

Essays in Asset Pricing

Ilaria Piatti

A dissertation submitted for the degree of
Ph.D. in Economics

Institute of Finance
Faculty of Economics
Università della Svizzera Italiana



May 26, 2014

Thesis Committee:

Prof. Andrea Buraschi, Imperial College London
Prof. Patrick Gagliardini, University of Lugano
Prof. Antonio Mele, University of Lugano
Prof. Fabio Trojani, University of Lugano
Prof. Andrea Vedolin, London School of Economics

A Luca e alla mia famiglia

Abstract

My dissertation consists of three main chapters and focuses on two recent strands of research in asset pricing, namely *heterogeneous beliefs about rare event risk* and *present-value models for predictability of market returns and dividend growth*.

The first chapter studies the asset pricing implications of investor disagreement about the probability of a systemic disaster. I start from a structural economy with multiple assets and heterogeneous beliefs on systemic rare event risk, in order to understand how fear and risk sharing mechanisms affect excess returns on equity and pure variance positions and the relation between them, both at an aggregate level and in the cross section of stock returns. First, I find that risk premia increase with the consumption share of the pessimistic agent, as the supply of insurance against future market downturns decreases. Second, the relation between equity and variance risk premia is time varying and stronger when pessimists dominate. This prediction is tested empirically and I show that indeed the predictive power of the variance risk premium for future excess returns is time varying and stronger in phases of financial distress, which are characterized by large levels of the variance risk premium and large disagreement between investors. The model also has implications for the cross section of stocks: I find that, both in the model and in the data, the forecasting power of the index variance risk premium for future returns is stronger for small stocks, mainly when difference in beliefs is large.

The second chapter proposes a bootstrap methodology to test predictability hypotheses in the context of present-value models with latent expectation processes for returns and dividends. We show that the test is asymptotically valid and has good finite-sample properties while conventional tests strongly over-reject the null of no predictability in small samples, due to restrictive distributional assumptions. We apply the method to study stock market return and dividend growth predictability, using post-war US data. We find evidence of return predictability while evidence of dividend growth predictability is weaker than previously thought. We also show that large R^2 can arise by chance alone, also out-of-sample, even under the null hypothesis of no predictability.

The third chapter proposes a new class of tractable present-value models with latent dividend and return processes and time-varying cash flow and discount rate risks, to study the joint predictability features of dividends, returns and their second moments. We first find a relatively large return predictability but no statistical evidence of dividend growth predictability. Second, we find a volatile and counter-cyclical market Sharpe ratio and highly time varying term structures of both expectations and volatilities of long horizon dividend growth and returns, which have the flexibility to take different shapes, consistent with recent empirical evidence.

Acknowledgements

The completion of my PhD has been a long journey, an endless series of highs and lows, but I enjoyed every moment. For this I have to acknowledge all the amazing people that I met on the way, who helped me grow personally and professionally. I could not have succeeded without their support, especially of those I'm about to mention.

First and foremost I have to thank my PhD advisor, Fabio Trojani, for his invaluable support and enthusiasm. He is the reason why I decided to pursue a career in research and I'm really glad he offered me the opportunity to start a PhD under his supervision long time ago. Fabio is always an inspiration and an example, and I'm still convinced I could not have had a better advisor.

Special thanks go to Andrea Vedolin, for sponsoring my visit at the London School of Economics, which has been a wonderful research and personal experience, and for insightful discussions, constant guidance and her helpfulness. She also encouraged me to spend some time at Duke (and made it possible together with Tim Bollerslev, who I also have to thank) and she told me all I needed to know about the job market. Thanks Andrea!

Of course I am very grateful also to the remaining members of my dissertation committee, Andrea Buraschi, Antonio Mele and Patrick Gagliardini, for their academic support and personal cheering.

Likewise, I want express my gratitude to all my current and former colleagues and friends in Lugano: Elisa, Diego, Davide, Hakim, Nicola, Tamara, Peter, Alexandru, Paul, Piotr, Andras, Filippo, Giuseppe, Cecilia, and all the others (sorry for not mentioning you all but you know who you are!), because I can always count on you, for a research or technical issue as well as for a smile, coffee or soccer match. I'm happy to start a new exciting journey, but I know that I will miss you all and it will be tough to find such friendly environment anywhere else!

This thesis is dedicated to Luca, to my parents, Damiana and Sergio, and to my brothers and sister, Andrea, Alberto, Francesco e Anna, for their unconditional love and support, and for reminding me that working hard is important but not the most important thing in life. I love you!

Contents

1	Introduction	1
2	Heterogeneous Beliefs about Rare Event Risk in the Lucas Orchard	5
2.1	Literature Review	7
2.2	An Economy with Multiple Trees and Heterogeneous Beliefs about Systemic Disasters	10
2.2.1	Dividend shares and consumption dynamics	11
2.2.2	Agent optimization problem	13
2.2.3	Price processes and market completeness	15
2.2.4	Market prices of risk	16
2.3	Results and Analysis	17
2.3.1	Equity and variance risk premia	18
2.3.2	Stock return correlation and correlation risk premium	26
2.3.3	Relation between the equity and the variance risk premium	29
2.3.4	The case of a large economy: $N \rightarrow \infty$	33
2.3.5	Consumption share dynamics and survival	34
2.4	Empirical Analysis	36
2.4.1	Data	36
2.4.2	Predictive regressions for the market	40
2.4.3	Predictive regressions in the cross section	47
2.5	Conclusion	53
3	Dividend Growth Predictability and the Price-Dividend Ratio	54
3.1	Present-Value Approach	57
3.1.1	The benchmark model	57
3.1.2	Estimation results	58
3.1.3	Dividend and return predictability	59
3.2	Testing Predictability Hypotheses	60
3.2.1	Asymptotic tests	60
3.2.2	Time-varying expectations	60
3.2.3	Expectation persistence and EIV-problem	61
3.2.4	Finite-sample reliability of asymptotic tests	64
3.2.4.1	Nonparametric Monte Carlo bootstrap	64
3.2.4.2	Constant return or dividend expectations	64
3.2.4.3	Equal expectation persistence	66
3.2.4.4	Why a nonparametric bootstrap?	66

3.3	Bootstrap Tests in the Present-Value Model	68
3.3.1	State-space representation	68
3.3.2	Nonparametric Monte Carlo bootstrap	68
3.3.3	Nonparametric Monte Carlo bootstrap likelihood ratio test	69
3.3.4	Finite-sample reliability of bootstrap likelihood ratio tests	71
3.3.5	The empirical evidence revisited	71
3.3.6	How much predictability?	72
3.3.7	Out-of-sample predictability	74
3.4	Broader Specifications of the Predictive Information Set	76
3.4.1	The present-value model with additional predictive variables: estimation results	77
3.4.2	Tests of constant dividend and return expectations	77
3.4.3	Variability of estimated R-squared values	80
3.5	A Tale of Two Periods	82
3.6	Conclusion	84
4	Predictable Risks and Predictive Regression in Present-Value Models	86
4.1	Present-Value Model	88
4.1.1	Price-dividend ratio	89
4.1.2	Time-varying risks in the present-value model	90
4.2	Data and Estimation Strategy	91
4.2.1	Data	91
4.2.2	State space representation	93
4.3	Predictability Features	94
4.3.1	Estimation results	94
4.3.2	Dividend and return predictability	95
4.3.3	Testing predictability hypotheses	98
4.3.4	Consistency with Predictive Regression Results	100
4.3.4.1	Joint dividend-return predictability features	100
4.3.4.2	Long-horizon predictive regressions	101
4.3.4.3	Out-of-sample predictability	101
4.4	Time-varying Risk Features	102
4.4.1	Correlation between expected and realized returns and dividends	102
4.4.2	Conditional Sharpe ratio dynamics	103
4.4.3	Term structure of long-horizon expectations	106
4.4.4	Term structure of risks	108
4.5	Conclusion	115
A		117
A.1	Proofs	117
A.1.1	Proof of Proposition 1	117
A.1.2	Proof of Proposition 2	117
A.1.3	Proof of Proposition 3	118
A.1.3.1	Stock prices in an economy with $N = 2$ stocks	118
A.1.3.2	General stock price expressions in an economy with N stocks	121
A.1.3.3	The case of a large economy: $N \rightarrow \infty$	121
A.2	Variance risk premium due to large jumps	121

B	125
B.1 Price-dividend ratio	125
B.2 Estimation Methodology	126
B.2.1 Benchmark model	126
B.2.2 Extended Model	128
B.3 Asymptotic EIV bias in standard predictive regressions	130
B.4 Asymptotic Validity of the Bootstrap Likelihood Ratio Test	131
B.5 Bootstrap Distribution of out-of-sample R-squares	133
C	135
C.1 Present-value model	135
C.1.1 Main notation	135
C.1.2 Price-dividend ratio	136
C.1.3 Asymptotic bias in standard predictive regressions	137
C.1.4 Term structure of conditional variances	138
C.2 Kalman Filter	139

List of Figures

2.1	Jump premium	19
2.2	Equity premium	20
2.3	Jumps in stock returns	22
2.4	Stock variance risk premium	23
2.5	Equity and variance risk premium for the market	25
2.6	Conditional stock return correlation in an economy with $N = 2$ assets	27
2.7	Conditional stock return correlation under the risk-neutral measure in an economy with $N = 2$ assets	28
2.8	Instantaneous correlation risk premium, in an economy with $N = 2$ assets	29
2.9	Simulated predictive regressions for different levels of the consumption share	30
2.10	Risk premia in a large economy	35
2.11	Jump in stock returns in a large economy	36
2.12	Observed variance risk premium	37
2.13	Predictive regressions for different levels of the variance risk premium	43
2.14	Predictive regressions for different levels of the difference in beliefs	44
2.15	Kernel regression	45
2.16	Rolling regressions	46
2.17	Variance risk premium regimes	46
2.18	Return predictive regressions for Fama and French portfolios	50
2.19	Return predictability of cap-based portfolios for different levels of difference in beliefs	51
3.1	Model-implied EIV bias	62
3.2	Model-implied R-squared as a function of relative persistence	63
3.3	Quantiles of the LR statistics	65
3.4	Normality test	67
3.5	Distribution of R-squares under H_0	73
3.6	Distribution of out-of-sample R-squares under H_0	75
4.1	Realized volatilities and correlation of returns and dividends	92
4.2	Expected vs Realized yearly returns and dividend growth	97
4.3	Correlation between expected and realized returns and dividends	104
4.4	Risk-Return tradeoff	105
4.5	Conditional Sharpe ratio	106
4.6	Term Structure of conditional expected long-horizon returns and dividends	107
4.7	Level and slope of the term structure of the dividend growth expectation	109
4.8	Term Structure of conditional return and dividend growth volatility	111

4.9	Decomposition of the term structure of conditional variance	112
4.10	Observed vs simulated variance ratios	113
4.11	Decomposition of the term structure of conditional variance of dividend growth .	114
A.1	Variance risk premium due to jumps	123
A.2	Market return predictive regressions using jumps variance risk premium	123
A.3	Market return predictive regressions using jumps variance risk premium for dif- ferent values of the difference in beliefs	124

List of Tables

2.1	Model parameters	18
2.2	Simulated market variance risk premium	24
2.3	Simulated market return predictive regressions	31
2.4	Simulated stock return predictive regressions	32
2.5	Survival	35
2.6	Summary statistics	39
2.7	Market return predictability by variance risk premium	41
2.8	Stock returns predictive regressions	47
2.9	Return predictive regressions for CRSP cap-based portfolios	49
2.10	Return predictability of cap-based portfolios for different levels of difference in beliefs	52
2.11	Size premium for different levels of difference in beliefs	52
3.1	Estimation Results	59
3.2	Standard Likelihood Ratio tests	61
3.3	p-values of the bootstrap tests	72
3.4	Estimation results for the extended model	78
3.5	Dividend growth predictability test for the extended model	79
3.6	Return predictability test for the extended model	81
3.7	Predictability tests assuming two regimes	83
4.1	Summary statistics	93
4.2	Estimation results	95
4.3	R-squares of returns and dividends	96
4.4	Price-dividend ratio decomposition	99
4.5	Simulated R-squares of returns and dividends	102

Introduction

My PhD thesis consists of three papers and focuses on two recent strands of research in asset pricing, namely (i) the equilibrium implications of disagreement about rare event risk and (ii) present-value models for predictability of market returns and dividend growth.

The first chapter, *Heterogeneous Beliefs about Rare Event Risk in the Lucas Orchard*, studies the asset pricing implications of investor disagreement about the probability of a systemic disaster. I start from a structural economy with multiple assets and heterogeneous beliefs on systemic rare event risk, in order to understand how fear and risk sharing mechanisms affect excess returns on equity and pure variance positions and the relation between them, both at an aggregate level and in the cross section of stock returns. Disaster risk, meant as the potential presence of infrequent adverse events of extreme magnitude, has always been a primary concern for academics and investors, even more since the outbreak of the global financial crisis in 2008. Several studies show that a small probability of an extreme event in economic fundamentals can have significant effects on asset prices and that equity and variance risk premia are largely due to a compensation for infrequent jump risk. However, since extreme events are rare by definition, it is difficult to accurately estimate their likelihood, which is a natural source of disagreement on perceived tail risk. Such heterogeneity of beliefs about disasters naturally suggests that a belief-driven risk sharing can be important in explaining the pricing of rare event risk. Intuitively, disaster risk premia are large when pessimists hold a large fraction of the aggregate endowment. Empirically, equity and variance risk premia are also concentrated in these phases, and the effect of disaster risk on risk premia is different for small and large stocks.

The main findings of the paper are the following. First, I find that risk premia increase with the consumption share of the pessimistic agent, as the supply of insurance against future market downturns decreases. Second, the relation between equity and variance risk premia is time varying and systematically linked to the degree of risk sharing. The relation is stronger when pessimists dominate, in phases of financial distress. This prediction is tested empirically and I show that indeed the predictive power of the variance risk premium for future excess returns is time varying and stronger in phases of financial distress, which are characterized by large (absolute) levels of the variance risk premium and large disagreement between investors. Third, the model implied correlation risk premium also increases with the consumption share of the pessimist and drives a large fraction of the aggregate variance risk premium, mainly in large economies, consistent with recent empirical evidence (Driessen, Maenhout, and Vilkov (2012)). The model also has implications for the cross section of stocks: I find that, both in the model and in the data, the forecasting power of the index variance risk premium for future returns is stronger for small stocks, mainly when difference in beliefs is large. Moreover, the variance risk

premium of sufficiently small stocks can switch sign if the consumption share of the optimist is large enough, consistent with recent empirical evidence of positive variance risk premia in the cross section of stocks (see, e.g., Carr and Wu (2009)).

There is a recent and growing literature on asset pricing with multiple trees. My paper is the first to consider a collection of Lucas trees with rare disasters and heterogeneous beliefs about systemic rare events. Other papers have studied disagreement about disaster risk. In particular, Chen, Joslin, and Tran (2012) show that the relation between disaster risk premium and disagreement about disaster risk is highly nonlinear. I contribute to this literature along several dimensions: First I study the effect of disagreement on the variance risk premium and its predictive power for future returns. Second, using a multiple trees setting I study the cross-sectional implications of heterogeneous rare event risk. Third, I test empirically the main models predictions, and finally I use a specification of disagreement that is consistent with empirical regularities of difference in beliefs (see, e.g., Patton and Timmermann (2010) and Buraschi, Trojani, and Vedolin (2013)). The predictive relation between market variance risk premia and future excess returns was first observed by Bollerslev, Tauchen, and Zhou (2009), who also give a theoretical motivation for this link, based on a long-run risk model with stochastic volatility of consumption growth volatility. In my model, the link is due to the presence of a disaster risk premium, motivated by the recent empirical evidence that risk premia largely reflect a premium for large unexpected market downturns (Bollerslev and Todorov (2011)). I also show, theoretically and empirically, that the relation is time varying and concentrated in periods of financial distress. I also contribute to the literature on the size effect, showing that exposure to aggregate variance risk can partly explain the size premium, but only when pessimists dominate.

The following two chapters are based on joint works with Fabio Trojani and build on the present-value model literature. Campbell and Shiller (1988) show that the price-dividend ratio contains information about future expected returns and expected dividend growth. Therefore, if the price-dividend ratio varies, it must forecast either dividend growth or returns, or both. This link has motivated a vast literature that studies *predictive regressions* of returns or dividend growth on the price-dividend ratio. Results are mixed since these regressions lead to different results, according to the sample used in the estimation. Moreover, a large literature studies the econometric properties of standard predictive regressions and outlines several statistical issues. Cochrane (2008b) argues that despite these econometric issues, more powerful tests could give stronger evidence in favour of return predictability. In particular, he suggests that returns and dividend growth have to be studied jointly in order to understand their predictability features, while most previous studies focus only on return predictability. Standard predictive regressions ignore the link between the price-dividend ratio, expected returns and expected cash flow growth, which is instead revealed using present-value models. Recent studies using present-value models to uncover market expectations for returns and dividends include e.g. Rytchkov (2012), Cochrane (2008a,b) and Binsbergen and Koijen (2010), among others.

Predictive regression results typically imply some economically significant evidence of return predictability, even if the statistical significance is weaker in some subperiods, and an almost constant expected dividend growth. This evidence suggests that the price-dividend ratio varies mainly because of discount rate shocks (see, e.g., Campbell (1991) and Cochrane (1992), among others). In contrast, the Kalman-Bucy filter estimation of a benchmark present-value model with hidden dividend and return expectations yields both a predictable return and a predictable dividend growth, indicating that the price-dividend ratio varies because of both dividend expectation and discount rate shocks; see, e.g., Binsbergen and Koijen (2010).

Similar to predictive regressions, in state-space models inference about relevant hypotheses is made tractable by the existence of an asymptotic theory, which under appropriate conditions implies consistency and asymptotic normality of parameter and latent state estimates; see, e.g.,

Liung and Caines (1979) and Spall and Wall (1984). However, the short or moderate length of time-series data available in many predictability studies can make the use of asymptotic inference methods potentially suspect for latent variable approaches in present-value models as well.

The second chapter, *Dividend Growth Predictability and the Price-Dividend Ratio*, proposes a bootstrap methodology to test predictability hypotheses in the context of present-value models with latent expectation processes for returns and dividends. We show that the test is asymptotically valid and has good finite-sample properties while conventional tests strongly over-reject the null of no predictability in small samples, due to restrictive distributional assumptions. We apply the method to study stock market return and dividend growth predictability, using post-war US data. We find evidence of return predictability but no evidence of dividend growth predictability, thus reconciling the diverging predictability conclusions in the literature. We also show that large R-squared values can arise by chance alone, also out-of-sample, even under the null hypothesis of no predictability. Our method could be used more generally to test parametric hypotheses in models estimated by a latent variable approach.

Standard present-value models, as well as forecasting regressions, counterfactually assume constant risks. The third chapter, *Predictable Risks and Predictive Regression in Present-Value Models*, proposes a new class of tractable present-value models with latent dividend and return processes and time-varying cash flow and discount rate risks, to study the joint predictability features of dividends, returns and their second moments. We first find a relatively large return predictability but no statistical evidence of dividend growth predictability. Second, we study the time-varying risk features and we uncover useful additional model implications for the estimated dynamics of market Sharpe ratios and the term structure of expectations and risks: We find quite volatile market Sharpe ratios, which are more countercyclical than under the assumption of constant risks, and highly time-varying term structures of both expectations and volatilities of long-horizon dividend growth and returns, which have the flexibility to take different shapes.

