =& \mu_D + \frac{1}{\rho x_t + \rho_v(1 - x_t)} \left(\rho^2 x_t + \rho_v^2(1 - x_t) - \theta_t^2 \sigma_R^2(x_t)\rho x_t - \theta_t \sigma_R(x_t)\rho_v \bar{\sigma}(1 - x_t) \right). \end{aligned}$$

■

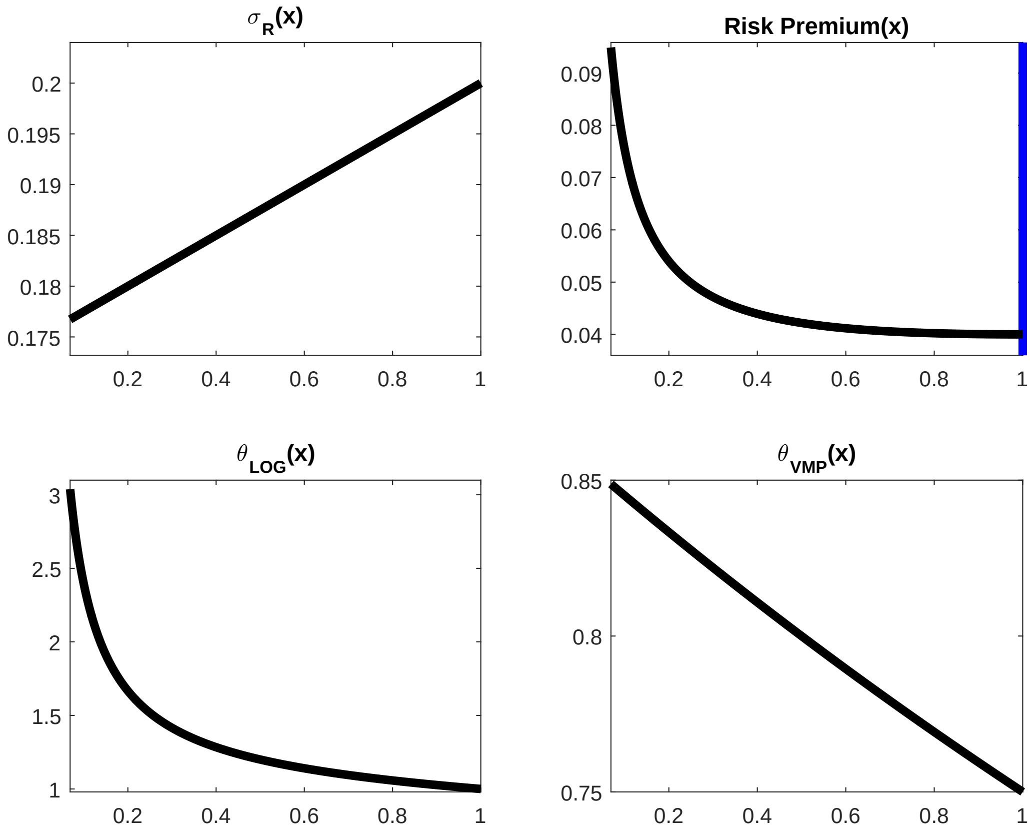

Figure 1.1: **Model Solution, Case** $\rho > \rho_v$
This figure plots equilibrium quantities of interest for the case in which $\rho > \rho_v$. In all sub-plots, the x-axis is the state variable of the model, x_t, which is the wealth share of the unconstrained investor. By construction, the wealth share takes values between zero and one. In this figure, however, I plot the volatility and the risk premium for values of x_t that are not in the immediate vicinity of the lower boundary of the state space. Please see the figure on the next page for the case in which $x \in (0, 1]$. The top-left panel depicts the volatility of the total return process, $\sigma_R(\cdot)$. The top-right panel depicts the risk premium on the risky asset, $\mathbb{E}_t(dR_t - r_t dt)/dt$. The bottom-left panel plots the portfolio policy of the unconstrained investor. Finally, the bottom-right panel plots the portfolio policy of the risk-neutral investor (volatility timer). The vertical blue line in the top-right panel passes through the value of x_t for which the risk premium attains a minimum. I plot the figure for $\bar{\sigma} = 0.15$, $\sigma_D = 0.20$, $\rho_v = 0.01$, and $\rho = 0.02$.

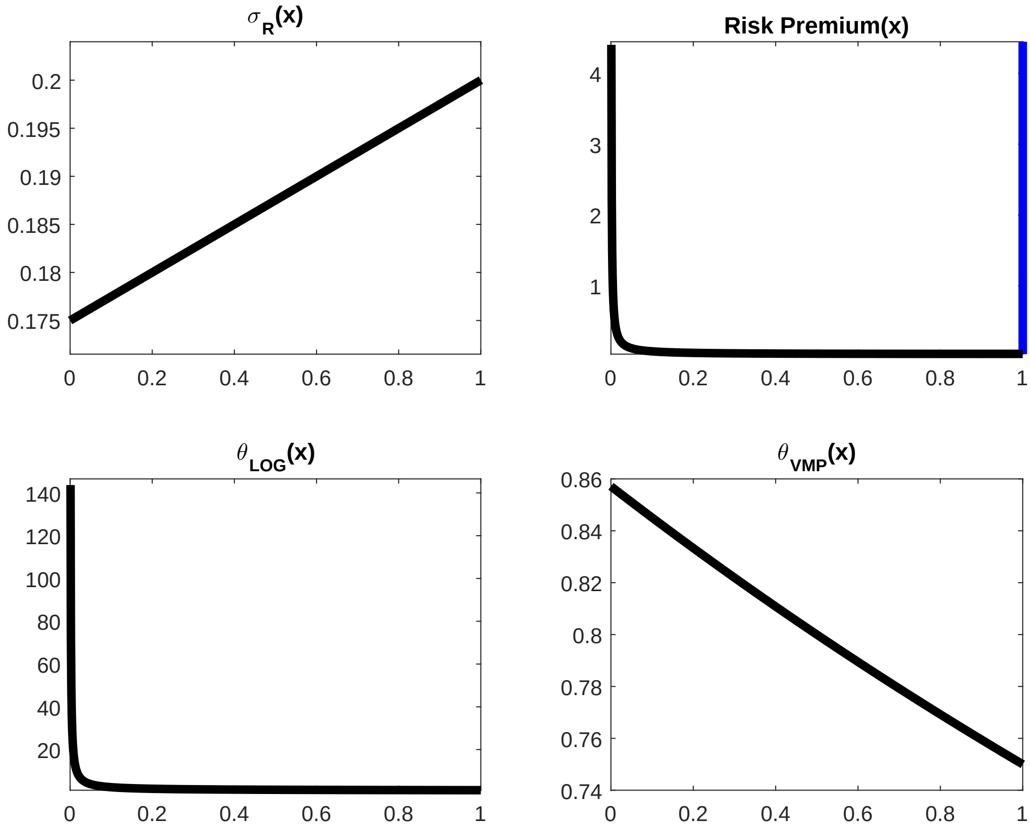

Figure 1.2: **Model Solution, Case** $\rho > \rho_v$
This figure plots equilibrium quantities of interest for the case in which $\rho > \rho_v$. In all sub-plots, the x-axis is the state variable of the model, x_t, which is the wealth share of the unconstrained investor. By construction, the wealth share takes values between zero and one. The top-left panel depicts the volatility of the total return process, $\sigma_R(\cdot)$. The top-right panel depicts the risk premium on the risky asset, $\mathbb{E}_t(dR_t - r_t dt)/dt$. The bottom-left panel plots the portfolio policy of the unconstrained investor. Finally, the bottom-right panel plots the portfolio policy of the risk-neutral investor (volatility timer). The vertical blue line in the top-right panel passes through the value of x_t for which the risk premium attains a minimum. I plot the figure for $\bar{\sigma} = 0.15$, $\sigma_D = 0.20$, $\rho_v = 0.01$, and $\rho = 0.02$.

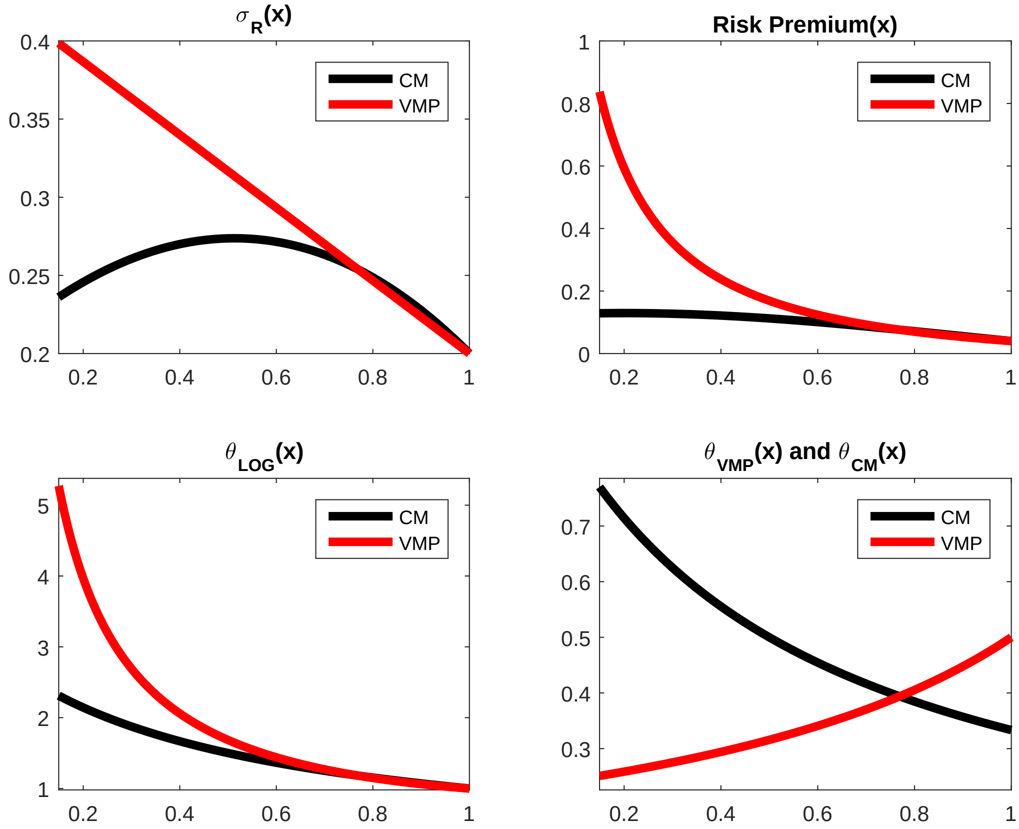

Figure 1.3: **Model Solution, Case** $\rho < \rho_v$
This figure plots equilibrium quantities of interest for the case in which the unconstrained investor is endowed with *logarithmic* utility over intertemporal consumption and $\rho < \rho_v$. In all sub-plots, the x-axis is the wealth share of the unconstrained investor, x_t. By construction, the wealth share takes values between zero and one. In the figures, however, I plot the volatility and the risk premium for values of x_t that are not in the immediate vicinity of the lower boundary of the state space. Please see the figure on the next page for the case in which $x_t \in (0, 1]$. The top-left panel depicts the volatility of the total return process, $\sigma_R(\cdot)$. The top-right panel depicts the risk premium on the risky asset, $\mathbb{E}_t(dR_t - r_t dt)/dt$. The bottom-left panel plots the portfolio policy of the unconstrained investor. Finally, the bottom-right panel plots the portfolio policies of the risk-neutral investor (volatility timer) and of the second investor in the benchmark model. In all sub-plots, the equilibrium quantities implied by the heterogeneous agent model featuring volatility timers are in red. The equilibrium quantities implied by the benchmark model are in black. I plot the figure for $\bar{\sigma} = 0.1$, $\sigma_D = 0.2$, $\gamma_{CM} = 3$, $\rho_v = 0.1$, and $\rho = 0.03$.

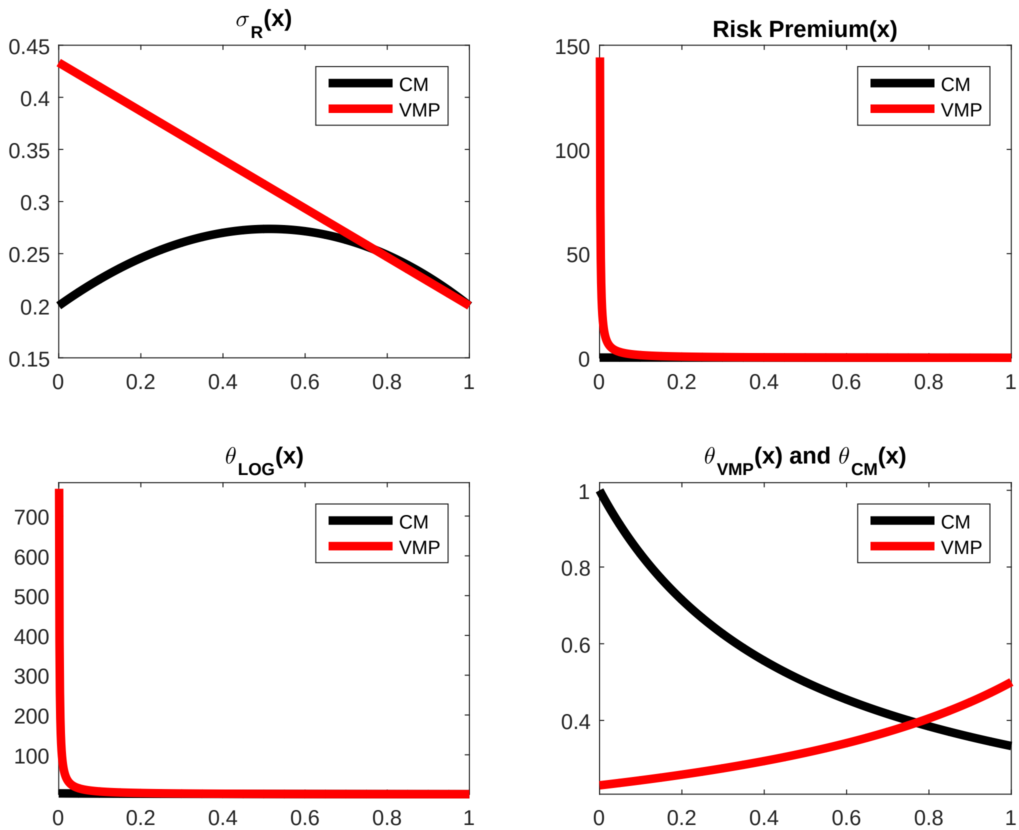

Figure 1.4: **Model Solution, Case $\rho < \rho_v$**
This figure plots equilibrium quantities of interest for the case in which the unconstrained investor is endowed with *logarithmic* utility over intertemporal consumption and $\rho < \rho_v$. In all sub-plots, the x-axis is the wealth share of the unconstrained investor, x_t. By construction, the wealth share takes values between zero and one. The top-left panel depicts the volatility of the total return process, $\sigma_R(\cdot)$. The top-right panel depicts the risk premium on the risky asset, $\mathbb{E}_t(dR_t - r_t dt)/dt$. The bottom-left panel plots the portfolio policy of the unconstrained investor. Finally, the bottom-right panel plots the portfolio policies of the risk-neutral investor (volatility timer) and of the second investor in the benchmark model. In all sub-plots, the equilibrium quantities implied by the heterogeneous agent model featuring volatility timers are in red. The equilibrium quantities implied by the benchmark model are in black. I plot the figure for $\bar{\sigma} = 0.1$, $\sigma_D = 0.2$, $\gamma_{CM} = 3$, $\rho_v = 0.1$, and $\rho = 0.03$.

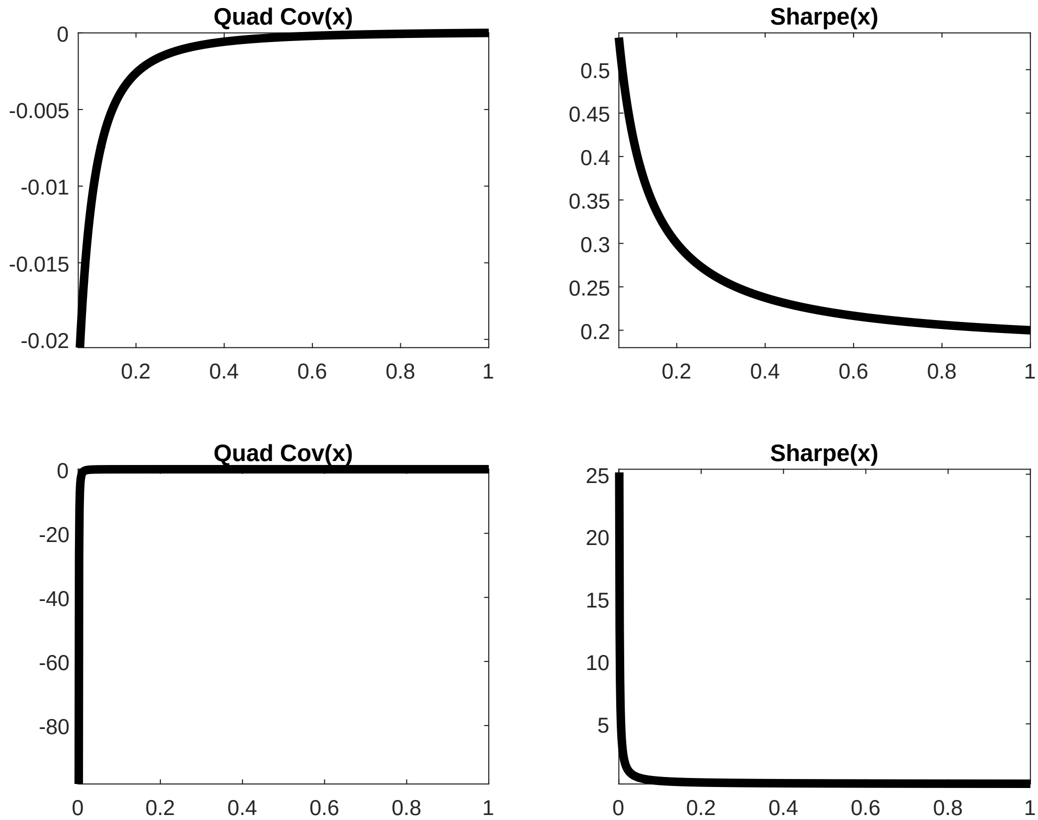

Figure 1.5: **Quadratic Co-variation and Sharpe Ratios, Case $\rho > \rho_v$**
In all sub-plots, the x-axis is the wealth share of the unconstrained investor, x_t. The top-left panel of the figure plots the quadratic co-variation between the risk premium, $\mathbb{E}_t(dR_t - r_t dt)/dt$, on the risky asset and its instantaneous volatility, $\sigma_R(\cdot)$. The top-left panel excludes the region of the state space that is in the immediate vicinity of the lower boundary of the state space. In the bottom-left panel of the figure, I plot the quadratic co-variation over the entire state space. Similarly, the top-right and bottom-right panels plot the Sharpe ratio over different regions of the state space. I plot the figure for $\bar{\sigma} = 0.15$, $\sigma_D = 0.20$, $\rho_v = 0.01$, and $\rho = 0.02$.

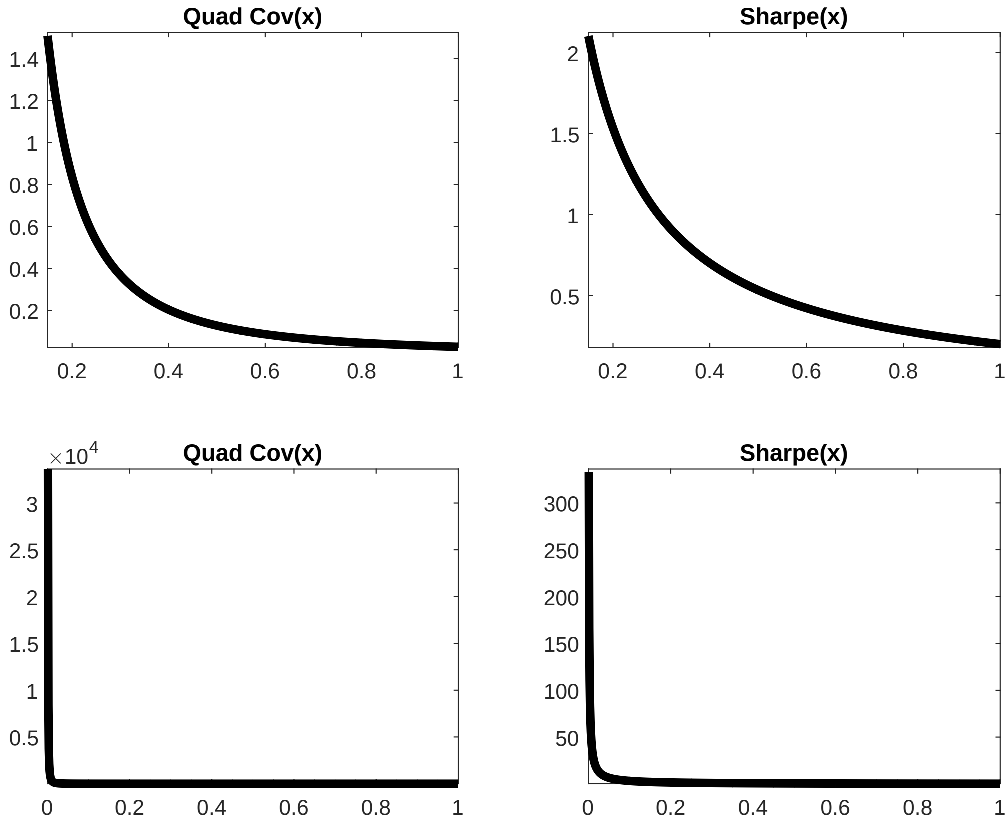

Figure 1.6: **Quadratic Co-variation and Sharpe Ratios, Case $\rho < \rho_v$**
In all sub-plots, the x-axis is the wealth share of the unconstrained investor, x_t. The top-left panel of the figure plots the quadratic co-variation between the risk premium, $\mathbb{E}_t(dR_t - r_t dt)/dt$, on the risky asset and its instantaneous volatility, $\sigma_R(\cdot)$. The top-left panel excludes the region of the state space that is in the immediate vicinity of the lower boundary of the state space. In the bottom-left panel of the figure, I plot the quadratic co-variation over the entire state space. Similarly, the top-right and bottom-right panels plot the Sharpe ratio over different regions of the state space. I plot the figure for $\bar{\sigma} = 0.1$, $\sigma_D = 0.2$, $\rho_v = 0.1$, and $\rho = 0.03$.

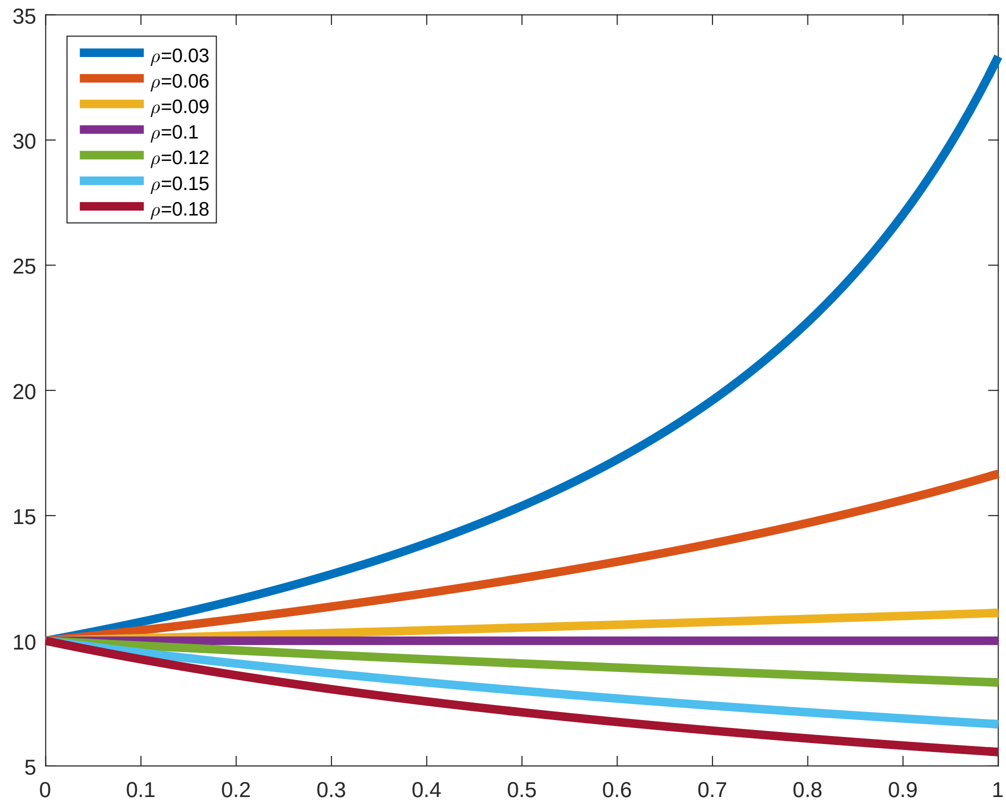

Figure 1.7: **Price-Dividend Ratio**
The figure plots the price-dividend ratio, S_t/D_t, of the risky asset. The x-axis is the wealth share of the unconstrained investor, x_t. By construction, the wealth share takes values between zero and one. I plot the figure for $\rho_v = 0.1$ and $\rho \in \{0.03, 0.06, 0.09, 0.10, 0.12, 0.15, 0.18\}$.

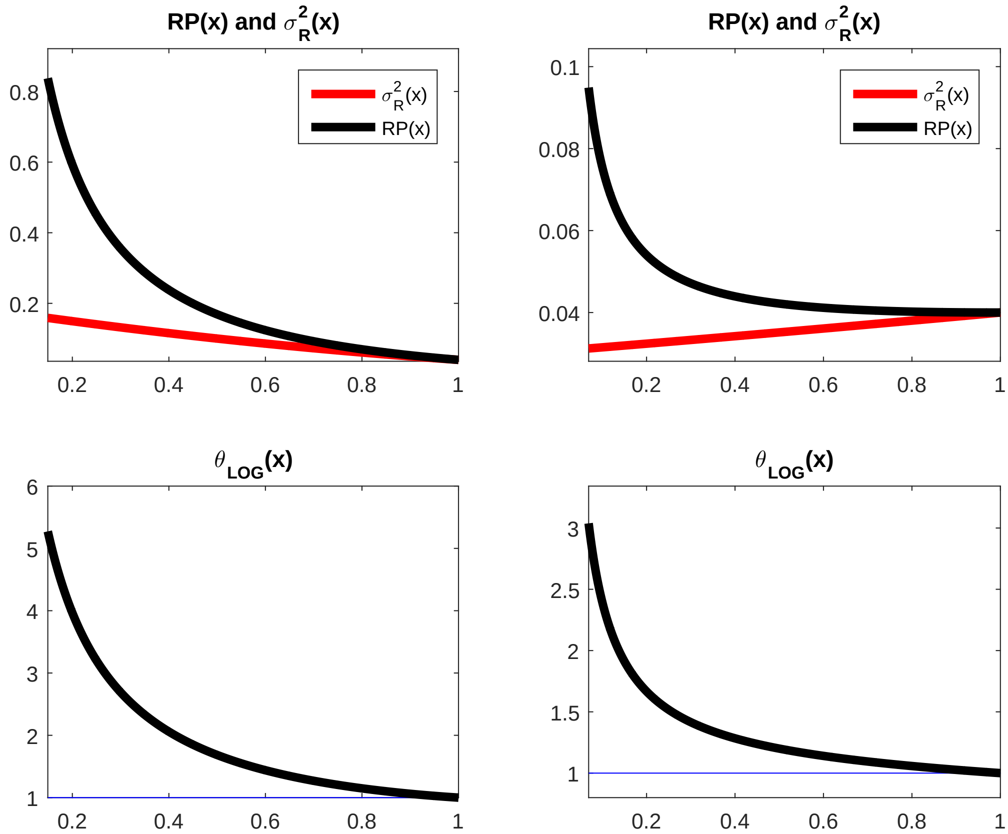

Figure 1.8: **Relation to Martin (2017)**
The top-left panel plots the risk premium, $\mathbb{E}_t(dR_t - r_t dt)/dt$, on the risky asset (black line) and the variance of the risky asset, $\sigma_R^2(x_t)$, (red line) for the case in which $\rho < \rho_v$. The bottom-left panel plots the portfolio choice of the unconstrained investor for the case in which $\rho < \rho_v$. The blue horizontal line passes through $\theta_t = 1$. The top-right panel plots the risk premium, $\mathbb{E}_t(dR_t - r_t dt)/dt$, on the risky asset (black line) and the variance of the risky asset, $\sigma_R^2(x_t)$, (red line) for the case in which $\rho > \rho_v$. The bottom-right panel plots the portfolio choice of the unconstrained investor for the case in which $\rho > \rho_v$. The blue horizontal line passes through $\theta_t = 1$. In all sub-plots, the x-axis is the wealth share of the unconstrained investor, x_t. By construction, the wealth share takes values between zero and one. In the figures, however, I plot the equilibrium quantities for values of x_t that are not in the immediate vicinity of the lower boundary of the state space. Please see the figure on the next page for the case in which $x \in (0, 1]$.

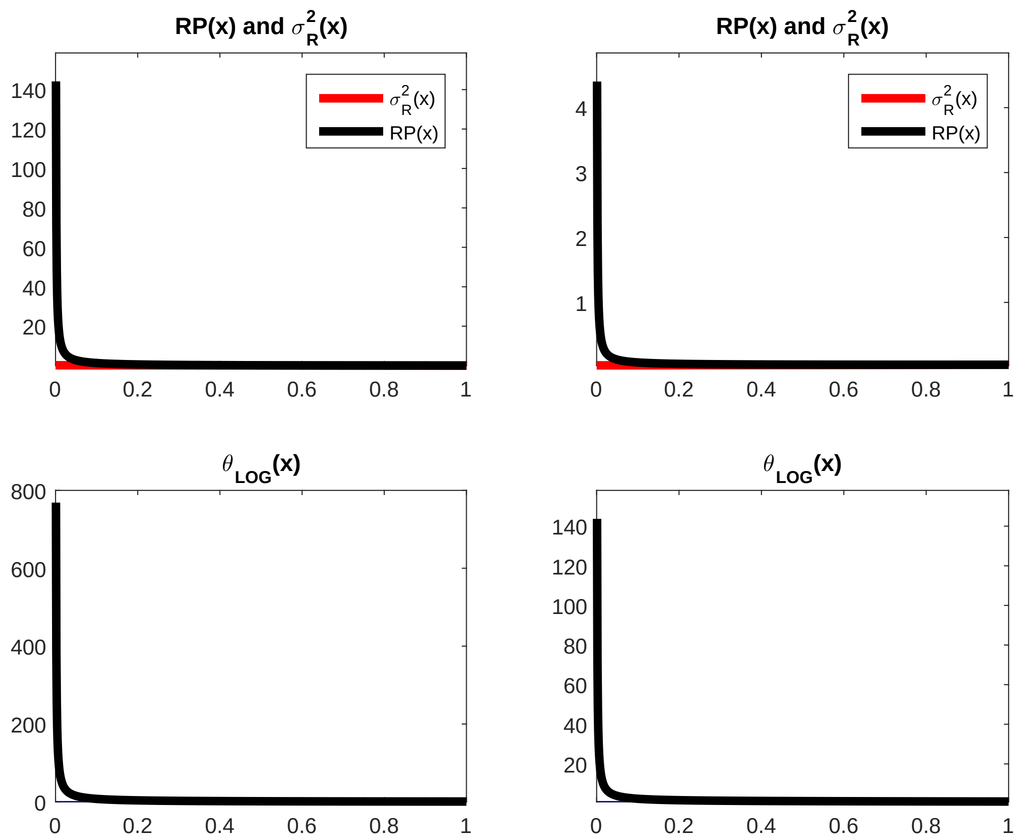

Figure 1.9: **Relation to Martin (2017)**
The top-left panel plots the risk premium, $\mathbb{E}_t(dR_t - r_t dt)/dt$, on the risky asset (black line) and the variance of the risky asset, $\sigma_R^2(x_t)$, (red line) for the case in which $\rho < \rho_v$. The bottom-left panel plots the portfolio choice of the unconstrained investor for the case in which $\rho < \rho_v$. The blue horizontal line passes through $\theta_t = 1$. The top-right panel plots the risk premium, $\mathbb{E}_t(dR_t - r_t dt)/dt$, on the risky asset (black line) and the variance of the risky asset, $\sigma_R^2(x_t)$ (red line) for the case in which $\rho > \rho_v$. The bottom-right panel plots the portfolio choice of the unconstrained investor for the case in which $\rho > \rho_v$. The blue horizontal line passes through $\theta_t = 1$. In all sub-plots, the x-axis is the wealth share of the unconstrained investor, x_t. By construction, the wealth share takes values between zero and one.

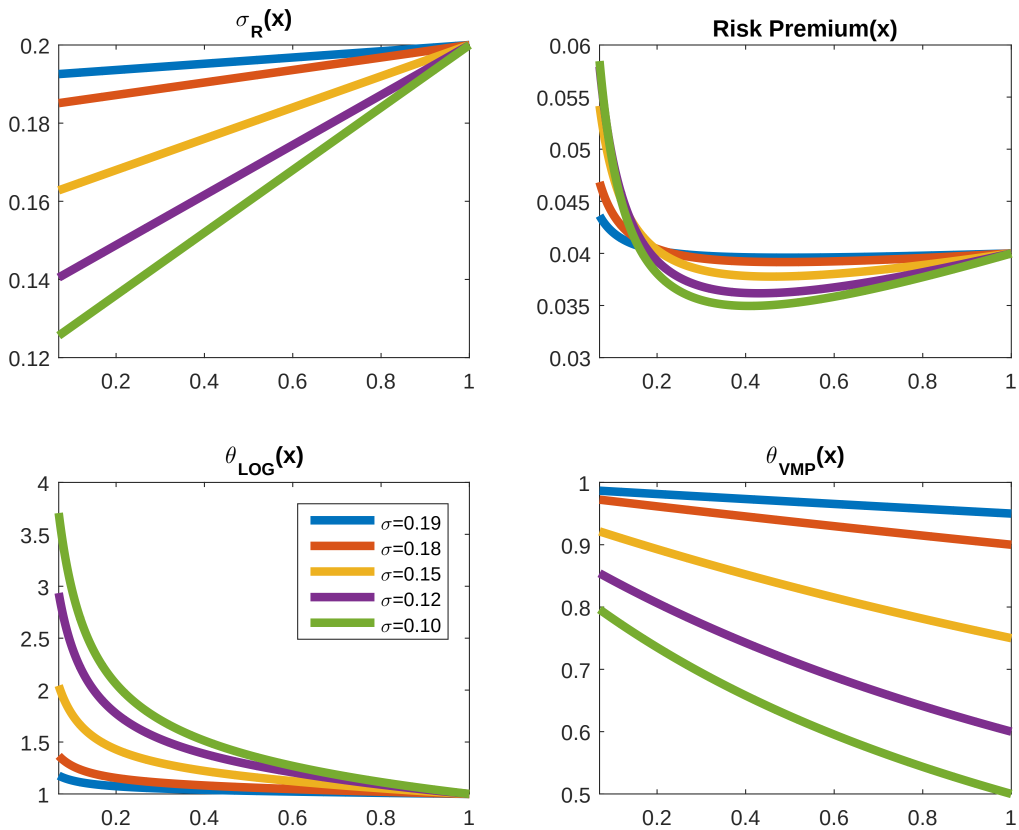

Figure 1.10: **Comparative Statics**
This figure plots equilibrium quantities of interest for the case in which $\rho > \rho_v$. In all sub-plots, the x-axis is the state variable of the model, x_t, which is the wealth share of the unconstrained investor. By construction, the wealth share takes values between zero and one. In this figure, however, I plot the volatility and the risk premium for values of x_t that are not in the immediate vicinity of the lower boundary of the state space. Please see the figure on the next page for the case in which $x \in (0, 1]$. The top-left panel depicts the volatility of the total return process, $\sigma_R(\cdot)$. The top-right panel depicts the risk premium on the risky asset, $\mathbb{E}_t(dR_t - r_t dt)/dt$. The bottom-left panel plots the portfolio policy of the unconstrained investor. Finally, the bottom-right panel plots the portfolio policy of the risk-neutral investor (volatility timer). I plot the figure for $\sigma_D = 0.20$, $\rho_v = 0.01$, $\rho = 0.05$, and $\bar{\sigma} = \{0.19, 0.18, 0.15, 0.12, 0.10\}$.

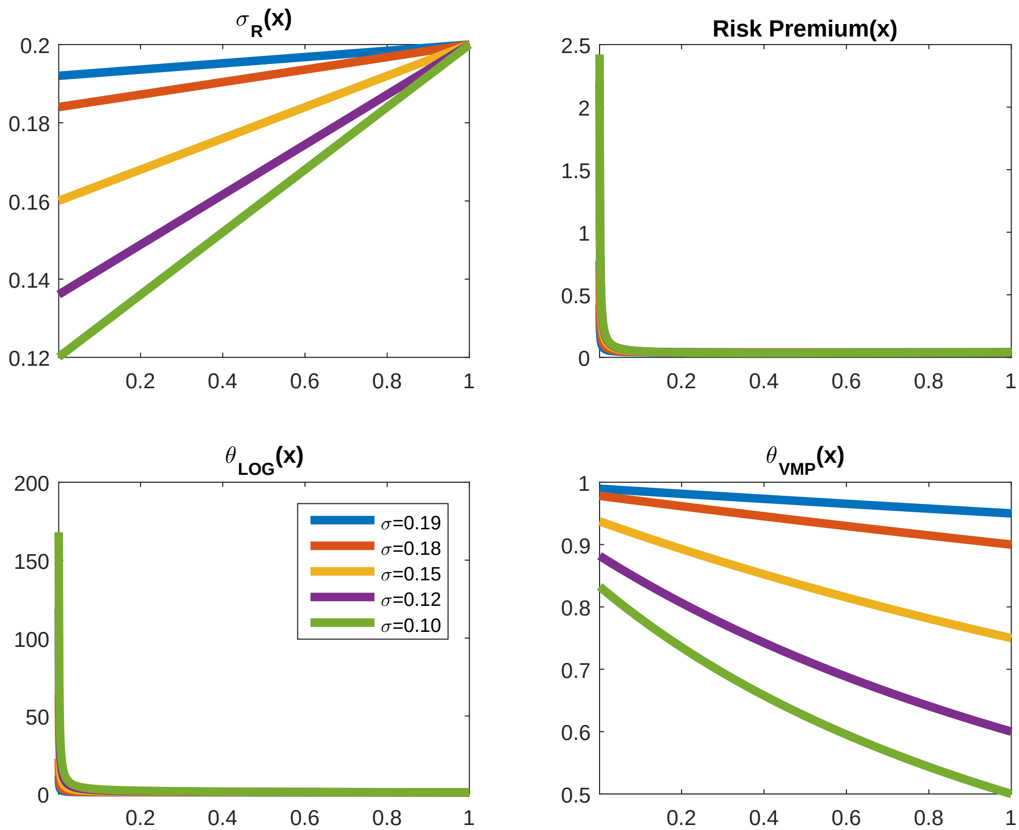

Figure 1.11: **Comparative Statics**
This figure plots equilibrium quantities of interest for the case in which $\rho > \rho_v$. In all sub-plots, the x-axis is the state variable of the model, x_t, which is the wealth share of the unconstrained investor. By construction, the wealth share takes values between zero and one. The top-left panel depicts the volatility of the total return process, $\sigma_R(\cdot)$. The top-right panel depicts the risk premium on the risky asset, $\mathbb{E}_t(dR_t - r_t dt)/dt$. The bottom-left panel plots the portfolio policy of the unconstrained investor. Finally, the bottom-right panel plots the portfolio policy of the risk-neutral investor (volatility timer). I plot the figure for $\sigma_D = 0.20$, $\rho_v = 0.05$, $\rho = 0.01$, and $\bar{\sigma} = \{0.19, 0.18, 0.15, 0.12, 0.10\}$.

Chapter 2

Intermediary Asset Pricing with Heterogeneous Financial Intermediaries

2.1. Introduction

The stochastic discount factor (SDF) takes a central place in asset pricing theory. The stochastic discount factor implied by standard equilibrium asset pricing models is proportional to the marginal value of wealth. Consequently, adverse states of the world, states in which the marginal value of wealth is high, are characterized by high values of the SDF, and good states of the world, states in which the marginal value of wealth is low, are characterized by low values of the SDF. Assets with payoffs that positively covary with the SDF are valuable hedges and are characterized by low expected returns. On the other hand, assets with payoffs that negatively covary with the SDF are very risky, carry a positive risk premium, and are characterized by high expected returns.

The standard consumption-based asset pricing literature focuses on the household sector of the economy and assumes that the aggregate household, a theoretical construct, is the marginal market participant in financial markets. Under this assumption, the SDF of the representative household prices all assets in the economy. Given that the SDF of the representative household is proportional to the marginal value of aggregate household wealth, asset returns depend on their covariance with aggregate household wealth. Notable extensions of the consumption-based literature, such as the habit formation model of Campbell and Cochrane (1999), the long-run risk model, Bansal and Yaron (2004), the multi-tree Lucas orchard model, Martin (2013a), and the heterogeneous agent framework in Chabakauri (2013), go a long way towards capturing some salient futures of financial markets. In all these models, however, the household sector occupies the driving seat, and the intermediary sector does not play any role in the price formation process.

Today, the direct ownership of financial assets on the part of the household sector is at historically low levels, and the very assumption that the representative household is the marginal player in financial markets is somewhat problematic. Even markets for plain-vanilla instruments are heavily dominated by institutional investors (financial intermediaries). For example, Koijen and Yogo (2015) report that, as of 2014, institutional investors hold 63% of the stock market. The assumption that the representative household is marginal in financial markets for complex financial instruments is even more problematic. This is because the share of institutional investors in highly specialized markets is even larger.

The standard literature on intermediary asset pricing, exemplified by He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014), shifts the focus of the attention from the household sector towards the intermediary sector and elevates financial intermediaries to the central stage of asset pricing. In particular, this strand of the literature assumes that sophisticated and resourceful financial intermediaries are better placed to trade in complex financial securities and posits that they are the marginal players in the market. Consequently, the SDF of the financial intermediaries, as opposed to the SDF of the aggregate household sector, constitutes a valid stochastic discount factor that prices all assets in the economy. The intermediary SDF is proportional to the level of financial intermediary capital. Upon the arrival of an adverse shock, the capital of the intermediary takes a hit and its capital position deteriorates. This leads to an increase in the marginal value of intermediary capital (wealth) and to a corresponding increase in the SDF. Consequently, assets with payoffs that negatively covary with negative shocks to intermediary capital are very risky, carry a positive risk premium, and offer high expected returns.

A casual observation of the composition of the financial services industry in its current form suffices to conclude that the financial sector exhibits a high degree of heterogeneity. Namely, a very broad variety of different types of financial institutions face different constraints and operate under very different objectives. While a narrow definition of the intermediary sector would only include commercial banks and large broker-dealers, a broader definition would encompass shadow banks, hedge funds, and some more traditional parts of the asset management industry. As a result, the notion of a *representative* financial intermediary is a very elusive concept. Early papers on intermediary asset pricing tend to focus on a particular type of financial intermediaries and largely ignore financial sector heterogeneity. For example, He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014) focus on *equity-constrained* financial intermediaries. This particular type of intermediaries can only issue a limited amount of outside equity, but can, in principle, borrow unlimited amounts of debt in fixed income markets. Models featuring *equity-constrained* intermediaries imply pro-cyclical intermediary capital dynamics (countercyclical leverage dynamics, because leverage is inversely proportional to capital). In contrast, models featuring *debt-constrained* intermediaries, for example Brunnermeier and Pedersen (2009), imply countercyclical intermediary capital dynamics and pro-cyclical leverage dynamics. In the latter class of models, the intermediary is unconstrained in the amount of equity capital it can issue (in Brunnermeier and Pedersen (2009) they can only issue capital with a lag, but the lag disappears in continuous time). It can, however, only issue a limited amount of debt.

Notwithstanding that these path-breaking papers offer many valuable insights and illuminate the links between intermediary capital and asset prices, they oversimplify the intermediary sector in a very extreme fashion. This is because each of these models focuses on a particular subset of the intermediary sector and implicitly assumes that the intermediaries in focus are the marginal intermediaries in the market. If the intermediary sector was homogeneous, or if, at the very least, it perfectly comprises vertically integrated intermediaries with operational internal capital markets, this approach would have been without loss of generality. There is, however, ample evidence to the contrary. In particular, He, Khang, and Krishnamurthy (2010) and Ang, Gorovyy, and Van Inwegen (2011), among others, report that different types of financial institutions exhibit very different behaviors and the intermediary sector is anything but homogeneous. They further show that financial institutions with comparatively stable funding bases, such as commercial banks, behave in a countercyclical way, and institutions dependent on repo and short-term financing, such as hedge funds, shadow banks, and broker-dealers, behave in a pro-cyclical way. Namely, upon the arrival of an adverse shock, the latter group faces binding funding constraints and finds itself forced to partially offload its risky asset exposure to the former group of intermediaries at fire-sale prices. This implies that different types of financial institutions are marginal in different states of the world (at different times). Consequently, any model featuring a homogeneous intermediary sector, populated by a single type of financial intermediary, would inevitably lead to a misspecification of the stochastic discount factor (SDF).

Additionally, models featuring a homogeneous intermediary sector are very unrealistic in the sense that there are not any countervailing mechanisms that can attenuate the effect of a hit to the capital position of the intermediary. Contrarian investors willing to support prices in adverse states of the world are completely absent from this type of models. In both He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014), shocks to capital lead to fire sales and to severe and prolonged recessions. The only path towards recovery is for intermediaries to very slowly beef up their capital bases. This mechanism is very counterfactual. In reality, different groups of intermediaries are hit at different times and complex financial assets never leave the aggregate balance sheet of the financial system. The bankruptcy of Lehman Brothers in September 2008 is an important case in point. In the aftermath of the bankruptcy, most of the assets previously held on the balance sheet of Lehman Brothers were sold off to stronger financial institutions, and the assets in question never left the balance sheet of the financial system.

Here, I propose an intermediary asset pricing model featuring a heterogeneous intermediary sector populated by *equity-constrained* and *shadow* financial intermediaries. I nest the heterogeneous intermediary sector in an otherwise standard dynamic general equilibrium Lucas (1978) economy with constrained market participation. The *shadow* intermediaries of the model map well to financial institutions with pro-cyclical leverage, such as broker-dealers, hedge funds, and shadow banks, Ang, Gorovyy, and Van Inwegen (2011). In general, these institutions rely on short-term repo and wholesale capital market funding. On the other hand, the *equity-constrained* intermediaries map to the group of financial institutions that are characterized by countercyclical leverage dynamics. These institutions tend to rely on more stable sources of funding, such as demand (term) deposits and accommodative central bank backdrops. Commercial banks constitute a good example of *equity-constrained* intermediaries. Whereas the *equity-constrained* financial intermediary is constrained in the amount of outside equity it can issue (intermediation constraint), the *shadow* financial intermediary operates under a leverage constraint. In the following sections, I describe the leverage constraint and the intermediation constraint in greater detail.

The explicit modeling of intermediary sector heterogeneity allows me to address some of the deficiencies inherent to the first generation of intermediary asset pricing models exemplified by Brunnermeier and Pedersen (2009), He and Krishnamurthy (2013), and Brunnermeier and Sannikov (2014) and featuring a homogeneous intermediary sector. In particular, I devote particular attention to the interplay between *shadow* and *equity-constrained* intermediaries in the price formation process. The main theoretical contribution of this chapter is threefold.

First, I show that when the leverage constraint of the *shadow* financial intermediary binds, intermediary sector heterogeneity reduces the risk premium on the intermediated risky asset in the constrained region. The reduction is relative to the case in which the intermediary sector is close to homogeneous. Even through the risk premium on the risky asset is decreasing in the wealth share of the *equity-constrained* intermediary (and to a lesser extent in the wealth share of the *shadow* financial intermediary) over most of the state space, the decrease is particularly large in the constrained region, where the intermediation constraint binds. The economic intuition behind the decrease in the risk premium is as follows. As I show in the chapter, the *equity-constrained* intermediary follows a contrarian (with respect to its wealth share) trading strategy. Namely, the intermediary increases leverage in states in which its wealth share is low and decreases leverage when its wealth share is high. In contrast, so long as the leverage constraint binds, the *shadow* financial intermediary follows a pro-cyclical trading strategy. It increases leverage in low-volatility states and decreases leverage in high-volatility states. Over the constrained region of the state space, volatility is decreasing in the wealth share of the *equity-constrained* intermediary. In other words, volatility is countercyclical with respect to its wealth share. Adverse states, where the wealth share of the *equity-constrained* financial intermediary is low and the intermediation constraint is tight are high-volatility states. When the wealth share of the *equity-constrained* intermediary is low, the risk premium should increase in order to induce the *equity-constrained* intermediary to increase its leverage and the risky asset market to clear. At the same time, the *shadow* financial intermediary reduces its leverage (because the state in which the wealth share of the *equity-constrained* intermediary is low is a high-volatility state), but does not entirely exit the market. Consequently, the risk premium should increase by less compared to the case in which the *equity-constrained* intermediary is the captive buyer of the risky asset. On the other hand, states in which the wealth share of the *equity-constrained* intermediary is high are low-volatility states. In these states, the contrarian *equity-constrained* intermediary decreases its leverage, and the *shadow* financial intermediary increases its risky asset position in accordance with the pro-cyclical trading strategy it follows. Therefore, when the wealth share of the *equity-constrained* intermediary is high, the *equity-constrained* intermediary partially offloads its risky asset holdings onto the *shadow* financial intermediary. This dampens the risk premium relative to the case in which the *shadow* financial intermediary, the agent who plays a price-supporting role in low-volatility states, is absent. In summary, the interplay between *equity-constrained* and *shadow* financial intermediaries is key to understanding the dynamics of the risk premium over different regions of the state space.

Second, I show that my model, featuring a heterogeneous intermediary sector, implies a pro-cyclical capital dynamics for the aggregate financial sector, where the aggregate financial sector combines *equity-constrained* and *shadow* financial intermediaries. This feature of my model is consistent with He and Krishnamurthy (2013) and He, Kelly, and Manela (2017) and is also on track with reality as the market capitalization of the aggregate financial sector is pro-cyclical. Even through the model implies countercyclical leverage for the aggregate financial sector, *equity-constrained* and *shadow* financial intermediaries exhibit very distinct and rich leverage dynamics. Whereas, over the constrained region of the state space, the model-implied leverage of the *equity-constrained* intermediary is countercyclical, the model-implied leverage of the *shadow* financial intermediary is pro-cyclical (when the leverage constraint binds). The latter result is consistent with Adrian, Etula, and Muir (2014). My model resolves the apparent contradiction between Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017); the former paper claims that the market price of risk of equity capital is negative, and the latter paper claims that it is positive. I show that Brunnermeier and Pedersen (2009), the theoretical precursor of Adrian, Etula, and Muir (2014), applies *within* the financial sector and He and Krishnamurthy (2013), the theoretical precursor of He, Kelly, and Manela (2017), applies at the level of the aggregate intermediary sector.

Finally, I prove that the constrained region shrinks relative to the benchmark model in which the intermediary sector is homogeneous. This results is consequential. It now takes a bigger adverse shock, relative to the benchmark model, for the economy to enter the constrained region, the region over which the intermediation constraint of the *equity-constrained* financial intermediaries binds. As I explain in detail below, this is mainly due to the shock absorption role (price supporting role) of the *shadow* financial intermediary.

This chapter relates to two different strands of the literature. The first studies intermediary asset pricing models. In a pioneering paper, He and Krishnamurthy (2011) proposed a dynamic general equilibrium intermediary asset pricing model. In their paper, optimal contracting considerations endogenously give rise to intermediation. In a closely related paper, He and Krishnamurthy (2013), the authors model risk premium dynamics in times of crises. Brunnermeier and Sannikov (2014) study the full equilibrium dynamics of an economy with financial frictions and show that their economy is prone to instability and occasionally enters volatile crisis episodes.

Endogenous risk, driven by asset illiquidity, plays a central role in their paper. In all these models, the net worth of the financial intermediary is a major determinant of asset prices. All these models, however, fail to take into account intermediary sector heterogeneity.

This chapter also relates to the literature on general equilibrium asset pricing in continuous time. In particular, it is related to models with segmented markets and limited market participation, Basak and Cuoco (1998) and Alvarez, Atkeson, and Kehoe (2002). My model is also related to the literature studying the equilibrium effects of constraints. Detemple and Murthy (1997), Basak and Croitoru (2000), and Kogan, Makarov, and Uppal (2007), among others, study the equilibrium effects of constraints.

The remainder of this chapter is organized as follows. Section 2.2 discusses the economic setup and elaborates on the equilibrium conditions. In Section 2.3, I outline the model solution approach and analytically solve the model. In particular, I solve for the boundary separating the constrained region from the unconstrained region, fully characterize the equilibrium, and derive equilibrium quantities of interest. All equilibrium quantities depend on the state variables of the model, the wealth share of the *equity-constrained* financial intermediary, and the wealth share of the *shadow* financial intermediary. I conclude the section with a discussion of the results and offer economic intuition. Finally, Section 2.4 concludes. In the mathematical Appendix, I prove all propositions from the main body of the paper and provide detailed derivations.

2.2. Model

The top panel of Figure 2.1 summarizes the main facets of my model. The model that I propose features a heterogeneous intermediary sector that is populated by two distinct types of financial intermediaries, an *equity-constrained* intermediary and a *shadow* financial intermediary. Whereas the *equity-constrained* intermediary is constrained in the amount of outside equity it can issue, it can freely borrow in the market for the risk-free asset. It is instructive to compare the augmented model (top panel of the figure) to the baseline model, He and Krishnamurthy (2013), that features a homogeneous intermediary sector. In the latter model, in the bottom panel of Figure 2.1, the intermediary sector of the economy solely comprises *equity-constrained* financial intermediaries,[1] and they are the only party allowed to trade the intermediated risky asset. To obtain my model, I augment He and Krishnamurthy (2013) by adding a second type of financial intermediaries, *shadow* financial intermediaries. In the augmented model, both types of intermediaries have access to the risky asset.

[Insert Figure 2.1]

There is a single (complex) risky asset and a risk-free bond. Below, I refer to the complex risky asset as a risky asset, or an intermediated risky asset. The risky asset is complex in the sense that trading this asset requires a certain degree of sophistication above the level of sophistication typical for the representative retail investor (household). Proprietary hedge fund strategies, structured equity products, and broad market indexes, all of which track difficult to trade and/or illiquid securities, constitute good examples of complex risky assets that are not directly investable by the general public.

In the model, I only allow financial intermediaries, which are representative of sophisticated investors, to trade in the risky asset. This assumption is without loss of generality. Today, the direct ownership of financial assets is at historically low levels. For example, as of 2014, institutional investors (intermediaries) hold 63% of the stock market, Koijen and Yogo (2015). In more specialized markets for complex financial instruments, the share of institutional investors is even higher. For example, Edwards, Harris, and Piwowar (2007) argue that only 2% of corporate bond trades are initiated by retail investors, and Siriwardane (2016) reports that the credit default swap (CDS) market in the United States is dominated by five big institutional players.

[1] specialists, or intermediaries in the language of He and Krishnamurthy (2013).

2.2.1. Assets

I model an infinite-horizon pure exchange Lucas (1978) economy cast in continuous time. A single perishable consumption good serves as a numeraire. Following convention, I normalize the total supply of the risky asset to one unit. The risk-free asset is in zero net supply. The holder of the risky asset is entitled to the dividend stream, $\{D_t\}$. The dividend process, $\{D_t\}$, follows the drift-diffusion

$$\frac{dD_t}{D_t} = gdt + \sigma dB_t,$$

where $g \in \mathbb{R}_{++}$ and $\sigma \in \mathbb{R}_{++}$ are exogenous constants. The diffusion coefficient, σ, admits the interpretation of fundamental risk in the economy. $\{B_t\}$ is a standard one-dimensional Wiener process.[2] The total return process, $\{R_t\}$, follows

$$dR_t = \frac{D_t dt + dP_t}{P_t},$$

where P_t is the price of the risky asset. As I will show below, $\{R_t\}$ follows a drift-diffusion process, and its diffusion coefficient is a function of fundamental risk, σ, and of the endogenously generated risk. The equilibrium risk-free interest rate, r_t, and the process thereof, $\{r_t\}$, are determined in equilibrium.

For the reasons that I outline above, the household sector of the economy can only invest in the risk-free asset and in the equity of the *equity-constrained* financial intermediary. *Equity-constrained* and the *shadow* financial intermediaries can invest in both the risky and the risk-free assets. For simplicity, the *shadow* financial intermediary cannot issue outside equity and is not allowed to invest in the equity of the *equity-constrained* financial intermediary.

[2] I define the one-dimensional Wiener process, $\{B_t\}$, on the filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$. I denote by $\mathbf{F}$ the augmented filtration generated by the Wiener process. The filtered probability space satisfies the *usual conditions*, i.e., the filtration is complete and right-continuous. Here, a process in this paper is by definition a stochastic process; that is progressively measurable with respect to $\{\mathcal{F}_t\}$.

2.2.2. *Equity-Constrained Financial Intermediary*

Brunnermeier and Sannikov (2014) and He and Krishnamurthy (2013) exemplify the class of intermediary models featuring *equity constraints* (capital constraints). The defining characteristic of this class of models is that they impose restrictions on the maximum amount of outside equity that agents can rise. There are, however, no any restrictions on borrowing in the risk-free asset. The representative *equity-constrained* financial intermediary maximizes its expected utility from instantaneous consumption,

$$\mathbb{E}\left(\int_0^\infty e^{-\rho t} u(C_t) dt\right) \quad \rho > 0,$$

where C_t is instantaneous consumption and ρ is the subjective discount factor of the intermediary. I consider a logarithmic utility function,

$$u(C_t) = \ln(C_t).$$

The total capital of the intermediary is the sum of its inside (own) equity, W_t, and the capital that the representative household decides to allocate to the intermediary, H_t. In the sequel, I follow the terminology of Brunnermeier and Sannikov (2014) and call H_t outside equity. The intermediary invests its total capital and is unconstrained in its portfolio choice. Let α_t^I be the risky asset position of the *equity-constrained* intermediary expressed as a fraction of total capital. Then, the return on total intermediary capital, $\tilde{R}_t$, follows the process

$$d\tilde{R}_t = r_t dt + \alpha_t^I (dR_t - r_t dt),$$

where R_t is the total return on the risky asset defined above. It is instructive to note that there is no limit on net borrowing. The intermediary can finance a leveraged position, $\alpha_t^I > 1$, in the risky asset by shorting $(\alpha_t^I - 1)(W_t + H_t)$ worth of the risk-free asset.

Following the extant literature on equity constraints, I assume a capital constraint of the form

$$H_t \leq mW_t,$$

where $m \in \mathbb{R}_{++}$ is a positive constant parameterizing the intermediation constraint. The constraint admits a very intuitive interpretation. Namely, the representative household is unwilling to invest more than a fraction of what the *equity-constrained* intermediary invests, mW_t. Whereas W_t is endogenously determined in equilibrium, I can exogenously control the tightness of the capital constraint, for any given value of W_t, by varying m.

The intermediation constraint is tight for low values of m and it loosens for high values of m. For a given m, the supply of intermediation, H_t, is increasing in the wealth of the intermediary, W_t. Consequently, in adverse states, characterized by low levels of *equity-constrained* intermediary capital, the ability of the household to indirectly participate in the risky asset market through its outside equity stake in the intermediary is severely impaired. Below, I show that the tightness of the intermediation constraint has profound implications for the risk premium on the risky asset.

As He and Krishnamurthy (2013) elucidate, capital constraints of the form $H_t \leq mW_t$ link net worth and external financing and are widely used in the financial frictions literature pioneered by the classical work of Holmstrom and Tirole (1997) and Kiyotaki and Moore (1997), among others. Agency and informational frictions open a possible avenue to the micro-foundation of the intermediation constraint. They also admit the interpretation of a *skin-in-the-game* requirement, one that is instrumental in the alignment of the incentives of households, on the one hand, and financial intermediaries, on the other hand.

The representative *equity-constrained* financial intermediary solves the consumption and portfolio choice problem

$$\max_{\{C_t,\alpha_t^I\}} \quad \mathbb{E}\left(\int_0^\infty e^{-\rho t}u(C_t)dt\right) \quad \rho > 0,$$

subject to the dynamic constraint

$$dW_t = -C_t dt + r_t W_t dt + \alpha_t^I W_t (dR_t - r_t dt), \tag{2.1}$$

where the first term on the right-hand side is instantaneous consumption and the second term is the risk-free return. $\alpha_t^I W_t$ is the dollar size of the risky asset position and $dR_t - r_t dt$ is the excess return on the risky asset. It is instructive to note that it is up to the representative household to decide whether and how much to invest in the outside equity of the *equity-constrained* intermediary. The intermediary passively accepts the outside equity supplied by the household and adds it to its inside equity. Therefore, the amount of outside equity capital that the *equity-constrained* financial intermediary raises is outside of its control. The equity capital raised from the household also admits the interpretation of a separately managed (wealth management) account; that is, the intermediary manages the account on behalf of the household in return for a management fee. Given that there are no any agency frictions between the intermediary and the representative household and the return on inside equity is equal to the return on outside equity, I can normalize the management fee to zero without loss of generality. The return on total intermediary capital, $W_t + H_t$, follows the diffusion process

$$d(W_t + H_t) = -\,C_t dt + r_t (W_t + H_t) dt + \alpha_t^I (W_t + H_t)(dR_t - r_t dt).$$

It is immediate to see that the evolution of total wealth admits the unique decomposition

$$dW_t = -\,C_t dt + r_t W_t dt + \alpha_t^I W_t (dR_t - r_t dt),$$

$$dH_t = r_t H_t dt + \alpha_t^I H_t (dR_t - r_t dt),$$

where the processes $\{W_t\}$ and $\{H_t\}$ govern the evolution of inside and outside intermediary equity, respectively. The amount of outside equity is pivotal for asset prices as risky asset demand (in terms of number of shares) is increasing in H_t. This claim directly follows from the fact that

$$\tilde{\alpha}_t^I = \alpha_t^I \frac{W_t + H_t}{S_t},$$

where $\tilde{\alpha}_t^I$ is the risky asset demand of the *equity-constrained* financial intermediary in terms of number of shares.