We add to the present-value model literature by incorporating the latent time-varying features of return and dividend risks. Notably, we include time-varying risks maintaining the same level of tractability of present-value models with constant risks. We show that this inclusion helps us to better reconcile under a single common framework a number of predictability findings in the literature, including the very weak post-war evidence of dividend growth predictability at yearly horizons and the low real-time predictability of stock returns. Second, our model allows for a negative correlation between expected and realized dividend growth, which has been shown (see Lettau and Wachter (2007)) to play a crucial role in explaining the value premium and the decreasing term structure of zero-coupon equity volatility documented by Binsbergen, Brandt, and Koijen (2012). In our setting, the estimated correlation between dividend growth and expected cash flow growth is highly time varying and it can also switch sign: it becomes positive in periods in which the conditional covariance between returns and cash flow growth is large. These periods can be often linked to financial turmoil such as the 1989 saving and loans crisis or the beginning of the recent financial crisis in 2007. Interestingly, a countercyclical covariance between returns and dividend growth could help explain the documented countercyclical variation in expected returns within an equilibrium model in which the stochastic discount factor prices dividend growth shocks, as in Lettau and Wachter (2007). Third, conditional Sharpe ratios estimated by our model are often countercyclical, consistently with the empirical evidence, and quite volatile, which is a useful implication with respect to the “Sharpe ratio volatility puzzle” highlighted in Lettau and Ludvigson (2010), among others. Fourth, our model is consistent with Binsbergen, Hueskes, Koijen, and Vrugt (2013) observation of time variation in the slope of the term structures of expected dividend growth and with the empirical evidence reported in Binsbergen, Brandt, and Koijen (2012), which suggests a

decreasing term structure of volatilities on dividend strips (i.e. claims to dividends paid over some specified future time interval), but has also the flexibility to let this term structure change over time, for example with the business cycle.

Heterogeneous Beliefs about Rare Event Risk in the Lucas Orchard

Introduction

Since the outbreak of the global financial crisis in 2008, *tail* or *disaster* risk—understood as the potential presence of infrequent adverse events of extreme magnitude—has been a concern for academics and investors alike. For instance, Hoang Le Huy, head of fixed income and event strategies at Schroders NewFinance Capital, a London-based fund of funds states:¹ “You need to hedge against disaster scenarios. Black swan events are at the forefront for a lot of investors right now. It is not something that people take lightly. A lot of tail risk funds were built on the back of the 2008 disaster.” Numerous studies show that even a small probability of an extreme event in economic fundamentals can have significant effects on asset prices. These extreme events are rare by definition and so accurately estimating their likelihood is difficult, which is a natural source of investor disagreement over perceived tail risk. Such heterogeneity of beliefs about disasters suggests that belief-driven risk sharing could explain the pricing of rare event risk. Compensation for disaster risk contributes to a significant fraction of expected returns on equity and pure variance positions. Thus, a better understanding of disaster risk premia should help explain the dynamics of equity and variance risk premia and the nature of their comovement. This paper studies, both theoretically and empirically, how agent disagreement about disaster risk affects excess return dynamics and the relation between the equity and the variance risk premia both for the market portfolio and the cross section of stocks.

I develop a general equilibrium Lucas (1978) economy with multiple assets and heterogeneous beliefs in which the premia for equity and variance positions and their comovement are endogenously driven by investor disagreement and the cross-sectional distribution of consumption. The model suggests a stronger predictive power of variance risk premium for future excess returns in periods during which pessimists have a relatively large consumption share—that is, in bad states of the economy, which are also characterized by higher (absolute) values of the variance risk premium. Accordingly, I find empirically that variance risk premia and their predictive power for future excess returns are concentrated in phases of substantial disagreement among investors. In these phases, regression coefficients and R^2 are particularly large for small stocks, whose returns are more dependent on the compensation for systemic rare event risk. Therefore, exposure to aggregate variance risk could partially explain the *size effect* (i.e. the observation that smaller firms have higher returns on average), which actually seems to be most

¹See [Risk.net](#).

pronounced during periods in which pessimists hold a large fraction of the aggregate endowment. The time variation in the sign and strength of the predictive power of the aggregate variance premium—both for future excess market returns and for individual stocks—is a challenge for existing consumption-based asset pricing models. The model I posit addresses these empirical challenges, and I provide a structural explanation based on the role of risk sharing between agents who disagree.

The main ingredients of the model are the following. First, I consider an endowment economy with a single consumption good but multiple trees.² The endowment processes follow a geometric Brownian motion with the addition of idiosyncratic and systemic jump components. The presence of multiple trees allows me to study the relation between equity premia in the cross section and the aggregate variance premium, together with the determinants of comovement between trees. Second, two groups of investors, who have constant relative risk aversion (CRRA) preferences over consumption, have different beliefs about the likelihood of a rare systemic event. Disagreement is an important source of non continuous variation in the variance risk premium dynamics. The observed market variance premium is, in fact, highly time varying; periods of small and smooth premium alternate with periods in which the variance premium is larger (in absolute value) and more volatile. The presence of disagreement about the intensity of disasters also allows the variance risk premium to switch sign in certain phases, mainly for small individual stocks; such switching is consistent with the empirical evidence. Third, I assume that the intensity of the systemic jump process is time varying and proportional to an exogenous state variable that can be interpreted as a continuous signal reflecting the state of the economy. The two agents disagree on the coefficient of proportionality, so that the absolute difference in perceived expected growth rates is also proportional to the exogenous state variable and never switches sign. This simple specification of disagreement can be considered as a reduced-form way to capture several empirical regularities of differences in opinion which have been recently documented.

Borrowing from the solution methods proposed for the Lucas Orchard by [Martin \(2013\)](#) and from methods used in the single-asset difference in beliefs model of [Chen, Joslin, and Tran \(2012\)](#), I derive semi-closed-form expressions for the stock prices in my multiple trees economy with heterogeneous beliefs. Price-dividend ratios of individual stocks and parameters in the price process dynamics depend on the consumption share of the two agents, on the state variable driving time-varying intensities, and on the dividend share distribution.

Using the model solution, I derive a number of testable predictions. First, the equity (variance) risk premium of an individual stock tends to increase (decrease) with its dividend share and with the consumption share of the pessimistic agent; these phenomena can be explained by the risk-sharing behavior of disagreeing investors. Moreover, as noted above, the variance risk premium can switch sign—in particular for small stocks—when optimists consume a large fraction of the aggregate endowment and disagreement is large enough. In line with the data, the variance premium is time varying; it alternates phases of small and smooth premia with periods in which the variance premium is larger (in absolute value) and more volatile, where the change in regime is driven by an abrupt change in the cross-sectional distribution of agent consumption. Second, the model-implied correlation risk premium inherits these features because, consistently with the empirical findings of [Driessen, Maenhout, and Vilkov \(2012\)](#), the index variance risk premium is largely due to a covariance premium, mainly when assets are relatively evenly distributed or the number of stocks in the economy is large. While the cross-sectional distribution of agent consumption mainly affects the risk-neutral stock return correlation, the physical correlation is relatively insensitive to it, which leads to a countercyclical correlation risk premium. Third, rare event risk implies a tight link between the equity and the variance risk

²In this context, *trees* are assets and a collection of trees is an *orchard*. See e.g. [Martin \(2013\)](#).

premium, both for the market and for the cross section of stock returns. This link provides the basic intuition for the role of the variance premium in predicting future excess returns. However, standard predictive regressions imply an unconditionally linear relation between equity and variance risk premia, whereas in the model the regression coefficients are stochastic and depend on the asset's dividend share and the agents' consumption share. I show by simulation that the aggregate variance premium's power to predict future excess returns is stronger when the consumption share of the pessimist is larger, i.e., in bad states of the economy. At a disaggregate level, the predictive power of the variance risk premium is especially great for small stocks. Fourth, I consider the special case of a large diversified economy in which the number of stocks approaches infinity and all stocks have the same dividend share, which approaches zero. In this case, only systemic risk is priced and the relation between equity and variance risk premia is conditionally linear. Moreover, infinitely small assets still earn a risk premium owing to the presence of systemic rare event risk.

The model's main predictions are tested using the aggregate *S&P500* composite index as a proxy for the aggregate market and the return time series of all its single constituents, as well as CRSP cap-based portfolio returns, to analyze cross-sectional implications and the differential effects of small versus big stocks, based on monthly data from January 1990 through December 2011. The test results confirm that the index variance premium's ability to predict future excess returns is (a) time varying for the market and for single stocks or stock portfolios and (b) stronger during periods of financial distress. Such periods are characterized by large (absolute) variance premia and substantial investor disagreement, which is proxied by the dispersion in one-year-ahead forecasts of real GDP growth from the BlueChip Economic Indicator. The predictive power of the variance premium is stronger (on average) for small stocks, which have returns that depend more on the compensation for systemic rare event risk. For example, the adjusted R^2 of a standard predictive regression of excess six-month returns on the aggregate variance premium is about 63% larger for the small-cap portfolio than for big caps. The difference between small- and big-cap portfolios is particularly evident in periods of high disagreement. Intuitively, investors will require higher return from assets that are more sensitive to systemic disaster risk.³ However, this reasoning holds only when the perceived premium for systemic jumps is sufficiently large. The model suggests that the systemic jump premium component can even have a negative effect on a stock's excess returns if the pessimists' consumption share is low enough. Thus the size premium could move in opposite directions depending on what agent type dominates the market. This finding is consistent with the mixed results reported in the empirical literature on size premia.

The rest of the paper is organized as follows. Section 2.1 provides a literature review. Section 2.2 introduces the basic model setup as well as the optimal consumption allocation, market prices of risk, and equilibrium market prices. Section 2.3 analyzes the properties of the equity and variance risk premia, of their relationship, and the correlation risk premium. It also studies the case of an infinitely large and diversified economy. Section 2.4 describes the data and tests empirically the main implications of the model. Section 2.5 concludes and discusses possible model extensions and directions for future research. All proofs can be found in Appendix A.

2.1 Literature Review

This paper is related to several different strands of the literature. The first is the growing research on asset pricing with multiple trees. [Cochrane, Longstaff, and Santa-Clara \(2008\)](#)

³The empirical literature on the size premium identifies several reasons why small stocks are more sensitive to systemic risk. One possible explanation is that small firms are more affected by tight credit conditions.

highlight the asset pricing implications of a two-trees [Lucas \(1978\)](#) economy with a log-utility representative investor. [Martin \(2013\)](#) introduces multiple Lucas trees (Lucas orchard) following jump-diffusion processes and a representative agent with power utility. [Chabakauri \(2013\)](#) considers two trees and two CRRA investors with heterogeneous risk aversions and portfolio constraints, and he looks at the effects on return correlations and volatilities. [Buraschi, Trojani, and Vedolin \(2014\)](#) specify a diffusive two-trees model with heterogeneity in beliefs; they characterize the relation between the difference in opinions, volatility and correlation risk premia of index and individual options. In contrast to previous papers, I specify a collection of Lucas trees with rare disasters and heterogeneous beliefs about the intensity of systemic rare events, with the goal of studying the implications for equilibrium risk premia and for the relation between the market variance risk premia and excess returns. Multiple trees allow me to analyze the contribution of a premium for covariance risk to this predictive relation. This insight is motivated by the empirical evidence in [Driessen, Maenhout, and Vilkov \(2012\)](#) that the index variance risk premium is largely due to the high price of correlation risk and that option-implied correlations have remarkable predictive power for future stock market returns. In my model, covariance risk can contribute to a large fraction of the aggregate variance premium when the economy is dominated by pessimistic agents. In such states, fear of systemic disasters requires large compensation for both equity and variance risks, leading to a strong comovement between equity and variance risk premia.

The predictive relation between the market variance risk premium and excess returns was first observed by [Bollerslev, Tauchen, and Zhou \(2009\)](#), who provide empirical evidence that the variance risk premium accounts for a nontrivial fraction of the time-series variation in post-1990 aggregate stock market returns at short horizons. They motivate this link theoretically in a long-run risk model with stochastic volatility of consumption growth volatility. [Londono \(2011\)](#) extends that model to an international setting and provides evidence that the US variance premium predicts local and foreign equity returns. Consistently with the implications of a consumption-based model along the lines of [Bollerslev, Tauchen, and Zhou \(2009\)](#), [Bali and Zhou \(2011\)](#) find that the variance risk premium explains both the time-series and cross-sectional variation in stock returns. In such models, recursive preferences are crucial to generate a premium for stochastic volatility of consumption growth volatility, which then drives both the equity and variance risk premia. Yet, [Wu \(2012\)](#) observes that there is no empirical correlation between the variance risk premium and the volatility of consumption growth volatility. Moreover, [Bollerslev and Todorov \(2011\)](#) find that more than half of the variance risk premium is driven by disaster risk and suggest that equilibrium-based asset pricing models should accommodate large and time-varying compensation for rare disasters. A step in this direction is taken by [Drechsler \(2013\)](#) and [Drechsler and Yaron \(2011\)](#), who incorporate time-varying uncertainty into a long-run risk model—with jumps in expected growth and growth volatility processes—and are able to fit sample moments of the variance risk premium.⁴ Intuitively, explaining the existence and properties of variance risk premia requires that equilibrium models endogenously generate time-varying non normality in returns. Following this idea, [Bekaert and Engstrom \(2010\)](#) propose a consumption-based model with preferences as in [Campbell and Cochrane \(1999\)](#) but with nonlinear consumption growth. This model posits two types of shocks (good and bad) that are drawn from potentially skewed and fat-tailed distributions and have a time-varying relative importance. However, none of the cited attempts to explain variance risk premia in a general equilibrium setting truly accounts for rare disasters, defined as jumps in

⁴Other recent papers using long-run risk models to explain variance risk premia include [Zhou \(2010\)](#) and [Zhou and Zhu \(2010\)](#). [Jin \(2013\)](#) compares several calibrated specifications of long-run risk models, with and without jumps in log-run expected consumption growth and consumption volatility, and argues that jumps in volatility are crucial for explaining variance risk premia and for generating a degree of return predictability that is consistent with the data.

realized (not expected) endowment growth.

The idea that the possibility of sudden downward jumps in the endowment may help explain the equity premium puzzle dates back to Rietz (1988). More recently, Gabaix (2012) and Wachter (2013) resolve several asset pricing puzzles by including a time-varying risk of disasters in otherwise standard models. The literature on rare disasters does not seek to explain the variance risk premium’s puzzling dynamics or its ability to predict future excess returns.⁵ This paper seeks to fill that gap starting from a general equilibrium model in which two sets of agents have different beliefs about the probability of a disaster occurring.

Previous papers have studied the disagreement that surrounds assessments of disaster risk. Dieckmann (2011) provides an equilibrium model in which log-utility investors have heterogeneous beliefs about the likelihood of rare events; he explores the asset pricing implications of this setup in an incomplete capital market as well as the effects of market completion. Chen, Joslin, and Tran (2012) consider a complete market setting and assume that two CRRA agents disagree about rare event risk. They show that the relation between the disaster risk premium and the extent of disagreement about disaster risk is highly nonlinear; a small proportion of optimistic investors can greatly attenuate the impact of disaster risk on stock prices. I contribute to this literature along several dimensions. First, I study the effects of disagreement on variance risk premia and its predictive power for excess returns. Second, using the multiple trees setting I study the cross-sectional implications of heterogeneous rare event risk. Third, I test empirically the model’s main predictions. Finally, I use a specification of disagreement that is consistent with several empirical regularities of differences in opinion. Patton and Timmermann (2010) and Buraschi, Trojani, and Vedolin (2013) show that differences in beliefs are highly time varying and countercyclical. Moreover, Patton and Timmermann (2010) suggest that there is a strong negative correlation between dispersion and consensus forecast on GDP growth. They also find that forecasters’ view are persistent—in other words, they tend to be consistently optimistic or pessimistic. In my model, disagreement is countercyclical whereas the average belief about expected consumption growth is procyclical; these dynamics lead to a perfect negative correlation between consensus and dispersion, whose persistence is guaranteed by a positive exogenous state variable.

The relation between the equity and variance risk premium reflects properties of the equilibrium risk–return trade-off. Instead of taking the physical expectation and variance of returns, the volatility risk premium considers the difference between physical and risk-neutral expectation of future volatility, thus isolating the priced component of volatility risk. Hence this paper relates also to the large literature on a time-varying risk–return trade-off. For example, Brandt and Wang (2010) estimate the monthly market risk–return relationship from the cross section of equity returns and show that this relationship is usually positive but varies considerably over time. They also show that the coefficient relating the market risk premium to the conditional market volatility exhibits a countercyclical pattern. Interestingly, Yu and Yuan (2011) find a positive risk–return trade-off when sentiment is low, but no relation when it is high.⁶ These authors argue that this result is a challenge for traditional asset pricing theories. Similarly, I find that the relation between equity and variance risk premium is evident mainly when pessimists account for a large fraction of aggregate consumption. Hong and Sraer (2012) also study the link between the risk–return relationship and divergence of opinion in the cross section; they show that the risk–return relationship for single stocks can flip from positive to negative when investor disagreement over the asset value is large enough. Anderson, Ghysels, and Juergens

⁵A recent exception is Kim (2013), who uses multiple regimes to model an endowment economy with time-varying likelihood of disasters.

⁶Yu and Yuan (2011) use the sentiment index by Baker and Wurgler (2006) to identify low and high sentiment periods.

(2009) augment the typical risk–return trade-off model with a measure of uncertainty that is based on the level of disagreement among professional forecasters. They find stronger empirical evidence for an uncertainty–return trade-off than for the traditional risk–return trade-off.

Finally, my results on the connection between the cross section of excess stock returns and the aggregate variance premium are related to the literature on the size effect. Lemmon and Portniaguina (2006) demonstrate a negative relation between the size premium and consumer confidence. In fact, the *size effect* (whereby smaller firms have higher returns on average) seems to be concentrated in periods characterized by large disagreement. Intuitively, investors tend to require higher returns from assets that are more sensitive to systemic disaster risk. This intuition is well explained by Jagannathan and Wang (1996), who argue that the market beta of firms with a greater likelihood of financial distress (e.g., small firms) is more sensitive to changes in the business cycle. Investor sentiment is thus related to time variation in the expected returns of those firms because such sentiment forecasts future business conditions. However, this reasoning holds only when the perceived systemic jump premium is high. My model indicates that the jump premium component can actually have a negative effect on the stock’s excess returns if the consumption share of the pessimists is sufficiently low. Therefore, the size premium can move in opposite directions, depending on which agent type dominates the market, consistently with the mixed results of the later empirical research on the size effect.⁷

2.2 An Economy with Multiple Trees and Heterogeneous Beliefs about Systemic Disasters

This section introduces the model, which is a simple, continuous-time generalization of the standard Lucas (1978) endowment economy. The model incorporates rare disasters and heterogeneous beliefs about the probability of a common jump in N Lucas trees. For notational convenience, vectors and matrices are denoted by bold symbols.

Two agents ($i = A, B$) observe the dividend stream produced by each tree, D_j , with the following exogenous dynamics:

$$\frac{dD_j(t)}{D_j(t)} = \mu_j dt + \sigma_j dW_{jt} + k_j dN_{jt} + k_j dN_{ct}, \quad j = 1, \dots, N, \quad (2.1)$$

where $\mathbf{W}_t = (W_{1t}, W_{2t}, \dots, W_{Nt})'$ is an N -dimensional standard Brownian motion driving regular economic risk, rare event risk enters through the Poisson processes $\mathbf{N}_t = (N_{1t}, N_{2t}, \dots, N_{Nt})'$ and N_{ct} with respective intensities $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)'$ and $\lambda_c(t)$. Thus each stock has both an idiosyncratic and a systemic event risk component. Namely, the jump in the dividend growth of stock j is idiosyncratic if driven by a jump in N_{jt} whereas jumps in N_{ct} are common to all stocks. For simplicity, and since the goal is to understand the asset pricing implications of heterogeneous beliefs on the probability of a common jump, I assume that the intensity of the systemic Poisson process is time varying while all other parameters, including the idiosyncratic jump intensities, are constant.⁸ Furthermore, the coefficients μ_j , σ_j , k_j , and λ_j —which represent, respectively, the expected growth rate and volatility of dividend growth without jumps, the jump size, and the idiosyncratic jump intensity—are assumed to be identical for all trees in the economy (hence I will suppress their subscript j). The jump size k is restricted to be

⁷See e.g. Crain (2011) for a review of the size effect.

⁸Wachter (2013), among others, underlies the importance of taking into account time variation in the probability of rare disasters to help explain, e.g., time variation in the equity premium and the excess volatility puzzle, while Berkman, Jacobsen, and Lee (2011) provide empirical support for time-varying rare disaster intensity.

negative and strictly less than 1 in absolute value; this ensures that dividend processes are positive.⁹

To focus on the effects of heterogeneous systemic rare event risk on risk premia, I assume that agent beliefs differ only with respect to the systemic rare event intensity $\lambda_c(t)$, which is a function of an exogenous affine state variable $X(t)$. In particular, agent i believes that the common jump frequency is given by¹⁰

$$\lambda_c^i(t) = \beta^i X(t), \quad (2.2)$$

for $i = A, B$, where $X(t)$ follows a CIR process,

$$dX(t) = \varphi[1 - X(t)]dt + \sigma_X \sqrt{X(t)} dW_t^X, \quad (2.3)$$

for W_t^X a standard Brownian motion that is independent of \mathbf{W}_t . This assumption ensures positivity of the intensity of a common jump under each agent's beliefs, which also follows a CIR process.¹¹ I assume that the long-term mean of X is equal to 1, so that β^i represents the expected systemic rare event intensity perceived by agent i .

The probability measures of the two agents are equivalent because they agree on null sets. Hence, their Radon–Nikodym derivative $\phi(t) = d\mathbb{P}^B/d\mathbb{P}^A$ exists and has been shown by [Chen, Joslin, and Tran \(2010\)](#) to have the following dynamics:

$$\frac{d\phi(t)}{\phi(t)} = (\beta^A - \beta^B)X(t)dt + \left[\frac{\beta^B}{\beta^A} - 1 \right] dN_{ct}. \quad (2.4)$$

If the agents observe a common jump (i.e. if $dN_{ct} = 1$) then the likelihood ratio jumps by a factor of β^B/β^A . If agent B is optimistic—which means he believes that the probability of a systemic (negative) jump is lower (i.e., $\beta^B < \beta^A$, as I assume throughout the paper)—then ϕ 's jump in response to systemic disaster is a downward one. That being said, the absence of systemic jumps over a period of time is more consistent with the optimist's beliefs and so the likelihood ratio increases deterministically at a rate $\lambda_c^A - \lambda_c^B$. Note that, even in the case of constant systemic rare event intensity (i.e., X constant) the state variable ϕ varies over time and decreases dramatically following a disaster.

2.2.1 Dividend shares and consumption dynamics

I consider an endowment economy in which all trees produce the same perishable consumption good; therefore, aggregate consumption equals the sum of all dividends:

$$C(t) = \sum_{j=1}^N D_j(t). \quad (2.5)$$

⁹The assumption of constant jump size could be relaxed, but it helps to maintain tractability and to isolate the effect of disagreement about rare event intensity. In single-asset models, [Wachter \(2013\)](#) assumes a lognormal distribution for the jump size, [Drechsler and Yaron \(2011\)](#) and [Jin \(2013\)](#) use Gamma distributions, and [Gabaix \(2012\)](#) and [Tsai and Wachter \(2013\)](#) assume a power law distribution—in line with [Bollerslev and Todorov \(2011\)](#)'s nonparametric evidence that the tails of the risk-neutral distribution of returns decay according to a power law.

¹⁰[Benzoni, Collin-Dufresne, Goldstein, and Helwege \(2012\)](#) assume a similar dynamics for a country's default intensity and use a β parameter that depends on the state of the world. They then employ learning to capture contagion effects in the perceived default intensities of different countries.

¹¹[Wachter \(2013\)](#) and [Chen, Joslin, and Tran \(2012\)](#) also assume CIR processes for the rare event intensity. [Chen, Joslin, and Tran \(2012\)](#) include disagreement directly in the long-run average jump intensity whereas here the proportionality coefficient β^i is used to introduce disagreement.

Let s_j be the share of consumption contributed by stock j ,

$$s_j = \frac{D_j(t)}{C(t)}. \quad (2.6)$$

An application of Itô's lemma to Equation (2.1) gives its dynamics:

$$ds_j = \sigma^2 s_j \left(\sum_{i=1}^N s_i^2 - s_j \right) dt + \sigma s_j \left(dW_{jt} - \sum_{i=1}^N s_i dW_{it} \right) + s_j \frac{k(1-s_j)}{ks_j+1} dN_{jt} - s_j \sum_{i \neq j} \frac{ks_i}{ks_i+1} dN_{it}. \quad (2.7)$$

Intuitively, the dividend share of asset j increases when there is a positive Brownian shock to its dividend growth dynamics or an idiosyncratic disaster involving any of the other dividend processes; the share decreases in response to an idiosyncratic jump in its own dividend growth. Systemic jumps do not affect dividend share dynamics because such jumps are assumed to have the same impact on all dividend processes. The drift in Equation (2.7) is zero when $s_j = 0$, $1/N$, or 1, and the dividend share distribution is not stationary because one asset ultimately becomes dominant in the market; that is, $ds_j = 0$ for $s_j = 0, 1$.¹²

By construction, the dividend shares s_j sum to 1 and the $N - 1$ state variables s_j for $j = 2, \dots, N$ are enough to describe the relative size of the N trees. Asset prices will depend on these $N - 1$ dividend shares, but [Martin \(2013\)](#) argues that it is often more convenient to use a monotonic transformation of these state variables,

$$u_j = \ln \frac{s_j}{s_1}, \quad (2.8)$$

which measures the size of asset j relative to asset 1. As s_j ranges from 0 to 1, u_j can take all values on the real line. Applying Itô's lemma to the definition in Equation (2.8) while assuming symmetric assets, we obtain the dynamics

$$du_j = d \ln D_j - d \ln D_1 = \sigma(dW_{jt} - dW_{1t}) + \ln(k+1)(dN_{jt} - dN_{1t}) \quad (2.9)$$

for $j = 2, \dots, N$. In matrix notation, the dynamics of $\mathbf{u} = (u_2, \dots, u_N)'$ is given by

$$d\mathbf{u} = \sigma \mathbf{U} d\mathbf{W}_t + \ln(k+1) \mathbf{U} d\mathbf{N}_t, \quad (2.10)$$

where \mathbf{U} is a $(N - 1) \times N$ matrix:

$$\mathbf{U} \equiv \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -1 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

From Equation (2.5), the dynamics of aggregate consumption growth is given by

$$\frac{dC(t)}{C(t)} = \mu dt + \sigma \sum_{j=1}^N s_j dW_{jt} + k \sum_{j=1}^N s_j dN_{jt} + k dN_{ct}. \quad (2.11)$$

¹²This feature is shared by many of the literature's general equilibrium models that involve two or more trees. See for example [Cochrane, Longstaff, and Santa-Clara \(2008\)](#), who discuss properties of the dividend share dynamics in the case of two trees.

Observe that even if agents agree on μ , the growth rate of consumption in *normal* times, disagreement about the systemic rare event intensity leads to disagreement about the total expected growth rate,

$$\mu_C^i = E_t^i \left[\frac{dC(t)}{C(t)} \right] = \mu + k(\lambda + \lambda_c^i(t)). \quad (2.12)$$

Here $E_t^i(\cdot)$ denotes conditional expectation under the probability measure \mathbb{P}^i , which summarizes agent i 's beliefs. Thus, the difference in expected growth rates can be expressed in terms of the dispersion in beliefs,

$$\mu_C^B - \mu_C^A = k(\lambda_c^B(t) - \lambda_c^A(t)) = -k(\beta^A - \beta^B)X(t), \quad (2.13)$$

which is linear in the exogenous state variable X .

This simple specification of disagreement is a parsimonious way to capture several empirical regularities of differences in opinion that have been reported recently. [Patton and Timmermann \(2010\)](#) and [Buraschi, Trojani, and Vedolin \(2013\)](#) show that differences in beliefs are highly time varying and countercyclical. Moreover, [Patton and Timmermann \(2010\)](#) suggest that (a) there is a strong negative correlation between belief dispersion and a consensus forecast of GDP growth and (b) forecasters' views tend to be consistently optimistic or consistently pessimistic. In the model, disagreement is countercyclical if the state variable X is interpreted as an exogenous continuous signal about the state of the economy. The average belief as regards expected consumption growth is a decreasing function of X , while the absolute difference in perceived expected growth rates is increasing in X ; the result is a perfect negative correlation between consensus forecast and the dispersion in forecasters' beliefs, the persistence of which is guaranteed by the positivity of X .¹³

2.2.2 Agent optimization problem

Agents have a constant relative risk aversion (CRRA) utility over consumption with finite horizon T :

$$U^i(C^i(t)) = \frac{C^i(t)^{1-\gamma}}{1-\gamma} \quad (2.14)$$

for $i = A, B$; here γ is the coefficient of relative risk aversion, which is assumed to be identical across agents.¹⁴ If we assume complete markets and use martingale techniques (see e.g. [Cox and Huang \(1989\)](#)) then agent i 's optimization problem can be written in static form as

$$J^i = \max_{C^i} E^i \left[\int_0^T e^{-\delta t} U^i(C^i(t)) dt \right], \quad \text{s.t.} \quad E^i \left[\int_0^T \eta^i(t) C^i(t) dt \right] \leq W^i(0). \quad (2.15)$$

Here δ is the time preference rate and $\eta^i(t)$ is the state price density of agent i , whose dynamics is given by

$$\frac{d\eta^i(t)}{\eta^i(t)} = -r(t)dt + \left[\sum_{j=1}^N (\lambda - \lambda_j^Q(t)) + (\lambda_c^i(t) - \lambda_c^Q(t)) \right] dt - \boldsymbol{\theta}(t)' d\mathbf{W}_t + \sum_{j=1}^N \left(\frac{\lambda_j^Q(t)}{\lambda} - 1 \right) dN_{jt} + \left(\frac{\lambda_c^Q(t)}{\lambda_c^i(t)} - 1 \right) dN_{ct}. \quad (2.16)$$

¹³Disagreement could instead switch sign in the single-asset belief disagreement model of [Chen, Joslin, and Tran \(2012\)](#), even if disaster intensities follow CIR processes, because these authors introduce disagreement directly in the long-run average jump intensity.

¹⁴[Dieckmann and Gallmeyer \(2005\)](#) introduce heterogeneity only through different levels of relative risk aversion and study equilibrium allocations. [Chabakauri \(2013\)](#) considers two trees and two CRRA investors with heterogeneous risk aversions and portfolio constraints, and he examines the effects on return correlations and volatilities. [Chen, Joslin, and Tran \(2012\)](#) argue that combining heterogeneous beliefs about disasters and different risk aversions can amplify the effects of risk sharing but does not qualitatively change basic asset pricing results.

In this expression, the N -vector $\boldsymbol{\theta}$ is the market price of regular economic risk associated with Brownian motion $\mathbf{W}(t)$, while λ_j^Q and λ_c^Q are the risk-neutral rare event intensities associated with the respective Poisson processes $N_j(t)$ and $N_c(t)$.¹⁵ Agents are assumed to be initially endowed with a fraction x_s^i of each stock; that is, $W^i(0) = x_s^i \sum_{j=1}^N S_j(0)$. The standard optimality condition now yields

$$C^i(t) = I^i \left(y^i \eta^i(t) e^{\delta t} \right) = \left(y^i \eta^i(t) e^{\delta t} \right)^{-1/\gamma},$$

where $I^i(\cdot)$ is the inverse marginal utility function of agent i and y^i is the Lagrange multiplier that solves the following budget constraint:

$$E^i \left[\int_0^T \eta^i(t) I^i \left(y^i \eta^i(t) e^{\delta t} \right) dt \right] = W^i(0).$$

The equilibrium allocations can be characterized by solving the optimization problem of a representative agent whose utility function is a weighted sum of the two agents' utilities,

$$U(C(t), \phi(t)) \equiv \max_{C^A + C^B = C} \{U^A(C^A(t)) + \phi(t) U^B(C^B(t))\}, \quad (2.17)$$

where the weight ϕ is stochastic and is driven by the difference in beliefs (see Equation (2.4)).¹⁶ Hence the planner's problem (under \mathbb{P}^A , the pessimist's probability measure) is as follows:

$$J = \max_{C^A, C^B} E^A \left[\int_0^T e^{-\delta t} \left(\frac{C^A(t)^{1-\gamma}}{1-\gamma} + \phi(t) \frac{C^B(t)^{1-\gamma}}{1-\gamma} \right) dt \right] \quad \text{s.t.} \quad C^A(t) + C^B(t) = C(t). \quad (2.18)$$

The equilibrium consumption allocations are obtained from the first-order condition of the representative agent's problem while using individual agents' optimality conditions as just described.

Proposition 1 *Equilibrium consumption allocations are*

$$C^A(t) = \frac{1}{1 + \phi(t)^{1/\gamma}} C(t) \quad \text{and} \quad C^B(t) = \frac{\phi(t)^{1/\gamma}}{1 + \phi(t)^{1/\gamma}} C(t), \quad (2.19)$$

and investors' state price densities are

$$\eta^A(t) = e^{-\delta t} \frac{(1 + \phi(t)^{1/\gamma})^\gamma}{y^A C(t)^\gamma} \quad \text{and} \quad \eta^B(t) = \eta^A(t) \frac{\phi(0)}{\phi(t)} = e^{-\delta t} \frac{(1 + \phi(t)^{1/\gamma})^\gamma}{y^B C(t)^\gamma \phi(t)}. \quad (2.20)$$

Here $\phi(0)$ solves either agent's individual budget constraint,¹⁷ and the stochastic weighting process $\phi(t) = y^A \eta^A(t) / y^B \eta^B(t)$ follows the dynamics given in Equation (2.4) with jump intensity $\lambda_c^A(t)$.

¹⁵The market prices of diffusion and jump risk are not agent specific if the market is complete, since the agents have to agree on the observed price paths, see e.g. Dieckmann (2011). Completeness of the market is discussed in Section 2.2.3.

¹⁶The approach to formulating a representative agent problem with state-dependent weight was introduced by Cuoco and He (1994); more recent examples can be found in Basak and Cuoco (1998) and Buraschi and Jiltsov (2006). In a complete markets setting with heterogeneous beliefs the weight is stochastic and equal to the Radon–Nikodym derivative $\phi(t) = d\mathbb{P}^B/d\mathbb{P}^A$.

¹⁷The budget constraints of agents determine only the ratio y^A/y^B . I set $y^A = U'(C(0), \phi(0))$ without loss of generality, so that $\eta^A(0) = \eta^B(0) = 1$.

Proposition 1 characterizes the dependence of individual state price densities and consumption policies on C and ϕ , which represent aggregate endowment and belief disagreement risk in the economy, respectively. In this N -trees setting, the aggregate endowment C depends on the exogenous single dividend growth processes and on the dividend shares. Under homogeneous beliefs, the disagreement risk vanishes because ϕ is constant and depends only on initial wealth: $\phi = (x_s^B/x_s^A)^\gamma$. This means that, under homogeneous beliefs, the investors who are initially more wealthy consume more in all future states and times. In contrast, consumption differences can change sign if agents are heterogeneous. Namely, if a systemic disaster occurs, the consumption share of the pessimist (agent A) increases as ϕ jumps down.

For convenience, I define explicitly the consumption shares of the pessimistic and optimistic agents as

$$c^A(t) \equiv \frac{C^A(t)}{C(t)} \quad \text{and} \quad c^B(t) \equiv \frac{C^B(t)}{C(t)},$$

respectively. These shares will drive the market prices of risk and risk premia.

2.2.3 Price processes and market completeness

Assume the existence of a capital market that allows agents to share risk and finance consumption. The market consists of N risky assets with price vector $\mathbf{S}(t) = (S_1(t), S_2(t), \dots, S_N(t))'$, each in unit net supply, as well as a riskless asset of price $B(t)$ in zero net supply. Then, for $j = 1, \dots, N$, the price process dynamics are as follows:

$$\frac{dS_j(t) + D_j(t)dt}{S_j(t)} = \mu_{S_j}(t)dt + \sigma_{S_j}(t)[d\mathbf{W}_t' d\mathbf{W}_t^X]' + \mathbf{k}_{S_j}(t)d\mathbf{N}_t + k_{S_j}^c(t)dN_{ct}, \quad (2.21)$$

$$\frac{dB(t)}{B(t)} = r(t)dt. \quad (2.22)$$

To simplify notation, let me define the $N \times 1$ vector $\boldsymbol{\mu}_S$ of expected returns in normal times, the $N \times N + 1$ matrix $\boldsymbol{\sigma}_S$ of diffusion volatilities, the $N \times N$ matrix \mathbf{k}_S of return jump sizes related to idiosyncratic jumps, and the $N \times 1$ vector \mathbf{k}_S^c of return jump sizes related to systemic jumps

$$\boldsymbol{\mu}_S = \begin{bmatrix} \mu_{S_1} \\ \mu_{S_2} \\ \vdots \\ \mu_{S_N} \end{bmatrix}, \quad \boldsymbol{\sigma}_S = \begin{bmatrix} \sigma_{S_1} \\ \sigma_{S_2} \\ \vdots \\ \sigma_{S_N} \end{bmatrix}, \quad \mathbf{k}_S = \begin{bmatrix} k_{S_1} \\ k_{S_2} \\ \vdots \\ k_{S_N} \end{bmatrix}, \quad \mathbf{k}_S^c = \begin{bmatrix} k_{S_1}^c \\ k_{S_2}^c \\ \vdots \\ k_{S_N}^c \end{bmatrix}. \quad (2.23)$$

All the vectors and matrices in (2.23), as well as the riskless rate r , are determined endogenously in equilibrium. However, with only N risky securities the market is incomplete, since they only span the uncertainty driven by the Brownian motions.¹⁸ Hence I assume agents can also trade in $N + 1$ rare event insurance products $P_j(t)$, $j = 1, \dots, N$ and $P_c(t)$, which are in zero-net supply, do not pay dividends, and have price processes

$$\frac{dP_j(t)}{P_j(t)} = \mu_{p_j}(t)dt + k_{p_j}(t)dN_{jt}, \quad (2.24)$$

$$\frac{dP_c(t)}{P_c(t)} = \mu_{p_c}(t)dt + k_{p_c}(t)dN_{ct}, \quad (2.25)$$

¹⁸More precisely, there are $N + 1$ Brownian shocks in the economy; however, the risk of changes in the disaster probability (i.e., shocks to W_t^X) are not priced in the power utility setting although they would be if agents had recursive preferences. See also Wachter (2013).

where μ_{p_j} and μ_{p_c} are determined in equilibrium, whereas jump sizes can be freely chosen and need only be different from zero in order to complete the market. These assets can be interpreted as insurance products against rare event risk because they do not contain any continuous source of uncertainty. The buyer of asset P_j , $j = 1, \dots, N, c$, is rewarded in the amount μ_{p_j} every moment of time, but runs the risk that the asset's value drops to $(1 + k_{p_j})P_j$ when the corresponding Poisson process N_{jt} jumps. Therefore, selling assets P_j , $j = 1, \dots, N$, provides insurance against idiosyncratic jumps, P_c is a form of insurance against systemic disasters.¹⁹ In general, any set of $N + 1$ assets spanning all jump components would complete the market, but the choice of disaster insurances is the most appealing since it isolates the impact of the different rare events.

2.2.4 Market prices of risk

Market prices of risk are obtained by applying Itô's lemma to Equation (2.20) and then comparing the resulting dynamics with Equation (2.16). They are summarized in the following proposition.

Proposition 2 *The market prices of normal economic risk, both risk-neutral rare event intensities, and the short rate are given by*

$$\theta_j(t) = \gamma s_j \sigma, \quad (2.26)$$

$$\lambda_j^Q = \lambda(s_j k + 1)^{-\gamma}, \quad (2.27)$$

$$\lambda_c^Q(t) = \left(c^A(t) \lambda_c^A(t)^{1/\gamma} + c^B(t) \lambda_c^B(t)^{1/\gamma} \right)^\gamma (k + 1)^{-\gamma}, \quad (2.28)$$

$$r(t) = \delta + \gamma \mu - \frac{1}{2} \gamma (\gamma + 1) \sum_{j=1}^N s_j^2 \sigma^2 - c^B(t) (\lambda_c^A(t) - \lambda_c^B(t)) + \sum_{j=1}^N (\lambda - \lambda_j^Q(t)) + (\lambda_c^A(t) - \lambda_c^Q(t)). \quad (2.29)$$

The market price of economic risk has the standard solution as extended to the case of N trees. The risk-neutral intensity λ_j^Q of an idiosyncratic jump in the dividend process D_j depends only on the dividend share of asset j given that agents agree on the physical idiosyncratic jump intensities. For any dividend share distribution, the idiosyncratic jump risk premia $\lambda_j^Q(t)/\lambda$ are constant and always greater than 1, and they tend to unity as s_j approaches zero. Thus the risk of idiosyncratic jumps in small assets is not priced, and in general the price associated with idiosyncratic jump risk is small when the number of stocks in the economy, N , is large. The risk-neutral common disaster frequency λ_c^Q is a nonlinear function of the two agents common jump intensities weighted by their consumption shares; it could be smaller than the physical intensity when the optimist's consumption share is large, leading to a systemic jump premium of less than 1. The riskless interest rate follows the standard expression in Lucas economies with multiple trees (see e.g. [Cochrane, Longstaff, and Santa-Clara \(2008\)](#)), with the addition of three components related to disagreement and jump premia. The equilibrium short rate is generally decreasing with the consumption share of the pessimistic agent A , as in the aftermath of a disaster.