2.2.3. Households

In the interest of tractability, I model the household sector as overlapping generations (OLG) of agents. A measure one of identical households optimize over

$$\rho_h \ln(C_t^H)dt + (1 - \rho_h dt)\mathbb{E}_t(\ln W_{t+dt}^H),$$

where ρ_h is the subjective discount factor of the household, C_t^H is household consumption, and W_t^H is household wealth. Given that utility is logarithmic, it is optimal for the agent to consume a constant fraction, ρ_h, of its wealth, $C_t^H = \rho_h W_t^H$. In this chapter, I assume that $\rho_h > \rho$; that is, households are less patient than the *equity-constrained* intermediary.

I divide the household sector into two distinct sub-sectors. A measure $\lambda \in [0, 1]$ of households can invest in the risk-free asset only. This modeling assumption creates baseline demand for the risk-free asset and ensures that the aggregate financial sector is leveraged at all times. I allow the complementary fraction, $1 - \lambda$, of households to invest in the risky equity of the *equity-constrained* intermediary. For expositional simplicity, I aggregate the latter group of households into a representative household. The OLG structure of the household sector gives rise to a myopic portfolio optimization problem. Namely, the representative household maximizes the mean-variance objective

$$\max_{\alpha_t^H} \quad \alpha_t^H \mathbb{E}_t(d\tilde{R}_t - r_t dt) - \frac{1}{2}(\alpha_t^H)^2 \mathrm{Var}_t(d\tilde{R}_t - r_t dt),$$

where the size of the position in the outside equity of the *equity-constrained* intermediary, α_t^H, is the only choice variable. The optimization problem of the representative household is subject to a dynamic constraint:

$$\alpha_t^H (1-\lambda) W_t^H \leq m W_t.$$

The household earns an instantaneous return of $\tilde{R}_t$ on its risky holdings. This is because the household invests in the equity of the *equity-constrained* financial intermediary and not in the risky asset directly. A brief clarification of the dynamic constraint is in order. W_t^H is total household wealth and $1-\lambda$ is the measure of households allowed to invest in the risky intermediary equity. Consequently, $(1-\lambda)W_t^H$ is the maximum amount of capital that the representative household can invest in the equity of the intermediary, absent any portfolio constraints. The imposition of the capital constraint, however, caps the dollar amount of the risky intermediary equity position of the representative household at mW_t. The stochastic differential equation

$$dW_t^H = (lD_t - \rho_h W_t^H)dt + r_t W_t^H dt + \alpha_t^H (1-\lambda) W_t^H (d\tilde{R} - r_t dt)$$

governs the intergenerational evolution of household wealth. The first term in the drift coefficient is the labor income of the household. It flows at the rate of lD_t per unit of time, dt. The labor income ameliorates the risk that the household sector vanishes, e.g., Dumas (1989), thus rendering the general equilibrium analysis uninteresting. The second component of the drift coefficient, $\rho_h W_t$, is the optimal consumption of the household. Finally, $\alpha_t^H (1-\lambda) W_t^H$ is the dollar size of its risky position, and $d\tilde{R}_t - r_t dt$ is the excess return that the household earns on its risky investment.

2.2.4. *Shadow Financial Intermediary*

The *shadow* financial intermediary solves the consumption and portfolio choice problem

$$\max_{\{C_t^D, \alpha_t^D\}} \quad \mathbb{E}\left(\int_0^\infty e^{-\rho_d t} u(C_t^D) dt\right) \quad \rho_d > 0,$$

subject to the dynamic budget constraint

$$dW_t^D = -C_t^D dt + r_t W_t^D dt + \alpha_t^D W_t^D (dR_t - r_t dt),$$

and to the leverage constraint

$$\alpha_t^D \leq \frac{\bar{\sigma}}{\sigma_R(\cdot)}.$$

I denote the instantaneous consumption of the *shadow* financial intermediary by C_t^D and its subjective discount factor by ρ_d, where $\rho > \rho_d$. The *shadow* financial intermediary is endowed with a logarithmic utility function, $u(C_t^D) = \ln(C_t^D)$. $\alpha_t^D W_t^D$ is the dollar size of its risky asset position and $dR_t - r_t dt$ is the excess return on the risky asset. It is instructive to note that I express α_t^D as a fraction of total wealth and α_t^D admits the interpretation of leverage. The leverage constraint caps the amount of leverage that the *shadow* financial intermediary can take. Over the region of the state space where the leverage constraint binds,

$$\alpha_t^D = \frac{\bar{\sigma}}{\sigma_R(\cdot)}.$$

Please note that I solve for the region where the leverage constraint binds and derive an expression for the endogenous volatility, $\sigma_R(\cdot)$, in the following sections. $\bar{\sigma} \in \mathbb{R}_{++}$

admits the interpretation of a risk limit. Clearly, over the region where the leverage constraint binds, the agent trades in a pro-cyclical fashion. It scales up its risky asset exposure in states in which volatility is low and exits the market at times of high volatility.

Notwithstanding that the *shadow* financial intermediary of my model is not subject to a debt constraint, it shares many similarities with the intermediaries implied by intermediary asset pricing models featuring a debt constraint, for example Brunnermeier and Pedersen (2009). Consequently, the *shadow* financial intermediary of my model admits the interpretation, albeit with a slight abuse of terminology, of what the literature calls *debt-constrained* intermediaries.

2.2.5. Equilibrium Conditions

In this subsection, I formally define the equilibrium concept used to solve the model. After enlisting all equilibrium conditions, I outline the model solution strategy that I follow in the following sections.

Definition 2. *An equilibrium is a set of price processes and investment policies* $\{\alpha^I(\cdot), \alpha^D(\cdot), \alpha^H(\cdot)\}$ *such that the investment policies solve the dynamic portfolio optimization problems of the household, of the equity constrained intermediary, and of the shadow financial intermediary.*

1. *Given the price process, the household, the equity-constrained intermediary, and the shadow financial intermediary solve their respective portfolio optimization problems.*

2. *Portfolio decisions are constrained by the intermediation constraint,* $H_t \leq mW_t$*.*

3. *The equity-constrained intermediary is unconstrained in its portfolio choice.*

4. *The goods market clears*

$$C_t + C_t^H + C_t^D = D_t(1 + l).$$

5. *The market for the risky asset clears*

$$\alpha_t^I(W_t + \alpha_t^H(1-\lambda)W_t^H) + \alpha_t^D W_t^D = P_t.$$

6. *The market for the risk-free asset clears by Walras' law.*

Given that I model a pure exchange Lucas (1978) economy, it should be the case that in equilibrium total wealth equals the price of the risky asset,

$$W_t + W_t^H + W_t^D = P_t.$$

A further elaboration on the market clearing condition,

$$\alpha_t^I(W_t + \alpha_t^H(1-\lambda)W_t^H) + \alpha_t^D W_t^D = P_t,$$

is in order. The aggregate household wealth is W_t^H, and only a measure $1-\lambda$ of households can invest in the intermediary equity. Therefore, $(1-\lambda)W_t^H$ is the total capital of the group of households that can invest in the equity of the *equity-constrained* intermediary. Consequently, $\alpha_t^H(1-\lambda)W_t^H$ is the outside equity of the *equity-constrained* intermediary that households supply and $W_t + \alpha_t^H(1-\lambda)W_t^H$ is its total capital (inside and outside equity). Finally, $\alpha_t^I(W_t + \alpha_t^H(1-\lambda)W_t^H)$ is the dollar size of the risky asset position of the *equity-constrained* intermediary.

Equipped with the above equilibrium conditions, I solve for the equilibrium outcome in terms of the state variables of the model, $x_t \in [0,1]$ and $\beta_t \in [0,1]$. The former is the wealth share of the *equity-constrained* financial intermediary, and the latter is the wealth share of the *shadow* financial intermediary.

2.2.6. Stochastic Discount Factor (SDF)

Whereas the household sector and the *shadow* financial intermediary face portfolio constraints, the *equity-constrained* financial intermediary is unconstrained in its portfolio choice. Consequently, the *equity-constrained* intermediary is always marginal in the market for the intermediated risky asset and its stochastic discount factor (SDF) prices all assets in the economy. The risk premium on the risky asset,

$$\mathbb{E}_t(dR_t) - r_t dt = \mathrm{Cov}_t\left(\frac{dC_t}{C_t}, dR_t\right),$$

directly follows from the Euler equation of the *equity-constrained* intermediary that is endowed logarithmic utility over instantaneous consumption,

$$-\rho dt - \mathbb{E}_t\left(\frac{dC_t}{C_t}\right) + \mathrm{Var}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{E}_t(dR_t) = \mathrm{Cov}_t\left(\frac{dC_t}{C_t}, dR_t\right).$$

Please refer to the mathematical Appendix for detailed derivations of the Euler equation and of the risk premium on the intermediated risky asset.

2.3. Model Solution

I solve the model in terms of the two state variables,

$$x_t = \frac{W_t}{P_t}, \quad \beta_t = \frac{W_t^D}{P_t}.$$

Whereas x_t is the wealth share of the *equity-constrained* financial intermediary, β_t is the wealth share of the *shadow* financial intermediary. The model is tractable. All equilibrium quantities of interest admit closed-form representations.

Here, I define some terminology for further reference. By constrain region I refer to the region of the state space over which the intermediation constraint binds. By unconstrained region I refer to the region of the state space over which the intermediation constraint does not bind. The dynamics of all equilibrium quantities in the constrained region are markedly different from their counterparts in the unconstrained region. For this reason, I first derive the boundary separating the two regions and then solve for the variables of interest by region.

2.3.1. Separating Boundary

The boundary separating the constrained region from the unconstrained region, x_c, solves

$$\begin{aligned} \alpha_t^{I,\text{constrained}} &= \alpha_t^{I,\text{unconstrained}}, \\ \frac{1-\alpha_t^D \beta_t}{x_t(1+m)} &= \frac{1-\alpha_t^D \beta_t}{x_t+(1-\lambda)(1-x_t-\beta_t)}. \end{aligned}$$

The second equality comes from the expressions for $\alpha_t^{I,\text{constrained}}$ and $\alpha_t^{I,\text{unconstrained}}$ that I derive in the mathematical Appendix. Solving the above expression for x_t, I obtain the cutoff, x_c,

$$x_c = \frac{(1-\lambda)(1-\beta_t)}{1+m-\lambda}.$$

For any given β_t, the intermediation constraint binds for $x_t \in [0, x_c)$, and it is slack for $x_t \in [x_c, 1-\beta_t]$. Some economic intuition is in order. For any given β_t, the wealth share of the household sector, $1 - x_t - \beta_t$, is inversely related to the wealth share of the *equity-constrained* intermediary, x_t. When x_t is high (to the right of x_c), $1 - x_t - \beta_t$ is low and the household sector has only a limited amount of capital to invest. As a result, the intermediation constraint is slack.

Clearly, for any given β_t, the absolute value of the cutoff, x_c, is increasing in the tightness of the intermediation constraint, m. In other words, an increase in m (loosening of the constraint) pushes x_c to the left, and the constrained region of the state space shrinks. This is because for high values of m, households can invest in the equity of the *equity-constrained* intermediary even in states in which their inside equity is low. The absolute value of x_c is decreasing in the wealth share of the *shadow* financial intermediary, β_t. This is because for high values of β_t the sum of the wealth shares of the household sector and of the *equity-constrained* intermediary is low. The risk-taking capacity of the households is severely impaired, and the constraint is unlikely to bind. Finally, the cutoff x_c is increasing in the measure of households allowed to invest in the risky asset, $1-\lambda$. For the extreme case in which households are not allowed to invest in the risky intermediary equity, $\lambda = 1$, the economy is unconstrained over the entire state space. As the measure of households allowed to invest in the risky asset increases, hitting the constraint becomes easier, and the boundary separating the two regions moves to the right (x_c increases).

It is instructive to compare the separating boundary implied by my model, featuring a heterogeneous intermediary sector, to its counterpart, x_c^H, implied by He and Krishnamurthy (2013), a model with homogeneous intermediary sector

$$x_c^H = \frac{1-\lambda}{1+m-\lambda}.$$

In the following proposition, I summarize some of the most important differences between the two results.

Proposition 8. (Separating Boundary)

- *Intermediary sector heterogeneity loosens the intermediation constraint. For any given value of β_t, the separating boundary, x_c, moves to the left (towards the lower boundary of the state space), and the intermediation constraint does not bind for lower levels of equity-constrained financial intermediary equity capital.*

- *The magnitude of the shift in x_c depends on β_t, the wealth share of the shadow financial intermediary.*

A brief note is in order. For $\beta_t = 0$, the model reduces to a model with homogeneous intermediary sector populated by *equity-constrained* intermediaries. For $\beta_t = 1$, the model reduces to a model with homogeneous intermediary sector populated by *shadow* financial intermediaries.

2.3.2. Risk Premium

Given that the *equity-constrained* financial intermediary is endowed with logarithmic utility over instantaneous consumption, it is optimal for it to consume a constant fraction, ρ, of its wealth between time t and time $t+dt$. As a result, consumption growth equals wealth growth,

$$\frac{dC_t}{C_t} = \frac{dW_t}{W_t}.$$

Consequently, in the logarithmic utility case, the risk premium on the intermediated risky asset simplifies to

$$\mathbb{E}_t(dR_t) - r_t dt = \mathrm{Cov}_t\left(\frac{dW_t}{W_t}, dR_t\right), \tag{2.2}$$

$$\mathbb{E}_t(dR_t) - r_t dt = \alpha_t^I \mathrm{Var}_t(dR_t), \tag{2.3}$$

where the last equation follows from the dynamic budget constraint of the *equity-constrained* financial intermediary. The portfolio choice of the *equity-constrained* financial intermediary is key to understanding the risk premium. In the constrained region, α_t^I is given by

$$\alpha_t^I = \underbrace{\frac{1}{x_t(1+m)}}_{\substack{\text{He and Krishnamurthy (2013)} \\ \text{Homogeneous Intermediaries}}} - \underbrace{\frac{\beta_t \alpha_t^D}{x_t(1+m)}}_{\text{Heterogeneous Intermediaries}}.$$

Please see the mathematical Appendix for a derivation. The expression for α_t^I admits a very intuitive decomposition. Namely, it decomposes into the portfolio weight, $1/(x_t(1+m))$, implied by the standard model with homogeneous intermediary sector, He and Krishnamurthy (2013), and a second term, $-\beta_t\alpha_t^D/(x_t(1+m))$, which takes into account intermediary sector heterogeneity. Importantly, both α_t^I from the above decomposition and the risk premium on the risky asset depend on the portfolio choice of the *shadow* financial intermediary. High risky asset demand on the part of the *shadow* financial intermediary has to be compensated by low risky asset demand on the part of the *equity-constrained* financial intermediary in order for the market for the risky asset to clear.

2.3.3. Characterization of Equilibrium

First, I consider the region of the state space over which both the intermediation constraint and the leverage constraint bind, i.e., the economy is in the constrained region of the state space, and the leverage constraint binds. In this region,

$$\alpha_t^D = \min\left(\alpha_t^I, \frac{\bar{\sigma}}{\sigma_R(\cdot)}\right) = \frac{\bar{\sigma}}{\sigma_R(\cdot)}.$$

Namely, the *shadow* financial intermediary follows a trading strategy that is inversely proportional to the level of volatility. I then derive an expression for the volatility, $\sigma_R(\cdot)$, of the total return process, $\{R_t\}$, and show that it is a function of the state variables x_t and β_t. The following proposition reports the equilibrium outcome.

Proposition 9. (Equilibrium, Both Constraints Bind)
Over the region of the state space in which both the intermediation constraint and the leverage constraint bind, the equilibrium volatility, the portfolio policy functions, and the risk premium are given by

$$\begin{aligned}
\sigma_R(x_t, \beta_t) &= \left(1 - \frac{A}{1+m}\right)^{-1} \left(\sigma(1 - Ax_t - B\beta_t) + B\bar{\sigma}\beta_t - \frac{A\bar{\sigma}\beta_t}{1+m}\right), \\
\alpha_t^I &= \frac{1 - \alpha_t^D \beta_t}{(1+m)x_t} = \frac{1 - \frac{\bar{\sigma}}{\sigma_R(x_t,\beta_t)}\beta_t}{(1+m)x_t}, \\
\alpha_t^H &= \frac{mx_t}{(1-\lambda)(1 - x_t - \beta_t)}, \\
\alpha_t^D &= \frac{\bar{\sigma}}{\sigma_R(x_t, \beta_t)}, \\
\mathrm{RP}_t &= \alpha_t^I \sigma_R^2(x_t, \beta_t),
\end{aligned}$$

where

$$A \doteq 1 - \frac{\rho}{\rho_h}, \quad B \doteq 1 - \frac{\rho_d}{\rho_h}.$$

Please see the mathematical Appendix for a detailed proof of the proposition. For any given $\beta_t > 0$, the intermediary sector is close to homogeneous for $x_t \to 0$, and it is heterogeneous for $x_t > 0$. This is because for $x_t \to 0$, the intermediary sector solely comprises the *shadow* financial intermediaries, and for $x_t > 0$ both *equity-constrained* and *shadow* financial intermediaries populate the intermediary sector. Clearly, for the case in which both the intermediation constraint and the leverage constraint bind, the volatility of the total return process is linear in the wealth shares x_t and β_t, the state variables of the model. In the Appendix, I show that, for any given value of β_t, the boundary, $\hat{x}_c$, that separates the constrained region over which the leverage constraint binds from the constrained region over which it does not bind is given by

$$\hat{x}_c = \frac{\beta_t \bar{\sigma}(1-B) - \sigma(1-B\beta_t)}{A(\bar{\sigma}-\sigma) - \bar{\sigma}(1+m)}.$$

Next, I consider the region of the state space over which the intermediation constraint binds, but the leverage constraint does not bind, i.e., the economy is in the constrained region of the state space, and the leverage constraint does not bind. The following proposition summarizes the equilibrium outcome.

Proposition 10. (Equilibrium, Only the Intermediation Constraint Binds) *Over the region of the state space in which only the intermediation constraint binds, the equilibrium volatility, the portfolio policy functions, and the risk premium are given by*

$$\sigma_R(x_t, \beta_t) = \left(1 - \frac{Ax_t + B\beta_t}{(1+m)x_t + \beta_t}\right)^{-1} \sigma(1 - Ax_t - B\beta_t),$$
$$\alpha_t^I = \frac{1}{(1+m)x_t + \beta_t},$$

$$\begin{aligned}
\alpha_t^H =& \frac{m x_t}{(1-\lambda)(1-x_t-\beta_t)},\\
\alpha_t^D =& \alpha_t^I,\\
\mathrm{RP}_t =& \alpha_t^I \sigma_R^2(x_t,\beta_t).
\end{aligned}$$

Please see the mathematical Appendix of the paper for a detailed proof of the proposition. Next, I consider the region of the state space over which the intermediation constraint does not bind, but the leverage constraint binds, i.e., the economy is in the unconstrained region of the state space, and the leverage constraint binds. The following proposition summarizes the equilibrium outcome.

Proposition 11. (Equilibrium, Only the Leverage Constraint Binds) *Over the region of the state space in which only the leverage constraint binds, the equilibrium volatility, the portfolio policy functions, and the risk premium are given by*

$$\begin{aligned}
\sigma_R(x_t,\beta_t) =& \left(1-\frac{A x_t}{x_t\lambda+(1-\lambda)(1-\beta_t)}\right)^{-1}\\
&\times\left(\sigma(1-Ax_t-B\beta_t)+B\bar{\sigma}\beta_t-\frac{A\bar{\sigma}\beta_t x_t}{\lambda x_t+(1-\lambda)(1-\beta_t)}\right),\\
\alpha_t^I =& \frac{1-\alpha_t^D\beta_t}{x_t\lambda+(1-\lambda)(1-\beta_t)} = \frac{1-\frac{\bar{\sigma}}{\sigma_R(x_t,\beta_t)}\beta_t}{x_t\lambda+(1-\lambda)(1-\beta_t)},\\
\alpha_t^H =& 1,\\
\alpha_t^D =& \frac{\bar{\sigma}}{\sigma_R(x_t,\beta_t)},\\
\mathrm{RP}_t =& \alpha_t^I\sigma_R^2(x_t,\beta_t).
\end{aligned}$$

In the Appendix, I show that for any given value of β_t, the boundary, $\hat{x}_u$, that separates the unconstrained region over which the leverage constraint binds from the unconstrained region over which it does not bind is given by

$$\hat{x}_u = \frac{\bar{\sigma}(\beta_t(\lambda-B)+1-\lambda)-\sigma(1-B\beta_t)}{A(\bar{\sigma}-\sigma)-\bar{\sigma}\lambda}.$$

Finally, I consider the region of the state space over which neither the intermediation constraint nor the leverage constraint binds, i.e., the economy is in the unconstrained region of the state space, and the leverage constraint does not bind. The following proposition summarizes the equilibrium outcome.

Proposition 12. (Equilibrium, Neither Constraint Binds)

Over the region of the state space in which neither the intermediation constraint nor the leverage constraint binds, the equilibrium volatility, the portfolio policy functions, and the risk premium are given by

$$\begin{aligned}
\sigma_R(x_t, \beta_t) &= \left(1 - \frac{Ax_t + B\beta_t}{x_t\lambda + \beta_t\lambda + 1 - \lambda}\right)^{-1} \sigma(1 - Ax_t - B\beta_t), \\
\alpha_t^I &= \frac{1}{x_t\lambda + \beta_t\lambda + 1 - \lambda}, \\
\alpha_t^H &= 1, \\
\alpha_t^D &= \alpha_t^I, \\
\mathrm{RP}_t &= \alpha_t^I \sigma_R^2(x_t, \beta_t).
\end{aligned}$$

For realistic values of the risk limit, $\bar{\sigma}$, the leverage constraint is likely to bind over the constrained region of the state space (the region in which the intermediation constraint binds). As a result, the case in which the economy is in the constrained region and the leverage constraint binds is of particular interest. For this reason, below I devote more attention to the study of the equilibrium outcome in the aforementioned case. In the following proposition, I study the properties of the total return process volatility, the portfolio policy functions, and the risk premium on the intermediated risky asset. The analysis of the properties of the diffusion coefficient is important, because it is the main driver of most of the equilibrium quantities of interest, such as the risk premium on the risky asset and the leverage dynamics.

Proposition 13. Equilibrium Quantities, Properties

Consider the region of the state space in which both the intermediation constraint and the leverage constraint bind. For $\bar{\sigma} < \sigma + A\sigma/(B(1+m) - A)$,

- *The total return process volatility, $\sigma_R(x_t, \beta_t)$, is decreasing in the wealth share of the equity-constrained financial intermediary, x_t, and in the wealth share of the shadow financial intermediary, β_t,*

$$\frac{\partial \sigma_R(x_t, \beta_t)}{\partial x_t} < 0, \quad \frac{\partial \sigma_R(x_t, \beta_t)}{\partial \beta_t} < 0.$$

- *The portfolio choice of the shadow financial intermediary is pro-cyclical with respect to the wealth shares of the intermediaries, x_t and β_t,*

$$\frac{\partial \alpha_t^D}{\partial x_t} > 0, \quad \frac{\partial \alpha_t^D}{\partial \beta_t} > 0.$$

- *The portfolio choice of the equity-constrained financial intermediary is countercyclical with respect to x_t and β_t,*

$$\frac{\partial \alpha_t^I}{\partial x_t} < 0, \quad \frac{\partial \alpha_t^I}{\partial \beta_t} < 0.$$

- *The risk premium on the intermediated risky asset is decreasing in the degree of intermediary sector heterogeneity.*

So long as the risk limit is not too loose, $\bar{\sigma} < \sigma + A\sigma/(B(1+m) - A)$, the volatility of the total return process exhibits countercyclical dynamics with respect to x_t and β_t over the constrained region of the state space. As a result of this, the leverage of the *shadow* financial intermediary exhibits pro-cyclical dynamics, and the leverage of the *equity-constrained* intermediary exhibits countercyclical dynamics. More importantly, the risk premium on the intermediated risky asset is decreasing in the degree of intermediary sector heterogeneity. For any given value of β_t (x_t), the intermediary sector is close to homogeneous for $x_t \to 0$ ($\beta_t \to 0$) and heterogeneous for $x_t > 0$ ($\beta_t > 0$). As the wealth share of the *equity-constrained* (*shadow* financial) intermediary moves away from its lower boundary, the degree of financial sector heterogeneity increases and the risk premium on the risky asset decreases.

Consequently, the risk premium is decreasing in the degree of intermediary sector heterogeneity. Intuitively, as x_t increases the intermediation constraint loosens and households can invest more in the outside equity of the intermediary. This leads to an increase in the risky asset demand on the part of the *equity-constrained* intermediary and to a corresponding decrease in the risk premium. In the knife-edge case, where $\beta_t \to 0$, the model reduces to a model with homogeneous intermediary sector populated by *equity-constrained* intermediaries, and I recover the He and Krishnamurthy (2013) results.

2.3.4. Price-Dividend Ratio

I impose market clearing in the goods market and obtain an expression for the price of the risky asset as a function of exogenous dividends, D_t, the endogenous wealth of the *equity-constrained* financial intermediary, W_t, and the endogenous wealth of the *shadow* financial intermediary, W_t^D,

$$P_t = \frac{1+l}{\rho_h} D_t + \left(1 - \frac{\rho}{\rho_h}\right) W_t + \left(1 - \frac{\rho_d}{\rho_h}\right) W_t^D.$$

Please see the mathematical Appendix for a detailed derivation. The first term on the right-hand side admits the interpretation of liquidation value of the risky asset. When W_t, the wealth of the *equity-constrained* financial intermediary, and W_t^D, the wealth of the *shadow* financial intermediary, approach zero, (near-complete disintermediation of the intermediary sector) the price of the risky asset approaches the liquidation value from above.

The price of the risky asset is increasing in the labor income of the household sector, l. This is because wealthier households invest more in the outside equity of the *equity-constrained* financial intermediary. This leads to higher risky asset demand on the part of the *equity-constrained* financial intermediary as its risky asset demand is increasing, in dollar terms, in total capital. Under the parametric restriction $\rho_h > \rho$; that is, households are impatient relative to the *equity-constrained* intermediary, $1 - \rho/\rho_h > 0$, and the price of the risky asset is increasing in *equity-constrained* intermediary wealth, W_t. The larger the differential between the two, ρ and ρ_h, the stronger the effect.

Whereas the price of the risky asset is a function of D_t, W_t, and W_t^D, the price-dividend ratio only depends on the state variables of the model (and on exogenous parameters), which are the wealth share of the *equity-constrained* financial intermediary x_t, and the wealth share of the *shadow* financial intermediary β_t. The price-dividend ratio is given by

$$\frac{P_t}{D_t} = \frac{1 + l}{\rho_h - (\rho_h - \rho_d)\beta_t - (\rho_h - \rho)x_t}.$$

The price-dividend ratio exhibits pro-cyclical dynamics, with respect to x_t, in the sense that it goes up in good states (when the wealth share of the *equity-constrained* financial intermediary is high) and down in bad states. This is because the tightness of the intermediation constraint is decreasing in x_t.

2.3.5. Model-Implied Leverage Dynamics

Models featuring debt constraints, for example, Brunnermeier and Pedersen (2009), imply pro-cyclical leverage dynamics (countercyclical capital dynamics). In other words, the price of risk of intermediary capital is negative. On the other hand, models featuring equity constraints, for example, Brunnermeier and Sannikov (2014) and He and Krishnamurthy (2013), imply counter-cyclical leverage dynamics (pro-cyclical capital dynamics). In these models, the price of risk of intermediary capital is positive.

In a recent paper, He, Kelly, and Manela (2017) empirically show that intermediary capital is strongly priced in a broad cross-section of assets and the price of risk of intermediary capital is positive. This implies a negative price of risk for leverage. On the other hand, Adrian, Etula, and Muir (2014) report a negative price of risk for intermediary capital (positive price of risk of leverage). Thus, the results related to the sign of the market price of risk in Adrian, Etula, and Muir (2014) are in direct contradiction to the results that He, Kelly, and Manela (2017) report.

He, Kelly, and Manela (2017) devote an entire chapter of their paper to the reconciliation of their results with Adrian, Etula, and Muir (2014). Even though they offer numerous different explanations, the authors conclude that the results are most likely different because the two papers analyse different types of financial institutions. Whereas Adrian, Etula, and Muir (2014) define intermediaries as stand-alone U.S. broker-dealers and broker-dealer subsidiaries of conglomerates, He, Kelly, and Manela (2017) calculate equity capital using holding company-level data.

Below, I study the leverage dynamics implied by my model. I first derive the leverage dynamics at the level of the aggregate financial sector. Let W_t^F be the capital (wealth) of the aggregate financial sector and let α_t^F be the risky asset position thereof. Given that only financial intermediaries are allowed to hold the risky asset directly, market clearing implies

$$\alpha_t^F W_t^F = P_t.$$

The wealth of the aggregate financial sector is

$$W_t^F = W_t + \alpha_t^H (1-\lambda) W_t^H + W_t^D.$$

The first term on the right-hand side of the expression is the inside equity of the *equity-constrained* financial intermediary. The outside equity that households supply to the *equity-constrained* financial intermediary is $\alpha_t^H(1-\lambda)W_t^H$, where $(1-\lambda)W_t^H$ is the total wealth of the group of households that are allowed to invest in the risky asset. Finally, W_t^D is the capital of the *shadow* financial intermediary. I then solve for the risky asset position of the aggregate intermediary sector. It admits the representation

$$\alpha_t^F = \frac{P_t}{W_t^F} = \frac{1}{(x_t + \beta_t)\lambda + 1 - \lambda}$$

in the unconstrained region of the state space (the region in which the intermediation constraint does not bind). Similarly, the risky asset position of the aggregate financial sector is

$$\alpha_t^F = \frac{P_t}{W_t^F} = \frac{1}{x_t(1+m)+\beta_t}$$

in the constrained region. Please see the mathematical Appendix for a detailed derivation.

Proposition 14. *The leverage of the aggregate intermediary sector (combining equity-constrained and shadow financial intermediaries) is countercyclical.*

In Proposition 13, I show that when both the leverage constraint and the intermediation constraint bind, the leverage of the *shadow* financial intermediary is pro-cyclical with respect to x_t and β_t (the state variables), and the leverage of the *equity-constrained* intermediary is countercyclical with respect to x_t and β_t.

Consequently, He and Krishnamurthy (2013) applies at the level of the aggregate financial sector, because it implies counter-cyclical leverage dynamics. In contrast, Brunnermeier and Pedersen (2009) implies pro-cyclical leverage and it only describes the leverage dynamics of a particular type of a financial intermediary, that of the *shadow* financial intermediary. In other words, Brunnermeier and Pedersen (2009) applies within the financial sector. My model, featuring a heterogeneous intermediary sector, generates rich leverage dynamics, and, as I show below, allows me to study flows between different types of intermediaries within the intermediary sector.

2.3.6. Discussion of Results and Economic Intuition

In this section, I calibrate the model and plot the equilibrium quantities of interests as functions of the state variables x_t and β_t. Notwithstanding that in the logarithmic-utility case all equilibrium quantities of interest admit closed-form expressions, the graphical representation of the results enables me to convey the underlying economic intuition in a more concise and accessible way. For expositional simplicity, I focus on a model calibration in which the leverage constraint binds over the entire state space.

[Insert Figure 2.2]

In Figure 2.2, I plot the boundary separating the constrained region from the unconstrained region for the different values of x_t and β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sector sum to one, by construction. For this reason, the region in dark blue is not attainable. The line separating the yellow region from the green region is the separating boundary. The economy is constrained in the yellow region (both the intermediation constraint and the leverage constraint bind) and unconstrained in the green region (the intermediation constraint does not bind, but the leverage constraint binds). For $\beta_t = 0$, the intermediary sector solely comprises the *equity-constrained* intermediary, and the model reduces to the benchmark model in He and Krishnamurthy (2013). On the other extreme, the case in which $\beta_t = 1$, the intermediary sector is again homogeneous, but this time it is solely populated by the *shadow* financial intermediary. The main region of interest is the region where $\beta_t \in (0, 1)$ and $x_t \in (0, 1)$, and the intermediary sector is heterogeneous. As the wealth share of the *shadow* financial intermediary, β_t, goes up the constrained region shrinks relative to the size of the unconstrained region. This is because the relative importance of the *equity-constrained* intermediary is decreasing in the wealth share of the *shadow* financial intermediary.

[Insert Figure 2.3]

[Insert Figure 2.4]

Figures 2.3 and 2.4 depict the risk premium on the risky asset. Whereas Figure 2.3 plots the risk premium over the entire state space, Figure 2.4 excludes the region that is in the immediate vicinity of the lower boundaries of the state space. Therefore, Figure 2.4 is a zoomed-in version of Figure 2.3. Notwithstanding that all equilibrium quantities are well defined over the entire state space, they take very extreme values in the vicinity of the lower boundaries of the state space, where $x_t \to 0$, or $\beta_t \to 0$, or $x_t \to 0$ and $\beta_t \to 0$. For this reason, Figure 2.4 is particularly useful in analyzing my results.

The risk premium is decreasing in x_t, the wealth share of the *equity-constrained* financial intermediary, and in β_t, the wealth share of the *shadow* financial intermediary. In the constrained region, a decrease in the wealth share of the *equity-constrained* intermediary, x_t, leads to a sharp increase in the risk premium. As the *equity-constrained* intermediary follows a countercyclical (countercyclical with respect to x_t) or contrarian trading strategy, it increases leverage as x_t falls. The lower the wealth share, the more aggressively the intermediary trades. Even though the *equity-constrained* intermediary follows an admissible trading strategy and not a doubling down strategy, its trading strategy is reminiscent of a doubling down strategy. As time goes by, the aggressive contrarian trading strategy starts bearing fruits. The *equity-constrained* intermediary earns a very high risk premium on a massively leveraged position. Its wealth recovers, and its wealth share rises. This mechanism pushes x_t towards the separating boundary (away from the boundaries of the state space).

As it is clear from the figure, the risk premium is decreasing in both x_t and β_t. It is instructive to note, however, that the risk premium is more sensitive to x_t. The amplification mechanism that is at play is key to understanding the sensitivity of the risk premium with respect to x_t. Consider a drop in the wealth of the *equity-constrained* intermediary, W_t. Any drop in W_t tightens the intermediation constraint, $H_t \leq mW_t$. In other words, it potentially reduces the maximum amount of outside equity that households can contribute in the constrained region. If the intermediation constraint is binding, a one-unit decrease in W_t leads to a reduction of outside equity by $m\Delta W_t = m$ units. Consequently, the total capital of the *equity-constrained* intermediary decreases by $1 + m$ units. Keeping α_t^I constant, the dollar amount of the risky asset position of the *equity-constrained* intermediary decreases by $(1+m)\alpha_t^I$ units. As is evident from the top panel of Figure 2.7, however, the *equity-constrained* intermediary effectively follows a contrarian investment strategy; that is, α_t^I goes up as x_t falls. Consequently, the trading strategy of the intermediary acts as a countervailing force to the above amplification mechanism and partially attenuates its effect. On balance, the risk premium should be high for low values of x_t in order to induce the intermediary to invest and the market for the risky asset to clear.

[Insert Figure 2.5]

[Insert Figure 2.6]

When x_t is low, volatility is high (please refer to Figures 2.5 and 2.6) and the *shadow* financial intermediary sells the risky asset. The risk premium should increase in order to induce the contrarian *equity-constrained* intermediary to increase its leverage and clear the market. For low values of x_t, the *shadow* financial intermediary reduces its risky asset position, but does not entirely exit the market. As a result, the risk premium should increase by less, compared to the case in which the intermediary sector is homogeneous and the *equity-constrained* intermediary is the only party that is allowed to trade in the risky asset market. In the unconstrained region, volatility is low and $x_t + \beta_t$ is high. As the *shadow* financial intermediary follows a pro-cyclical (with respect to x_t and β_t) trading strategy, it increases its leverage and α_t^D goes up. At the same time, the contrarian *equity-constrained* intermediary reduces its leverage. The *equity-constrained* intermediary sells, and the *shadow* financial intermediary buys, and by doing so, it supports prices and lowers the risk premium. The incremental demand on the part of the *shadow* financial intermediary dampens the risk premium, compared to the case in which the *equity-constrained* intermediary is the only buyer. In summary, it is not the case that a particular investor type determines the risk premium over the entire state space. On the contrary, the risk premium is governed by the interplay between *equity-constrained* and *shadow* financial intermediaries.

When $x_t \to 0$, the *equity-constrained* intermediary becomes fully dis-intermediated, and the heterogeneous intermediary sector reduces to a homogeneous intermediary sector populated by *shadow* financial intermediaries. On the other hand, when $\beta_t \to 0$, the *shadow* financial intermediary becomes fully dis-intermediated, and the heterogeneous intermediary sector reduces to a homogeneous intermediary sector populated by *equity-constrained* financial intermediaries. Importantly, over the constrained region of the state space, the risk premium on the risky asset is decreasing in intermediary sector heterogeneity. This is because for any given value of β_t, the risk premium is decreasing in x_t, and for any given value of x_t it is decreasing in β_t. The risk premium attains a maximum in the vicinity of the state in which $x_t \to 0$ and $\beta_t \to 0$. This is the state in which the household sector owns most of the wealth in the economy.

[Insert Figure 2.7]

The top panel of Figure 2.7 depicts the risky asset position of the *equity-constrained* intermediary, α_t^I, over the entire state space, and the bottom panel of the figure plots α_t^D, the portfolio policy function of the *shadow* financial intermediary.

Consistent with the theoretical results from the previous section, α_t^I exhibits countercyclical behavior with respect to x_t and β_t over the constrained region of the state space. In other words, when the wealth share of the *equity-constrained* intermediary is low and the intermediation constraint is binding, the *equity-constrained* intermediary increases its leverage. The *shadow* financial intermediary follows a procyclical (with respect to x_t and β_t) trading strategy (bottom panel of Figure 2.7). As volatility is decreasing in x_t (Figures 2.5 and 2.6), the trading strategy of the *shadow* financial intermediary also admits the interpretation of a volatility timing strategy. It increases the size of the risky asset position in times of low volatility (high x_t states) and reduces exposure in high-volatility states (low x_t states). Importantly, the effect of the introduction of the *shadow* financial intermediary is not limited to a shift in the boundary that separates the constrained region from the unconstrained region. The introduction of the *shadow* financial intermediary profoundly affects the trading behavior of the *equity-constrained* intermediary over the entire state space.

2.4. Conclusion

I propose a general equilibrium intermediary asset pricing model featuring a heterogeneous financial sector comprising *equity-constrained* and *shadow* financial intermediaries.

I show that, in the constrained region, the risk premium on the risky asset is decreasing in the degree of intermediary sector heterogeneity. This is because *shadow* financial intermediaries trade in a pro-cyclical fashion and *equity-constrained* intermediaries follow a contrarian trading strategy. At times of high volatility, the *shadow* financial intermediary offloads its risky asset holdings onto the *equity-constrained* intermediary. Thus, in high-volatility states, the trading behavior of the *equity-constrained* intermediary ameliorates the pricing pressure and reduces the risk premium. The opposite holds true in low-volatility regimes. Being a contrarian, the *equity-constrained* intermediary wants to sell. At the same time, the *shadow* financial intermediary wants to buy as it effectively follows a volatility timing trading strategy. Thus, in low-volatility states, the *shadow* financial intermediary ameliorates the pricing pressure on the risky asset and dampens the risk premium.

Additionally, intermediary sector heterogeneity allows for a very rich leverage dynamics within the intermediary sector.

2.5. Mathematical Appendix

2.5.1. Euler Equation of the Equity-Constrained Financial Intermediary

The *equity-constrained* financial intermediary is unconstrained in its portfolio choice and is always marginal in the market. Consequently, the stochastic discount factor of the *equity-constrained* intermediary constitutes a valid stochastic discount factor. The SDF, $\Lambda(\cdot)$, is proportional to the marginal utility of consumption of the *equity-constrained* financial intermediary,

$$\Lambda_t \propto e^{-\rho t}\frac{\partial u(C_t)}{\partial C_t}.$$

I then do Itô on Λ_t and note that

$$\frac{d\Lambda_t}{\Lambda_t} = -\rho dt + \frac{e^{-\rho t}}{\Lambda_t}\frac{\partial^2 u(C_t)}{\partial C_t^2}dC_t + \frac{1}{2}\frac{e^{-\rho t}}{\Lambda_t}\frac{\partial^3 u(C_t)}{\partial C_t^3}d\langle C, C\rangle_t.$$

The last step is to impose a drift restriction on $d\Lambda_t R_t$ and use the fact that

$$u(C_t) = \ln(C_t).$$

After a straightforward simplification I obtain the Euler equation,

$$-\rho dt - \mathbb{E}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{V}\mathrm{ar}_t\left(\frac{dC_t}{C_t}\right) + \mathbb{E}_t(dR_t) = \mathbb{C}\mathrm{ov}_t\left(\frac{dC_t}{C_t}, dR_t\right),$$

of the *equity-constrained* financial intermediary.

■

2.5.2. Portfolio Choice

In this subsection, I derive expressions for the optimal portfolio policy functions of the representative household, of the *equity-constrained* financial intermediary, and of the *shadow* financial intermediary. I do the derivations for the constrained and unconstrained regions of the state space separately.

- Constrained Region

  Note that in the constrained region $\alpha_t^H(1-\lambda)W_t^H = mW_t$. This is because in the constrained region the intermediation constraint, $H_t \leq mW_t$, binds and the household sector invests the maximum permissible amount, mW_t, in the outside equity of the *equity-constrained* intermediary. Solving for α_t^H, I obtain the portfolio policy function of the household,

$$\alpha_t^H = \frac{mx_t}{(1-\lambda)(1-x_t-\beta_t)}.$$

  The portfolio policy function of the *shadow* financial intermediary, α_t^D, takes the form

$$\alpha_t^D = \min\left(\alpha_t^I, \frac{\bar{\sigma}}{\sigma_R(\cdot)}\right).$$

  When the leverage constraint does not bind, the portfolio choice of the *shadow* financial intermediary is the same as the portfolio choice of the *equity-constrained* financial intermediary. This is because both intermediaries maximize logarithmic utility over instantaneous consumption. I derive the portfolio policy function of the *equity-constrained* financial intermediary from the market clearing condition,

$$\alpha_t^I(W_t + mW_t) + \alpha_t^D W_t^D = P_t.$$

  Consequently,

$$\alpha_t^I = \frac{1-\alpha_t^D\beta_t}{x_t(1+m)}.$$

  There are two cases to consider. For the case in which the leverage constraint binds,

$$\alpha_t^D = \frac{\bar{\sigma}}{\sigma_R(\cdot)}, \quad \alpha_t^I = \frac{1}{x_t(1+m)} - \frac{\beta_t}{x_t(1+m)} \frac{\bar{\sigma}}{\sigma_R(\cdot)}.$$

For the case in which the leverage constraint does not bind,

$$\alpha_t^D = \alpha_t^I, \quad \alpha_t^I = \frac{1}{(1+m)x_t + \beta_t}.$$

Below, I summarize the portfolio policy functions of all agents of the model in the constrained region.

- Region in which the leverage constraint binds

$$\begin{aligned}
\alpha_t^I =& \frac{1 - \alpha_t^D \beta_t}{(1+m)x_t} = \frac{1 - \frac{\bar{\sigma}}{\sigma_R(\cdot)}\beta_t}{(1+m)x_t}, \\
\alpha_t^H =& \frac{m x_t}{(1-\lambda)(1 - x_t - \beta_t)}, \\
\alpha_t^D =& \frac{\bar{\sigma}}{\sigma_R(\cdot)}.
\end{aligned}$$

- Region in which the leverage constraint does not bind

$$\begin{aligned}
\alpha_t^I =& \frac{1}{(1+m)x_t + \beta_t}, \\
\alpha_t^H =& \frac{m x_t}{(1-\lambda)(1 - x_t - \beta_t)}, \\
\alpha_t^D =& \alpha_t^I.
\end{aligned}$$

- Unconstrained Region

In the unconstrained region $\alpha_t^H = 1$. Below, I formally prove this claim. Households that are allowed to invest in the risky *equity-constrained* intermediary equity optimize over

$$\max_{\alpha_t^H} \quad \alpha_t^H \mathbb{E}_t(d\tilde{R}_t - r_t dt) - \frac{1}{2}(\alpha_t^H)^2 \mathrm{Var}_t(d\tilde{R}_t - r_t dt)$$

subject to their dynamic budget constraint. In the unconstrained region, the intermediation constraint is slack by construction. I obtain

$$\alpha_t^H = \frac{\mathbb{E}_t(d\tilde{R}_t) - r_t dt}{\mathrm{Var}(d\tilde{R}_t)}$$

from the FOC. Using the facts that $d\tilde{R}_t - r_t dt = \alpha_t^I(dR_t - r_t dt)$, and $(\mathbb{E}_t(dR_t - r_t dt))/dt = \alpha_t^I \sigma_R^2(\cdot)$ (please see the main body of the paper for a proof of the latter claim), I obtain

$$\alpha_t^H = \frac{\alpha_t^I(\mathbb{E}_t(dR_t) - r_t dt)}{(\alpha_t^I)^2 \mathrm{Var}(dR_t)} = \frac{(\alpha_t^I)^2}{(\alpha_t^I)^2} = 1.$$

The portfolio policy function of the *shadow* financial intermediary, α_t^D, takes the form

$$\alpha_t^D = \min\left(\alpha_t^I, \frac{\bar{\sigma}}{\sigma_R(\cdot)}\right).$$

When the leverage constraint does not bind, the portfolio choice of the *shadow* financial intermediary is the same as the portfolio choice of the *equity-constrained* financial intermediary. This is because both intermediaries maximize logarithmic utility over instantaneous consumption. I derive the portfolio policy function of the *equity-constrained* financial intermediary from the market clearing condition,

$$\alpha_t^I(W_t + \alpha_t^H(1-\lambda)W_t^H) + \alpha_t^D W_t^D = P_t.$$

Consequently,

$$\alpha_t^I = \frac{1 - \alpha_t^D \beta_t}{x_t + (1-\lambda)(1 - x_t - \beta_t)} = \frac{1 - \alpha_t^D \beta_t}{x_t\lambda + (1-\lambda)(1-\beta_t)}.$$

There are two cases to consider. For the case in which the leverage constraint binds,

$$\alpha_t^D = \frac{\bar{\sigma}}{\sigma_R(\cdot)}, \quad \alpha_t^I = \frac{1-\alpha_t^D\beta_t}{x_t\lambda + (1-\lambda)(1-\beta_t)} = \frac{1-\frac{\bar{\sigma}}{\sigma_R(\cdot)}\beta_t}{x_t\lambda + (1-\lambda)(1-\beta_t)}.$$

For the case in which the leverage constraint does not bind,

$$\alpha_t^D = \alpha_t^I, \quad \alpha_t^I = \frac{1}{x_t\lambda + \beta_t\lambda + 1 - \lambda}.$$

Below, I summarize the portfolio policy functions of all agents of the model in the unconstrained region.

- Region in which the leverage constraint binds

$$\begin{aligned}
\alpha_t^I =& \frac{1-\frac{\bar{\sigma}}{\sigma_R(\cdot)}\beta_t}{x_t\lambda + (1-\lambda)(1-\beta_t)}, \\
\alpha_t^H =& 1, \\
\alpha_t^D =& \frac{\bar{\sigma}}{\sigma_R(\cdot)}.
\end{aligned}$$

- Region in which the leverage constraint does not bind

$$\begin{aligned}
\alpha_t^I =& \frac{1}{x_t\lambda + \beta_t\lambda + 1 - \lambda}, \\
\alpha_t^H =& 1, \\
\alpha_t^D =& \alpha_t^I.
\end{aligned}$$

■

2.5.3. Price-Dividend Ratio

The household, the *equity-constrained* financial intermediary, and the *shadow* financial intermediary are endowed with logarithmic utility over instantaneous consumption. Consequently, it is optimal for them to consume constant fractions of their wealth,

$$
\begin{aligned}
C_t^H =& \rho_h W_t^H, \\
C_t^D =& \rho_d W_t^D, \\
C_t =& \rho W_t.
\end{aligned}
$$

Using the fact that $W_t + W_t^D + W_t^H = P_t$, and imposing market clearing in the goods market,

$$
\begin{aligned}
C_t + C_t^H + C_t^D =& D_t(1+l), \\
\rho W_t + \rho_h W_t^H + \rho_d W_t^D =& D_t(1+l).
\end{aligned}
$$

After a straightforward simplification, I obtain

$$
P_t = \frac{1+l}{\rho_h} D_t + \left(1 - \frac{\rho_d}{\rho_h}\right) W_t^D + \left(1 - \frac{\rho}{\rho_h}\right) W_t.
$$

To obtain an expression for the equilibrium price-dividend ratio, I divide both sides of the above equation by D_t and solve for P_t/D_t. The price-dividend ratio is given by

$$
\frac{P_t}{D_t} = \frac{1+l}{\rho_h - (\rho_h - \rho_d)\beta_t - (\rho_h - \rho)x_t}.
$$

■

2.5.4. Volatility of the Total Return Process

I start from the expression for the total return on the intermediated risky asset

$$
dR_t = \frac{dP_t}{P_t} + \frac{D_t}{P_t} dt = \mu_R(\cdot)dt + \sigma_R(\cdot)dB_t.
$$

I then do Itô on P_t,

$$P_t = \frac{1+l}{\rho_h} D_t + \left(1 - \frac{\rho_d}{\rho_h}\right) W_t^D + \left(1 - \frac{\rho}{\rho_h}\right) W_t,$$

and obtain

$$dP_t = \frac{1+l}{\rho_h} dD_t + \left(1 - \frac{\rho}{\rho_h}\right) dW_t + \left(1 - \frac{\rho_d}{\rho_h}\right) dW_t^D.$$

The next step is to note that W_t, W_t^D, and D_t evolve according to

$$\begin{aligned}
dW_t &= - C_t dt + r_t W_t dt + \alpha_t^I W_t (dR_t - r_t dt), \\
dW_t^D &= - C_t^D dt + r_t W_t^D dt + \alpha_t^D W_t^D (dR_t - r_t dt), \\
dD_t &= g D_t dt + \sigma D_t dB_t.
\end{aligned}$$

Consequently, the SDE for P_t takes the form

$$\begin{aligned}
dP_t &= \frac{1+l}{\rho_h} dD_t + \left(1 - \frac{\rho}{\rho_h}\right) dW_t + \left(1 - \frac{\rho_d}{\rho_h}\right) dW_t^D, \\
\frac{dP_t}{P_t} &= (\cdot) dt + \frac{1+l}{\rho_h} \frac{1}{P_t} \sigma D_t dB_t + \left(1 - \frac{\rho}{\rho_h}\right) \frac{1}{P_t} \alpha_t^I W_t \sigma_R(\cdot) dB_t \\
&\quad + \left(1 - \frac{\rho_d}{\rho_h}\right) \frac{1}{P_t} \alpha_t^D W_t^D \sigma_R(\cdot) dB_t.
\end{aligned}$$

It is useful to note that

$$\frac{D_t(1+l)}{\rho_h} = P_t - \left(1 - \frac{\rho}{\rho_h}\right) W_t - \left(1 - \frac{\rho_d}{\rho_h}\right) W_t^D.$$

I then match the diffusion coefficients to obtain

$$\begin{aligned}
\sigma_R(\cdot) &= \sigma \frac{1}{P_t} \left(P_t - \left(1 - \frac{\rho}{\rho_h}\right) W_t - \left(1 - \frac{\rho_d}{\rho_h}\right) W_t^D\right) \\
&\quad + \left(1 - \frac{\rho}{\rho_h}\right) \alpha_t^I x_t \sigma_R(\cdot) + \left(1 - \frac{\rho_d}{\rho_h}\right) \alpha_t^D \beta_t \sigma_R(\cdot).
\end{aligned}$$

After a straightforward simplification, $\sigma_R(\cdot)$ simplifies to

$$\begin{aligned}\sigma_R(\cdot) =& \sigma - \sigma\left(1-\frac{\rho}{\rho_h}\right)x_t - \left(1-\frac{\rho_d}{\rho_h}\right)\sigma\beta_t + \left(1-\frac{\rho}{\rho_h}\right)\alpha_t^I x_t \sigma_R(\cdot) \\ &+ \left(1-\frac{\rho_d}{\rho_h}\right)\alpha_t^D \beta_t \sigma_R(\cdot), \\ =& \sigma(1 - Ax_t - B\beta_t) + A\alpha_t^I x_t\sigma_R(\cdot) + B\alpha_t^D\beta_t\sigma_R(\cdot),\end{aligned}$$

where I define

$$\begin{aligned}A &\doteq 1 - \frac{\rho}{\rho_h}, \\ B &\doteq 1 - \frac{\rho_d}{\rho_h},\end{aligned}$$

in order to simplify notation. Below, I derive expressions for volatility region by region.

2.5.5. Leverage Dynamics

The wealth of the aggregate financial sector is given by

$$W_t^F = W_t + \alpha_t^H(1-\lambda)W_t^H + W_t^D.$$

Market clearing in the market for the intermediated risky asset implies $\alpha_t^F W_t^F = P_t$. Consequently,

$$\begin{aligned}\alpha_t^F =& \frac{P_t}{W_t^F} = \frac{P_t}{W_t + \alpha_t^H(1-\lambda)W_t^H + W_t^D}, \\ =& \frac{1}{x_t + \alpha_t^H(1-\lambda)(1-x_t-\beta_t) + \beta_t}.\end{aligned}$$

Given that $\alpha_t^H = 1$ in the unconstrained region (please see above for a proof), α_t^F is given by

$$\alpha_t^F = \frac{P_t}{W_t^F} = \frac{1}{x_t\lambda + \beta_t\lambda + 1 - \lambda}$$

in the unconstrained region of the state space and by

$$\alpha_t^F = \frac{P_t}{W_t^F} = \frac{1}{x_t(1+m)+\beta_t}$$

in the constrained region of the state space. Clearly, in both regions α_t^F is countercyclical with respect to x_t and β_t.

■

2.5.6. *Proof of Proposition 8 (Separating Boundary)*

The boundary that separates the constrained region from the unconstrained region of the state space is decreasing in β_t. The claim of the proposition follows immediately.

■

2.5.7. *Proof of Proposition 9 (Equilibrium, Both Constraints Bind)*

Please refer to section 2.5.2 of the mathematical Appendix for the derivation of α_t^I, α_t^D, and α_t^H. I start from the expression for $\sigma_R(\cdot)$,

$$\sigma_R(\cdot) = \sigma(1 - Ax_t - B\beta_t) + A\alpha_t^I x_t \sigma_R(\cdot) + B\alpha_t^D \beta_t \sigma_R(\cdot),$$

that I derived above, and substitute out the expressions for α_t^I and α_t^D. After a straightforward simplification, I obtain

$$\begin{aligned}
\sigma_R(\cdot) &= \sigma(1 - Ax_t - B\beta_t) + A\alpha_t^I x_t \sigma_R(\cdot) + B\alpha_t^D \beta_t \sigma_R(\cdot), \\
\sigma_R(\cdot) &= \left(1 - \frac{A}{1+m}\right)^{-1} \left(\sigma(1 - Ax_t - B\beta_t) + B\bar{\sigma}\beta_t - \frac{A\bar{\sigma}\beta_t}{1+m}\right).
\end{aligned}$$

By substituting the expressions for $\sigma_R(\cdot)$ and α_t^I into $\text{RP}_t = \alpha_t^I \sigma_R^2(\cdot)$, I obtain the equilibrium risk premium over the region in which both the intermediation constraint and the leverage constraint bind.

■

2.5.8. *Proof of Proposition 10 (Equilibrium, Only the Intermediation Constraint Binds)*

Analogous to the proof of Proposition 9 (Equilibrium, Both Constraints Bind).

■

2.5.9. *Proof of Proposition 11 (Equilibrium, Only the Leverage Constraint Binds)*

Analogous to the proof of Proposition 9 (Equilibrium, Both Constraints Bind).

■

2.5.10. *Proof of Proposition 12 (Equilibrium, Neither Constraint Binds)*

Analogous to the proof of Proposition 9 (Equilibrium, Both Constraints Bind).

■

2.5.11. *Proof of Proposition 13 (Equilibrium Quantities, Properties)*

Over the region of the state space in which both the intermediation constraint and the leverage constraint bind, the volatility of the total return process is given by

$$\sigma_R(x_t, \beta_t) = \left(1 - \frac{A}{1+m}\right)^{-1} \left(\sigma(1 - Ax_t - B\beta_t) + B\bar{\sigma}\beta_t - \frac{A\bar{\sigma}\beta_t}{1+m}\right).$$

Please refer to Proposition 9 for a proof of this claim. It is immediate to see that

$$\frac{\partial \sigma_R(x_t, \beta_t)}{\partial x_t} = -\sigma A \left(1 - \frac{A}{1+m}\right)^{-1}.$$

By assumption, $\rho_h > \rho$. Therefore, $A \in (0, 1)$. Consequently, the total return process volatility is decreasing in x_t. Similarly,

$$\frac{\partial \sigma_R(x_t, \beta_t)}{\partial \beta_t} = \left(1 - \frac{A}{1+m}\right)^{-1} \left(-B\sigma + B\bar{\sigma} - \frac{A\bar{\sigma}}{1+m}\right).$$

Under the parameter restriction,

$$\bar{\sigma} < \sigma + \frac{A\sigma}{B(1+m) - A},$$

the total return process volatility is decreasing in β_t. Therefore,

$$\frac{\partial \sigma_R(x_t, \beta_t)}{\partial x_t} < 0, \quad \frac{\partial \sigma_R(x_t, \beta_t)}{\partial \beta_t} < 0.$$

The portfolio policy function of the *shadow* financial intermediary is given by

$$\alpha_t^D = \frac{\bar{\sigma}}{\sigma_R(x_t, \beta_t)}.$$

Given that volatility is decreasing in x_t and β_t, α_t^D is increasing in x_t and β_t. I then prove the claim related to the leverage dynamics of the *equity-constrained* intermediary. The expression for the leverage of the *equity-constrained* intermediary is given by

$$\alpha_t^I = \frac{1 - \alpha_t^D \beta_t}{(1+m)x_t} = \frac{1 - \frac{\bar{\sigma}}{\sigma_R(\cdot)}\beta_t}{(1+m)x_t}.$$

The result follow immediately from the fact that volatility is decreasing in x_t and β_t. Finally, the risk premium on the intermediated risky asset is given by

$$\mathrm{RP}_t = \alpha_t^I \sigma_R^2(x_t, \beta_t).$$

Given that both α_t^I and $\sigma_R^2(x_t, \beta_t)$ are decreasing in x_t and β_t, the risk premium is decreasing in the state variables. This completes the proof of the proposition.

■

2.5.12. Separating Boundary (Leverage Constraint)

In this sub-section, I derive the separating boundary that separates the region over which the leverage constraint binds from the region over which it does not bind. I do the derivations separately for the constrained and unconstrained regions of the state space (The constrained region is the region over which the intermediation constraint binds).

- Constrained region

  The portfolio policy function of the *shadow* financial intermediary is given by

$$\alpha_t^D = \min\left(\alpha_t^I, \frac{\bar{\sigma}}{\sigma_R(\cdot)}\right).$$

  At the boundary,

$$\alpha_t^I = \frac{\bar{\sigma}}{\sigma_R(\cdot)},$$

  where

$$\begin{aligned}
\alpha_t^I =& \frac{1}{(1+m)x_t + \beta_t}, \\
\sigma_R(\cdot) =& \sigma(1 - Ax_t - B\beta_t) + A\alpha_t^I x_t \sigma_R(\cdot) + B\alpha_t^D \beta_t \sigma_R(\cdot) \\
=& \sigma(1 - Ax_t - B\beta_t) + Ax_t\sigma_R(\cdot)\frac{\bar{\sigma}}{\sigma_R(\cdot)} + B\beta_t\sigma_R(\cdot)\frac{\bar{\sigma}}{\sigma_R(\cdot)} \\
=& \sigma(1 - Ax_t - B\beta_t) + Ax_t\bar{\sigma} + B\beta_t\bar{\sigma}.
\end{aligned}$$

  Consequently,

$$\begin{aligned}
\alpha_t^I =& \frac{\bar{\sigma}}{\sigma_R(\cdot)}, \\
\frac{1}{(1+m)x_t + \beta_t} =& \frac{\bar{\sigma}}{\sigma(1 - Ax_t - B\beta_t) + Ax_t\bar{\sigma} + B\beta_t\bar{\sigma}}.
\end{aligned}$$

  I then solve for x_t to determine the boundary,

$$\hat{x}_c = \frac{\beta_t\bar{\sigma}(1 - B) - \sigma(1 - B\beta_t)}{A(\bar{\sigma} - \sigma) - \bar{\sigma}(1+m)}.$$

- Unconstrained region

At the boundary,

$$\alpha_t^I = \frac{\bar{\sigma}}{\sigma_R(\cdot)},$$

where

$$\begin{aligned}\alpha_t^I =& \frac{1}{x_t\lambda + \beta_t\lambda + 1 - \lambda},\\ \sigma_R(\cdot) =& \sigma(1 - Ax_t - B\beta_t) + Ax_t\bar{\sigma} + B\beta_t\bar{\sigma}.\end{aligned}$$

Consequently,

$$\begin{aligned}\alpha_t^I =& \frac{\bar{\sigma}}{\sigma_R(\cdot)},\\ \frac{1}{x_t\lambda + \beta_t\lambda + 1 - \lambda} =& \frac{\bar{\sigma}}{\sigma(1 - Ax_t - B\beta_t) + Ax_t\bar{\sigma} + B\beta_t\bar{\sigma}}.\end{aligned}$$

I then solve for x_t to determine the boundary,

$$\hat{x}_u = \frac{\bar{\sigma}(\beta_t(\lambda - B) + 1 - \lambda) - \sigma(1 - B\beta_t)}{A(\bar{\sigma} - \sigma) - \bar{\sigma}\lambda}.$$

■

2.6. Appendix (Figures)

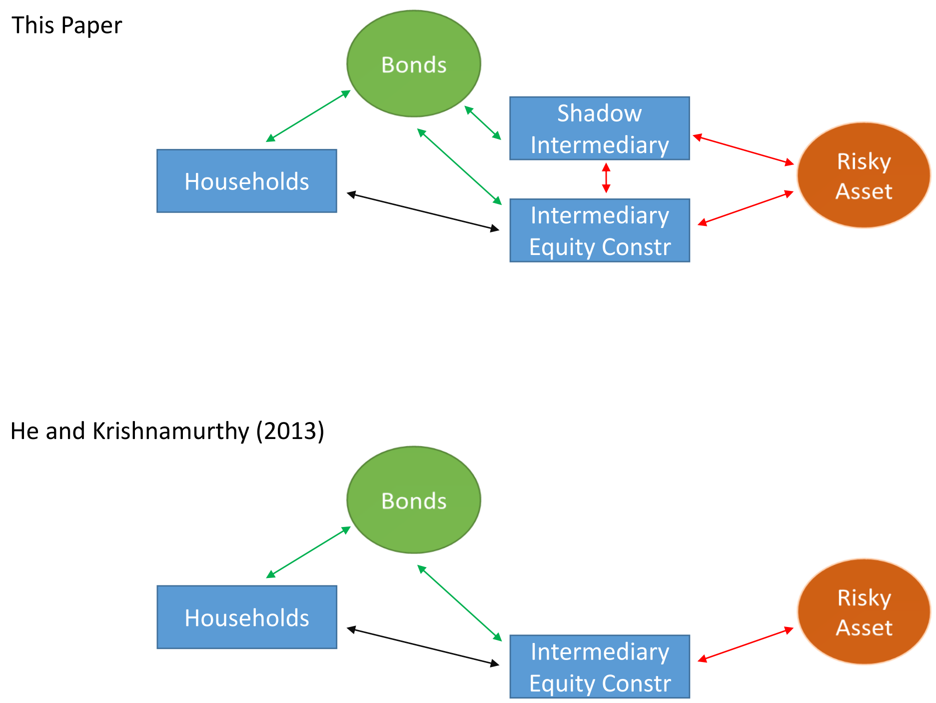

Figure 2.1: **Model Diagram**
This diagram summarizes the main facets of my model (top panel), which features a heterogeneous intermediary sector, and compares it to the baseline model (bottom panel) with a homogeneous intermediary sector, He and Krishnamurthy (2013). The top panel shows the three different types of agents (blue rectangles): an *equity-constrained* intermediary, a *shadow* financial intermediary, and an aggregate household. The risk-free asset (green circle) and the risky asset (orange circle) are the only financial assets in the economy. Green arrows refer to trading in the risk-free asset. All three agents of the model can trade the risk-free asset. The black arrow refers to trading in the outside equity of the *equity constrained* intermediary. Only households can invest in the outside equity of the *equity-constrained* financial intermediary. Red arrows refer to trading in the intermediated risky asset. Both the *equity-constrained* intermediary and the *shadow* financial intermediary can trade the intermediated risky asset. Households cannot directly trade the intermediated risky asset, but can get exposure to it by investing in the outside equity of the *equity-constrained* intermediary. In the baseline model, in the bottom panel, the intermediary sector solely comprises *equity-constrained* intermediaries.