Using Proposition 2, it is possible to derive explicitly the risk premia on the risky assets once the volatilities and jump sizes of the stock price processes (i.e., $\sigma_S(t)$, $k_S(t)$ and $k_S^c(t)$) are known. This can be done by applying Itô's lemma to the stock prices. For integer risk aversion γ , the resulting equation can be solved in semi-closed-form as summarized in the next proposition.

¹⁹Catastrophe bonds can be viewed as the real-world counterpart to these theoretical securities.

Proposition 3 *The price of stock j is given by*

$$S_j(t) = E_t^A \left[\int_t^T \frac{\eta^A(s)}{\eta^A(t)} D_j(s) ds \right] = D_j(t) g_j(\phi(t), X(t), \mathbf{u}(t), t). \quad (2.30)$$

Here the price-dividend ratio g_j depends on time t , on the stochastic weighting process ϕ , on the state variable X that drives time-varying systemic disaster intensity, and on the dividend share distribution through the $(N-1)$ -dimensional state variable \mathbf{u} :

$$g_j(\phi, X, \mathbf{u}, t) = e^{-\gamma \sum_{j=2}^N u_j/N} (1 + e^{u_2} + \dots + e^{u_N})^\gamma \sum_{k=0}^{\gamma} a_k(\phi) \int \mathcal{F}_\gamma^N(\mathbf{z}) e^{i\mathbf{u}'\mathbf{z}} b_{jk}(X, t, \mathbf{z}) d\mathbf{z}, \quad (2.31)$$

where the integral is evaluated on \mathbb{R}^{N-1} , $\mathcal{F}_\gamma^N(\mathbf{z})$ is given by Equation (A.24) in Appendix A.1.3 and

$$\begin{aligned} a_k(\phi) &= \binom{\gamma}{k} \frac{\phi(t)^{k/\gamma}}{(1 + \phi(t)^{1/\gamma})^\gamma}, \\ b_{jk}(X, t, \mathbf{z}) &= \int_t^T e^{(\tau-t)[-\delta + (\mu - \frac{1}{2}\sigma^2)\mathbf{1}'_N(\mathbf{e}_j - \gamma/N + i\mathbf{U}'\mathbf{z}) + \frac{1}{2}\sigma^2(\mathbf{e}_j - \gamma/N + i\mathbf{U}'\mathbf{z})'(\mathbf{e}_j - \gamma/N + i\mathbf{U}'\mathbf{z})] + \alpha_{0,k}^N(\tau-t) + \alpha_{2,k}^N(\tau-t)X(t)} d\tau \end{aligned}$$

Here \mathbf{e}_j is the N -vector with a 1 in the j th entry and 0s elsewhere, $\mathbf{1}_N$ is an N -dimensional vector of 1s and $\alpha_{0,k}^N(\tau)$ and $\alpha_{2,k}^N(\tau)$ satisfy the system of Riccati equations given in Appendix A.1.3.

Semi-closed-form expressions²⁰ for diffusion volatilities and jump sizes of stock j 's return process follow after application of Itô's lemma for jump-diffusion processes, using dividend growth, stochastic weight process, dividend shares and exogenous state variable dynamics in Equations (2.1), (2.4), (2.10) and (2.3), respectively:

$$\sigma_{S_j}(t) = \left[\sigma \left(\mathbf{e}'_j + \frac{g'_{j\mathbf{u}}}{g_j} \mathbf{U} \right) \frac{g_{jX}}{g_j} \sigma_X \sqrt{X(t)} \right], \quad k_{S_j}^c(t) = (k+1) \frac{g_j \left(\phi^{\beta^B/\beta^A}, X, \mathbf{u}, t \right)}{g_j(\phi, X, \mathbf{u}, t)} - 1. \quad (2.32)$$

The i th component of vector $\mathbf{k}_{S_j}(t)$ is given by

$$k_{S_j,i}(t) = \begin{cases} \frac{g_j(\phi, X, \mathbf{u} + \ln(k+1)\mathbf{U}\mathbf{e}_i, t)}{g_j(\phi, X, \mathbf{u}, t)} - 1 & \text{if } i \neq j, \\ (k+1) \frac{g_j(\phi, X, \mathbf{u} + \ln(k+1)\mathbf{U}\mathbf{e}_i, t)}{g_j(\phi, X, \mathbf{u}, t)} - 1 & \text{if } i = j, \end{cases} \quad (2.33)$$

where $g_{j\mathbf{u}}$ and g_{jX} are the derivatives of the price-dividend ratio g_j with respect to \mathbf{u} and X , respectively, which can also be obtained in semi-closed form. Time-varying disaster risk and disagreement endogenously generate time variation in the diffusion volatilities and jump sizes of stock returns, even if the parameters in the dividend growth processes are constant.

2.3 Results and Analysis

In this section I study the properties of the risk premia and other asset pricing implications of the model presented in Section 2.2.

²⁰Up to the solution of the ordinary differential equations for $\alpha_{0,k}(\tau)$ and $\alpha_{2,k}(\tau)$, which is easily obtained numerically after evaluating an $(N-1)$ -dimensional integral that is well-behaved but can be computationally intensive for large N .

Instead of attempting to estimate the model, I analyze its main qualitative implications and the mechanisms behind them by means of a simple numerical illustration for a symmetric economy with two stocks, $N = 2$. In the baseline calibration, dividend growth processes have a drift $\mu = 2.5\%$ and a diffusion volatility $\sigma = 5\%$.²¹ Rare events have an impact of $k = -0.41$, consistently with the estimates reported in Dieckmann and Gallmeyer (2005) and in Barro (2006). Idiosyncratic jumps have a constant intensity $\lambda = 1\%$ and systemic jumps occur with a long-term frequency $\beta^A = 1\%$, so jumps in individual dividend processes occur on average each fifty years. The optimistic agent believes that the long-term mean of the frequency of systemic disasters is smaller than does the pessimistic agent; that is, $\beta^B < \beta^A$. In most cases I fix $\beta^B = 0.01\%$ but various levels of the difference $\beta^A - \beta^B$ are also considered. The two agents have the same CRRA preferences along with a time horizon T of 50 years, a time preference rate $\delta = 4\%$, and a risk aversion parameter $\gamma = 4$. The parameters of the X process are $\varphi = 0.142$ and $\sigma_X = 0.05$. Preference parameters are taken from Chen, Joslin, and Tran (2012), while the parameters in the X process are chosen to match the properties of Chen, Joslin, and Tran (2012)'s calibrated time-varying disaster intensity. Model parameters are summarized in Table 2.1.

2.3.1 Equity and variance risk premia

From agent i 's perspective, the risk premium for any security is defined as the difference between the expected return under \mathbb{P}^i and under the risk-neutral measure \mathbb{Q} . I report risk premia relative to agent A 's beliefs, \mathbb{P}^A . Define the cum-dividend instantaneous return of stock j as

$$dR_{jt} = \frac{dS_j(t) + D_j(t)dt}{S_j(t)}.$$

The instantaneous conditional equity risk premium of the individual stock j , ERP_j , is thus

$$\begin{aligned} ERP_{jt} &= E_t^A(dR_{jt}) - E_t^Q(dR_{jt}) \\ &= \gamma\sigma^2 \left(\mathbf{e}'_j + \frac{g'_{ju}}{g_j} \mathbf{U} \right) \mathbf{s} + \sum_{i=1}^N k_{S_j,i}(t)(\lambda - \lambda_i^Q(t)) + k_{S_j}^c(t)(\lambda_c^A(t) - \lambda_c^Q(t)) \\ &= \gamma\sigma^2 \left(\mathbf{e}'_j + \frac{g'_{ju}}{g_j} \mathbf{U} \right) \mathbf{s} - \lambda \sum_{i=1}^N k_{S_j,i}(t)(JP_{it} - 1) - \lambda_c^A(t)k_{S_j}^c(t)(JP_{ct} - 1), \end{aligned} \quad (2.34)$$

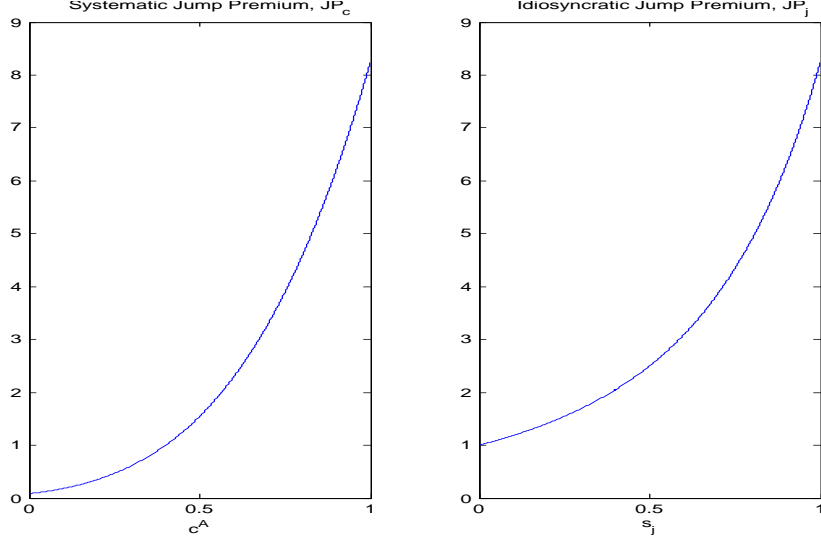
where $\mathbf{s} = (s_1, s_2, \dots, s_N)'$ is the vector of dividend shares, $JP_{it} = \lambda_i^Q(t)/\lambda$ is the jump premium related to an idiosyncratic jump in the dividend growth of asset i , and $JP_{ct} = \lambda_c^Q(t)/\lambda_c^A(t)$ is the jump premium related to a common jump. The first term in (2.34) is the compensation for diffusion risk; the other terms represent a premium for bearing idiosyncratic and systemic disaster risk, respectively.

Table 2.1: Model parameters

Preferences	$\delta = 0.04$	$\gamma = 4$	
Dividends	$\mu = 2.5\%$	$\sigma = 5\%$	$k = -0.41$
Intensities	$\beta^A = 1\%$	$\beta^B = 0.01\%$	$\lambda = 1\%$
	$\varphi = 0.142$	$\sigma_X = 0.05$	

Figure 2.1: Jump premium

The left panel plots the systemic jump premium, JP_c , as a function of agent A 's consumption. The right panel plots the idiosyncratic jump premium, JP_j , as a function of the dividend share of asset j .



As mentioned in Section 2.2.4, the jump premium for idiosyncratic event risk, $JP_{it} = (s_i k + 1)^{-\gamma}$, is always greater than 1 and it is also close to 1 for small stocks (see right panel of Figure 2.1). We can use Equation (2.28) to write the jump premium for systemic event risk as

$$\frac{\lambda_c^Q(t)}{\lambda_c^A(t)} = \left[\frac{1 + \left(\phi(t) \frac{\beta^B}{\beta^A} \right)^{1/\gamma}}{(1 + \phi(t)^{1/\gamma})} \right]^\gamma (k + 1)^{-\gamma} = (k_{CA}(t) + 1)^{-\gamma}, \quad (2.35)$$

where

$$k_{CA}(t) = \frac{(1 + \phi(t)^{1/\gamma})(k + 1)}{1 + \left(\phi(t) \frac{\beta^B}{\beta^A} \right)^{1/\gamma}} - 1$$

is the size of the jump in equilibrium consumption of agent A in response to a systemic disaster. This jump size varies depending on the level of disagreement and the consumption share distribution, due to risk sharing between agents. Since agent B (the optimist) thinks systemic disasters are highly unlikely, he is willing to give up consumption in future systemic disaster states in exchange for higher consumption in all other future states. This mechanism reduces the consumption loss of agent A in the event of a systemic disaster and lowers the corresponding jump risk premium. The more wealth the optimist has, the more disaster insurance he is able to sell. So when the wealth share of the optimist is high, consumption of agent A can even increase at a disaster. That scenario would lead to a jump premium lower than 1 (see left panel of Figure 2.1)—in other words, to a risk-neutral intensity λ_c^Q lower than the physical intensity

²¹The values for the diffusion component of the dividend dynamics, μ and σ , are within the ranges considered in the literature. See, among others, Campbell (2003) and Cochrane, Longstaff, and Santa-Clara (2008).

λ_c^A .²² A higher level of relative risk aversion γ would lead to a much faster rise in the systemic jump premium, although the qualitative implications would remain unchanged.

Figure 2.2: Equity premium

Instantaneous equity premium of stock 1, under agent A 's beliefs, as a function of the dividend share of asset 1, s_1 , when the consumption share of the pessimistic agent is $c^A = 0.1$ (left panel) or $c^A = 0.9$ (right panel).

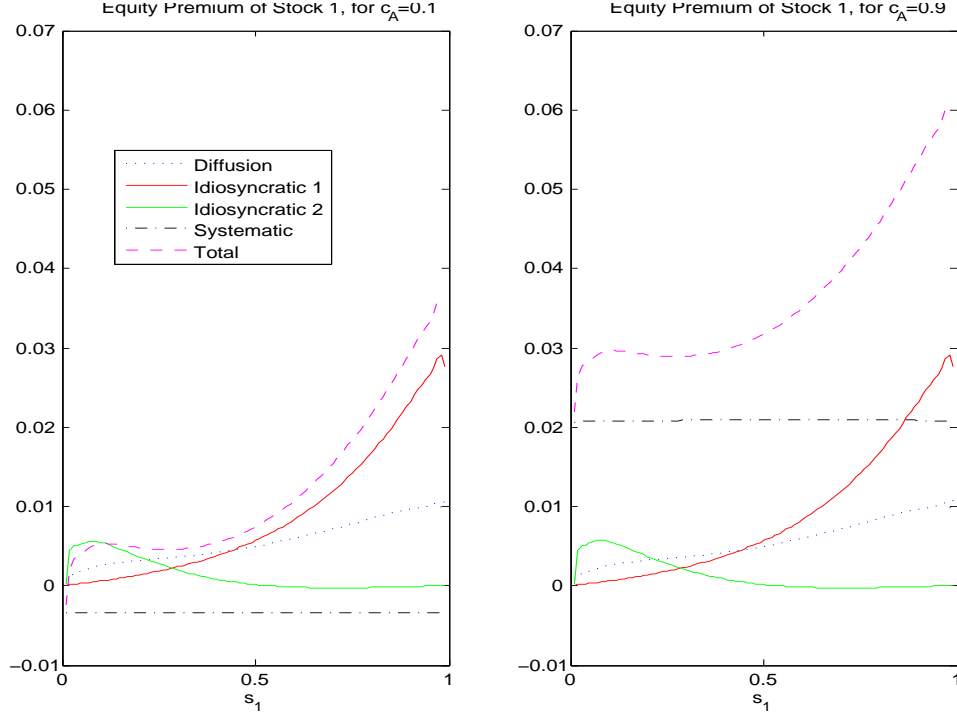


Figure 2.2 shows the conditional instantaneous equity premium of stock 1 at time $t = 0$ and its components, as a function of the dividend share s_1 , for two possible values of the initial wealth share of the pessimistic agent A ($c^A = 0.1$ in the left panel and $c^A = 0.9$ in the right panel). The equity premium is first slightly decreasing and then increasing in the dividend share of the asset; a pattern that is due to the behavior of the compensation for diffusion risk (see Martin (2013)) and to the fact that the overall premium for idiosyncratic risk is lower for intermediate values of the dividend share. Note that the compensation for diffusion and idiosyncratic rare event risk does not change with the consumption share of the two agents, since they disagree only with

²²More precisely, the systemic jump premium is less than 1 when the ratio of the consumption shares is large:

$$\frac{c^B(t)}{c^A(t)} > \frac{-k}{k + 1 - \left(\frac{\beta^B}{\beta^A}\right)^{1/\gamma}}$$

if the disagreement is large enough, that is, if

$$\frac{\beta^B}{\beta^A} < (k + 1)^\gamma.$$

In the calibration this condition is satisfied when the consumption share of the optimist, $c^B(t)$, is at least 60%.

respect to the systemic disaster intensity. The contribution to stock 1's equity premium of its own idiosyncratic jump risk starts at zero but increases substantially with its dividend share, as the asset becomes more *systemic*. The compensation due to idiosyncratic rare event risk in asset 2's dividends is small unless the second stock contributes to a large fraction of aggregate consumption. On the other hand, the component of asset 1's equity premium that is due to systemic rare event risk is basically flat with the dividend share but depends on disagreement risk and reflects risk sharing between agents. The compensation for systemic jump risk is negative for small consumption shares of the pessimist but increases rapidly, and for large values of c^A that compensation accounts for a large fraction of the individual equity premium (mainly when dividends are evenly distributed between the two stocks). This effect is primarily driven by the jump premium for systemic disasters, JP_c , in the left panel of Figure 2.1. Besides the jump risk premia, the equity premium is also a function of the jump sizes of stock returns k_s and k_s^c , which depend on the dividend loss and on changes in the price-dividend ratios, as shown in Equations (2.33) and (2.32). The left panel of Figure 2.3 plots the jump size in the return of a stock at a systemic disaster, $k_{S_j}^c$, as a function of the consumption share of the pessimistic agent, c^A , for a small stock ($s = 0.1$) and a large stock ($s = 0.9$). Under CRRA utility, the drop in the risk-free rate following a systemic disaster can dominate the effect of a rising risk premium, which would lead to a higher price-dividend ratio. That higher price-dividend ratio partially offsets the drop in dividends, making the return less sensitive to systemic disasters. The variation of systemic jump size in stock returns with the pessimist's consumption share is stronger for a small stock (blue line in the figure) and depends crucially on the assumption of difference in beliefs. In fact, if there is no disagreement then the systemic jump size $k_{S_j}^c$ is constant and equal to the loss in dividend growth k (see the black dotted line in the left panel of Figure 2.3). The right panel displays the jump size in stock 1's return at an idiosyncratic jump in its dividend growth process (blue line), $k_{S_1,1}$, and in the dividend growth process of the second asset (red line), $k_{S_1,2}$. Observe that $k_{S_1,2}$ can become slightly positive for large s_1 , which means that idiosyncratic jump risk in the dividend growth of the small stock can have a negative effect on the equity premium of the large stock. This effect can be interpreted as a flight to safety from the small to the large stock.

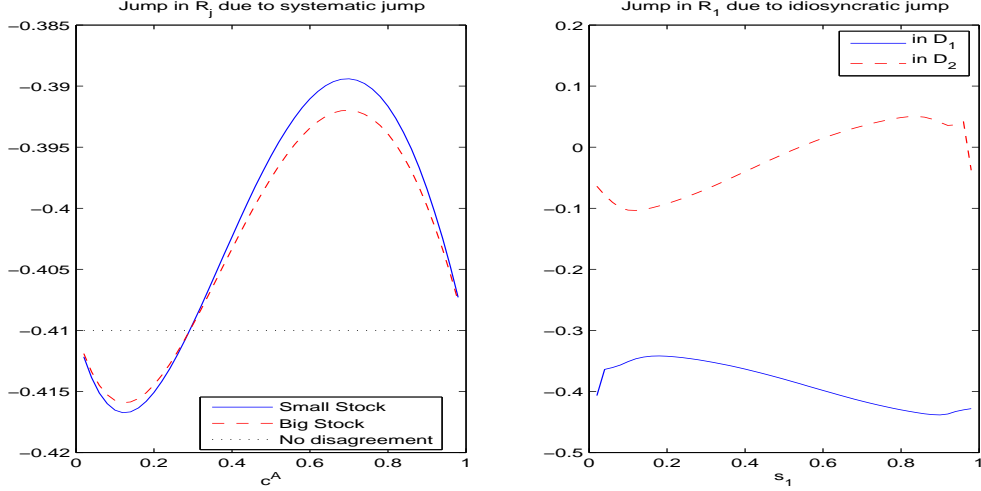
In the same way, the instantaneous variance risk premium of stock j , VRP_j , can be computed as the difference between objective and risk-neutral expectations of the return variance:

$$\begin{aligned}
VRP_{jt} &= E_t^A[(dR_{jt})^2] - E_t^Q[(dR_{jt})^2] \\
&= \sum_{i=1}^N k_{S_j,i}(t)^2(\lambda - \lambda_i^Q(t)) + k_{S_j}^c(t)^2(\lambda_c^A(t) - \lambda_c^Q(t)) \\
&= -\lambda \sum_{i=1}^N k_{S_j,i}(t)^2(JP_{it} - 1) - \lambda_c^A(t)k_{S_j}^c(t)^2(JP_{ct} - 1). \tag{2.36}
\end{aligned}$$

Given the assumption of constant dividend growth volatilities, the variance risk premium depends only on the jump risk components. Yet empirical evidence reported in [Bollerslev and Todorov \(2011\)](#) and [Ait-Sahalia, Karaman, and Mancini \(2012\)](#) shows that compensation for rare events actually accounts for a large fraction of variance risk premia. The instantaneous variance premium VRP_j is usually negative, as expected, but it can become positive when $JP_{ct} < 1$ and large enough to balance out the contribution of the idiosyncratic jump components, which is always negative (as discussed previously; see Figure 2.4). The variance risk

Figure 2.3: Jumps in stock returns

The left panel plots the jump size in the return of a stock at a systemic disaster, $k_{S_j}^c$, as a function of the consumption share of the pessimistic agent, c^A , for a small stock ($s = 0.1$) and a large stock ($s = 0.9$), as well as in the case of no disagreement. The right panel displays the jump size in stock 1's return at an idiosyncratic jump in its dividend growth process (blue line), $k_{S_1,1}$, and in the dividend growth process of the second asset (red line), $k_{S_1,2}$, respectively.



premium is negatively related to the systemic jump premium: it decreases with agent A 's consumption share, and it is either decreasing or hump-shaped with respect to a stock's dividend share (depending on the value of the calibrated parameters). As for the individual equity premium, the compensation due to idiosyncratic rare event risk in asset 2's dividends is nearly zero; however, the contribution of idiosyncratic rare event risk in its own dividend process is increasing (in absolute value) in the dividend share.

The model relates the correlation between individual variance premia to the systemic rare event risk. This systemic component is stronger when the consumption share of the pessimist is higher. In the case of a two-stocks economy, the average model-implied correlation between variance premia ranges from -0.4 (when the consumption share of agent A is 10%) to about 0.75 (when the pessimist consumes 90% of the aggregate dividend).

Now let me define the instantaneous return on the stock market index as the weighted sum of all individual asset returns:²³

$$dR_t = \sum_{j=1}^N s_j dR_{jt}. \quad (2.37)$$

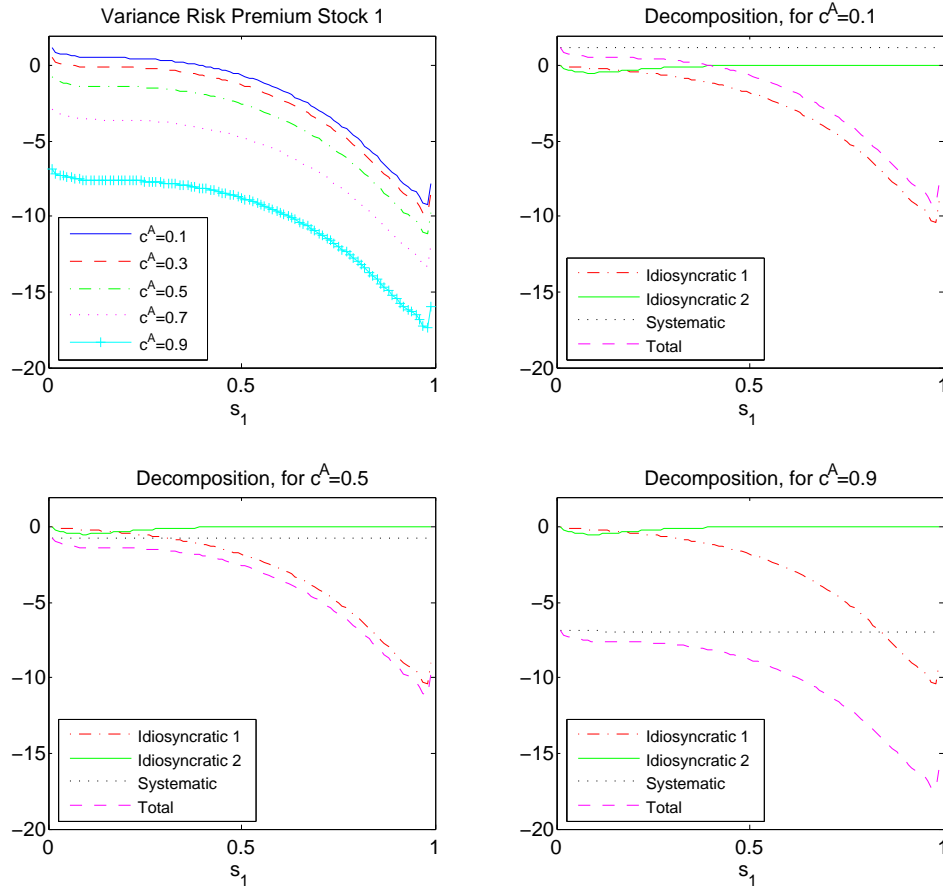
The instantaneous equity premium on the index is then

$$\begin{aligned} ERP_t &= \sum_{j=1}^N s_j ERP_{jt} \\ &= \gamma \sigma^2 \mathbf{s}' \left(\mathbf{I}_N + \frac{\mathbf{g}' \mathbf{u}}{\mathbf{g}} \mathbf{U} \right) \mathbf{s} + \mathbf{s}' \mathbf{k}_s(t) (\lambda - \boldsymbol{\lambda}^Q(t)) + \mathbf{s}' \mathbf{k}_s^c(t) (\lambda_c^A(t) - \lambda_c^Q(t)), \end{aligned} \quad (2.38)$$

²³The stock market index can also be viewed as a claim on the aggregate endowment $C = D_1 + \dots + D_N$.

Figure 2.4: Stock variance risk premium

Instantaneous variance risk premium of stock 1, under agent A 's beliefs, in monthly squared percentage, as a function of the dividend share of asset 1, s_1 , for different values of the consumption share of the pessimistic agent is c^A . The second, third and fourth panels show the decomposition of the individual variance risk premium in its idiosyncratic and systemic jump components when the consumption share of the pessimistic agent is $c^A = 0.1$, $c^A = 0.5$ and $c^A = 0.9$, respectively.



where \mathbf{I}_N is the identity matrix of dimension N , $\boldsymbol{\lambda}^Q = (\lambda_1^Q, \lambda_2^Q, \dots, \lambda_N^Q)'$, $\frac{\mathbf{g}'_{\mathbf{u}}}{\mathbf{g}} = \left(\frac{g_{1\mathbf{u}}}{g_1}, \frac{g_{2\mathbf{u}}}{g_2}, \dots, \frac{g_{N\mathbf{u}}}{g_N} \right)'$, and the instantaneous index variance risk premium is given by

$$\begin{aligned} VRP_t &= \sum_{j=1}^N s_j^2 VRP_{jt} + \sum_{j=1}^N \sum_{i \neq j} s_j s_i CRP_{jit} \\ &= \mathbf{s}' [\mathbf{k}_s(t) \text{diag}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^Q(t)) \mathbf{k}_s(t)' + \mathbf{k}_s^c(t) \mathbf{k}_s^c(t)' (\boldsymbol{\lambda}_c^A(t) - \boldsymbol{\lambda}_c^Q(t))] \mathbf{s}. \end{aligned} \quad (2.39)$$

Here $CRP_{jit} = E_t^A[dR_{jt}dR_{it}] - E_t^Q[dR_{jt}dR_{it}]$ is the premium associated with the covariance between returns of assets j and i , and $\text{diag}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^Q(t))$ is an $N \times N$ matrix with the elements of the N -vector $\boldsymbol{\lambda} - \boldsymbol{\lambda}^Q(t)$ on the diagonal and with 0s elsewhere.

Figure 2.5 plots the instantaneous equity (upper panels) and variance (lower panels) risk premium of the market, under agent A 's beliefs, as a function of the dividend share of asset 1, s_1 , and their decomposition in terms of individual equity and variance premia for different values of the consumption share of the pessimistic agent. The market equity premium increases with the consumption share of the pessimist, and it is lower when the two assets contribute in the same way to the aggregate dividend because the equity premium of individual stocks grows more than linearly with the dividend share. The same reasoning holds for the absolute value of the aggregate variance premium—which includes, however, an additional component reflecting the priced covariance between stock returns. The covariance premium can contribute to a large portion of the aggregate variance premium when the economy is dominated by the pessimistic agent, mostly when the number of assets increases and they are relatively evenly distributed.

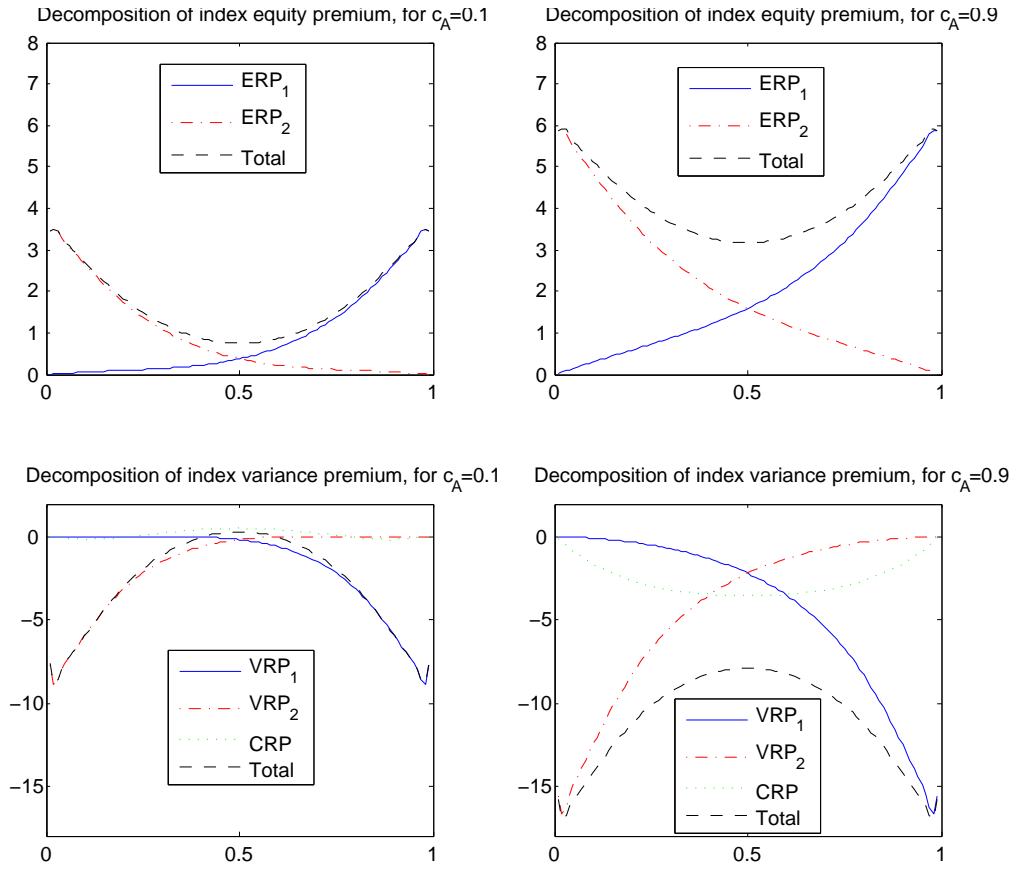
Apart from the aggregate variance premium's dependence on the relative dividend and consumption shares, its dynamic properties are worth examining. I simulate 30-year paths of the variance risk premium at a monthly frequency from the model while using calibrated parameters for different values of the initial wealth share of the pessimistic agent, c^A . Table 2.2 shows that, as the consumption share of the pessimist increases, the VRP is both larger (in absolute value) and more volatile. A systemic disaster induces an upward jump in the consumption share of the pessimist. That leads to a downward jump in the variance risk premium, which is then followed by more negative and volatile premia. Despite the setting's simplicity, the dynamics of model-implied premia resembles the behavior of observed variance risk premia (see Section 2.4 and Figure 2.12), in which periods of low and smooth premia seem to be followed by larger and more volatile values. Empirically a regime switch often corresponds to the beginning of a crisis, so it could be linked to a systemic jump in the endowment process.

Table 2.2: Simulated market variance risk premium

c^A	0.1	0.3	0.5	0.7	0.9
mean VRP	−4.70	−5.48	−6.76	−8.85	−12.74
	(1.13)	(1.33)	(1.50)	(1.52)	(1.33)
std VRP	0.55	0.66	0.82	0.99	0.95
	(0.27)	(0.47)	(0.65)	(0.71)	(0.42)

Figure 2.5: Equity and variance risk premium for the market

Instantaneous equity (upper panels, in percentage) and variance (lower panels, in monthly squared percentage) risk premium of the market, under agent A 's beliefs, as a function of the dividend share of asset 1, s_1 , and their decomposition in terms of individual equity and variance premia, for different values of the consumption share of the pessimistic agent is c^A . The first and third panels use $c^A = 0.1$, while the second and the fourth are for $c^A = 0.9$.



2.3.2 Stock return correlation and correlation risk premium

From Equation (2.21), the instantaneous conditional correlation between returns of stock i and stock j is given by

$$\text{Corr}_t^A(dR_{it}, dR_{jt}) = \frac{\sigma_{S_i}(t)\sigma'_{S_j}(t) + \mathbf{k}_{S_i}(t)\mathbf{k}'_{S_j}(t)\lambda + k_{S_i}^c(t)k_{S_j}^c(t)\lambda_c^A}{\sqrt{(\sigma_{S_i}(t)\sigma'_{S_i}(t) + \mathbf{k}_{S_i}(t)\mathbf{k}'_{S_i}(t)\lambda + k_{S_i}^c(t)^2\lambda_c^A)(\sigma_{S_j}(t)\sigma'_{S_j}(t) + \mathbf{k}_{S_j}(t)\mathbf{k}'_{S_j}(t)\lambda + k_{S_j}^c(t)^2\lambda_c^A)}}. \quad (2.40)$$

The first panel in Figure 2.6 shows the conditional stock return correlation in a symmetric economy with $N = 2$ stocks as a function of the first tree's dividend share s_1 and the pessimistic investor's consumption share c^A while using the model parameters in Table 2.1. The other panels in Figure 2.6 display the same correlation for special cases of the model. The second panel considers the case of no disagreement ($\beta^A = \beta^B = 0.01$), in the fourth panel I assume there are no idiosyncratic disasters ($\lambda = 0$), and the third panel combines these last two cases. Comparing the first and second (or the third and fourth) panels reveals that disagreement reduces stock return correlation on average and in particular when risk sharing is stronger—that is, when the consumption shares of the two agents are similar. The possibility of idiosyncratic disasters in the dividend growth processes also reduces the average correlation (compare the first and third panels of Figure 2.6), albeit mainly when dividend shares are relatively evenly distributed. Overall, however, the correlation under the pessimistic agent's objective measure is relatively flat: it has values between 35% and 43% and an average across all states of about 39%.

The correlation risk premium is defined as the difference between the instantaneous conditional correlation computed under the physical and the risk-neutral measure,

$$\text{CorrRP}_{ij,t} = \text{Corr}_t^A(dR_{it}, dR_{jt}) - \text{Corr}_t^Q(dR_{it}, dR_{jt}). \quad (2.41)$$

Here the risk-neutral correlation is computed as in (2.40) but using the risk-neutral idiosyncratic and systemic rare event intensities λ^Q and λ_c^Q , respectively; see Figure 2.7. The model-implied risk-neutral correlation (first panel) increases substantially with the consumption share of the pessimistic agent and ranges approximately between 11% and 75%. This result is consistent with the empirical findings of Driessen, Maenhout, and Vilkov (2012), who show that the implied correlation for the S&P500 is highly countercyclical and fluctuates between 0.2 and 0.8 for the period 1996–2010. The other panels in Figure 2.7 show that the dynamics of the risk-neutral correlation is almost entirely driven by disagreement between agents about the probability of a systemic disaster. The average risk-neutral correlation for the full model is about 46%, which corresponds to an average instantaneous correlation risk premium of about -7% ; this value, too, is consistent with the empirical findings reported by Driessen, Maenhout, and Vilkov (2012). However, the model-implied correlation premium (see Figure 2.8) can be much larger in absolute value when the pessimist accounts for a large part of the aggregate consumption, and it can also become positive when the pessimist's consumption share is relatively low—mainly when the dividend shares of the two assets are similar.

Figure 2.6: Conditional stock return correlation in an economy with $N = 2$ assets

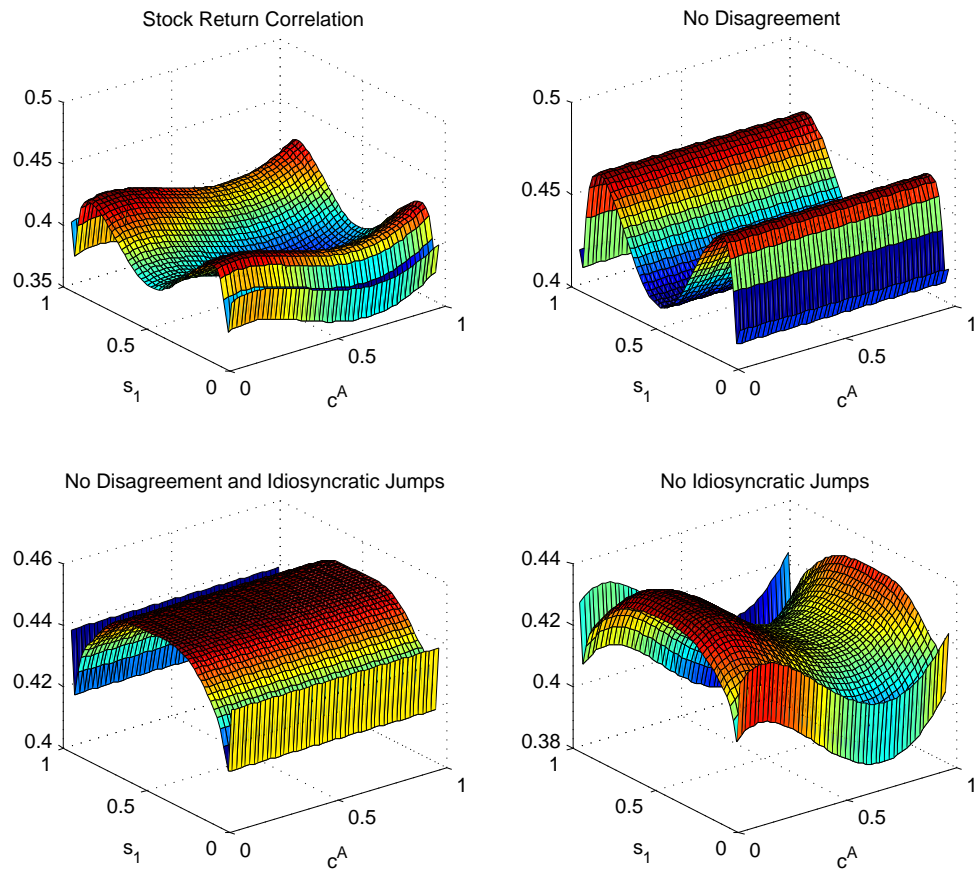


Figure 2.7: Conditional stock return correlation under the risk-neutral measure in an economy with $N = 2$ assets