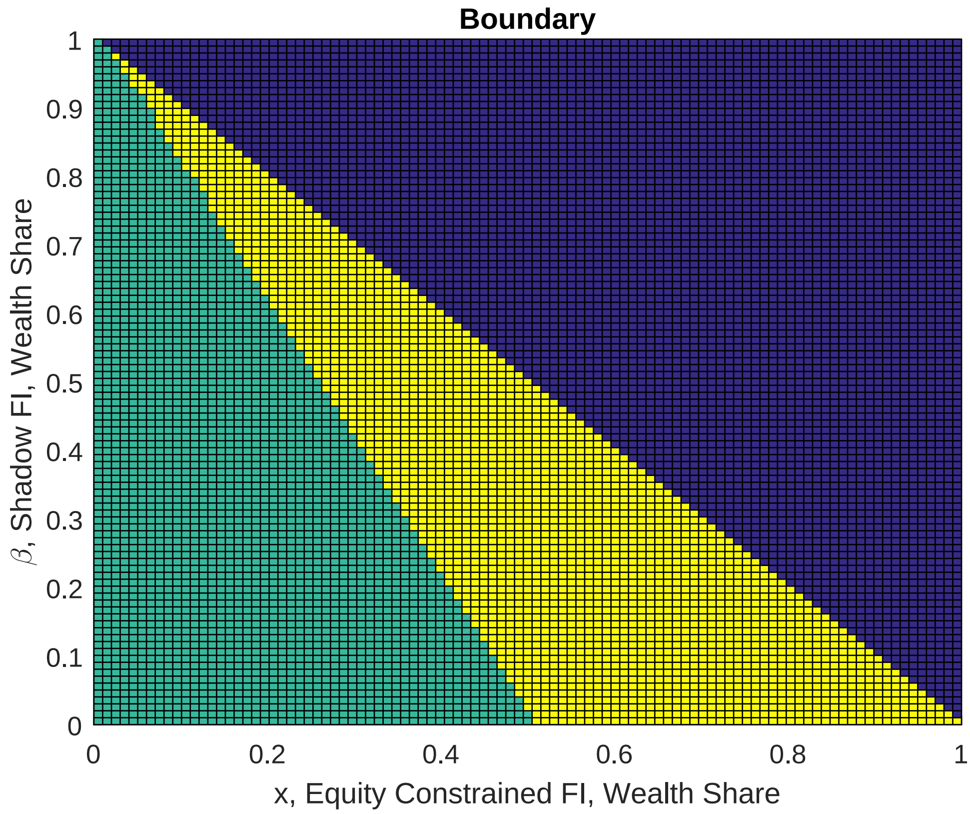

Figure 2.2: **Separating Boundary**
This figure depicts the boundary separating the constrained region (the region of the state space over which the intermediation constraint binds) from the unconstrained region. On the x-axis, I have the wealth share of the *equity-constrained* intermediary, x_t. On the y-axis, I have the wealth share of the *shadow* financial intermediary, β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sum to one. Consequently, the region in dark blue is not attainable. The line separating the yellow region from the green region is the separating boundary. The model economy is unconstrained in the green region and, is constrained in the yellow region. Please see the table in the Appendix for parameter values.

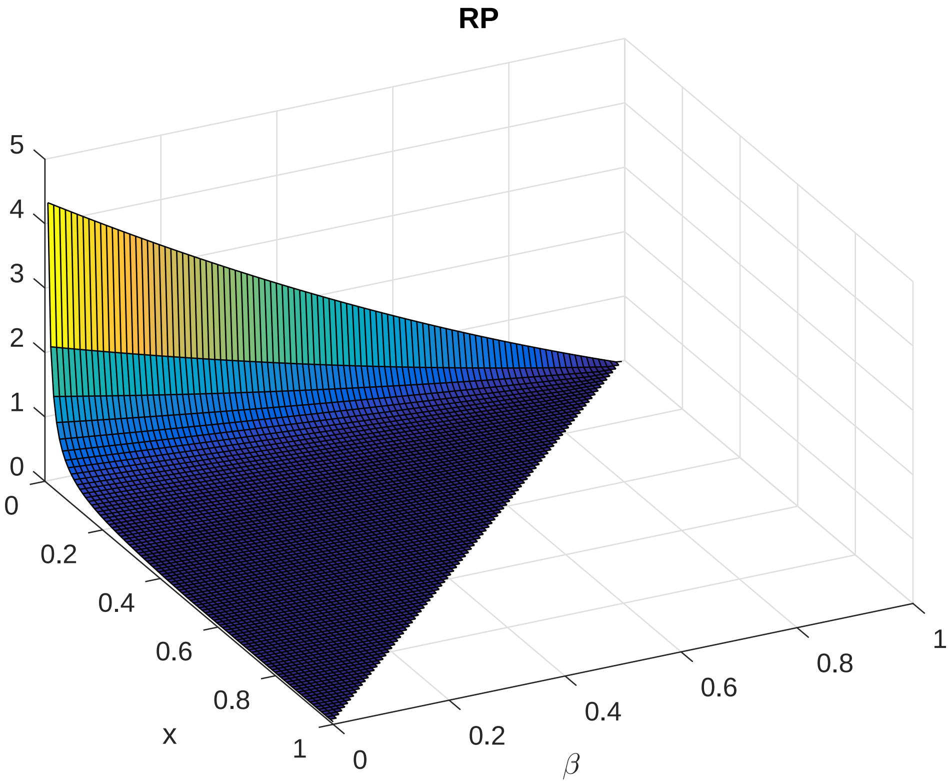

Figure 2.3: **Risk Premium**
This figure depicts the risk premium on the risky asset. On the x-axis, I have the wealth share of the *equity-constrained* intermediary, x_t. On the y-axis, I have the wealth share of the *shadow* financial intermediary, β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sum to one. I plot the risk premium on the risky asset over the entire state space. Please see the table in the Appendix for parameter values.

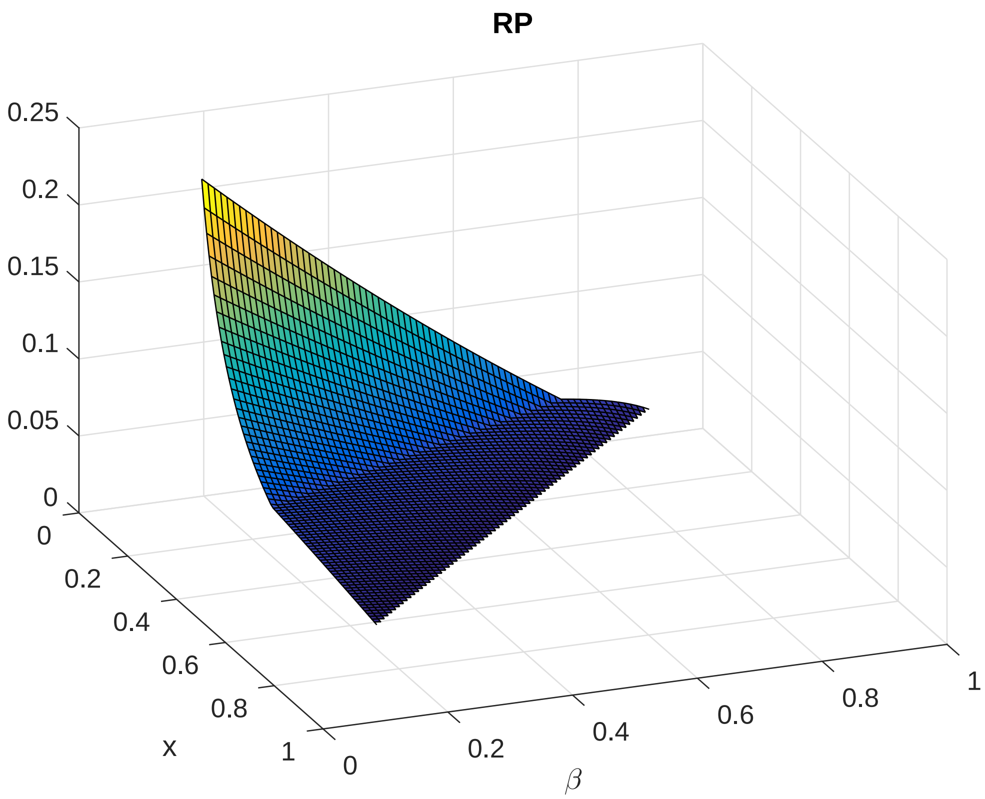

Figure 2.4: **Risk Premium (zoomed-in version)**
This figure depicts the risk premium on the risky asset. On the x-axis, I have the wealth share of the *equity-constrained* intermediary, x_t. On the y-axis, I have the wealth share of the *shadow* financial intermediary, β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sum to one. I plot the risk premium for values of β_t and x_t that are not in the immediate vicinity of the boundaries of the state space. Please see the table in the Appendix for parameter values.

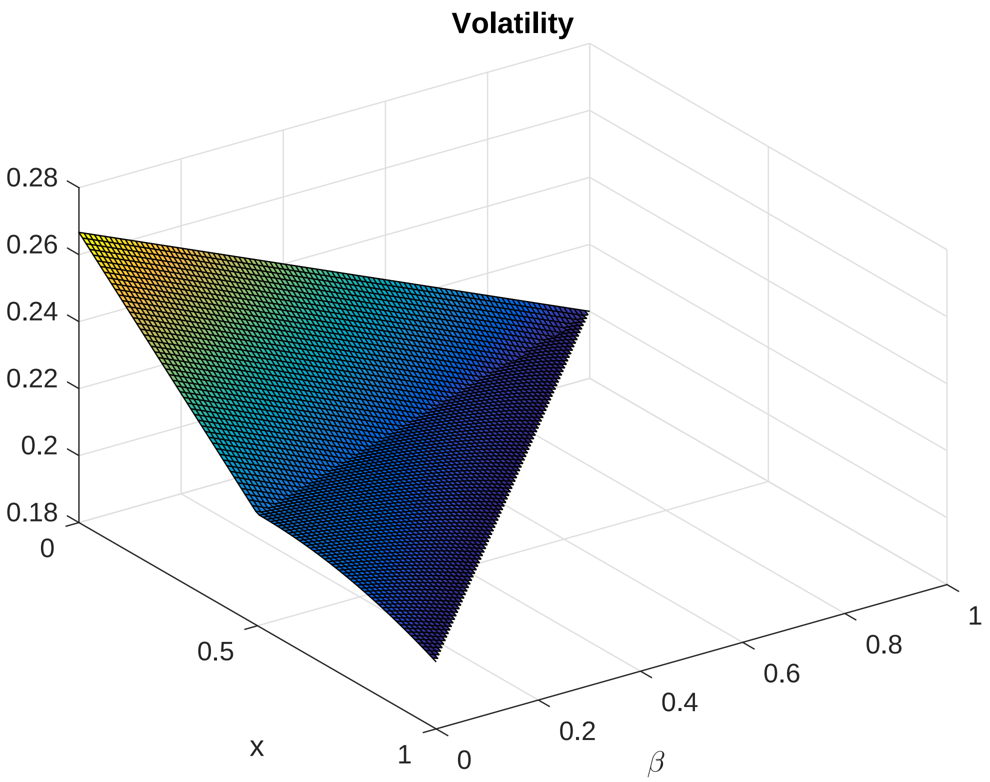

Figure 2.5: **Total Return Process Volatility**
This figure depicts the volatility of the total return process, $\sigma_R(\cdot)$. On the x-axis, I have the wealth share of the *equity-constrained* intermediary, x_t. On the y-axis, I have the wealth share of the *shadow* financial intermediary, β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sum to one. Here, I plot the volatility over the entire state space. Please see the table in the Appendix for parameter values.

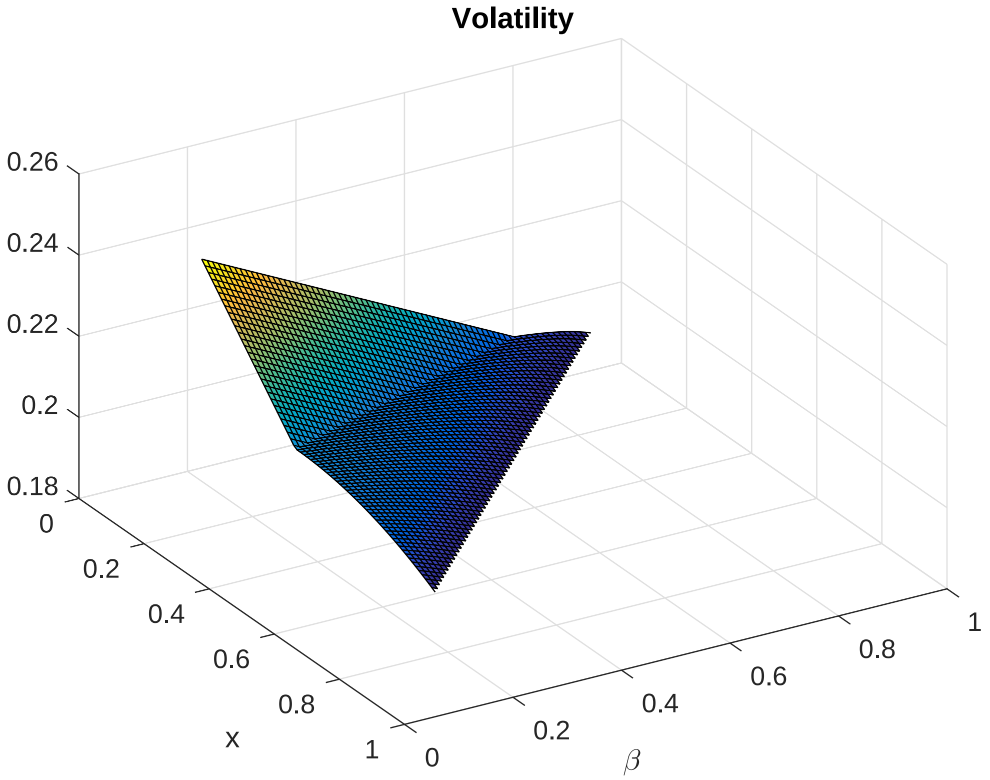

Figure 2.6: **Total Return Process Volatility (zoomed-in version)**
This figure depicts the volatility of the total return process, $\sigma_R(\cdot)$. On the x-axis, I have the wealth share of the *equity-constrained* intermediary, x_t. On the y-axis, I have the wealth share of the *shadow* financial intermediary, β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sum to one. I plot the volatility for values of β_t and x_t that are not in the immediate vicinity of the boundaries of the state space. Please see the table in the Appendix for parameter values.

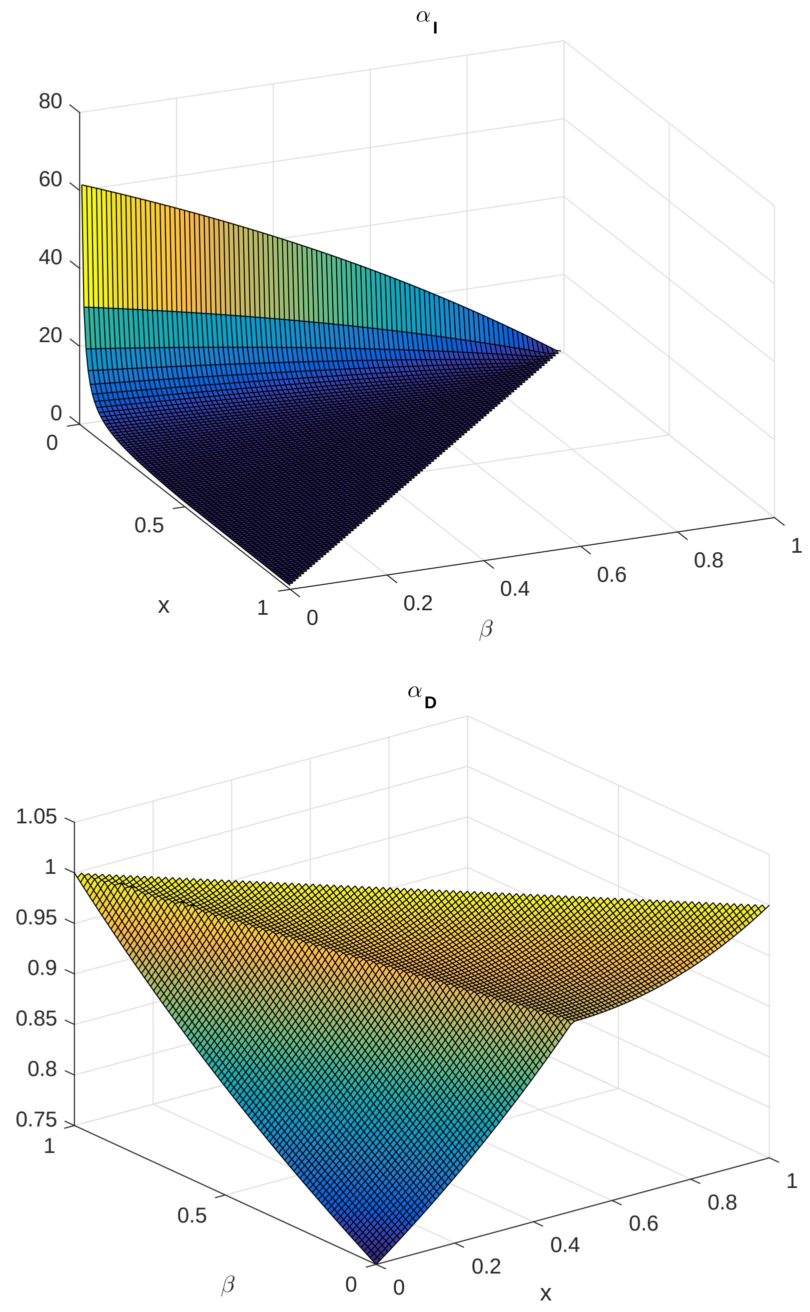

Figure 2.7: **Portfolio Policy Functions**
This figure depicts the portfolio choice of the *equity-constrained* intermediary, α_t^I, (top panel of the figure) and the portfolio choice of the *shadow* financial intermediary, α_t^D, (bottom panel of the figure). On the x-axis, I have the wealth share of the *equity-constrained* intermediary, x_t. On the y-axis, I have the wealth share of the *shadow* financial intermediary, β_t. The wealth share of the *equity-constrained* intermediary, the wealth share of the *shadow* financial intermediary, and the wealth share of the household sum to one. I plot the portfolio policy functions over the entire state space. Please see the table in the Appendix for parameter values.

2.7. Appendix (Tables)

Table 2.1: **Parameter Values**
This table summarizes the parameter values that I use in the calibration of the model.

Parameter		Parameter Value
Subjective discount factor of the representative household	ρ_h	0.05
Subjective discount factor of the *shadow* financial intermediary	ρ_d	0.01
Subjective discount factor of the *equity-constrained* intermediary	ρ	0.03
Tightness of the intermediation constraint	m	0.6
Fraction of households that are not allowed to invest in the outside equity of the *equity-constrained* intermediary	λ	0.4
Risk limit of the *shadow* financial intermediary	$\bar{\sigma}$	0.2
Volatility of the dividend growth process	σ	0.2
Labor income of the household sector	l	0.1

Chapter 3

Variance Risk Premia on Stocks and Bonds

3.1. Introduction

Recent episodes in both the United States and Europe allow us to underscore an important link between expected interest rates, volatility, and bond risk premia. For example, between mid-2014 and 2015, we observe a 60% increase in option-implied bond market volatility. During this period, implied volatility increased in anticipation of monetary tightening on the part of the Federal Reserve and fueled fears of liquidity squeezes in the bond market. Similarly, in the summer of 2015, Mario Draghi, President of the European Central Bank, commented that "we should get used to periods of higher volatility" in an era of low interest rates. Market participants tend to associate heightened volatility with bad economic states, and a plethora of research documents the negative impact of volatility shocks onto the real economy. It is common to measure the compensation for volatility shocks as the difference between physical expected variance and risk-neutral expected variance. The literature refers to this differential as the variance risk premium (VRP). Surprisingly, the extant literature that studies variance risk premia almost exclusively focuses on the equity market and largely ignores the Treasury market. Here, we seek to fill this gap in the literature.

Here, we document a set of novel facts that relate to the Treasury variance risk premia (TVRP), the co-movement between Treasury and equity variance risk premia, and the link between the EVRP and the TVRP and expected stock and bond returns. First, the premium that investors are willing to pay to hedge against changes in variance in the bond market is smaller in absolute terms than the equity variance risk premium. However, accounting for variance risk premium volatility, we document that the TVRP are economically comparable to the EVRP. Second, the sign of the conditional correlation between stock returns and bond yields switches more often than the sign of the conditional correlation between equity and Treasury variance risk premia. Third, both the equity and the Treasury variance risk premia predict equity and bond excess returns. We document predictability at both short

(3 months) and long (12 months) horizons. In particular, short maturity TVRP predict excess returns on short maturity bonds. Long maturity TVRP and the EVRP predict excess returns on long maturity bonds. Interestingly, whereas the EVRP predicts equity returns for horizons of up to 6 months, the 30-year TVRP is a formidable predictor at longer horizons. Finally, we present evidence that expected inflation is a powerful common denominator of the findings we document. Below, we further elaborate on these points.

The first contribution of this chapter pertains to the quantification of the ex-ante variance risk premia for 5-, 10-, and 30-year Treasuries, as well as the equity variance risk premium for the S&P 500 index. The ex-ante variance risk premium admits the definition of the spread between expected physical and expected risk-neutral variance. Although it is possible to calculate the latter quantity from the cross section of option prices, in model-free fashion, the calculation of the objective expectation requires some mild auxiliary modeling assumptions. A priori, it is not immediately clear what the best proxy for the objective expectation should be. For example, Andersen, Bollerslev, and Diebold (2007) show that simple autoregressive models, which are directly estimated from realized returns, tend to outperform parametric approaches geared towards forecasting integrated volatility. Thus, in the calculation of our benchmark bond variance risk premia, we use an extension of the HAR-TCJ model for realized variance that Corsi, Pirino, and Renò (2010) propose. In particular, we augment the original Corsi, Pirino, and Renò (2010) model by including lagged implied variances as additional regressors.[1]

Using data covering the period from 1991 to 2014, we obtain the following results. First, in absolute terms the size of the Treasury variance risk premium is orders of magnitude smaller than the size of its equity market counterpart. More specifically, whereas the average equity VRP is -13.48 (monthly and expressed as a squared

[1]In a recent paper, Bollerslev, Sizova, and Tauchen (2012) use a simple heterogeneous autoregressive RV model to construct a proxy for the equity variance risk premium. In a closely related paper, Busch, Christensen, and Nielsen (2011) estimate the augmented HAR-RV model featuring lagged IVs with the objective of forecasting realized volatility. Bekaert and Hoerova (2014) evaluate a series of different models to obtain the "best" estimate of the ex-ante equity risk premium.

percentage), the bond market counterparts are -3.24, -1.59, and -0.54 for the 30-, 10-, and 5-year Treasury futures, respectively. However, standardizing by their volatility to study their relative economic magnitude, we find that TVRP(5) $<$ TVRP(30) $<$ TVRP(10) $\sim$ EVRP.[2]

Second, bond market variance risk premia are particularly large during periods of distress that are unique to the bond market and exhibit less extreme variation in times of distress attributable to the equity market. To better understand the difference between variance risk premia in the bond and the equity markets, we study their conditional correlation. The study of the conditional correlation is important, because conditional correlations are a key input into any asset allocation decision. We first consider the correlations between Treasury variance risk premia of different maturities. On average, the correlations between Treasury variance risk premia are very high. We further document that correlations *decrease* at the onset of recessions, when market participants expect the Federal Reserve to loosen monetary policy. On the other hand, correlations *increase* during recoveries, when the Federal Reserve tends to take a less accommodative stance on monetary policy. The greater the wedge between maturities, the more pronounced the correlations pattern. Next, motivated by the broad empirical literature that studies the correlation between returns on the S&P 500 index and long-term Treasury yields and establishes that the correlation between equity returns and bond yields has changed signs multiple times in the past three decades, we study the co-movement between the EVRP and the TVRP at different maturities. We find that the conditional correlation between bond and equity variance risk premia is more stable over time. Unconditionally, the correlation is positive, but low. Economically, this implies that relative hedging demands against shocks to bond and equity variance are driven by distinct factors. Namely, the price attached to equity variance risk can be high (low) at times when the price attached to bond variance risk is low (high). Moreover, we find that the conditional correlation between bond and equity variance risk premia is more volatile

[2] TVRP(τ) denotes the 1-month Treasury variance risk premium on a τ-year bond.

than the correlation between equity returns and bond yields.

Third, we study the predictive power of bond and equity variance risk premia for Treasury and equity futures excess returns. In a set of univariate predictability regressions, we find that Treasury variance risk premia significantly predict Treasury and equity futures returns for a wide range of horizons of up to 12 months. In addition to being highly statistically significant, our point estimates are also economically relevant. For example, at the 3-month horizon, TVRP(5) and the TVRP(10) forecast 5- and 10-year futures excess returns with factor loadings significant at the 1% level and point estimates that imply a 0.2 standard deviation change in expected excess returns for a 1-standard-deviation shock to the TVRP.

We further document that the predictive power of the TVRP(5) and the TVRP(10) is particularly strong for shorter maturity futures and shorter horizons. At the same time, 30-year futures excess returns are only marginally predictable by the TVRP(5) and appear unrelated to other variance risk premia. We confirm the well-documented short-run (holding periods of up to 6 months) predictive power of the equity VRP for equity excess returns and note that the predictive power of TVRP for equity excess returns is much more nuanced. Surprisingly, TVRP(5) and TVRP(10) are only weakly related to equity excess returns. On the other hand, the TVRP(30) contains substantial information about expected equity returns for horizons in excess of 6 months. For instance, at the 12-month horizon we obtain a factor loading of -0.35, a t-statistic of -4.36, and a R^2 of 12%.

Finally, considering a representative set of multivariate regressions, we find that the predictive power of the TVRP(5) and the TVRP(10) for 5- and 10-year futures excess returns remains essentially unchanged when we add the equity VRP as a second predictor variable. More interestingly, when including both the TVRP(30) and the EVRP when predicting equity excess returns, we find that the EVRP remains highly significant up to a 6 month forecasting horizon, after which time the forecasting power is driven by the TVRP(30). We argue that this is an important finding for the literature given the attention devoted thus far to studying short-run EVRP predictability of equity returns. We also study the predictive power of the equity and the bond variance risk premia in multivariate regressions with standard control variables. For equity returns, we find the predictive power of the EVRP and the TVRP to be virtually unchanged when we add standard predictors to the regression. On the other hand, we find that for bond returns the predictive power of the TVRP is subsumed by well-established predictors, such as the slope of the term structure and the Cochrane and Piazzesi (2005) factor. In other words, the compensation for variance risk is a spanned determinant of term structure dynamics.

Exploiting real versus nominal Treasury dynamics, we present reduced-form evidence that expected inflation is a powerful common determinant of variance risk premia and their co-movement. In particular, we show that short maturity break-even inflation rates (a proxy for expected inflation) explain between 36% and 53% of the variation in variance risk premia across stock and bonds. At the same time, break-even rates are important determinants of stock bond correlation and of the correlation between variance risk premia. The economic and statistical magnitudes of these findings are large. We claim that these results are consistent with the signaling role of inflation in deflationary economies. Namely, negative inflation shocks serve as a strong signal about future growth, a signal which raises the price agents are willing to pay to insure against volatility shocks. At the same time, negative inflation shocks drive stock returns and bond yields in opposite directions. This is because stock returns are low through the expected cash flow channel, and bonds serve as deflation hedges. Finally, we focus on the link between break-even inflation and the correlation between variance risk premia. The loading on break-even inflation is negative, large in magnitude, and highly statistically significant. Intuitively, deflationary shocks drive a positive hedging demand against future return volatility in states in which agents are willing to pay more to insure against variance shocks.

Related Literature

This chapter relates to two different strands of the literature. The first studies variance risk premia in reduced form. Carr and Wu (2009) use option portfolios to approximate variance swaps on individual stocks. Martin (2013b) studies simple variance swaps that are robust to jumps in the price process of the underlying instrument and that can be perfectly replicated. Bondarenko (2014) empirically documents that the S&P 500 carries a large and negative variance risk premium. While the above papers focus on variance risk in the equity space only, a different strand of literature looks at compensation for volatility risk in the fixed income market. For example, Trolle (2009) reports that shorting variance swaps in the Treasury futures market generates Sharpe ratios that are about two to three times larger than the Sharpe ratios of the underlying Treasury futures. Choi, Mueller, and Vedolin (2017) empirically document economically large and negative variance risk premia and argue that there are significant returns to variance trading in Treasury markets that are comparable to those earned in the equity variance market. However, none of these papers studies the joint dynamics of bond and equity variance risk premia.