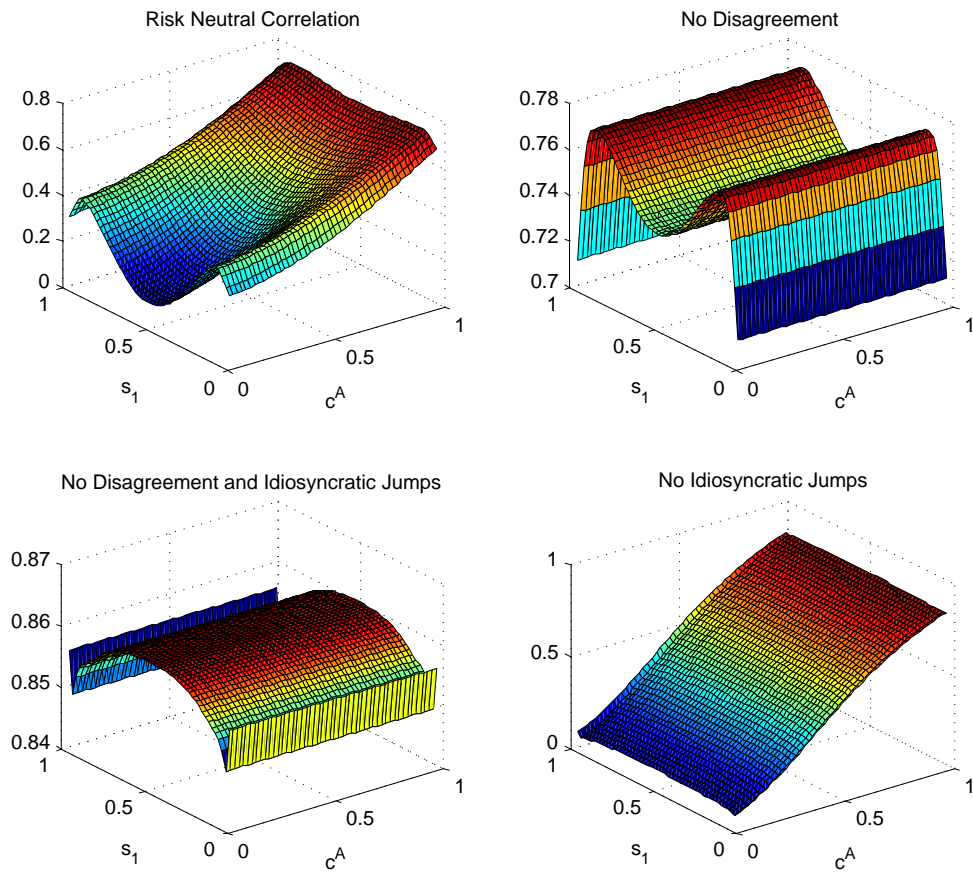
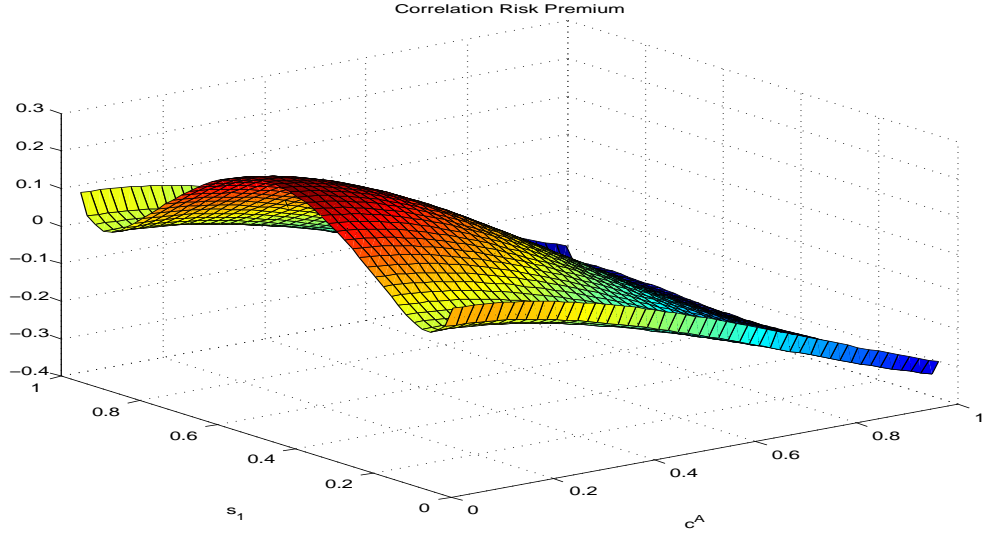


Figure 2.8: Instantaneous correlation risk premium, in an economy with $N = 2$ assets

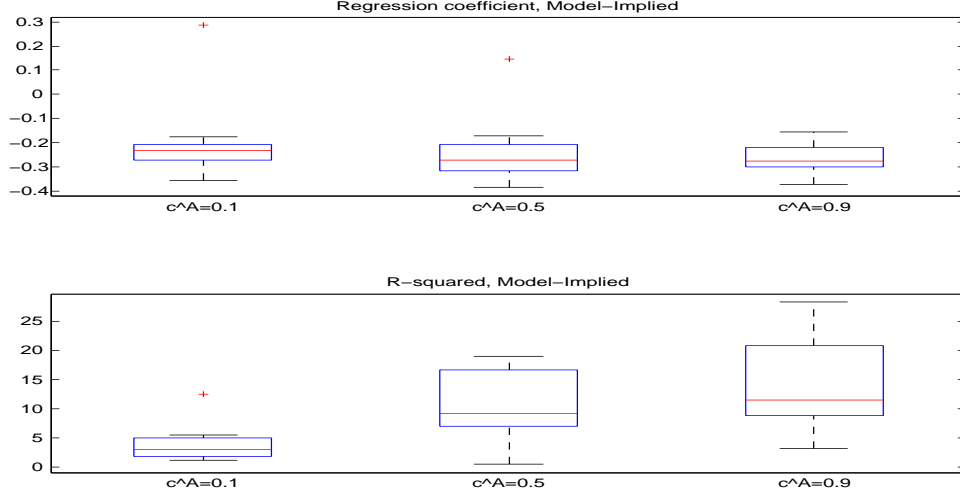


2.3.3 Relation between the equity and the variance risk premium

Comparing the expressions for the variance premium (Equations (2.36) and (2.39)) with those for the equity premium (Equations (2.34) and (2.38)) shows that rare event risk implies a tight link between the two, both for the market and for the cross section of stock returns. This link provides our basic intuition for the role of the variance premium in predicting future excess returns, which is consistent with Bollerslev, Tauchen, and Zhou (2009)'s empirical finding that aggregate variance risk premium can explain a nontrivial fraction of the time-series variation in post-1990 aggregate stock market returns. Premia that are high (in absolute value) predict high future returns—though mainly over short horizons, when the compensation for rare events accounts for a large portion of the empirical equity and variance risk premia (see e.g. Bollerslev and Todorov (2011) and Ait-Sahalia, Karaman, and Mancini (2012)). Yet standard predictive regressions imply an unconditionally linear relation between equity and variance risk premia, whereas the model's relation is conditional on the information set at time t . The idiosyncratic and systemic event risk components of the equity and variance risk premium are linearly related, but the regression coefficients are stochastic and given by the inverse of the corresponding jump size. Depending on which of the jump risk components dominates, the relation can be either weaker or stronger. The importance of the idiosyncratic and systemic rare event risk contribution to the risk premia, as well as the jump sizes in stock returns, are both functions of the asset's dividend share and the agents' consumption share. Thus, time variation in share distributions leads to a time-varying relation between equity and variance risk premia, both at an aggregate level and for individual stocks. For individual stocks, however, empirical estimates of the variance premium are noisy owing to lack of reliable high-frequency data for computing the realized variance. Moreover, there is evidence of a large systematic component in the cross section of variance risk premia (see e.g. Carr and Wu (2009)). Hence this paper also explores the relation between the instantaneous equity premium of individual stocks and the market's variance risk premium.

Figure 2.9: Simulated predictive regressions for different levels of the consumption share

Standard OLS regression of simulated excess market returns at the six-month horizon on the simulated lagged instantaneous variance risk premium, for different levels of the initial share of consumption of the pessimistic agent, c^A . Upper panel display the distribution of simulated regression coefficients and lower panel of percentage R^2 .



To develop a better understanding of the model implications that concern the predictive power of aggregate variance premium for market and individual stock excess returns, I run regressions on simulated data. This involves simulating 30 years of monthly excess stock and market returns and the instantaneous variance risk premium from the model in Section 2.2 while assuming a symmetric economy with $N = 2$ stocks and using the baseline model parameters. The purpose of these simulations is to investigate the model's qualitative implications for the interaction between the aggregate variance premium and the excess stock and market returns. This is a natural step between the model and the empirical evidence presented in Section 2.4. In addition to the monthly return horizon, I also consider multi-period return regressions of the form

$$r_{i,t+h}^e = \alpha_i + \beta_i VRP_t + \varepsilon_{i,t+h},$$

where $r_{i,t+h}^e$ is the simulated log excess return of the two stocks ($i = 1, 2$) or of the market ($i = M$) and where h is the return horizon in months. The excess return is given as an annualized percentage and the variance premium is given as a monthly squared percentage for consistency with the literature (see e.g. Bollerslev, Tauchen, and Zhou (2009) and Drechsler (2013)). Table 2.3 reports the average regression coefficient and adjusted R^2 (with standard errors in parentheses) for the market at horizons $h = 1, 6$, and 12 months and for different values of the initial consumption share of the pessimistic agent, $c^A = 0.1, 0.5$ and 0.9. The predictive coefficient is generally negative. The predictive power increases with the horizon and with the initial wealth share of the pessimistic agent, which is also associated with larger (absolute) values of the variance risk premium and of its volatility (again, see Table 2.2). Note that the average estimated regression coefficient is quite close to what is found in the data (see Section 2.4), even if the model is not estimated or calibrated to match the observed VRP moments. Figure 2.9 presents box plots of the predictive regression coefficients (upper panel) and of the

adjusted R^2 (lower panel) at the 6-month horizon. The regression coefficient is significantly different from zero only when the pessimist holds a large fraction of the aggregate endowment.

Turning now to the cross section, the first two panels of Table 2.4 display results of the same predictive regressions for the two individual stocks in the economy. Initially, the small stock (Panel A) has dividend share $s_1 = 0.1$ and the big stock (Panel B) has a share $s_2 = 0.9$. The regression coefficient for the small stock is often positive and not significant. The reason is that, with only two stocks in the economy, the fear of idiosyncratic disasters in the dividend growth of the large stock has a strong effect on the equity premium of the small stock; the corresponding jump size in the small stock return, $k_{S_1,2}(t)$, can be positive and thereby lead to a weak positive relation between the equity premium of the small stock and the market variance risk premium. This effect holds also after eliminating idiosyncratic jump risk ($\lambda = 0$) for small values of the consumption share of the pessimist, because in that case the variance risk premium of the small stock is negatively correlated with the market variance premium. In contrast, the predictive regression results for the large stock (Panel B) are in line with those discussed for the market return regression, since the big stock contributes to a large fraction of the aggregate dividend.

An economy consisting only of two stocks, one of which accounts for 90% of aggregate consumption, is clearly not realistic. It would be interesting to run cross-sectional predictive regressions for an economy with many assets and relatively small values of the dividend share, since these features better characterize real-world markets. However, it is not computationally feasible to simulate the model for large N because the solution would require numerical evaluation of a high-dimensional integral at each time step. However, Section 2.3.4 investigates theoretically the special case of a large and diversified economy and demonstrates that, as N increases, the idiosyncratic jump premium contribution to the both equity and variance risk premia vanishes and the aggregate variance risk premium becomes due almost entirely to a covariance premium. So in order to mimic the case of a large economy without the need to simulate it, I look at the predictive power of the simulated covariance risk premium for the excess return of individual stocks. Panels C and D of Table 2.4 display results of the regression

$$r_{i,t+h}^e = \alpha_i + \beta_i CRP_t + \varepsilon_{t+h},$$

where $r_{i,t+h}^e$ is the simulated log excess return of the small and the big stock and CRP is the covariance risk premium (in monthly squared percentage). For the small asset (Panel C), the predictive coefficient is negative and significantly different from zero. Its average value is similar across horizons and consumption shares of the pessimist, but the standard deviation of

Table 2.3: Simulated market return predictive regressions

Market return predictability by variance risk premium (VRP), from simulated monthly data, at horizons $h = 1, 6$ and 12 months, for different values of the initial consumption share of the pessimistic agent, $c^A = 0.1, 0.5$, and 0.9 . The table shows the average of the regression coefficient and adjusted R^2 over all simulations, with standard errors in parenthesis. Returns are in annualized percentage while VRP is in monthly squared percentage.

Horizon (months)	$c^A = 0.1$			$c^A = 0.5$			$c^A = 0.9$		
	1	6	12	1	6	12	1	6	12
VRP Coeff	-0.51 (0.19)	-0.20 (0.18)	-0.19 (0.17)	-0.38 (0.12)	-0.24 (0.15)	-0.24 (0.14)	-0.35 (0.08)	-0.26 (0.06)	-0.25 (0.07)
Adj R^2 (%)	2.15 (0.98)	4.04 (3.29)	7.56 (6.29)	3.15 (1.54)	10.12 (6.06)	18.34 (10.94)	4.26 (1.80)	14.32 (7.69)	23.97 (12.97)

Table 2.4: Simulated stock return predictive regressions

Predictability of excess returns of small (Panels A and C) and big (Panels B and D) stock by variance risk premium (*VRP*, Panels A and B) and by covariance risk premium (*CRP*, Panels C and D), from simulated monthly data, at horizons $h = 1, 6$, and 12 months, for different values of the initial consumption share of the pessimistic agent, $c^A = 0.1, 0.5$, and 0.9. The small (big) stock has an initial dividend share of $s = 0.1$ ($s = 0.9$). The table shows the average of the regression coefficient and adjusted R^2 over all simulations, with standard errors in parenthesis. Returns are in annualized percentage while *VRP* and *CRP* are in monthly squared percentage.

Panel A: Regression of small stock returns on <i>VRP</i>									
Horizon (months)	$c^A = 0.1$			$c^A = 0.5$			$c^A = 0.9$		
	1	6	12	1	6	12	1	6	12
<i>VRP</i> Coeff	0.41 (0.38)	0.04 (0.26)	-0.02 (0.25)	0.25 (0.34)	0.06 (0.24)	0.02 (0.22)	0.02 (0.21)	-0.07 (0.23)	-0.08 (0.23)
Adj R^2 (%)	1.57 (1.58)	4.02 (5.75)	7.13 (9.43)	1.30 (2.04)	4.13 (6.62)	6.67 (10.01)	0.72 (1.32)	5.86 (7.63)	9.61 (12.84)
Panel B: Regression of big stock returns on <i>VRP</i>									
Horizon (months)	$c^A = 0.1$			$c^A = 0.5$			$c^A = 0.9$		
	1	6	12	1	6	12	1	6	12
<i>VRP</i> Coeff	-0.54 (0.24)	-0.15 (0.20)	-0.14 (0.19)	-0.42 (0.14)	-0.24 (0.15)	-0.23 (0.14)	-0.37 (0.09)	-0.26 (0.07)	-0.25 (0.07)
Adj R^2 (%)	1.91 (1.06)	2.43 (3.62)	4.53 (5.02)	3.00 (1.48)	8.45 (5.77)	15.27 (10.50)	3.96 (1.74)	12.46 (7.24)	21.27 (12.48)
Panel C: Regression of small stock returns on <i>CRP</i>									
Horizon (months)	$c^A = 0.1$			$c^A = 0.5$			$c^A = 0.9$		
	1	6	12	1	6	12	1	6	12
<i>CRP</i> Coeff	-0.38 (0.11)	-0.37 (0.10)	-0.36 (0.11)	-0.31 (0.13)	-0.32 (0.13)	-0.32 (0.13)	-0.32 (0.06)	-0.30 (0.07)	-0.28 (0.08)
Adj R^2 (%)	1.95 (0.96)	11.37 (5.05)	19.09 (8.16)	1.58 (1.50)	9.78 (7.27)	16.28 (11.43)	4.14 (2.79)	18.84 (11.05)	28.22 (15.34)
Panel D: Regression of big stock returns on <i>CRP</i>									
Horizon (months)	$c^A = 0.1$			$c^A = 0.5$			$c^A = 0.9$		
	1	6	12	1	6	12	1	6	12
<i>CRP</i> Coeff	0.14 (0.15)	0.14 (0.15)	0.13 (0.16)	-0.07 (0.23)	-0.04 (0.23)	-0.04 (0.25)	-0.18 (0.19)	-0.16 (0.15)	-0.15 (0.14)
Adj R^2 (%)	0.36 (0.82)	3.43 (4.15)	7.41 (8.15)	0.51 (1.03)	4.03 (5.67)	8.57 (9.99)	1.19 (1.39)	6.04 (6.05)	10.78 (10.90)

the regression coefficient decreases with c^A and so leads to high adjusted R^2 when the pessimist accounts for a large share of aggregate consumption. On average, the R^2 values are even higher than those reported in Table 2.3 for the aggregate market. At the 6-month horizon, for example, the average adjusted R^2 for the regression of excess small asset returns on the instantaneous covariance premium is almost 19%, as compared with a 14% R^2 for the regression of market excess returns on the variance premium. For the large asset (Panel D), results are much weaker. Regression coefficients are even positive for small values of the pessimist's consumption share yet become negative (but only marginally significant) for large c^A . Figure 2.9 presents box plots of the predictive regression coefficients (upper panel) and of the adjusted R^2 (lower panel) obtained by regressing simulated 6-month excess returns of a small stock (starting from $s = 0.1$; blue box plots) and a big stock (starting from $s = 0.9$; red box plots) on the simulated lagged instantaneous covariance risk premium for different levels of the initial share of consumption of the pessimistic agent, c^A . These results indicate that, in a relatively large economy, the forecasting power of the aggregate variance risk premium for future excess returns should be stronger for small stocks because their returns are more dependent on the compensation for systemic rare event risk (though mainly for large values of the consumption share of the pessimistic agent). This model prediction is tested empirically in Section 2.4.3.

2.3.4 The case of a large economy: $N \rightarrow \infty$

Let me now consider analytically the case in which the number of assets in the economy, N , approaches infinity and dividends are evenly distributed across assets; that is, $s_j = 1/N$ for $j = 1, \dots, N$. In this case, the premium for idiosyncratic risk in individual assets vanishes because $JP_{jt} = (k/N + 1)^{-\gamma}$ converges to unity for all j . Moreover, the diffusion component in the equity premium for stock j reduces to $\gamma\sigma^2/N$, which tends to zero as the number of assets N increases. Hence the expressions for the equity risk premium of stock j and of the index in Equations (2.34) and (2.38) can be simplified as follows:

$$ERP_{jt} = -\lambda_c^A(t)k_{S_j}^c(t)(JP_{ct} - 1), \quad (2.42)$$

$$ERP_t = \gamma\sigma^2 - \lambda_c^A(t)k_S^c(t)(JP_{ct} - 1). \quad (2.43)$$

The equity premium is the same for any stock because in this special case, $k_S^c(t) \equiv k_{S_j}^c(t) = k_{S_i}^c(t)$ for all i and j . Furthermore, the market equity premium is equal to the equity premium of single stocks plus the standard (constant) compensation for diffusion risk, $\gamma\sigma^2$, that arises in economies where dividend growth follows a geometric Brownian motion. Note that even if stocks are negligibly small, they still earn a risk premium due to the presence of the systemic jump component, which does not depend on the dividend share (see the black dashed-dotted lines in Figure 2.2).

Similarly, the variance risk premium is the same for any individual stock,

$$VRP_{jt} = -\lambda_c^A(t)k_S^c(t)^2(JP_{ct} - 1), \quad (2.44)$$

and is equal to the premium for the covariance between stock i and stock j , $CRP_{jit} = VRP_{jt}$. Then by way of Equation (2.39), the variance premium for the market index becomes

$$\begin{aligned} VRP_t &= \sum_{j=1}^N s_j^2 VRP_{jt} + \sum_{j=1}^N \sum_{i \neq j} s_j s_i CRP_{jit} \\ &= \frac{VRP_{jt}}{N} + \frac{N-1}{N} CRP_{jit} \\ &= -\lambda_c^A(t)k_S^c(t)^2(JP_{ct} - 1), \end{aligned} \quad (2.45)$$

which is equal to the variance risk premium of any individual stock and also to the covariance premium. In particular, from the second line of Equation (2.45) it is evident that, as N increases, all the market variance risk premium is due to a premium for covariance. In accordance with this model-implied feature, Driessen, Maenhout, and Vilkov (2012) show empirically that the variance risk premium for the *S&P500* index can be largely attributed to the high price of correlation risk.