In a recent paper, Dew-Becker, Giglio, Le, and Rodriguez (2017) explore the term structure of equity variance risk premia. They find that it is difficult to reconcile the data-implied time-series dynamics of the variance risk premium with what is implied by standard consumption based asset pricing models and offer market segmentations as a possible explanation. In a similar vein, Barras and Malkhozov (2016) estimate the equity variance risk premia implied by option prices and stock returns. They find that although the same economic factors drive both variance risk premia, the two are significantly different. The authors interpret their findings as evidence of market frictions between equity and equity option markets. In line with these results, we confirm that the equity and the bond VRP are likely driven by similar economic determinants.

This chapter also relates to Adrian, Crump, and Vogt (2016), who study the predictive power of equity volatility for stock and bond returns. More specifically, Adrian, Crump, and Vogt (2016) investigate the non-linear relationship between risk and returns. They show that specifications that allow for non-linearities are superior to standard predictive regressions that use either the VIX index or realized volatility. Using the VIX as a proxy, Adrian, Crump, and Vogt (2016) find strong predictive power for both stock and bond portfolios at different horizons. Ghysels, Guérin, and Marcellino (2014) estimate a regime-switching model and report similar results.

We are not the first to study the effects of macroeconomic variables on the correlation between stocks and bonds. For example, Li (2002) documents that uncertainty about future inflation and, to a lesser extent, the real interest rate are the main drivers of time variation in the stock-bond correlation. Baele, Bekaert, and Inghelbrecht (2010) estimate a structural regime-switching model and find that uncertainty about future inflation, the equity variance risk premium, and stock market liquidity are important determinants of the stock-bond correlation. The results of Asgharian, Christiansen, and Hou (2016) point in the same direction. We are the first to show that macroeconomic variables drive the correlation between variance risk premia in equity and bond markets.

3.2. Estimation of Expected Variance

In this section, we describe the methods we use to estimate the expected physical variance, $\mathbb{E}_t^{\mathbb{P}}\left(\int_t^{t+\tau}\sigma_u^2 du\right)$, the expected risk neutral variance, $\mathbb{E}_t^{\mathbb{Q}}\left(\int_t^{t+\tau}\sigma_u^2 du\right)$, and the variance risk premium. Please note that σ_u stands for the conditional standard deviation of returns. We define the variance risk premium as the difference between expected physical variance and expected risk-neutral variance,

$$\text{VRP} = \mathbb{E}_t^{\mathbb{P}}\left(\int_t^{t+\tau}\sigma_u^2 du\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^{t+\tau}\sigma_u^2 du\right). \tag{3.1}$$

We calculate the expected risk-neutral and expected physical variance in real time, that is, over a forecasting horizon of 1-month ($\tau = 1$ month), and use these quantities to construct proxies for the equity and Treasury variance risk premia.

For Treasuries, we start in 1991 and compute a monthly TVRP measure sampled daily. In the process, we use options and futures on 5-, 10-, and 30-year Treasury notes and bonds. Following convention in the empirical literature on variance risk premia, in the equity space we use the squared VIX index as a proxy for expected variance under the risk-neutral measure. For the realized leg of the EVRP, we use high-frequency data on the S&P 500 index.[3]

3.2.1. Data

In the calculation of our proxies for implied and realized variance, we use futures and options data from the Chicago Mercantile Exchange (CME). In particular, we use high-frequency intra-day price data for 5-, 10-, and 30-year Treasury notes and bond futures, as well as high-frequency S&P 500 index data. Additionally, we use end-of-day option prices, where the options are written on the underlying indexes. The sample period starts in 1990 and runs until 2014.

At present, Treasury futures are only electronically traded on GLOBEX. However, historically, they were also traded by open outcry, and electronic trade data only became available in August 2000. To maximize our time span, we use data both from electronic and pit trading sessions. We only consider trades that occur during regular trading hours, that is, 7:20 am–2:00 pm Central Time (CT), when the products are traded side-by-side in both markets.[4] The contract months for the Treasury futures are the first five consecutive contracts in the March, June, September, and December quarterly cycle. Consequently, at any given point in time, we observe up to five contracts on the same underlying instrument. To obtain a single

[3] We can alternatively use high-frequency S&P 500 index futures data to calculate the realized variance and options written on index futures to calculate the implied variance. The difference between the implied variance extracted from options on index futures and the implied variance calculated from the VIX index is negligible. No matter what measure of implied variance we use, all results remain qualitatively the same. Nevertheless, to be consistent with the literature on equity variance risk premia, we use the VIX index for our benchmark results. As we will explain later, the approach that we follow in the estimation of the equity variance risk premium is slightly different from the approach that we follow in the estimation of the Treasury variance risk premia.

[4] Liquidity in the after-hours electronic market is significantly lower than the liquidity during regular trading hours.

time series, we roll over the futures on the 28th day of the month preceding the month when the futures contract expires.

In the equity market, it is trivial to construct an investable time series of returns (and of excess returns). For example, one can use the S&P 500 index or an alternative stock market index, see, e.g., Goyal and Welch (2008). On the other hand, in fixed income markets the construction of an investable index constitutes a herculean task, because we construct *hypothetical* bond excess returns series by interpolating zero coupon yields, see, e.g., Cochrane and Piazzesi (2005). An alternative approach to constructing excess bond return series would be to use daily changes in the smoothed zero-coupon yield curve. In the interest of consistency, we opt for a different approach. Namely, we use the returns on a fully collateralized futures position. These returns are investable because both Treasury and S&P 500 index futures are very liquid and easy to trade. This approach not only allows us to be consistent across fixed income and equity but it has the added advantage of using the same series we use in the calculation of the Treasury variance risk premia.

[Insert Table 3.1]

Table 3.1 reports summary statistics for the 1-month excess returns on Treasury and equity futures, as well as on the S&P 500 cash index. It is instructive to note that the summary statistics for the spot S&P 500 index are very similar to the summary statistics for the S&P 500 index futures. In particular, the correlation between equity cash index returns and futures returns is close to 99%. This is indicative of the fact that trading in the futures is a viable alternative to trading in the spot stock market index. Importantly, Treasury futures returns and volatilities are increasing in the tenor of the underlying. The same holds true for the maximum and for the minimum (in absolute terms) monthly excess returns. Not surprisingly, Treasury futures returns are less negatively skewed than are equity returns and are characterized by a slightly lower kurtosis.

The contract months for the options are the first 3 consecutive months (two serial expirations and one quarterly expiration) plus 4 months in the March, June, September, and December quarterly cycle. Whereas serials exercise into the first nearby quarterly futures contract, quarterlies exercise into futures contracts of the same delivery period. The procedure we apply to the rollover of the options is consistent with the procedure we apply to the rollover of the futures.[5]

3.2.2. Variance Trading and Variance Risk Premia

An abundance of over-the-counter and exchange-traded instruments deliver exposure to (or allow for the hedging of) variance shocks. The variance swap contract constitutes a good example. In a plain-vanilla variance swap, the buyer pays the variance swap strike price (or the expected variance under the risk-neutral measure), and the seller pays the realized variance at expiry, Allen, Einchcomb, and Granger (2006). For variance swaps on the S&P 500 index, it is standard practice to use the VIX squared as a measure of expected variance under the martingale measure. For the realized leg, it is common to use squared daily log returns.

The VIX squared is equal to the fair strike of a variance swap only under some very restrictive assumptions, such as the absence of jumps in the price process of the underlying instrument. In a recent paper, Martin (2013b) proposes *simple* variance swaps that are robust to jumps. Given that in the presence of jumps one can no longer use the VIX as a fair strike price for the swap contract, Martin (2013b) defines a new index, the SVIX, and shows that the squared SVIX is the fair strike of the *simple* variance swap. Whereas the VIX squared depends on all cumulants of log returns, the SVIX squared delivers clean exposure to the risk-neutral variance of simple returns. In summary, Martin (2013b) proposes a powerful approach that allows for the perfect replication of variance swap contracts.

Bondarenko (2014) takes a different approach in an attempt to ameliorate the difficulties related to the perfect replication of the standard variance contract. In particular, in addition to changing the pay-off function, he proposes a novel specifi-

[5] Please refer to the CME website (www.cmegroup.com) for detailed information about contract specifications for Treasury futures and options.

cation for *realized* variance,

$$\widetilde{RV}_{t,D} = 2\sum_{i=1}^{N} \left(x_{t,i} - \ln(1 + x_{t,i})\right), \tag{3.2}$$

where $x_{t,i} = P_i/P_{i-1} - 1$ denotes the simple return over $[t_{i-1}, t_i]$, and N is the number of observed intra-day returns. For comparison, the standard contract uses the sum of squared logarithmic returns, and Martin (2013b) uses the sum of squared simple returns. Bondarenko (2014) proceeds to show that any partition of the resultant payoff can be perfectly replicated even for the case in which the price of the underlying asset follows a jump-diffusion process. Consequently, the definition of realized variance that Bondarenko (2014) puts forward is particularly suitable for real-world applications, where it is infeasible to rebalance the replicating portfolio on a continuous basis and daily data form the basis for the calculation of the variance swap payoff.

Empirically, Bondarenko (2014) finds that variance risk is strongly priced in the equity market. Additionally, the variance risk premium is large in magnitude and is negative. Choi, Mueller, and Vedolin (2017) use the specification in Bondarenko (2014) and introduce a variance swap contract for Treasury volatility. In their paper, Choi, Mueller, and Vedolin (2017) allow for stochastic interest rates and show how to perfectly replicate the Treasury variance swap contract. The authors find that variance risk premia in the bond market are negative and economically significant.

3.2.3. Physical Variance

The variance risk premium, as we define it in equation (3.1), is a purely theoretical construct. This is because we do not observe integrated variance, $\int_t^{t+\tau} \sigma_u^2 du$. The insights from Bondarenko (2014) and Choi, Mueller, and Vedolin (2017) nudge us towards forecasting a measure of realized variance that is relevant from a practical point of view and that admits perfect replication under realistic assumptions. Given that the specification in equation (3.2) ticks both boxes, in our implementation we forecast $\widetilde{RV}$ as defined in equation (3.2).

Before we discuss the model, we define some useful notation for further reference. The daily realized variance (implied by the squared logarithmic returns), $\mathrm{RV}_{t,D}$, is given by

$$\mathrm{RV}_{t,D} = \sum_{i=1}^{N} r_{t,i}^2, \tag{3.3}$$

where $r_{t,i} = \ln P_i - \ln P_{i-1}$. Clearly, $r_{t,i}$ admits the interpretation of an intra-day logarithmic return over $[t_{i-1}, t_i]$ and P_i is the futures price at time t_i. On each trading day, we sample $r_{t,i}$ during CME pit trading hours, which is between 7:20 am and 2:00 pm Central Time.[6] Following Andersen, Bollerslev, and Diebold (2007), we calculate the daily realized variance, $\mathrm{RV}_{t,D}$, using prices at the 5-minute frequency.

Below, we also make use of the weakly, $RV_{t,W}$, as well as of the monthly, $RV_{t,M}$, realized variances. The daily realized variance serves as an input to the calculation of these quantities. More formally,

$$RV_{t,W} = \frac{1}{5} \times \sum_{j=0}^{4} \mathrm{RV}_{t-j,D}, \quad \text{and} \quad RV_{t,M} = \frac{1}{21} \times \sum_{j=0}^{21-1} \mathrm{RV}_{t-j,D}.$$

One possible way to obtain an estimate for the expected physical variance between time t and time $t+\tau$ is to use empirical projections of realized variance on some variables in the information set. Important insights that we draw from the large empirical literature that studies the salient properties of realized variance justify this approach. First, the literature documents that realized variance is very persistent. Consequently, the most recent variance estimates are high in informational content, Corsi (2009). Second, the informational content implicit in the jump component of realized variance is different from that implicit in the continuous component of realized variance, Andersen, Bollerslev, and Diebold (2007) and Corsi, Pirino, and Renò (2010). Moreover, Corsi, Pirino, and Renò (2010) show that jumps can have a highly significant impact on the estimation of future variance.

[6] For the S&P 500 index, we sample the returns during NYSE opening hours, which fall between 9:30 am and 4:00 pm Eastern Time.

In the estimation of variance, we account for jumps and follow the approach in Corsi, Pirino, and Renò (2010). The HAR-TCJ model, suitable for forecasting daily realized variance, admits the representation

$$RV_{t+1,D} = \alpha + \beta_D \widehat{TC}_{t,D} + \beta_W \widehat{TC}_{t,W} + \beta_M \widehat{TC}_{t,M} + \beta_J \widehat{TJ}_{t,D} + \varepsilon_{t+1}, \tag{3.4}$$

where $\widehat{TJ}_{t,D}$ and $\widehat{TC}_{t,D}$ are the jump and continuous components of daily realized variance, respectively, and $\widehat{TC}_{t,D} = \mathrm{RV}_{t,D} - \widehat{TJ}_{t,D}$. Please see Corsi, Pirino, and Renò (2010) for a detailed procedure related to the estimation of the jump component, $\widehat{TJ}_{t,D}$. Similarly, $\widehat{TC}_{t,W}$ stands for the continuous component of weakly realized variance, and $\widehat{TC}_{t,M}$ is the continuous component of monthly realized variance.

It is easy to extend the HAR-TCJ model by adding, for example, extra covariates that contain predictive power. In our benchmark specification, we augment the original HAR-TCJ model by adding current and lagged implied variances as additional predictor variables. More importantly, we resort to an expanding data sample of daily observations (expanding window regression) in the implementation of the forecasting regression. The expanding window regression suits well our objective of predicting 1-month realized variance in real time. We require at least 1 year (or 252 days) of data to make the first true out-of-sample prediction. Given data availability, the first variance forecast that we obtain is for July 1991. To ameliorate the risk of forecasting negative variances, we further modify the regression and run it in logs instead of levels. At each forecasting step, we add one-half of the mean squared error to the log prediction before taking the exponent.

Finally, since we follow Bondarenko (2014) and Choi, Mueller, and Vedolin (2017) in using the empirically relevant realized variance that is sampled at the daily frequency, as opposed to the unobservable integrated variance, we make a final adjustment to the regression specification in equation (3.4). Namely, we replace the 1-day-ahead daily realized variance calculated using high-frequency data, which is on the left-hand side of the regression equation, with $\widetilde{RV}_{t+21,M} = \sum_{j=1}^{21} \widetilde{RV}_{t+i,D}$, the 1-month-ahead monthly realized variance that we calculate using daily data, $\widetilde{RV}_{t,D}$, as implied by equation (3.2). The resultant model specification

$$\begin{aligned} \ln \widetilde{RV}_{t+21,M} &= \alpha + \beta_{C,D} \ln \widehat{TC}_{t,D} + \beta_{J,D} \ln(1 + \widehat{TJ}_{t,D}) + \beta_{C,W} \ln \widehat{TC}_{t,W} \\ &+ \beta_{C,M} \ln \widehat{TC}_{t,M} + \beta_{IV,0} \ln IV_t + \beta_{IV,1} \ln IV_{t-1} \end{aligned} \tag{3.5}$$

allows us to forecast quantities that matter empirically. Please note that $\ln IV_t$ stands for log implied variance. It is instructive to keep in mind that while we run the regression using daily data the forecasting horizon is 1 month.

[Insert Table 3.2]

In Panel A of Table 3.2, we present summary statistics for our monthly realized variance predictions. For ease of interpretation, we report the results in terms of annualized volatilities. To calculate the volatilities, we take the square root of the variance forecasts and multiply by the square root of 252. In line with earlier research, we find that the average realized volatility for the stock market is around 16% with a standard deviation of 7%. Realized volatilities of bonds are between 4% (for the 5-year Treasuries) and 9.3% (for the 30-year Treasuries). We note that all four realized volatility measures are highly persistent as indicated by the AR(1) coefficients, which are close to 0.85.

3.2.4. Implied Variance

We use the cross section of options on Treasury futures to calculate the expected Treasury risk-neutral variance. In particular, we follow Choi, Mueller, and Vedolin (2017), who show how to calculate the fair strike price of a Treasury variance swap. Their approach is robust to jumps and allows for stochastic interest rates. For the S&P 500 index, we follow the extant literature and use the squared VIX.[7]

[7] As we already have explained, in the interest of consistency we apply the same procedure we use for Treasury futures and options to S&P 500 index futures and options. The option-implied

In Panel B of Table 3.2, we present summary statistics for the annualized implied volatilities (to obtain the volatility, we take the square root of variance). Over the sample period, the average of implied volatility is 20% for the stock market. Implied Treasury volatilities are, on average, much lower. They fall in the range of 4.7% (5-year Treasuries) and 11% (30-year Treasuries). Similar to the realized volatility measures, implied volatilities are highly persistent with autocorrelation coefficients as high as 0.86. The averages in Panel B are very similar to the averages we report in Panel A. However, the averages in Panel B are higher across all instruments. As a result of this, we expect the variance risk premia to be negative on average.

[Insert Figure 3.1]

Figure 3.1 plots the time series of the expected physical and risk-neutral volatilities for 30-year Treasuries (top panel) and equities (bottom panel), respectively. Consistent with the extant literature, the magnitude of the spikes in equity volatility is much bigger. For example, during the LTCM crisis of 1998, the annualized realized volatility spiked considerably, whereas the spike in fixed income markets is much more subdued and around half the size of the spike observed in equity markets. The same holds true at the time of the Lehman default on 15 September 2008, when the implied volatility of the equity index spiked at around 60%, whereas fixed-income implied volatility was close to 24%.

3.3. Descriptive Analysis

We use the proxies for the expected physical variance and risk-neutral variance that we introduced in Section 3.2 to calculate the *ex-ante* variance risk premia for stocks and bonds.

estimates of implied variance are close to perfectly correlated with the squared VIX series that we use in the analysis.

3.3.1. Basic Properties

Panel C of Table 3.2 reports summary statistics for the equity and Treasury variance risk premia in percentage points. For conciseness, we will refer to the equity and the Treasury variance risk premia by EVRP and by TVRP, respectively. Unsurprisingly, variance risk premia are negative on average for both the equity index and the Treasuries. For the S&P 500 index, the average variance risk premium is -13.5. Treasury variance risk premia are increasing, in absolute terms, in the tenor of the underlying instrument. The variance risk premia increase, in absolute terms, from -0.54 for the 5-year Treasury bond futures to -3.24 for the 30-year Treasury bond futures. The *volatility* risk premium is, on average, -4.3% p.a. for the S&P 500 index and -0.7%, -1.3%, and -1.8% p.a. for the 5-, 10-, and 30-year Treasuries, respectively.

Table 3.2 also shows a substantial cross-sectional difference in the volatility of the variance risk premia. For Treasury futures, volatility is increasing in the maturity of the underlying. Given that equity admits the interpretation of a very long duration asset, it is not surprising that the volatility of the equity variance risk premium is higher than the volatility of the Treasury variance risk premia. To make the variance risk premia comparable in an economically meaningful way across different assets, we normalize the levels by volatility.

[Insert Figure 3.3]

Figure 3.3 plots the resultant statistics alongside standard error bounds that we compute using a bootstrap procedure. The relative economic ordering is as follows: TVRP(5) $<$ TVRP(30) $<$ TVRP(10) $\sim$ EVRP.

3.3.2. Time-Series Evidence

[Insert Figure 3.2]

It is instructive to compare the time-series dynamics of variance risk premia across stock and bonds. Figure 3.2 plots the 1-month variance risk premia sampled monthly. In the top panel of the figure, we plot the time series of bond variance risk premia, and, in the bottom panel of the figure, we plot the time series of the 30-year TVRP and the EVRP. Several important observations follow from the figure.

First, TVRP are increasing in the maturity of the underlying Treasury bonds. They are also increasing in magnitude over time. Additionally, bond variance risk premia exhibit counter-cyclical time-series dynamics. In other words, whereas the variance risk premia are large (in absolute value) in adverse states of the world, such as during the mortgage refinancing boom from 2002 to 2003 and the recent global financial crisis from 2008 to 2009, they are small in magnitude in normal times.[8]

It is also important to note that the EVRP is much bigger than the TVRP over most of the sample period. The LTCM crisis constitutes a particularly extreme example. During the LTCM debacle and the recent financial crisis, the equity variance risk reached -60 (in squared percentage), and the 30-year bond variance risk premium peaked at around -20 in August 2013 during the "Taper Tantrum." Second, whereas the equity variance risk premium and the 30-year TVRP are always negative, the 5- and the 10-year TVRP exhibit some positive realizations. Based on data sampled at the monthly frequency, whereas the 5-year TVRP is positive around 9% of the time, the 10-year TVRP is positive less than 1% of the time. Third, stock and bond variance risk premia display distinct dynamics around noteworthy episodes. The following discussion highlights this point.

- **Case Study 1: A Tale of Equity**

  The failure of Long Term Capital Management (LTCM) nearly caused a meltdown in the financial markets and almost triggered a catastrophe for the global economy. Testifying to this, Alan Greenspan, then the Chairman of the Federal Reserve, stated that "Had the failure of LTCM triggered the seizing up of markets, . . . , and could have potentially impaired the economies of many nations, including our own." Comparing the dynamics of equity to bond variance

[8]We formally estimate the correlation between variance risk premia and observable macroeconomic variables in one of the following sections.

risk premia, we document a surprising result. While the spike in the ex-ante equity variance risk premium is huge, a corresponding spike in Treasury variance risk is largely absent. The reason this is surprising is two-fold. First, the Federal Reserve facilitated negotiations between LTCM and primary dealers to take over the balance sheet of LTCM after the hedge fund was unable to meet its margin calls. As a result of this, broker-dealers absorbed a large amount of risk in the aftermath of LTCM failure. Second, LTCM failed largely because of *fixed income* statistical arbitrage trading. Indeed, after Russia announced restructuring of its Sovereign bond payments on 17 August 1998, and effectively defaulted on its debt, LTCM lost over USD 500 million in less than a week.

[Insert Figure 3.6]

- **Case Study 2: A Tale of Fixed Income**
  Around the LTCM episode shocks to variance risk premia were largely concentrated in the equity market, so the Taper Tantrum provides an informative counterexample. In the summer of 2013, the Taper Tantrum was largely precipitated by a string of comments on the part of Ben Bernanke, Chairman of the Federal Reserve at the time. In his testimony before the Congress in May and June 2013, Bernanke hinted that the Fed would likely start tapering the pace of its bond purchases later in the year, conditional on continuing robust economic data. The ensuing market reaction was dramatic. Long-term U.S. bond yields and dollar foreign exchange rates spiked substantially, as did realized stock and bond variance. What is much more interesting, however, is the dynamics of risk-neutral versus physical variance across stock and bonds. Figure 3.6 plots the EVRP and the 30-year TVRP in the aftermath of the Taper Tantrum. Between June and December 2013, the 30-year Treasury variance risk premium increased three-fold as wrong-footed market participants were desperate to hedge their volatility exposure in the Treasury market and were

willing to pay a high premium to do so. It is important to note that during this period spikes in variance risk premia are contained in the fixed income market. Indeed, while equity volatility spikes following Ben Bernanke's statement, the equity variance risk premium does not change much.

3.3.3. Co-movement

The above case studies motivate us to examine the co-movement between stock and bond variance risk premia in greater detail. To study the time-varying correlations between equity and Treasury variance risk premia, we estimate a dynamic conditional correlation (DCC) model. We proceed with the estimation in steps. First, we estimate a vector autoregression (VAR) jointly on all Treasury and Equity VRP levels at the daily frequency.[9] Second, we use the four residual time series from the VAR in the maximum likelihood estimation of the DCC(2,2) model.

In estimating the time-varying correlations of stock and bond returns, we use daily log changes in futures prices.

[Insert Figure 3.4]

[Insert Figure 3.5]

Figure 3.4 plots the return correlation series, and Figure 3.5 displays the dynamic correlation between variance risk premia. The top panels in both figures plot correlations between Treasury returns and Treasury VRP, respectively. The bottom panels plot the conditional correlations between equity returns or equity VRP, on the one hand, and Treasury returns or Treasury VRP, on the other hand.

The two figures exhibit distinctly different patterns. First, Treasury futures returns are highly correlated over the entire sample period. On average, correlations

[9] We opt for a specification with 120 lags. Given that there are 250 trading days per year, 120 days are approximately 6 months. The lags allow us to capture the autocorrelations and cross-autocorrelations that are implicit in daily data. It is instructive to note, however, that our specification is robust to the number of lags. That is to say, the results are not materially different in the case in which we use a smaller number of lags. In the step that precedes the estimation of the DCC model, we pre-filter the variance under the risk-neutral measure by taking monthly rolling averages. This exercise is useful in removing some of the noise that is implicit in the implied variance time series.

are as high as 90% for all three contracts. Second, the correlations between equity and fixed income returns are time varying with large swings from around +40% in the mid-1990s to −50% in the early and late 2000s. This pattern is largely in line with the results that Campbell, Sunderam, and Viceira (2013) report.

The patterns that emerge in the variance risk premia domain are very different. While correlations between TVRP are still positive over the entire sample period, they are considerably lower compared to the correlations of returns. We can see that TVRP correlations rarely exceed 80% (top panel of Figure 3.5). At the same time, correlations between TVRP and EVRP are also largely positive, but are lower on average. They hover around 20% and exhibit distinct spikes in both directions. Interestingly, the sign of the correlations occasionally turns negative. In the aftermath of the financial crisis, the EVRP/TVRP correlation reaches a maximum of 60% and a minimum of 0%. The swings are particularly large in the run-up to the Taper Tantrum episode and in its aftermath.

3.4. Predictive Regressions

In this section, we investigate the extent to which our measures of ex-ante variance risk premia contain predictive power for bond and stock excess returns. In particular, we study the in-sample predictive power of bond and equity variance risk premia for fixed income and equity excess returns. While the predictability regressions we run are in-sample, our proxies for the variance risk premia are constructed without any forward-looking bias. More importantly, the predictors we use are observable in real time.

First, we study predictability by running univariate regressions of excess returns on the variance risk premia. As we explained in the previous section, we use returns on a fully collateralized futures position in either Treasury or S&P 500 index futures in our predictive regressions. This ensures that both return series are not only investable but are also directly comparable. Second, we add commonly used control variables to our univariate regressions and conduct robustness checks.

3.4.1. Univariate Regressions

We run univariate predictability regressions for Treasury futures with underlying maturities between 5 and 30 years and for the S&P 500 index futures. The forecast horizons range between 1 and 12 months. We estimate the model:

$$xrt_{t+h}^{(i)} = \alpha_{i,h} + \beta_{i,h} VRP_t^{(i)} + \epsilon_{t+h}^{(i)},$$

where xrt_{t+h} denotes the h-period excess return, and i stands for the VRP (either 5-, 10-, or 30-year Treasuries or the S&P500 index futures).

All reported regression results are standardized; that is, we normalize all regressors and regressands to have a mean of zero and a standard deviation of one. As a result, constants appear nowhere in our tables. The normalization not only allows us to compare coefficients across different specifications but it also aids our interpretation of economic significance. We report t-statistics that are calculated using Newey and West (1987) standard errors with twelve lags. Table 3.3 summarizes the univariate regression results. Panels A through D contain the predictability regression results for the three Treasury variance risk premia and the equity variance risk premium, respectively.

[Insert Table 3.3]

Panel A of the table reports the regression results for the case in which we use S&P 500 excess returns as a regressand. In line with the extant literature exemplified by Bekaert and Hoerova (2014) and Bollerslev, Tauchen, and Zhou (2009), we find that the equity VRP has significant predictive power for equity returns at intermediate horizons. For example, at the 1-month horizon, a 1-standard-deviation (negative) shock to the VRP predicts a 0.13-standard-deviation increase in S&P 500 returns. As we increase the horizon, the slope coefficient doubles to 0.25 and 0.24 for the 3- and 6-month horizon, respectively. For the 1-year horizon, the coefficient decreases to 0.17.

Considering the impact of Treasury VRP on equity returns, we find that neither the 5-year nor the 10-year variance risk premia has any predictive power for the S&P 500 at any horizon. However, the 30-year VRP is a strong predictor of S&P 500 excess returns. The point estimate is not only highly statistically significant but it is also economically non-trivial in magnitude. For example, we find that while at the 1-month horizon there is no predictive power, for the 6- and 12-month horizon the t-statistics are -4.15 and -4.36, respectively. Comparing the economic impact of equity to 30-year VRP at the 6-month horizon, we find that a 1-standard-deviation (negative) shock to the 30-year bond variance risk premium raises equity returns by 0.3 standard deviations, which is in fact larger than the impact of the equity VRP. It is instructive to also check whether there is predictability for horizons that are longer than 1 year.

[Insert Figure 3.7]

[Insert Figure 3.8]

Figures 3.7 and 3.8 report R^2 and t-statistics, respectively, for horizons of up to 2 years. The R^2 for the Treasury VRP is hump shaped. This is consistent with what we find in the literature. When we use the 30-year bond VRP to forecast excess returns on equity, the R^2 is increasing in the length of the holding period. Predictability arising from variance risk premia on very long-term bonds more closely resembles price-dividend (long-run) predictability than equity variance risk premium (short-run) predictability. Below, we will further elaborate on this finding.

Panel B of Table 3.3 presents predictive regression results for the 5-year bond returns. We find that the equity variance risk premium has no predictive power for the 5-year bond returns at any horizon. However, we find that both the 5- and 10-year bond variance risk premia have strong predictive power for 5-year bond returns, and we observe a pattern similar to the one in Panel A. While there is no predictive power for the shortest horizon, estimated coefficients are highly significant starting at a horizon of 3 months. The estimated coefficients increase (in the horizon) for both the 5- and the 10-year variance risk premia. For the 30-year variance risk premium, we find no predictive power, except for the 1-year horizon, which has a t-statistic of -2.19.

The results that we report in Panel C of Table 3.3 mirror our findings for the 5-year bond returns (Panel B). In particular, we find that both the 5- and the 10-year bond variance risk premia have strong predictive power that is increasing in the horizon. At the same time, the 30-year bond variance risk premium is not statistically significant. Finally, in Panel D of Table 3.3, we report predictability results for the 30-year bond excess returns. We find that while the 5-year variance risk premium has significant predictive power for bond returns between 3 and 12 months, the TVRP(10) contains some power only at 12 months.

[Insert Figure 3.7]

[Insert Figure 3.8]

We present longer horizon t-statistics and R^2 for the bond returns in Figures 3.7 and 3.8. We note that both the 5- and 10-year bond VRP have the strongest predictive power at the 1-year horizon as we observe a U-shaped pattern for the t-statistics.

3.4.2. Multivariate Regressions

The multivariate regressions that we run in this section are a natural extension of the analysis from the previous section. The multivariate regressions allow us to study whether equity and bond variance risk premia can jointly predict equity and bond excess returns. Given that variance risk premia are highly correlated across different tenors, we only include one bond variance risk premium at a time.

[Insert Table 3.4]

Panel A of Table 3.4 reports the results. We find that the equity variance risk premium remains a highly statistically significant predictor even after adding the bond variance risk premia. As is the case in the univariate results, neither the 5-year nor the 10-year bond variance risk premia has predictive power for equity returns. However, the 30-year bond variance risk premium is highly statistically significant for 6- and 12-month horizons. Interestingly, we find that the 30-year bond variance risk premium drives out the predictive power of the equity variance risk premium at the 1-year horizon. This result is important, because it suggests that long-term bond variance risk premia and the equity variance risk premium capture different dimensions of the compensation for variance risk.