To clarify premia behavior in this special case as a function of the consumption share distribution, Figure 2.10 shows the equity premium of stock j and of the index, the systemic jump premium, and the index variance risk premium as functions of the consumption share of the pessimistic agent, c^A .²⁴ In this case the link between variance risk premia and excess stock returns, both for the index and for single stocks, is straightforward:

$$ERP_t = \gamma\sigma^2 + \frac{1}{k_s^c(t)} VRP_t, \quad (2.46)$$

$$ERP_{jt} = \frac{1}{k_s^c(t)} VRP_t. \quad (2.47)$$

and it is linear conditionally on the information set at time t . In particular, the regression coefficient $1/k_s^c(t)$ depends only on the consumption share of the two agents (as shown in Figure 2.11). This relation is negative and stronger for large values of the consumption share of the pessimist; the maximum is around $c^A = 0.7$, above which the relation becomes weaker for extreme values of c^A .

For a large and diversified index such as the *S&P500*, the model thus suggests a stronger predictive power of variance risk premium for future excess returns in periods during which pessimists have a relatively large consumption share—that is, in bad states of the economy, which are also generally linked to higher (absolute) values of the variance risk premium. I investigate this intuition empirically in Section 2.4.2.

2.3.5 Consumption share dynamics and survival

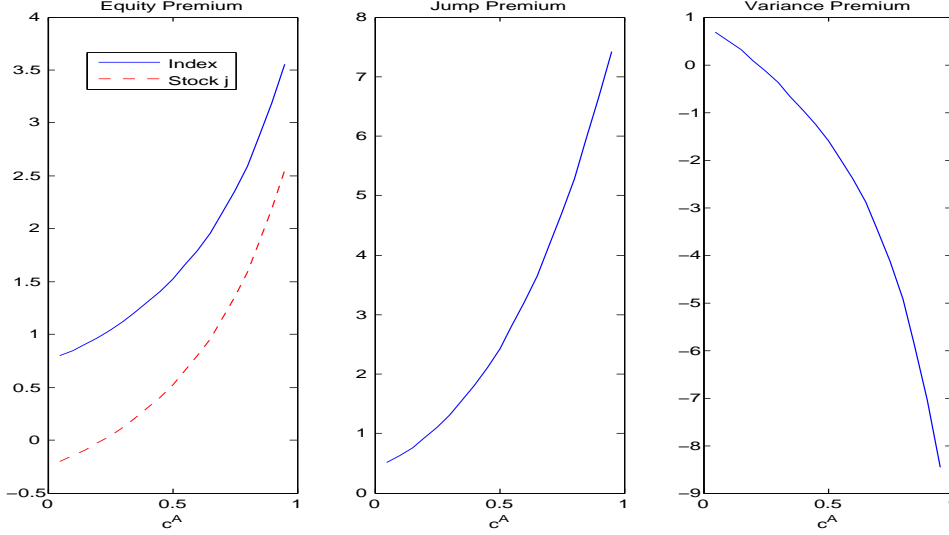
Agent survival is an important issue in complete markets models with heterogeneous beliefs and time-separable preferences (see e.g. Yan (2008) and Kogan, Ross, Wang, and Westerfield (2006)). Under most models in which agents have identical CRRA preferences, only those agents whose beliefs are closest to the truth will survive in the long run. If the irrational agent (optimist in the foregoing analysis) is quickly eliminated from the economy then the price effects generated by trading between agents disappear. It is therefore worth analyzing the survival of agents A and B , which is defined as their asymptotic share of consumption as the horizon goes to infinity (see e.g. Berrada (2009) and Dumas, Kurshev, and Uppal (2009)).

Table 2.5 shows the mean and standard deviation of the share of consumption of the optimistic agent, c^B , at horizon $T = 50, 100$, and 500 years; these values are obtained from 1,000 simulations starting from $c^B = 0.1, 0.5$, and 0.9. I find that the optimist can survive for long

²⁴Numerical results are obtained for the parameters in Table 2.1 and $N = 10$, which is not that large but already entails solving a 9-dimensional integral—even though in the special case of equal dividend shares, the expression for the price-dividend ratio is simpler than in the general case in Section 2.2 (see Appendix A.1.3.3). Nonetheless, already for $N = 8$ or 9 the results are nearly identical.

Figure 2.10: Risk premia in a large economy

Instantaneous equity premium (annualized and in percentage, first panel), systemic jump premium (second panel) and index variance risk premium (in monthly terms and squared percentage, third panel), under agent A 's beliefs, in the case of a large symmetric economy, as a function of the consumption share of the pessimistic agent, c^A .



periods and that his consumption share actually increases if there are no systemic disasters. Therefore, the risk-sharing dynamics documented previously are not likely to disappear quickly.

The dynamics of the consumption share of the pessimistic agent is obtained by applying Itô's lemma to $c^A = (1 + \phi(t)^{1/\gamma})^{-1}$ and using Equation (2.4):

$$dc^A(t) = -\frac{1}{\gamma}c^A(t)c^B(t)(\beta^A - \beta^B)X(t)dt + c^A(t)c^B(t)\frac{1 - \left(\frac{\beta^B}{\beta^A}\right)^{1/\gamma}}{c^A(t) + c^B(t)\left(\frac{\beta^B}{\beta^A}\right)^{1/\gamma}}dN_{ct}. \quad (2.48)$$

The drift is negative; thus c^A declines deterministically when there is no systemic disaster but increases in response to systemic disaster, and both effects are stronger as the level of disagreement increases.

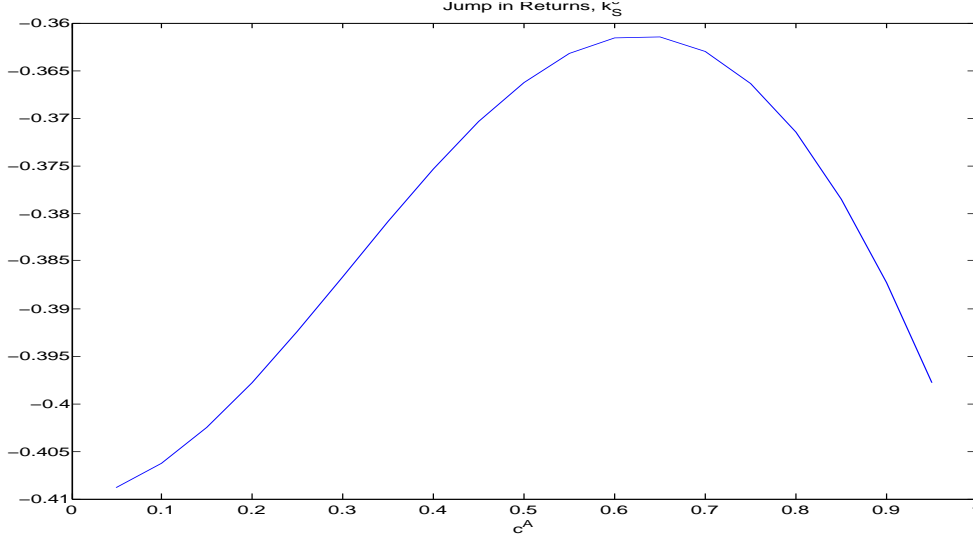
Table 2.5: Survival

This table displays the share of consumption of the optimistic agent, c^B , at horizon $T = 50, 100$, and 500 years, obtained from 1,000 simulations starting from $c^B = 0.1, 0.5$, and 0.9.

c^B	$T = 50$	$T = 100$	$T = 500$
0.1	0.08 (0.03)	0.07 (0.04)	0.02 (0.04)
0.5	0.44 (0.12)	0.38 (0.15)	0.14 (0.16)
0.9	0.86 (0.08)	0.81 (0.13)	0.44 (0.28)

Figure 2.11: Jump in stock returns in a large economy

Size of jumps in stock returns due to a systemic disaster, k_s^c , under agent A 's beliefs, in the case of a large symmetric economy, as a function of the consumption share of the pessimistic agent, c^A .



However, the consumption share's distribution is not stationary, which means that at infinite horizon one of the agents eventually disappears. Such nonstationarity could potentially be an issue in light of an estimation of the model. Possible solutions are provided by [Borovicka \(2012\)](#), who shows that recursive preference specifications lead to equilibria in which both agents survive, and by [Garleanu and Panageas \(2014\)](#), who propose an overlapping-generations framework to obtain a nondegenerate stationary equilibrium. I leave extensions of the model in these directions to future research.

2.4 Empirical Analysis

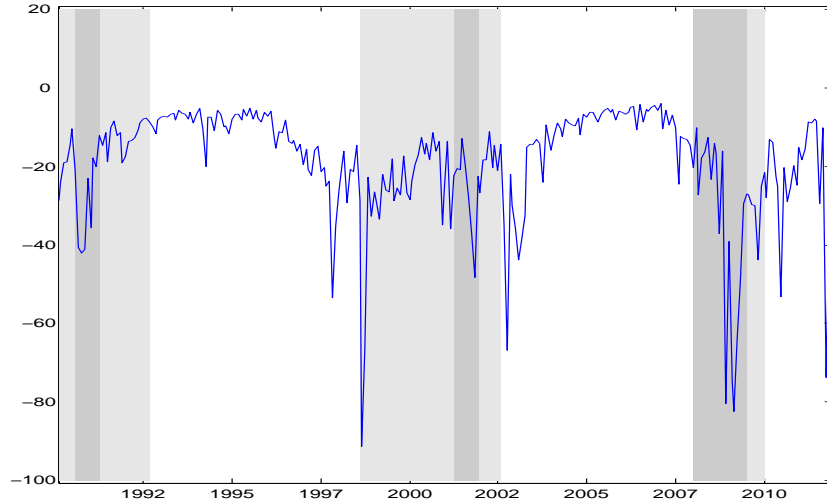
This section briefly introduces the data before testing empirically the model's main implications. In particular, I first analyze the link between equity and variance risk premia at an aggregate level via predictive regressions of market excess returns on the variance risk premium. Second, I investigate cross-sectional variations in the forecasting power of the aggregate variance premium.

2.4.1 Data

The empirical analysis is based on the aggregate *S&P500* composite index (a proxy for the aggregate market portfolio) and on returns for each constituent of the *S&P500* and for CRSP cap-based portfolios (to analyze cross-sectional implications and the differential effects of small versus big stocks). I use monthly data from January 1990 through December 2011 for a total of 264 monthly observations. Excess returns are constructed by subtracting the log 30-day T-bill yield to the monthly returns, all obtained from CRSP.

Figure 2.12: Observed variance risk premium

Time series of variance risk premium, in monthly squared percentage, where the physical expectation of the realized variance is computed from a projection of realized variance on the value of the lagged squared VIX and on lagged realized variance. Light gray shaded areas denote phases in which difference in beliefs, measured based on the dispersion of one-year-ahead forecasts on real GDP growth from the BlueChip Economic Indicator, is above average. Dark gray shaded areas denote NBER recessions.



The variance risk premium for any asset is defined (see also Section 2.3.1) as the difference between physical and risk-neutral expectations of total return variance for a given horizon. The Volatility Index (VIX), from the Chicago Board Options Exchange (CBOE) provides a model-free measure of the risk-neutral expectation of total market return variation over the subsequent 30 days and is based on the highly liquid *S&P500* index options.²⁵ The VIX is reported in terms of annualized percentage volatility; however, for consistency with the recent literature on variance risk premia in the stock market,²⁶ my measure of implied variance (*IV*) is given by VIX squared and then divided by 12 to obtain a monthly quantity. In order to measure empirically the market variance risk premium (*VRP*) one also needs a conditional forecast of total return variation under the physical measure. A measure of the realized variance (*RV*) of the market for a given month can be obtained by summing up *S&P500* squared five-minute log returns.²⁷ Bollerslev, Tauchen, and Zhou (2009) use the average *RV* over the previous month to approximate the physical expectation of the total return variation. The persistence of volatility renders this approximation a fairly accurate one in general; still, it can produce counterintuitive results in periods of high volatility (i.e., when the persistence of the volatility process is lower; see e.g. Fusari and Gonzalez-Perez (2012)). Thus I compute the expectations under the physical

²⁵See e.g. Carr and Wu (2009) for details on computing the risk-neutral return variation from a portfolio of options. The VIX is subject to some approximation error (see, e.g., the discussion in Jiang and Tian (2007)), but the CBOE procedure for calculating the VIX has emerged as the industry standard. Therefore, relying on the squared VIX as a measure of the risk-neutral expected variance facilitates comparison with other studies.

²⁶See Bollerslev, Tauchen, and Zhou (2009), Drechsler (2013), and Drechsler and Yaron (2011).

²⁷Several studies suggest that, for highly liquid assets such as the *S&P500* index, a five-minute sampling frequency provides a reasonable balance between increasing estimation precision and limiting microstructure noise (see, e.g., the discussion in Hansen and Lunde (2006)). I obtain a monthly time series of realized variance based on five-minute returns from Hao Zhou's webpage: <https://sites.google.com/site/haozhouspersonalhomepage/>.

measure of total stock market return variance by a simple projection of the realized variance measure on a set of predictor variables. As in Drechsler (2013) and Drechsler and Yaron (2011), the realized variance is projected on the value of the squared VIX at the end of the previous month and on a lagged realized variance measure.²⁸ The difference between the conditional forecast from the projections and the risk-neutral expectation, measured using the VIX, yields the series of one-month market variance premium estimates plotted in Figure 2.12.²⁹ The variance premium for the market is negative on average, which means that investors are willing to pay a premium to be insured against high-variance states; the premium is time varying, with periods of a small and smooth premium alternating with periods in which the variance premium is larger (in absolute value) and more volatile. These phases of high and volatile variance premia seem to coincide with periods of large disagreement between investors, denoted by the light gray shaded areas in Figure 2.12. Periods in which differences in beliefs are large include recessions (denoted by dark gray shaded areas) and other times of financial distress, such as the Long-Term Capital Management crisis and Russian default in 1998. Table 2.6 gives summary statistics for the estimated variance risk premium and its two constituents: the risk-neutral (implied) and physical expectation of realized variance, denoted respectively by IV and ERV . For the sake of evaluating robustness, Table 2.6 also shows summary statistics for the period prior to the recent financial crisis (until July 2007).³⁰ The expected variance measure and the risk premium display significant deviation from normality in both samples, with large skewness and kurtosis. The crisis period is characterized by unprecedented spikes in stock market variance, which is reflected in extremely large standard deviation, skewness, and kurtosis statistics for the ERV full-sample time series; variance risk premium statistics are similar in the two samples. Note also that the variance risk premium is less persistent than the two variance forecasts, with an autocorrelation coefficient of about 0.65.

Variance risk premia for individual stocks can also be computed by using option prices to approximate the risk-neutral expectation of the return variation and using the sum of squared daily returns to approximate realized variance. This procedure is followed by Carr and Wu (2009) for 35 individual stocks and by Buraschi, Trojani, and Vedolin (2014) for the constituents of the *S&P100* index. Yet Bollerslev, Tauchen, and Zhou (2009) underscore the importance of using realized variance based on high-frequency data, which are not easily available for single stocks, when estimating the predictive power of variance risk premia for future excess returns.³¹ Moreover, there is evidence of a large systematic component in the cross section of variance risk premia (see e.g. Carr and Wu (2009)). Hence I study empirically the forecasting power of the market variance risk premium for both the market excess return and the cross section of stock and portfolio returns. Because the model-implied variance risk premium includes compensation only for jump risk, as a robustness check I run the same predictive regressions using the time

²⁸I also implement the same regressions using an expanding window to rule out any look-ahead bias. Because the results are almost identical, I use the in-sample estimates to facilitate comparison with existing studies and to avoid losing observations at the beginning of the sample for the initial estimation.

²⁹Similar dynamics are obtained when using more sophisticated models for the realized variance forecasts, such as the heterogeneous autoregressive (HAR) model of Corsi (2009), but here I focus on the simplest measure because more complex models are difficult to identify using monthly data. Bekaert and Hoerova (2014) compare different volatility forecasting models and show that the projection I use is the best within the simple specifications and also performs relatively well in comparison with more sophisticated models.

³⁰The start of the recent financial crisis is usually considered to be August 2007, so my pre-crisis sample ends in July 2007. The results do not change qualitatively if instead I end the pre-crisis period in December 2007 (Jin (2013)), in June 2007 (Drechsler and Yaron (2011)), or even in August 2008—just before the height of the crisis and the two anomalous observations of realized variance for October and November 2008.

³¹An exception is the working paper by Han and Zhou (2011), who use high-frequency stock prices to compute the objective expectation of return variation and then estimate the variance risk premium of a large cross section of stocks.

Table 2.6: Summary statistics

Summary statistics of the risk-neutral (IV) and physical (ERV) expectation of market return variance, in monthly squared percentage, for the full and pre-crisis sample.

	Full Sample			Pre-crisis		
	IV	ERV	VRP	IV	ERV	VRP
Mean	40.33	21.74	-18.60	33.14	16.65	-16.49
Median	31.66	15.72	-14.42	24.62	12.05	-13.23
Std.dev	36.35	25.69	14.21	23.99	13.04	11.99
Max	298.90	282.68	-4.01	163.39	73.13	-4.01
Min	9.05	3.99	-91.16	9.05	3.99	-91.16
Skew	3.24	5.44	-2.27	2.01	1.97	-2.35
Kurt	18.18	47.31	9.63	8.98	7.61	11.75
AC(1)	0.81	0.76	0.63	0.79	0.81	0.68

series of market variance risk premium due to large jumps as computed by [Bollerslev and Todorov \(2011\)](#), although data are available only for the period 1996–2007.³² Details on the data and a summary of the results using this alternative measure of the variance risk premium are provided in [Appendix A.2](#) and are generally consistent with the results discussed in this section.

Proxies of belief disagreement are calculated using the mean absolute deviation of one-year-ahead forecasts on real GDP growth from the BlueChip Economic Indicator, which are available at a monthly frequency through December 2009.³³ Being consistent with the model presented in [Section 2.2](#) would normally require that I measure disagreement about the perceived probability of a systemic disaster, but this is proportional to disagreement about the total expected consumption growth provided agents do agree on the expected growth rate in normal times (see [Equation \(2.13\)](#)). Another fundamental variable arising from the model is the consumption share of optimistic and pessimistic investors. This consumption share is not observable; however, as proxies for investor optimism the literature has used survey-based measures of investor or consumer sentiment such as the University of Michigan Consumer Sentiment Index (MCSI), the measure of investor sentiment compiled by the American Association of Individual Investors (AAII), and Shiller’s Crash Confidence Index (CCI).³⁴ Some empirical evidence also suggests that these sentiment measures forecast market returns; see for example [Charoenrook \(2002\)](#), [Lemmon and Portniaguina \(2006\)](#), and [Edelen, Marcus, and Tehranian \(2010\)](#).

³²I thank Viktor Todorov for providing the data.

³³See [Buraschi and Whelan \(2011\)](#) for details on the database, disagreement measures, seasonal adjustment, and construction of forecasts at fixed one-year horizons. I am grateful to Andrea Vedolin and Paul Whelan for providing the time series of belief disagreement on GDP growth.

³⁴The MCSI is one of the most widely followed measures of consumer confidence and has been used extensively in academic research (see e.g. [Ludvigson \(2004\)](#), [Lemmon and Portniaguina \(2006\)](#) and [Berkman, Jacobsen, and Lee \(2011\)](#)). It is available on a monthly basis from January 1978 and is based on surveys conducted for a minimum of 500 households. The AAI asks respondents to classify themselves as bullish, bearish, or neutral on the stock market for the next six months (see e.g. [Fisher and Statman \(2003\)](#)). The CCI refers to the percentage of respondents who state that the probability of a stock market crash occurring within the next two quarters is less than 10% (see e.g. [Koulovatianos and Wieland \(2011\)](#)). Data and explanations can be found on the website <http://icf.som.yale.edu/stock-market-confidence-indices-united-states>.