In Panels B to D, we focus on a set of bivariate regressions, where we add the equity variance risk premium as a second predictor variable to the 5-, 10-, or 30-year bond variance risk premia. The regression model that we estimate takes the form of

$$xrt_{t+h}^{(i)} = \alpha_{i,h} + \beta_{i,h} TVRP_t^{(i)} + \beta_{EVRP,h} EVRP_t + \epsilon_{t+h}^{(i)},$$

where $EVRP_t$ stands for the equity variance risk premium, and $TVRP^{(i)}$ stands for the respective bond variance risk premium.

The results from the univariate regressions mostly carry over to the corresponding bivariate cases. While the 5- and the 10-year TVRP have strong predictive power for horizons above 1 month, neither the 30-year nor the equity variance risk premium is significant. In summary, the TVRP retain their predictive power when we pair them in a horse race with the equity variance risk premium. At the same time, the 30-year TVRP drives out the equity variance risk premium as a predictor of stock excess returns.

3.4.3. Controls

In this subsection, we study the extent to which the predictive ability of the variance risk premia is related to alternative forecasting factors commonly used in the literature. In the context of equity predictability, we consider the log dividend yield, DY, the log earnings to price ratio, EP, and the net equity expansion, NTIS, Goyal and Welch (2008). In the context of bond return predictability, we consider the slope, Slope, as proposed by Fama and Bliss (1987); the forward rate factor, CP, of Cochrane and Piazzesi (2005); and two macro factors that we extract from a large panel of variables related to macroeconomic growth, Ludvigson and Ng (2009).

[Insert Table 3.5]

[Insert Table 3.6]

Table 3.5 reports the regression results for equity, and Table 3.6 reports the regression results for bonds. In both tables, the top panels report our findings for holding periods of 6 months, and the bottom panels report our findings for holding periods of 12 months.

In the section in which we discussed the results related to Table 3.4, we noted that TVRP(30) predictability captures a dimension of risk that is different from the one implicit in the equity VRP. In terms of long-horizon R^2 patterns, the TVRP(30) predictability very much resembles DY predictability. Indeed, comparing the equity predictability columns of Table 3.4 with column 4 of Table 3.5, we see that the price-dividend ratio drives out the TVRP(30). This result admits a very intuitive interpretation. Namely, part of the variability in the price-dividend ratio comes from compensation for volatility risk. At the same time, compensation for volatility risk that shows up in long-term bonds is largely orthogonal to the equity variance risk premium itself. Finally, we note that the TVRP(30) predictability is robust to the EP and NTIS factors.

For the case of 10-year Treasury futures returns, we find that TVRP(10) predictability is not related to the conditional first moments of the macro variables LN1 and LN2. We find, however, that TVRP(10) predictability is correlated with Slope and CP predictability. In fact, Slope and CP drive out TVRP predictability. Consequently, we can infer that part of the variability in the Slope and CP factors that is related to volatility risk is due to volatility risk compensation. This result is interesting in the context of the seminal findings of Duffee (2002), who shows that

Slope does not proxy for volatility. The results in Duffee (2002) spawned a vast literature developing term structure models with flexible price of risk specifications.

Our findings show that while the Slope does not proxy for volatility itself, it does proxy for volatility risk *compensation*. This finding is, in fact, very intuitive. The underlying state variables that drive the variance risk premia are also common to date-t yields. The final rows of Table 3.6, where we run two-stage regressions, further clarify this point. In the first stage, we project TVRP(10) onto the Slope and CP factors

$$TVRP_t(10) = \alpha + \underbrace{(-0.49)}_{-4.48} Slope_t + error_t^{\text{Slope}}, \qquad R^2 = 24\%, \tag{3.6}$$

$$TVRP_t(10) = \alpha + \underbrace{(-0.31)}_{-3.19} CP_t + error_t^{\text{CP}}, \qquad R^2 = 10\%. \tag{3.7}$$

From the first stage, we obtain the fitted value, which is the component spanned by Slope/CP, and the residual, which is the component uncorrelated with Slope/CP. The second-stage regression results, which we report in the table, suggest that Treasury variance risk premia, the forecasters of expected excess bond returns, reside in the date-t information set that is common to bond yields.

3.5. Real Nominal Risks

The correlation between stock and bond returns has substantially varied over time. Between the early 1980s and the mid-1990s, the stock-bond correlation was positive, and bonds were considered risky. Figure 3.4 shows that around the LTCM crisis the sign of the stock-bond correlation changed from positive to negative. The post-dot-com bubble was a period of a near-zero stock-bond correlation that then turned very negative during the financial crisis and remains so since. Therefore, in the present regime bonds command a negative risk premium and are considered hedges.

Motivated by the empirical literature on inflation non-neutrality, a stream of the literature has tried to understand this phenomenon in terms of shocks to inflation being correlated with shocks to the real economy. For example, Piazzesi and Schneider (2007) assume that investors dislike shocks to inflation for two reasons. First, they lower the payoff on nominal bonds. Second, they are bad news for future consumption growth. The second effect can be large when investors have recursive utility with preference for early resolution of uncertainty à la Bansal and Yaron (2004). Hasseltoft (2009) considers the impact of this setting for dividend growth and studies the joint properties of the "Fed model" and the stock-bond covariance.[10] Campbell, Sunderam, and Viceira (2013) study real-nominal covariance in the context of a latent factor quadratic term structure model. In the context of the new Keynesian framework, Campbell, Pflueger, and Viceira (2014) argue that monetary policy shifts in reaction to supply shocks alter the relationship between stocks and bonds in a way that can rationalize the change of the sign of the stock-bond correlation.

At the same time, a large but separate literature has devoted its attention to studying variance risk premia. In reduced form, Carr and Wu (2009) and Bondarenko (2014) use different approaches to reach the common conclusion that equity variance risk premia are large, negative, and display substantial time variation, especially in periods of distress (financial and economic uncertainty). A smaller but growing literature studies Treasury variance risk premia. Trolle (2009) reports that shorting variance swaps in the Treasury futures market generate Sharpe ratios that are three times larger than the Sharpe ratios of the underlying Treasury futures. Choi, Mueller, and Vedolin (2017) empirically document large and negative Treasury variance risk premia and argue that there are significant returns to variance trading in Treasury markets that are comparable to the equity variance market. In the context of equity variance risk premia, Bollerslev, Tauchen, and Zhou (2009), Zhou and Zhu (2009), and Bollerslev, Sizova, and Tauchen (2012) study economies with long-run risks and recursive preferences. When preferences are time-non-separable, volatility risk is priced and gives rise to a natural structural

[10] The "Fed model" is a term that is commonly used to describe the positive relation between U.S. dividend yields and nominal interest rates.

understanding of variance risk premia.

Surprisingly, few attempts have been made to link these literatures. This section argues the existence of a common factor determining stock-bond correlation, variance risk premia, and the correlation between variance risk premia.

[Insert Table 3.7]

Table 3.7 reports contemporaneous OLS regression results. We regress variance risk premia (VRPs), the stock-bond correlation (SB Corr), and the correlation between variance risk premia (VRP Corr) on the first two principle components of the nominal term structure and on expected inflation. We use the four-quarter-ahead consensus forecasts for consumer price inflation from BlueChip Financial forecasts as a proxy for expected inflation. To keep the results manageable, we drop the 5-year Treasury variance risk premium. The results for this tenor are qualitatively the same as those for the 10-year tenor. In the calculation of the principle components, we use bonds with maturities ranging from 2 to 10 years. Additionally, we rotate the slope so that a positive shock to this factor raises long-term yields and lowers short-term yields. We report t-statistics in parentheses. The standard errors are Newey and West (1987) standard errors with 12 lags. Left- and right-hand side variables are standardized. The sample period is from 1992.1 to 2013.1 for all regressions.

Considering Treasury variance risk premia, the R^2 is decreasing in maturity (from 31% for TVRP(10) to 25% for TVRP(30)), but remains high. This implies that a significant amount of the variation in compensation for variance risk is spanned by the term structure. For the long-term bond VRP, the level of the term structure has a positive and highly significant loading. Given that the TVRP is negative on average, this implies that negative shocks to short-term interest rates, such as rate cuts by the Federal Reserve, raise (in absolute terms) the compensation required for holding volatility risk related to long-term bonds. The slope of the term structure loads with high significance on both the TVRP(10) and the TVRP(30) with a negative loading. This is interesting since a steep yield curve is often interpreted as a signal of increased risk through a term premium component. Considering the EVRP, while the term structure factors are virtually uncorrelated with the EVRP, expected inflation is a highly significant determinant, with a positive loading and a t-statistics of 4.56. In economic terms, the factor loading implies that a 1-standard-deviation *negative* shock to expected inflation *raises* the EVRP by 0.56 standard deviations. Below, we will offer some more intuition related to this result.

Next, considering the stock-bond correlation we obtain a very large, in both economic and statistical terms, implied relationship to the level of the term structure. Moreover, conditional on the level, expected inflation is also positive and significant at the 1% level. This finding is somewhat stronger, but consistent with the empirical evidence presented by Hasseltoft (2009) and David and Veronesi (2015). The theoretical interpretation of these papers is that the level of yields and expected inflation play a negative signalling role for future economic growth in inflationary environments (pre-2000) and a positive signalling role in deflationary environments (post-2000). The net result is a positive link between stock-bond correlation and these factors.

[Insert Table 3.8]

[Insert Table 3.9]

The literature documents that the correlation between real and nominal variables flipped signs in the late 1990s and early 2000s, see, for example, Campbell, Sunderam, and Viceira (2013). This structural shift in the economy is important for learning about real nominal risks in the context of the questions we ask. Motivated by the extant literature, we consider the post-2000 subsample, which contains two deflationary recessions (2001–2002, 2007–2008), the financial crisis, and the subsequent policy response by the Federal Reserve. An advantage of considering this period in isolation is that the U.S. Treasury Department began issuing inflation-protected securities (TIPS) in 1999. Using them allows us to consider a simple combination of real and nominal yields that spans real nominal risks with date-t tradable securities that allows for a (semi-)structural interpretation.

Table 3.8 considers the left-hand-side variables from above projected on the first two principle components of the *real term structure* and 2-year break-even inflation. The break-even inflation is equal to the difference between 2-year nominal yields and 2-year real yields.[11] The sample period is from 2000.1 to 2013.1. Table 3.9 repeats the exercise replacing real-term structure PCs with nominal term structure PCs. In Tables 3.8 and 3.9, consistent with Table 3.7, we find a positive and significant relationship between TVRP(30) and the level of yields. At the same time, TVRP(10) and the slope of the term structure are negatively related. In terms of VRPs, we find a remarkable relationship between break-even inflation and VRPs on stock and bonds. In Table 3.8, we obtain factor loadings ranging from 0.62 to 0.70, with p-values well below the 1% level. The explanatory power is also large, ranging from 36% on EVRP to 53% on TVRP(30). Estimates in Table 3.9 are quantitatively the same. We interpret this finding as follows.

Standard textbook algebra tells us that 2-year break-even inflation is equal to expected inflation plus the 2-year inflation risk premium. In a low-inflation environment, such as the post-2000 experience, break-even inflation is likely dominated by expected inflation. This suggests an interpretation consistent with the positive loading obtained for the full sample: expected inflation plays a positive signalling role about the real economy. This channel is often dubbed the "inflation proxy hypothesis."[12] Consider the financial crisis of 2007–2008 and the subsequent recession, during which time the U.S. economy experienced a deflationary episode. Consistent with the inflation proxy hypothesis, investors interpreted series of negative inflation shocks as bad news about future growth. These shocks raised the price agents were

[11] In the calculation of the principle components, we use bonds with maturities ranging from 2 to 10 years. Additionally, we rotate the slope in such a way so that a positive shock to this factor raises long-term yields and lowers short-term yields. We report t-statistics in parentheses. The standard errors are Newey and West (1987) standard errors with 12 lags. Left- and right-hand side variables are standardized.

[12] Fama (1981) was the first to propose the inflation proxy hypothesis. In his seminal paper, Fama (1981) investigates the empirical link between stock returns and inflation.

willing to pay to insure against future risk (volatility), thus increasing variance risk premia across stock and bonds. A subsequent series of positive shocks was interpreted as good news compressing variance risk premia. Next, we consider the link between break-even inflation, on the one hand, and the correlation between variance risk premia, on the other hand. The point estimate for break-even inflation is large, negative, and highly statistically significant. Please see Tables 3.8 and 3.9. Intuitively, this implies that deflationary shocks drive a positive correlation for volatility hedging across stocks and bonds at a time when agents are willing to pay more for this insurance.

3.6. Conclusion

We document a set of novel facts related to variance risk premia on stocks and bonds.

First, the premium that investors are willing to pay to hedge against changes in expected bond variance is smaller in absolute terms than is the equity variance risk premium. However, accounting for variance risk premium volatility, we document that bond variance risk premia are comparable in magnitude to the equity variance risk premium.

Second, the correlation between stock and bond variance risk premia is unconditionally positive ($\sim 20\%$), but conditionally displays high-frequency variation (in the range of 0% to 60%) and has been especially volatile since the financial crisis. These dynamics are distinct from the well-studied pattern of stock-bond return correlation and provide a novel channel through which to learn about the pricing of volatility risk.

Third, both the equity and bond variance risk premia predict equity and bond excess returns at both short horizons (3 months) and long horizons (12 months). In particular, short maturity TVRP predict excess returns on short maturity bonds. Long maturity TVRP and the EVRP predict excess returns on long maturity bonds. Finally, whereas the EVRP predicts equity returns for horizons of up to 6 months, the long maturity TVRP is a formidable predictor at longer horizons.

An investigation of the common economic determinants of variance risk premia, the stock-bond correlation, and the co-movement between variance risk premia concludes the paper. Using regression-based evidence on nominal Treasuries, real Treasuries, and survey data, we present reduced-form evidence that expected inflation is a powerful determinant of each of these quantities. We leave the structural interpretation of these findings to future research.

3.7. Appendix (Figures)

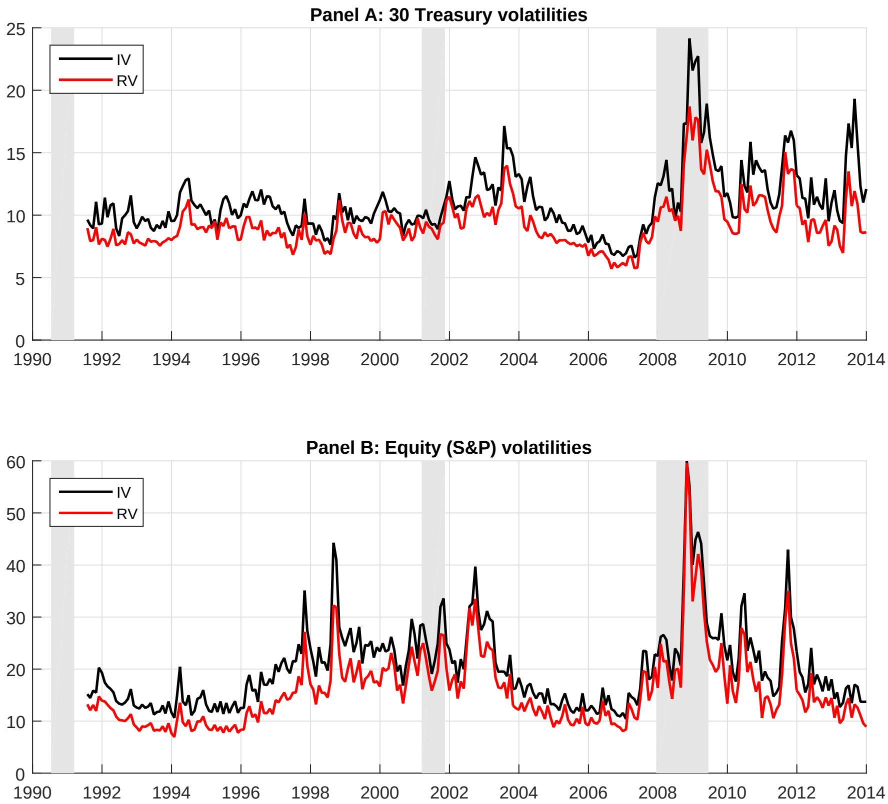

Figure 3.1: **Expected Physical and Risk Neutral Volatility**
Panel A plots the time series of the ex-ante physical and risk neutral 30-year Treasury volatility. Panel B plots the same time series for equity. For the Treasury volatilities, we use options on Treasury futures, as well as the high-frequency price data for the underlying security. For the ex-ante risk-neutral equity volatility, we use the VIX index, and to calculate the ex-ante physical volatility we use data on the S&P 500 index. Please refer to the main body of the paper for details. All series are annualized and expressed in percent. The data is monthly and runs from 1991 to 2014.

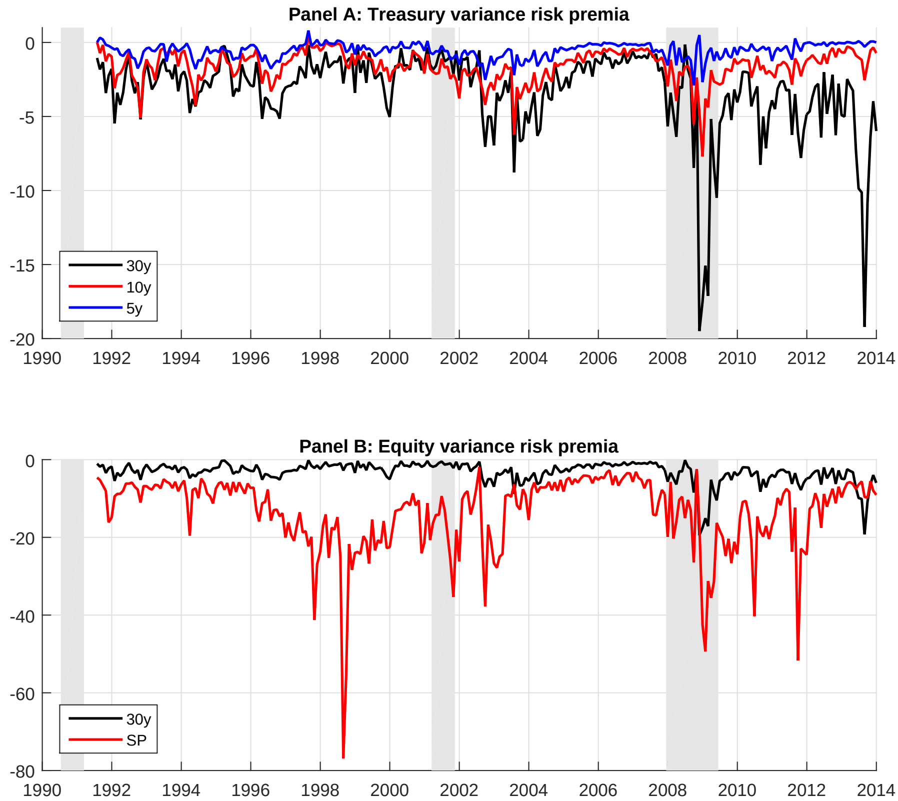

Figure 3.2: **Equity and Treasury Variance Risk Premia**
This figure plots the time series of the Treasury and the equity variance risk premia. Panel A compares Treasury variance risk premia on 5-, 10-, and 30-year Treasury futures. Panel B compares the variance risk premium on 30-year Treasury bond futures to the equity variance risk premium. We calculate the variance risk premium as the difference between the ex-ante physical and risk neutral variances. The time series are monthly and expressed in squared percent. The data is monthly and runs from 1991 to 2014.

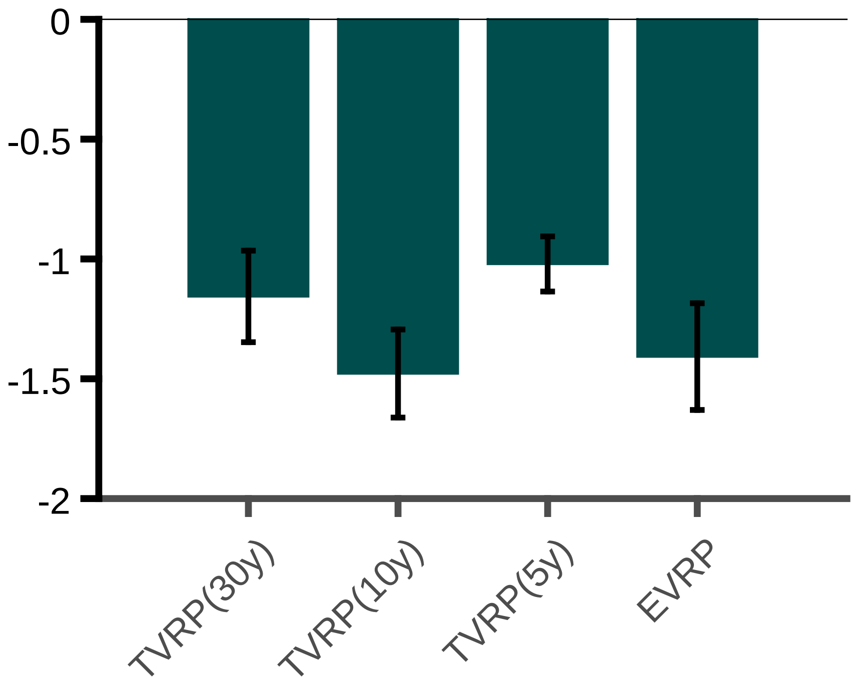

Figure 3.3: **Standardized Variance Risk Premia**

In this figure, we normalize each variance risk premium by its volatility. In the calculation of the standard errors (black lines), we resort to a bootstrap procedure with 1000 repetitions. The data runs from 1991 to 2014.

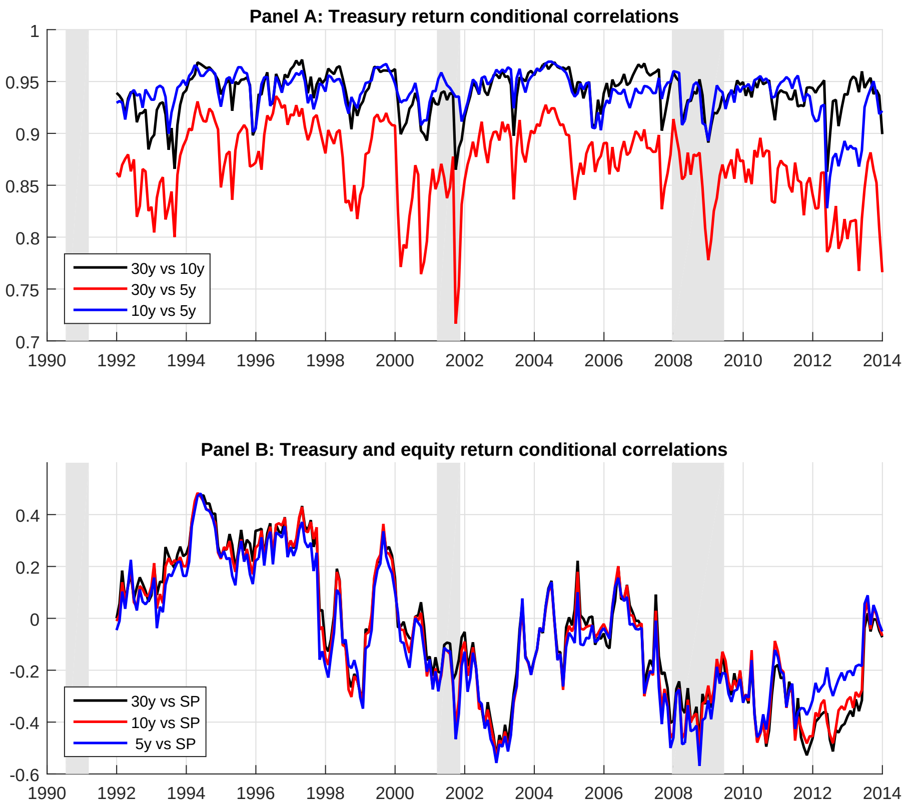

Figure 3.4: **Conditional Correlations between Equity and Treasury Returns** Panel A plots the conditional correlations between daily Treasury futures returns (5-, 10-, and 30-year). Panel B plots the conditional correlations between daily Treasury futures returns (5-, 10-, and 30-year) and returns on the S&P 500 futures. We estimate a DCC model on daily data in order to calculate the conditional correlations. We sample the data at the monthly frequency. The data runs from 1991 to 2014.

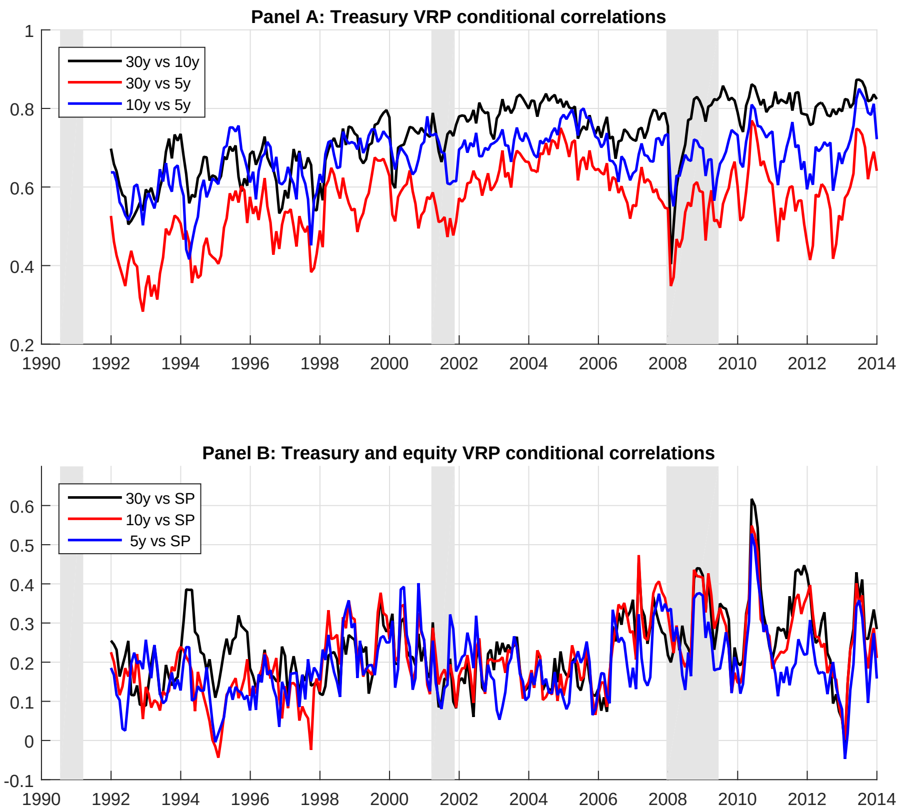

Figure 3.5: **Conditional Correlations between Treasury and Equity Variance Risk Premia**

Panel A plots the conditional correlations between daily Treasury variance risk premia (5-, 10-, and 30-year). Panel B plots the conditional correlations between daily Treasury variance risk premia (5-, 10-, and 30-year) and the variance risk premium on the S&P 500 index. We estimate a DCC model in order to calculate the conditional correlations. We sample the data at the monthly frequency. The data runs from 1991 to 2014.

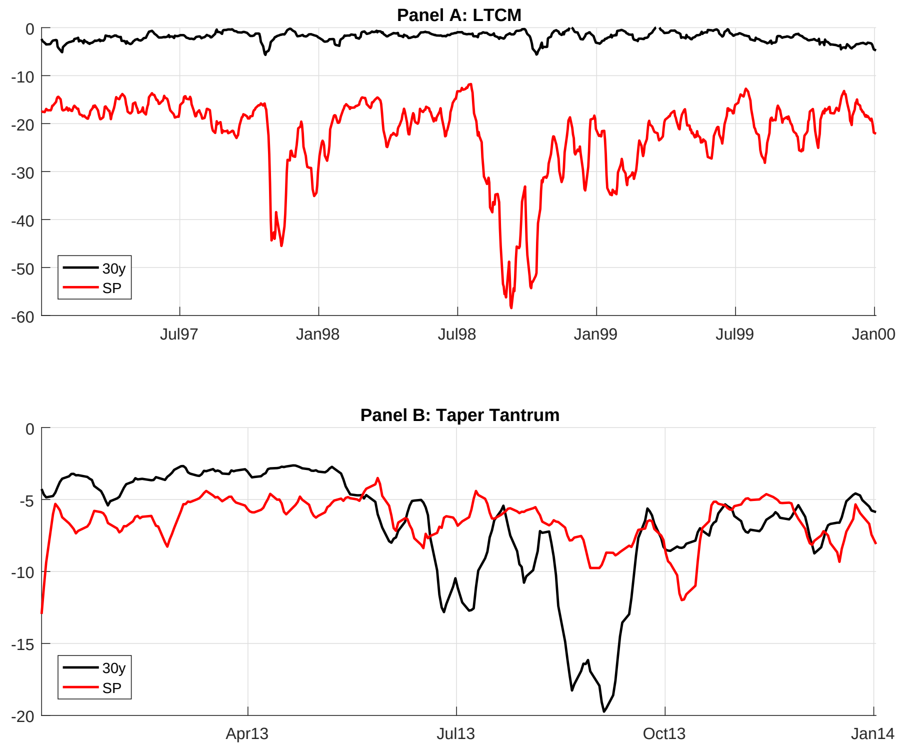

Figure 3.6: **Variance Risk Premia around LTCM and Tamper Tantrum**
Panel A plots the 30-year Treasury variance risk premium and the equity variance risk premium between January 1997 and December 1999. Panel B plots the 30-year Treasury variance risk premium and the equity variance risk premium between January 2013 and December 2013. We use daily data and smooth the time series by taking 5-day rolling averages.

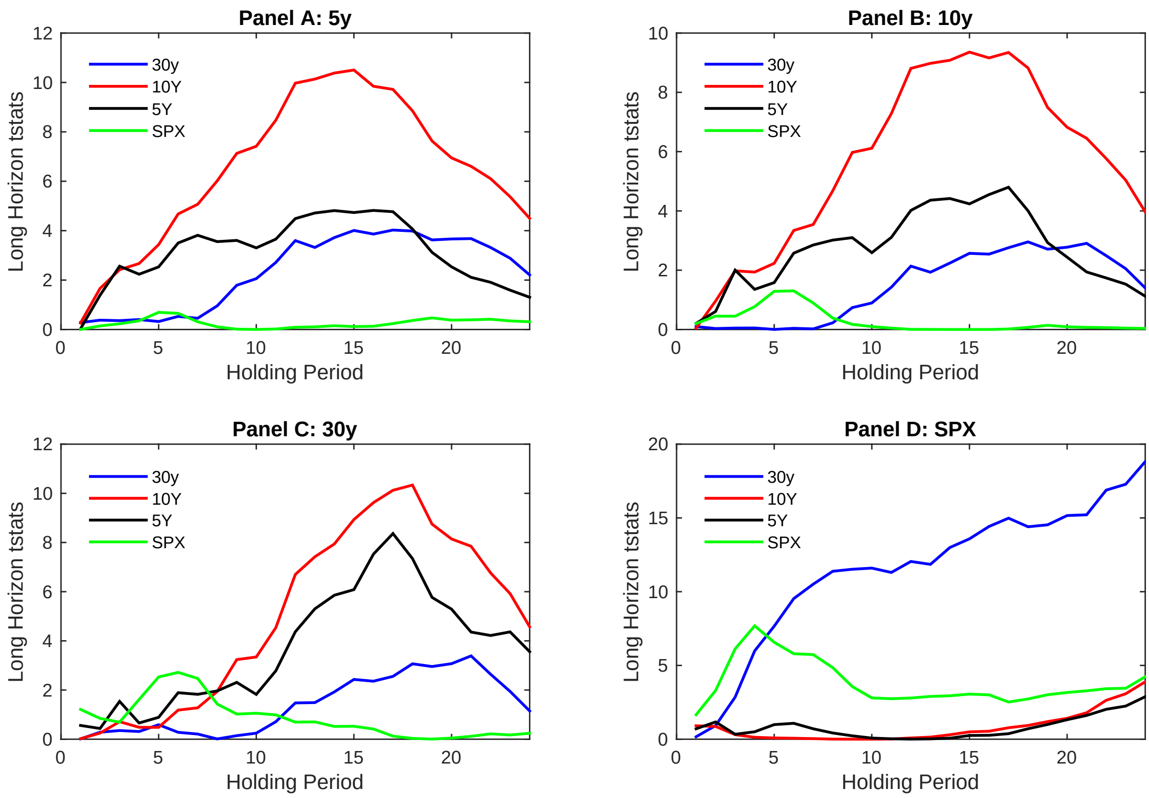

Figure 3.7: **Futures Excess Returns Long Horizon Predictability**
This figure plots R^2 obtained from h-period univariate predictability regressions of the form $xrt^j_{t+h} = \alpha_{i,h} + \beta_{i,h} VRP^{(i)}_t + \epsilon^{(i)}_{t+h}$, where i and j are 5-, 10-, 30-year, and SPX, respectively. For all regressions, the sample period is 1991 to 2014. Left and right hand side variables are standardized.

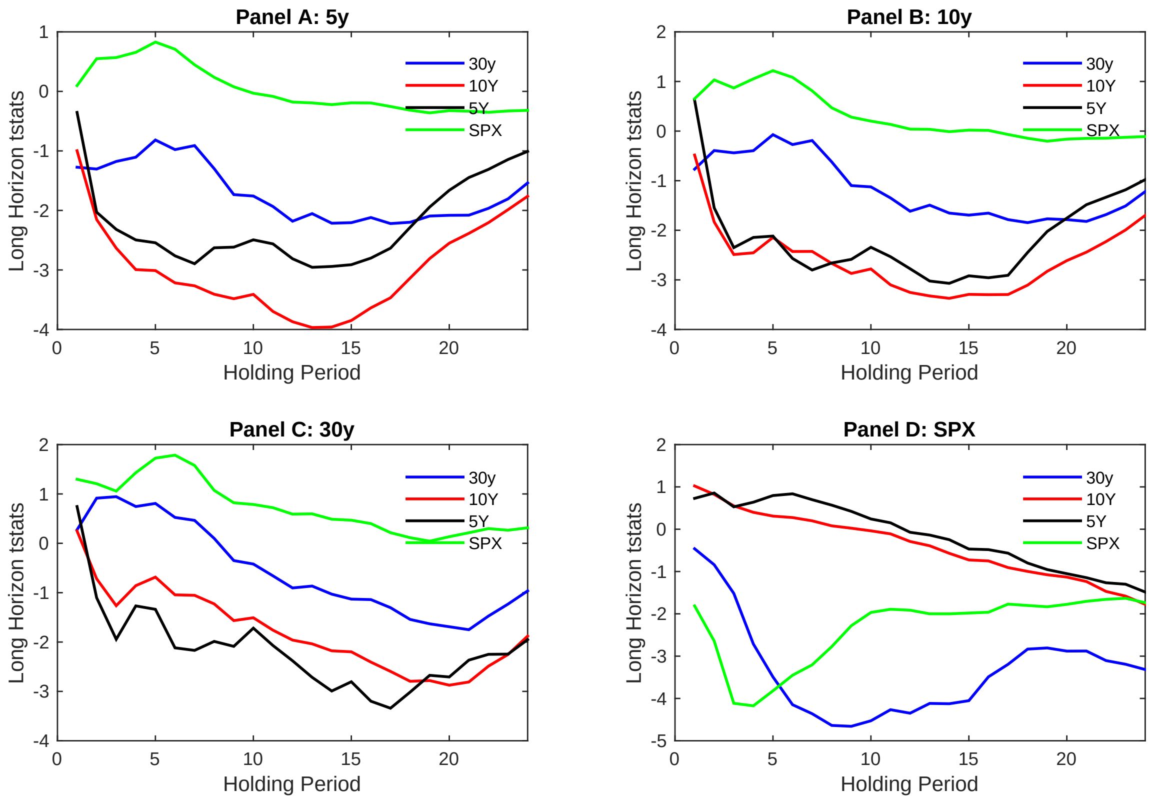

Figure 3.8: **Futures Excess Returns Long Horizon Predictability**
This figure plots t-statistics obtained from h-period univariate predictability regressions of the form $xrt^{j}_{t+h} = \alpha_{i,h} + \beta_{i,h} VRP^{(i)}_{t} + \epsilon^{(i)}_{t+h}$, where i and j are 5-, 10-, 30-year, and SPX, respectively. For all regressions, the sample period is 1991 to 2014. Left and right hand side variables are standardized.

3.8. Appendix (Tables)

Table 3.1: **Summary Statistics: Futures Excess Returns**
The table reports summary statistics for 1-month returns in excess of the 1-month Treasury Bill rate for the 5-, 10-, and 30-year Treasury bond futures and the S&P 500 index futures. For comparison, we also report the 1-month excess returns on the S&P 500 index. Means and standard deviations are annualized and expressed in percent. The data is monthly and runs from 1991 to 2014.

	30y Bonds	10y Bonds	5y Bonds	S&P 500 Futures	S&P 500 Index
Mean	1.52	1.51	0.09	3.90	3.91
Std Dev	9.47	6.17	4.15	14.86	14.90
Min	-10.62	-6.05	-3.42	-19.07	-18.64
Max	12.78	8.29	5.13	10.57	10.23
Skew	-0.01	0.11	0.09	-0.82	-0.80
Kurt	5.29	4.43	3.93	4.86	4.71
AR(1)	0.04	0.10	0.15	0.08	0.06

Table 3.2: **Summary Statistics: Implied Volatilities and Variance Risk Premia**

Panel A reports summary statistics for the 1-month physical volatilities. The underlying instruments are 5-, 10-, and 30-year Treasury notes and bond futures and the S&P 500 index. Panel B reports the risk-neutral volatilities. Panel C reports summary statistics for the variance risk premia, where the variance risk premium is equal to the difference between realized variance and risk-neutral variance. Volatilities are annualized and expressed in percent. Variance risk premia are monthly and expressed in squared percent. All series are sampled at the monthly frequency and the data runs from 1991 to 2014.

	30y Bond	10y Bond	5y Bond	S&P 500 Index
	Panel A: Physical Volatility			
Mean	9.28	5.95	4.01	15.84
Std Dev	2.04	1.25	0.93	7.30
Min	5.70	3.17	1.59	6.96
Max	18.69	10.62	7.68	59.64
Skew	1.59	0.96	0.32	2.10
Kurt	6.85	5.09	3.95	10.25
AR(1)	0.86	0.84	0.84	0.85
	Panel B: Risk Neutral Volatility			
Mean	11.04	7.29	4.66	20.11
Std Dev	2.69	1.70	1.29	7.85
Min	6.60	3.97	1.79	10.42
Max	24.15	14.33	9.69	59.89
Skew	1.72	0.80	0.25	1.65
Kurt	7.55	4.48	3.39	7.05
AR(1)	0.85	0.84	0.83	0.86
	Panel C: Variance Risk Premia			
Mean	-3.24	-1.59	-0.54	-13.48
Std Dev	2.80	1.08	0.53	9.58
Min	-19.50	-7.70	-2.90	-76.87
Max	-0.14	0.46	0.81	-1.92
Skew	-2.97	-1.60	-1.20	-2.31
Kurt	15.27	7.95	5.29	11.89
AR(1)	0.68	0.65	0.53	0.65

Table 3.3: **Return Predictability Regressions (Univariate)**
This table reports the univariate return predictability regression results. The regression model admits the representation

$$xrt^{(i)}_{t+h} = \alpha_{i,h} + \beta_{i,h} VRP^{(i)}_t + \epsilon^{(i)}_{t+h},$$

where $h = 1, 3, 6,$ and 12 months. We report the t-statistics in parentheses. The standard errors are Newey and West (1987) standard errors with 12 lags. The sample period is 1991 to 2014 for all regressions. Left and right hand variables are standardized.

h	1m	3m	6m	12m	1m	3m	6m	12m
				Panel A: S&P 500				
	S&P 500 VRP				5y VRP			
β	-0.13	-0.25	-0.24	-0.17	0.08	0.06	0.10	-0.01
t-stat	(-1.81)	(-4.11)	(-3.45)	(-1.90)	(0.73)	(0.53)	(0.84)	(-0.06)
R^2	1.66	6.12	5.79	2.76	0.71	0.33	1.08	0.00
	10y VRP				30y VRP			
β	0.10	0.06	0.03	-0.03	-0.04	-0.17	-0.31	-0.35
t-stat	(1.02)	(0.55)	(0.28)	(-0.27)	(-0.46)	(-1.52)	(-4.15)	(-4.36)
R^2	0.91	0.31	0.06	0.08	0.16	2.85	9.53	12.12
				Panel B: 5yr bond				
	S&P 500 VRP				5y VRP			
β	0.01	0.05	0.08	-0.03	-0.02	-0.16	-0.19	-0.21
t-stat	(0.10)	(0.57)	(0.71)	(-0.17)	(-0.35)	(-2.32)	(-2.76)	(-2.77)
R^2	0.00	0.23	0.65	0.09	0.03	2.56	3.50	4.38
	10y VRP				30y VRP			
β	-0.05	-0.16	-0.22	-0.31	-0.05	-0.06	-0.07	-0.19
t-stat	(-1.00)	(-2.63)	(-3.22)	(-3.85)	(-1.27)	(-1.18)	(-0.98)	(-2.19)
R^2	0.26	2.42	4.68	9.83	0.28	0.36	0.53	3.66
				Panel C: 10yr bond				
	S&P 500 VRP				5y VRP			
β	0.04	0.07	0.11	0.01	0.05	-0.14	-0.16	-0.20
t-stat	(0.65)	(0.87)	(1.08)	(0.05)	(0.62)	(-2.35)	(-2.57)	(-2.73)
R^2	0.20	0.45	1.30	0.01	0.21	2.00	2.58	3.88
	10y VRP				30y VRP			
β	-0.02	-0.14	-0.18	-0.29	-0.03	-0.02	-0.02	-0.15
t-stat	(-0.48)	(-2.49)	(-2.43)	(-3.24)	(-0.77)	(-0.44)	(-0.27)	(-1.63)
R^2	0.05	1.98	3.34	8.64	0.10	0.05	0.04	2.20
				Panel D: 30yr bond				
	S&P 500 VRP				5y VRP			
β	0.11	0.08	0.16	0.09	0.07	-0.12	-0.14	-0.20
t-stat	(1.30)	(1.06)	(1.79)	(0.61)	(0.74)	(-1.94)	(-2.12)	(-2.27)
R^2	1.21	0.69	2.72	0.74	0.56	1.54	1.89	4.02
	10y VRP				30y VRP			
β	0.01	-0.08	-0.11	-0.25	0.01	0.06	0.05	-0.13
t-stat	(0.25)	(-1.26)	(-1.04)	(-1.92)	(0.27)	(0.94)	(0.52)	(-0.93)
R^2	0.02	0.71	1.18	6.33	0.01	0.35	0.28	1.61

Table 3.4: **Return Predictability Regressions (Multivariate)**
This table reports the multivariate return predictability regression results. We regress S&P 500 and bond futures excess returns on the equity and Treasury variance risk premia. The t-statistics are in parentheses. The standard errors are Newey and West (1987) standard errors with h-lags. The sample period is 1991 to 2014 for all regressions. Left and right hand variables are standardized.

	3m	6m	12m	3m	6m	12m	3m	6m	12m
					Panel A: S&P 500				
VRP^{SPX}	-0.27	-0.27	-0.17	-0.28	-0.26	-0.17	-0.22	-0.17	-0.08
	(-4.42)	(-3.24)	(-1.92)	(-4.38)	(-3.45)	(-2.12)	(-3.34)	(-2.35)	(-0.91)
VRP^{5y}	0.10	0.15	0.02						
	(0.82)	(1.06)	(0.21)						
VRP^{10y}				0.12	0.09	0.01			
				(0.95)	(0.85)	(0.14)			
VRP^{30y}							-0.11	-0.26	-0.33
							(-0.87)	(-3.29)	(-3.65)
$\overline{R}^2$	6.83	7.64	2.45	7.16	6.17	2.41	6.89	11.85	12.35
					Panel B: 5yr bond				
VRP^{SPX}	0.08	0.12	0.01	0.09	0.14	0.05	0.07	0.11	0.02
	(0.95)	(1.09)	(0.05)	(1.09)	(1.37)	(0.32)	(0.84)	(0.97)	(0.15)
VRP^{5y}	-0.17	-0.21	-0.21						
	(-2.32)	(-2.67)	(-2.49)						
VRP^{10y}				-0.18	-0.25	-0.32			
				(-2.65)	(-3.42)	(-3.57)			
VRP^{30y}							-0.08	-0.10	-0.20
							(-1.59)	(-1.33)	(-1.85)
$\overline{R}^2$	2.80	4.48	4.03	2.82	6.18	9.71	0.43	1.26	3.35
					Panel C: 10yr bond				
VRP^{SPX}	0.09	0.15	0.04	0.11	0.17	0.08	0.08	0.13	0.05
	(1.21)	(1.46)	(0.27)	(1.32)	(1.71)	(0.56)	(1.00)	(1.20)	(0.31)
VRP^{5y}	-0.16	-0.19	-0.20						
	(-2.33)	(-2.55)	(-2.53)						
VRP^{10y}				-0.17	-0.22	-0.31			
				(-2.50)	(-2.81)	(-3.25)			
VRP^{30y}							-0.04	-0.06	-0.16
							(-0.85)	(-0.73)	(-1.49)
$\overline{R}^2$	2.51	4.30	3.71	2.69	5.62	8.94	0.25	1.22	2.09
					Panel D: 30yr bond				
VRP^{SPX}	0.11	0.19	0.12	0.11	0.20	0.15	0.07	0.16	0.13
	(1.25)	(2.10)	(0.94)	(1.29)	(2.27)	(1.24)	(0.92)	(1.73)	(0.95)
VRP^{5y}	-0.14	-0.17	-0.22						
	(-1.98)	(-2.41)	(-2.65)						
VRP^{10y}				-0.11	-0.16	-0.29			
				(-1.62)	(-1.62)	(-2.39)			
VRP^{30y}							0.04	0.01	-0.16
							(0.65)	(0.09)	(-1.24)
$\overline{R}^2$	2.31	5.22	5.18	1.48	4.68	8.24	0.47	2.36	2.82

Table 3.5: **Equity Return Predictability Regressions (Controls)**
This table reports the return predictability regression results. We regress equity futures excess returns on the 30-year Treasury variance risk premium, the equity variance risk premium, the log dividend yield (DY), the log earnings to price ratio (EP), and the net equity expansion (NTIS) from Goyal and Welch (2008). We report the t-statistics in parentheses. The standard errors are Newey and West (1987) standard errors with h-lags. The sample period is 1991 to 2014 for all regressions. Left and right hand variables are standardized.

	1	2	3	4	5	6
			Panel A: Holding Period: 6m			
VRP^{30y}				-0.08	-0.24	-0.31
				(-0.96)	(-2.75)	(-4.06)
VRP^{SPX}				-0.30	-0.27	-0.20
				(-4.30)	(-4.22)	(-2.65)
DY	0.28			0.30		
	(2.42)			(2.88)		
EP		0.02			0.19	
		(0.10)			(0.97)	
NTIS			0.31			0.39
			(1.61)			(2.08)
$\overline{R}^2$	7.58	0.04	9.47	18.10	14.07	25.63
			Panel B: Holding Period: 12m			
VRP^{30y}				-0.11	-0.32	-0.40
				(-1.29)	(-3.02)	(-4.72)
VRP^{SPX}				-0.24	-0.18	-0.10
				(-3.26)	(-3.17)	(-1.18)
DY	0.40			0.40		
	(3.05)			(3.50)		
EP		0.05			0.21	
		(0.23)			(1.22)	
NTIS			0.36			0.46
			(1.62)			(2.56)
$\overline{R}^2$	15.75	0.25	13.07	23.87	14.69	31.22

Table 3.6: **10y Treasury Futures Return Predictability Regressions (Controls)**

This table reports the return predictability regression results. We regress 10-year bond futures excess returns on the 10-year Treasury variance risk premium, the equity variance risk premium, Slope (annualized GSW slope, 10-year-1-year), the Cochrane and Piazzesi (2005) factor (the CP factor), and the first two Ludvigson and Ng (2009) macro factors, LN1 and LN2. $E[VRP^{10y}|\text{Slope}]$ and $E[VRP^{10y}|\text{CP}]$ are the fitted components of the 10-year Treasury variance risk premium on the Slope and Cochrane Piazzesi factors, respectively. $error^{\text{Slope}}$ and $error^{\text{CP}}$ are the residuals from the first stage regressions. We report the t-statistics in parentheses. Standard errors are Newey and West (1987) standard errors with h-lags. The sample period is 1991 to 2014 for all regressions. Left and right hand variables are standardized.

	1	2	3	4	5	6	7	8
				Panel A: Holding Period: 6m				
VRP^{30y}				-0.03	-0.09	-0.17		
				(-0.42)	(-1.12)	(-1.96)		
VRP^{SPX}				0.13	0.12	0.20		
				(1.59)	(1.50)	(2.17)		
Slope	0.38			0.37				
	(3.54)			(2.99)				
CP		0.42			0.39			
		(4.29)			(4.16)			
LN1			-0.17			-0.16		
			(-1.85)			(-1.93)		
LN2			0.12			0.13		
			(2.69)			(3.03)		
$E\left[VRP^{10y}\middle\|\text{Slope}\right]$							-0.38	
							(-3.54)	
$error^{\text{Slope}}$							0.00	
							(0.05)	
$E\left[VRP^{10y}\middle\|\text{CP}\right]$								-0.42
								(-4.27)
$error^{\text{CP}}$								-0.05
								(-0.65)
$\overline{R}^2$	14.58	17.74	4.04	15.51	18.75	8.67	14.26	17.71
				Panel B: Holding Period: 12m				
VRP^{30y}				-0.10	-0.14	-0.26		
				(-0.94)	(-1.64)	(-3.00)		
VRP^{SPX}				0.04	0.02	0.12		
				(0.32)	(0.19)	(0.87)		
Slope	0.46			0.41				
	(3.29)			(2.51)				
CP		0.54			0.49			
		(5.04)			(4.79)			
LN1			-0.25			-0.18		
			(-2.52)			(-1.92)		
LN2			0.13			0.14		
			(3.39)			(3.79)		
$E\left[VRP^{10y}\middle\|\text{Slope}\right]$							-0.46	
							(-3.29)	
$error^{\text{Slope}}$							-0.08	
							(-0.82)	
$E\left[VRP^{10y}\middle\|\text{CP}\right]$								-0.54
								(-4.93)
$error^{\text{CP}}$								-0.13
								(-1.51)
$\overline{R}^2$	21.28	28.68	7.53	21.49	29.94	12.96	21.63	30.15

Table 3.7: **Real versus Nominal Risks: Full Sample**
The table reports contemporaneous OLS regression results. We regress the variance risk premia (VRPs), the stock-bond correlation (SB Corr), and the correlation between variance risk premia (VRP Corr) on the first two principle components of the nominal term structure computed using 2- to 10-year maturities. The first principle component is the Level and the second principle component is the Slope. We rotate the Slope such that a positive shock to this factor raises long term yields and lowers short term yields. Expected inflation is the 4-quarter ahead consensus forecasts for consumer price inflation from BlueChip Financial forecasts. t-statistics are reported in parentheses and computed using Newey and West (1987) standard errors with 12 lags. The sample period is 1992.1 to 2013.1. Left and right hand variables are standardized.

	VRPs			SB Corr		VRP Corr	
	10y	30y	SP	10y	30y	10y	30y
Nominal Level	−0.00	0.37	−0.23	0.61	0.64	−0.07	−0.22
	(−0.02)	(3.53)	(−1.64)	(7.31)	(9.17)	(−0.36)	(−0.97)
Nominal Slope	−0.56	−0.31	−0.07	−0.03	−0.02	−0.06	−0.01
	(−6.41)	(−3.34)	(−0.51)	(−0.31)	(−0.25)	(−0.51)	(−0.05)
Exp. Inflation	0.20	0.04	0.56	0.25	0.25	−0.45	−0.09
	(1.43)	(0.39)	(4.56)	(2.52)	(2.79)	(−3.26)	(−0.53)
$\overline{R}^2$	31%	25%	15%	66%	71%	25%	8%

Table 3.8: **Real versus Nominal Risks: Post 2000 Sample TIPS regressions**
The table reports contemporaneous OLS regression results. We regress the variance risk premia (VRPs), the stock-bond correlation (SB Corr), and the correlation between variance risk premia (VRP Corr) on the first two principle components of the real term structure computed using 2- to 10-year maturities. The first principle component is the Level and the second principle component is the Slope. We rotate the Slope such that a positive shock to this factor raises long term yields and lowers short term yields. Break-even inflation (Break-Even) is the difference between 2-year nominal yields and 2-year real yields. t-statistics are reported in parentheses and computed using Newey and West (1987) standard errors with 12 lags. The sample period is 2000.1 to 2013.1. Left and right hand variables are standardized.

	VRPs			SB Corr		VRP Corr	
	10y	30y	SP	10y	30y	10y	30y
Real Level	−0.10	0.25	0.06	0.45	0.55	0.01	−0.10
	(−1.46)	(3.56)	(0.63)	(5.44)	(6.48)	(0.06)	(−0.57)
Real Slope	−0.39	−0.01	−0.17	−0.18	−0.14	−0.28	−0.24
	(−3.97)	(−0.07)	(−2.16)	(−1.52)	(−1.21)	(−2.29)	(−1.77)
Break-Even	0.64	0.70	0.62	0.47	0.48	−0.36	−0.39
	(7.68)	(5.90)	(5.19)	(3.29)	(3.68)	(−3.93)	(−3.74)
$\overline{R}^2$	48%	53%	36%	40%	50%	22%	23%

Table 3.9: **Real versus Nominal Risks: Post 2000 Sample Nominal regressions**
The table reports contemporaneous OLS regression results. We regress the variance risk premia (VRPs), the stock-bond correlation (SB Corr), and the correlation between variance risk premia (VRP Corr) on the first two principle components of the nominal term structure computed using 2- to 10-year maturities. The first principle component is the Level and the second principle component is the Slope. We rotate the Slope such that a positive shock to this factor raises long term yields and lowers short term yields. Break-even inflation (Break-Even) is the difference between 2-year nominal yields and 2-year real yields. t-statistics are reported in parentheses and computed using Newey and West (1987) standard errors with 12 lags. The sample period is 2000.1 to 2013.1. Left and right hand variables are standardized.

	VRPs			SB Corr		VRP Corr	
	10y	30y	SP	10y	30y	10y	30y
Nominal Level	−0.03	0.28	0.11	0.52	0.62	0.06	−0.07
	(−0.46)	(3.71)	(0.86)	(5.75)	(7.21)	(0.36)	(−0.37)
Nominal Slope	−0.44	−0.04	−0.08	−0.09	−0.06	−0.26	−0.20
	(−3.37)	(−0.33)	(−0.81)	(−0.65)	(−0.44)	(−1.90)	(−1.28)
Break-Even	0.42	0.57	0.51	0.19	0.18	−0.53	−0.47
	(3.81)	(3.57)	(3.06)	(1.05)	(1.09)	(−7.55)	(−6.26)
$\overline{R}^2$	50%	53%	34%	39%	50%	20%	20%

Bibliography

Adrian, Tobias, Richard Crump, and Erik Vogt, 2016, Nonlinearity and flight to safety in the risk-return trade-off for stocks and bonds, Working paper, Federal Reserve Bank of New York.

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Allen, Peter, Stephen Einchcomb, and Nicolas Granger, 2006, Variance swaps, Working paper, J.P. Morgan Investment Strategies.

Alvarez, Fernando, Andrew Atkeson, and Patrick J. Kehoe, 2002, Money, interest rates, and exchange rates with endogenously segmented markets, *Journal of Political Economy* 110, 73–112.

Andersen, Torben G., Tim G. Bollerslev, and Francis X. Diebold, 2007, Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics* 89, 701–720.

Ang, Andrew, Sergiy Gorovyy, and Gregory B. Van Inwegen, 2011, Hedge fund leverage, *Journal of Financial Economics* 102, 102–126.

Asgharian, Hossein, Charlotte Christiansen, and Ai J. Hou, 2016, Macro-finance determinants of the long-run stockbond correlation: The DCC-MIDAS specification, *Journal of Financial Econometrics* 14, 617–642.

Baele, Lieven, Geert Bekaert, and Koen Inghelbrecht, 2010, The determinants of stock and bond return comovements, *Review of Financial Studies* 23, 2374–2428.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Barras, Laurent, and Aytek Malkhozov, 2016, Does variance risk have two prices? evidence from the equity and option markets, *Journal of Financial Economics* 121, 79–92.

Basak, Suleyman, 1995, A general equilibrium model of portfolio insurance, *Review of Financial Studies* 8, 1059–1090.

——— , and Benjamin Croitoru, 2000, Equilibrium mispricing in a capital market with portfolio constraints, *Review of Financial Studies* 13, 715–748.

Basak, Suleyman, and Domenico Cuoco, 1998, An equilibrium model with restricted stock market participation, *Review of Financial Studies* 11, 309–341.

Basak, Suleyman, and Anna Pavlova, 2013, Asset prices and institutional investors, *American Economic Review* 103, 1728–1758.

Bekaert, Geert, and Marie Hoerova, 2014, The VIX, the variance premium and stock market volatility, *Journal of Econometrics* 183, 181–192.

Bollerslev, Tim, Natalia Sizova, and George Tauchen, 2012, Risks for the long run: A potential resolution of asset pricing puzzles, *Review of Finance* 16, 31–80.

Bollerslev, Tim G., George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

Bondarenko, Oleg, 2014, Variance trading and market price of variance risk, *Journal of Econometrics* 180, 81–97.

Brunnermeier, Markus K., and Lasse H. Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201–2238.

Brunnermeier, Markus K., and Yuliy Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104, 379–421.

Busch, Thomas, Bent J. Christensen, and Morten Ø. Nielsen, 2011, The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets, *Journal of Econometrics* 160, 48–57.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., Carolin Pflueger, and Luis M. Viceira, 2014, Monetary policy drivers of bond and equity risks, Working paper, NBER.

Campbell, John Y., Adi Sunderam, and Luis M. Viceira, 2013, Inflation bets or deflation hedges? the changing risks of nominal bonds, Working paper, Harvard Business School.

Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.

Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 523–546.

Chabakauri, Georgy, 2013, Dynamic equilibrium with two stocks, heterogeneous investors, and portfolio constraints, *Review of Financial Studies* 26, 3104–3141.

Chacko, George, and Luis M. Viceira, 2005, Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets, *Review of Financial Studies* 18, 1369–1402.

Choi, Hoyong, Philippe Mueller, and Andrea Vedolin, 2017, Bond variance risk premiums, *Review of Finance* 21, 987–1022.

Cochrane, John H., and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

Corsi, Fulvio, 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

———, David Pirino, and Roberto Renò, 2010, Threshold bipower variation and the impact of jumps on volatility forecasting, *Journal of Econometrics* 159, 276–288.

Danielsson, Jon, Hyun S. Shin, and Jean-Pierre Zigrand, 2010, Risk appetite and endogenous risk, Working paper, Financial Markets Group, London School of Economics.

David, Alexander, and Pietro Veronesi, 2015, The economics of the comovement of stocks and bonds, *Handbook of Fixed Income, John Wiley & Sons.*

Detemple, Jerome, and Shashidhar Murthy, 1997, Equilibrium asset prices and no-arbitrage with portfolio constraints, *Review of Financial Studies* 10, 1133–1174.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2017, The price of variance risk, *Journal of Financial Economics* 123, 225–250.

Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405–443.

Dumas, Bernard, 1989, Two-person dynamic equilibrium in the capital market, *Review of Financial Studies* 2, 157–188.

Edwards, Amy K., Lawrence E. Harris, and Michael S. Piwowar, 2007, Corporate bond market transaction costs and transparency, *Journal of Finance* 62, 1421–1451.

Fama, Eugene F., 1981, Stock returns, real activity, inflation, and money, *American Economic Review* 71, 545–565.

———, and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Ghysels, Eric, Pierre Guérin, and Massimiliano Marcellino, 2014, Regime switches in the risk-return trade-off, *Journal of Empirical Finance* 28, 118–138.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Grossman, Sanford J., and Zhongquan Zhou, 1996, Equilibrium analysis of portfolio insurance, *Journal of Finance* 51, 1379–1403.

Hasseltoft, Henrik, 2009, The 'fed model' and the changing correlation of stock and bond returns: An equilibrium approach, Working paper, SSRN.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

He, Zhiguo, In G. Khang, and Arvind Krishnamurthy, 2010, Balance sheet adjustments during the 2008 crisis, *IMF Economic Review* 58, 118–156.

He, Zhiguo, and Arvind Krishnamurthy, 2011, A model of capital and crises, *Review of Economic Studies* 79, 735–777.

———, 2013, Intermediary asset pricing, *American Economic Review* 103, 732–770.

Holmstrom, Bengt, and Jean Tirole, 1997, Financial intermediation, loanable funds, and the real sector, *Quarterly Journal of Economics* 112, 663–691.

Kiyotaki, Nobuhiro, and John Moore, 1997, Credit cycles, *Journal of Political Economy* 105, 211–248.

Kogan, Leonid, Igor Makarov, and Raman Uppal, 2007, The equity risk premium and the riskfree rate in an economy with borrowing constraints, *Mathematics and Financial Economics* 1, 1–19.

Koijen, Ralph, and Motohiro Yogo, 2015, An equilibrium model of institutional demand and asset prices, Workink paper, NBER.

Li, Lingfeng, 2002, Macroeconomic factors and the correlation of stock and bond returns, Working paper, Yale University.

Longstaff, Francis A., and Jiang Wang, 2012, Asset pricing and the credit market, *Review of Financial Studies* 25, 3169–3215.

Lucas, Robert E., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1445.

Ludvigson, Sidney C., and Serena Ng, 2009, Macro factors in bond risk premia, *Review of Financial Studies* 22, 5027–5067.

Martin, Ian, 2013a, The lucas orchard, *Econometrica* 81, 55–111.

———, 2013b, Simple variance swaps, Working paper, London School of Economics.

———, 2017, What is the expected return on the market?, *Quarterly Journal of Economics* 132, 367–433.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Moreira, Alan, and Tyler Muir, 2016, How should investors respond to increases in volatility?, Working paper.

———, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Piazzesi, Monika, and Martin Schneider, 2007, *Equilibrium yield curves* (MIT Press).

Rytchkov, Oleg, 2014, Asset pricing with dynamic margin constraints, *Journal of Finance* 69, 405–452.

Siriwardane, Emil, 2016, Concentrated capital losses and the pricing of corporate credit risk, Working paper, Harvard Business School.

Trolle, Anders B., 2009, The price of interest rate variance risk and optimal investments in interest rate derivatives, Working paper, EPFL.

Zhou, Guofu, and Yingzi Zhu, 2009, A long-run risks model with long- and short-run volatilities: explaining predictability and volatility risk premium, Working paper.