2.4.2 Predictive regressions for the market

The simple general equilibrium model in Section 2.2 implies a tight link between variance risk premia and excess returns of the stock market index, which is consistent with the empirical findings in Bollerslev, Tauchen, and Zhou (2009), Drechsler (2013), and Drechsler and Yaron (2011). Table 2.7 displays results from ordinary least-squares (OLS) estimation of standard return predictability regressions of the form

$$r_{t+h}^e = \alpha + \beta VRP_t + \varepsilon_{t+h}. \quad (2.49)$$

I regress monthly *S&P500* excess returns—at horizons h ranging from one month to one year—on the variance risk premium. The excess return series for $h > 1$ are overlapping, t -statistics are Newey–West corrected, and I report adjusted R^2 in percentage. Panel A reports regression estimates for the full sample of 264 monthly observations, while Panel B is limited to the pre-crisis sample. In line with results reported in the literature, there is a negative and significant relation between the variance premium and excess returns; also, the predictive power (measured either as the adjusted R^2 or as t -statistics of the regression coefficient) is highest between the three-month and the six-month horizon. In particular, for the pre-crisis sample the R^2 peaks at a quarterly horizon and the regression coefficients become insignificant at long horizons; these findings are consistent with those of Bollerslev, Tauchen, and Zhou (2009), whose sample ends in December 2007. Including the financial crisis (Panel A) yields stronger results and a variance risk premium that significantly predicts excess returns also at longer horizons, consistently with Fusari and Gonzalez-Perez (2012). Panels C and D report results from robust regressions that employ Huber-type weights to limit the influence of outliers. The robust regression estimates agree both in magnitude and sign with the OLS estimates, and in most cases the predictability evidence is even stronger.³⁵ Overall, these results indicate a considerable ability of the variance risk premium to predict future market excess returns.

³⁵ A more naive way to control for the effect of outliers is to run the OLS regression without the two potentially anomalous observations of October and November 2008 (at the peak of the financial crisis) when the realized variance experienced unprecedented levels. Results do not qualitatively change, and the estimates are just slightly more significant than in Panel A of Table 2.7.

Table 2.7: Market return predictability by variance risk premium

Panel A: Full sample (1990.1-2011.12)					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.553	-0.555	-0.476	-0.342	-0.272
t -stat	-2.048	-2.777	-3.705	-2.991	-2.702
Adj R^2 (%)	1.88	5.99	7.75	5.74	4.63
Panel B: Pre-crisis sample (1990.1-2007.7)					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.509	-0.557	-0.307	-0.176	-0.119
t -stat	-2.435	-3.160	-2.005	-1.191	-0.922
Adj R^2 (%)	1.19	5.83	3.59	1.48	0.56
Panel C: Robust, Full sample (1990.1-2011.12)					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.824	-0.649	-0.473	-0.353	-0.292
t -stat	-4.021	-5.896	-5.338	-5.060	-4.860
Adj R^2 (%)	1.71	6.16	8.11	6.10	4.98
Panel D: Robust, Pre-crisis sample (1990.1-2007.7)					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.694	-0.554	-0.332	-0.244	-0.193
t -stat	-2.774	-4.238	-3.363	-3.102	-2.804
Adj R^2 (%)	1.44	6.28	4.02	1.60	0.64

Stability analysis and flexible regression methods

Estimating the regression in Equation (2.49) implicitly imposes major restrictions on the relation between variance risk premium and future returns, since that regression assumes a monotone and linear structure. The theoretical asset pricing model presented in Section 2.2 suggests that this relation is only conditionally linear; unconditionally the relation need not be linear or even monotonic. Therefore, I study empirically potential instabilities or nonlinearities in the standard regression results. Introducing additional regressors does not qualitatively change the results (as shown by Bollerslev, Tauchen, and Zhou (2009)), so I rely on simple regressions to outline the properties of the relation between returns and variance premia. I also focus on the six-month horizon, for which significance of the standard predictive regressions seems to be stronger. The regression coefficient β for a large market index should vary with the distribution of the consumption share between agents (see Section 2.3.4). That distribution is not observable, but in the model it is directly linked to the level and volatility of the variance risk premium (see Table 2.2); hence I investigate the shape of the predictive relation for different levels of the premia.

First, I run regression (2.49) separately for different quantiles of the variance risk premium. Figure 2.13 plots the distributions of regression coefficients (upper panel) and R^2 (lower panel), which are obtained by applying a block bootstrap procedure. In both panels, the leftmost box plot corresponds to small absolute values of the premium ($VRP < q_{70\%}$), the rightmost one to large values ($VRP > q_{30\%}$), and the middle box plot to average values of the VRP . In accordance with the model, predictive power is increasing in the (absolute) level of the variance premium and the regression coefficient is significantly different from zero only for large values of the variance risk premium. From an empirical standpoint, changes in the variance premium are of course not exclusively related to changes in the cross-sectional consumption distribution of disagreeing agents. In order to relate more tightly the instability of standard predictive regressions to the extent of risk sharing among agents, I stratify regression (2.49) according to the level of difference in beliefs (DB). Figure 2.14 shows the distributions of regression coefficients (upper panel) and R^2 (lower panel) for small, average, and large DB values (once again a block bootstrap procedure is used). The regression coefficient increases in absolute value with the level of DB, from -0.2415 to -0.9725 ; the adjusted R^2 is 1.09% for small DB and increases to 4.02% and 20.91% for average and large DB, respectively.³⁶ The link between the level of the variance premium and measures of disagreement is confirmed by a simple OLS regression of monthly VRP on DB from January 1990 through December 2009. A change of one standard deviation in DB yields a change of 0.38 standard deviations in the variance premium; this result is strongly significant both statistically and economically, with a Newey–West corrected t -statistic of -4.319 and an adjusted R^2 of about 14% .³⁷

A second way to analyze the validity of a simple linear regression is to estimate a fully nonparametric regression of the form

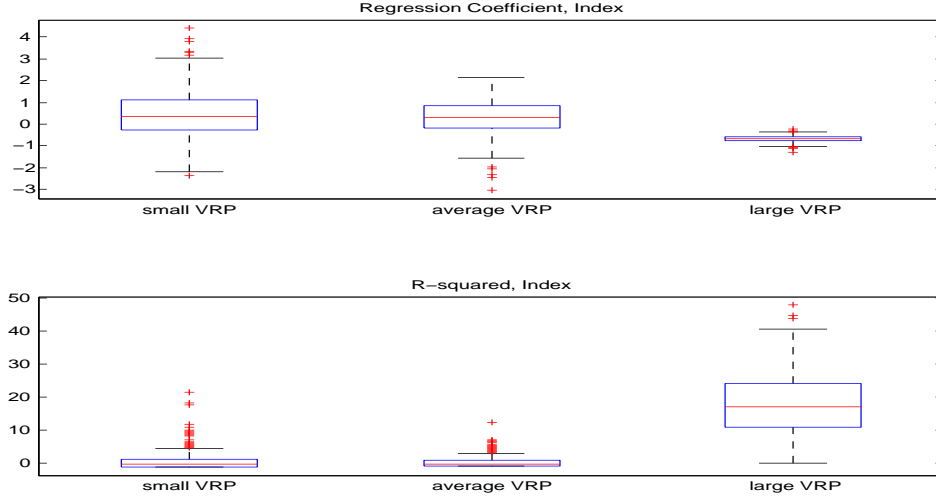
$$r_{t+h}^e = m(VRP_t) + \varepsilon_{t+h}, \quad (2.50)$$

³⁶The values of the regression coefficients and R^2 are not exactly comparable to the results obtained previously because DB is available only until December 2009 (see Section 2.4.1).

³⁷The level of the variance premium is positively linked also to measures of sentiment, such as the MCSI, with a correlation of almost 20% between the two measures. However, the results are less significant, probably because sentiment measures—which are based on surveys at the household level—are relatively noisy.

Figure 2.13: Predictive regressions for different levels of the variance risk premium

Standard OLS regression of excess market returns at the six-month horizon on the lagged variance risk premium, for different levels of the VRP . The first box plot corresponds to small absolute values of the premium ($VRP < q_{70\%}$), the last to large values ($VRP > q_{30\%}$) and the middle box plot to average values of the VRP . Upper panel display the distribution of regression coefficients and lower panel of percentage R^2 , both obtained applying a block bootstrap procedure.



where $m: \mathbb{R} \rightarrow \mathbb{R}$ is an arbitrary function fulfilling some smoothness conditions. An estimate of the function m can be simply obtained by using the Nadaraya–Watson kernel estimator with, in this univariate case, a Gaussian kernel; see Figure 2.15.³⁸ The number of observations is not large enough to draw strong conclusions, but a visual inspection clearly confirms the absence of any link between the two variables for small (absolute) values of the variance risk premia, although there is a stronger negative relation for more extreme values of those premia. Hence this simple nonparametric analysis supports the conclusion that the predictive power of variance risk premia for market returns is a time-varying and nonlinear phenomenon.

To avoid a fully nonparametric procedure, it is possible to model explicitly the regression coefficient's time variation. The conditional β could, for example, be a function of the variance premium itself:

$$\tilde{r}_{t+h}^e = \underbrace{(\beta_0 + \beta_1 \widetilde{VRP}_t)}_{\beta_t} \widetilde{VRP}_t + \varepsilon_{t+h}, \quad (2.51)$$

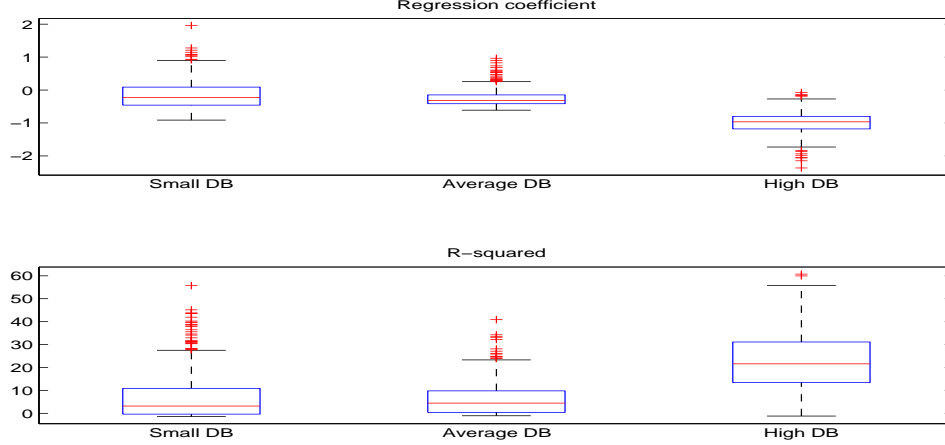
where the tilde marks variables that are standardized. Equation 2.51 is equivalent to a quadratic regression and can be estimated via standard OLS.³⁹ If β_1 proved to be insignificant then we could not reject a linear relation between excess returns and lagged variance premia, but β_1 is actually both positive and significant whereas β_0 (the linear term) is insignificant; the adjusted R^2 of this regression is 9.47%, which corresponds to a 22% increase over the linear regression's adjusted R^2 of 7.75% (at six-month horizon). In general, time variation in the regression

³⁸The optimal bandwidth, computed as suggested by Bowman and Azzalini (1997), is 4.77.

³⁹In other words, I am estimating Equation (2.50) while requiring that m be a quadratic function. One could, theoretically, employ other functional forms, but this is the most obvious alternative to a linear regression.

Figure 2.14: Predictive regressions for different levels of the difference in beliefs

Standard OLS regression of excess market returns at the six-month horizon on the lagged variance risk premium, for different levels of the difference in beliefs (DB). The first box plot corresponds to small values of disagreement ($DB < q_{30\%}$), the last to large values ($DB > q_{70\%}$) and the middle box plot to average values of DB. Upper panel display the distribution of regression coefficients and lower panel of percentage R^2 , both obtained applying a block bootstrap procedure.



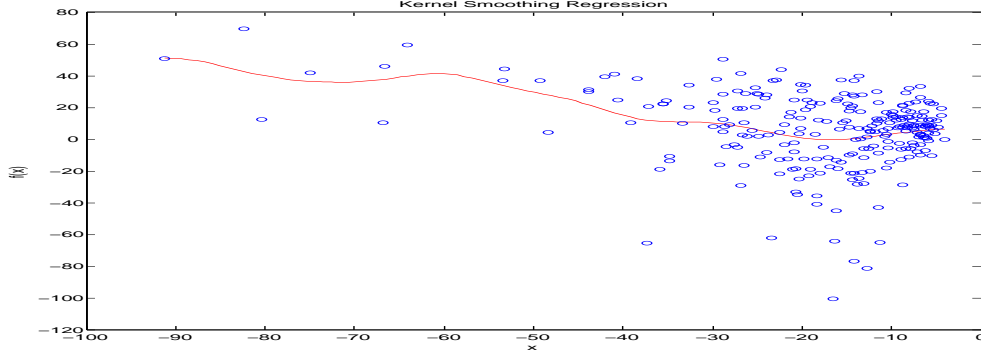
coefficient is modeled by introducing an interaction term. Apart from the level of VRP , other reasonable candidates worth exploring are the volatility of VRP and the level of disagreement or optimism. If we use DB as a conditioning variable, then the regression coefficient β_t becomes more negative with increasing disagreement (as expected), and the adjusted R^2 for the monthly 1990–2009 sample increases from 6.3% to 7.6%.

An alternative way of analyzing the time variation in the predictive power of the variance premium for future returns is to estimate a standard regression on a rolling window. Toward that end, I regress market excess returns at a 6-month horizon on lagged variance risk premium using a rolling window of 50 months. Figure 2.16 reports estimates of the slope coefficient and corresponding adjusted R^2 . Instability of the predictive relation is evident, and it is possible to relate the time variation in the slope coefficient to measures of disagreement. The correlation between the rolling regression coefficient estimate and a moving average of the difference in belief measure is equal to -51.18% , which means that the regression coefficient becomes more negative as disagreement increases. The relation between the regression's slope coefficient and the difference in beliefs suggests that the predictive power of variance risk premia for future excess returns is countercyclical, given that measures of disagreement are known to increase in bad times (see e.g. Patton and Timmermann (2010) and Buraschi, Trojani, and Vedolin (2013)).

A growing body of empirical evidence documents instabilities and nonlinearities in the strength of the return predictability by popular macroeconomic variables such as the dividend yield and short rate variables. For example, Henkel, Martin, and Nardari (2011) use a regime-switching model to show that standard aggregate return predictors are effective during business cycle contractions but practically useless during expansions. In the same way, I examine the

Figure 2.15: Kernel regression

Kernel regression of standardized excess market returns at the six-month horizon, in annualized percentage, on the lagged variance risk premium, in monthly squared percentage. Single dots represent the data, while the solid line is an estimated kernel regression using Nadaraya–Watson estimator with a Gaussian kernel.



dynamics in the predictive power of variance risk premia via estimation of a regime-switching model:

$$\tilde{r}_{t+h}^e = \beta_s \widetilde{VRP}_t + \varepsilon_{s,t+1}, \quad (2.52)$$

where $\varepsilon_s \sim N(0, \sigma_s^2)$ and the state $s \in \{1, 2\}$ follows a Markov chain with constant transition probabilities. I find that predictability is present only in state 2, which is characterized by more volatile and larger (on average, in absolute value) variance risk premia (see Figure 2.17). In state 1, the regression coefficient is positive and insignificant. The estimated transition probability matrix is

$$\begin{bmatrix} 0.97 & 0.03 \\ 0.07 & 0.93 \end{bmatrix};$$

thus states are persistent and with expected durations of about 3.27 and 1.24 years, respectively. It is most interesting that state 2 corresponds to periods of financial crisis. In particular: the first shaded area in Figure 2.17 corresponds to the US savings and loans crisis; the second, starting in June 1996, includes the Asian financial crisis, the Russian default, and the bursting of the dot-com bubble; and the last shaded area starts in September 2007 with the recent financial crisis. Therefore, the shift in regime of the variance premium (and of its predictive power) could be linked to a systemic disaster or to a jump in the consumption share of pessimistic agents, as would be implied by the simple model in Section 2.2.

Figure 2.16: Rolling regressions

Predictive regressions of market 6-month excess returns on lagged variance risk premium on a rolling window of 50 months. Upper panel shows regression coefficient estimates with 95% confidence bounds, while lower panel reports adjusted R^2 in percentage. The dashed red line in the upper panel denotes the regression coefficient estimated on the full sample.

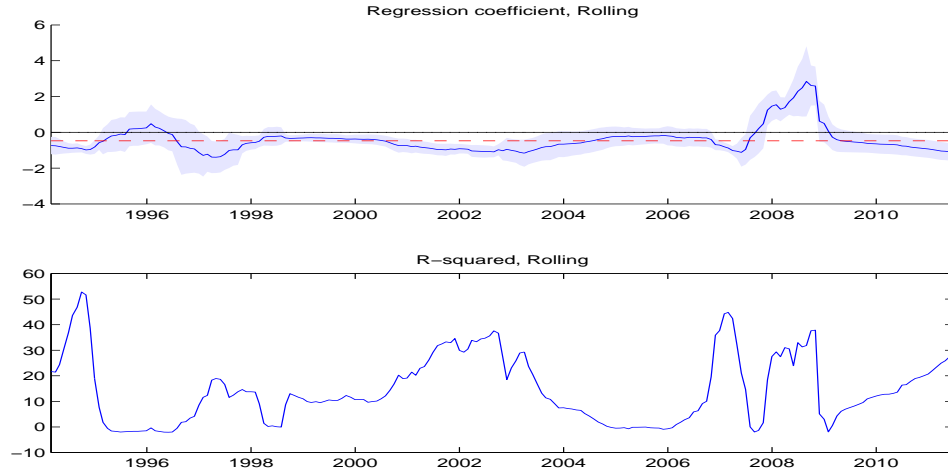
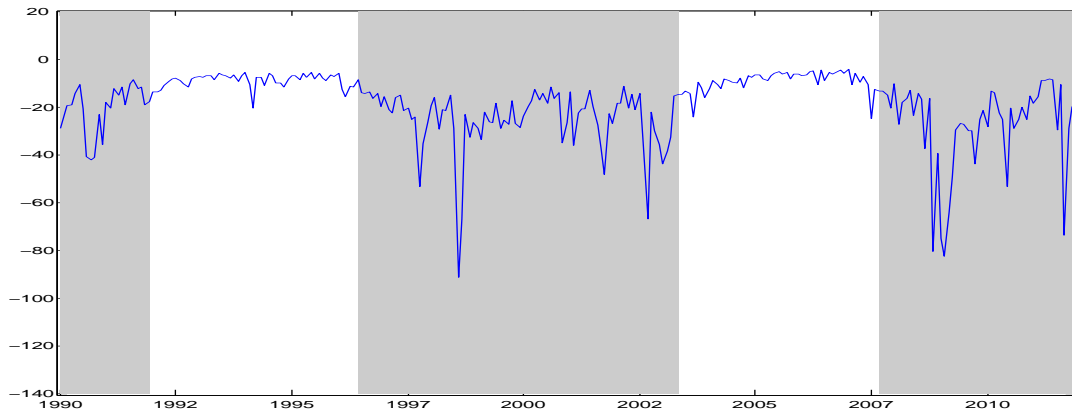


Figure 2.17: Variance risk premium regimes

Time series of variance risk premia with estimated regimes. Shaded areas correspond to state 2, which is characterized by stronger return predictability.



2.4.3 Predictive regressions in the cross section

I next test the hypothesis that market variance risk premia predict excess returns also for single stocks and portfolios, consistently with the model in Section 2.2 and with the empirical evidence that aggregate variance premium is a priced factor (see e.g. Carr and Wu (2009)). If investors are averse to variance risk, then stocks with high (negative) predictive *VRP* loadings will have higher expected returns and stocks with low or positive regression coefficients will serve as hedges and thus have lower expected returns. In line with the foregoing aggregate predictive analysis, I estimate regressions of the form

$$r_{i,t+h}^e = \alpha_i + \beta_i \text{VRP}_t + \varepsilon_{t+h}, \quad (2.53)$$

where $r_{i,t+h}^e$ denotes the monthly excess returns on stock i with horizon h . Average regression coefficients are similar to those obtained for the stock index in Panel A of Table 2.7, but on average they are not significant owing to a large cross-sectional variation. So as to understand better the cross-sectional dynamics, I sort stocks based on their estimated variance risk premium loading, β_i , at horizon $h = 6$ months.

Table 2.8: Stock returns predictive regressions

Single stock returns predictability by variance risk premium for the full sample (1990.1–2011.12). Stocks are sorted into quintile portfolios based on their estimated *VRP* loading, β_i . Panels A-E report average statistics in each quintile.

Panel A: First Quintile					
Horizon (months)	1	3	6	9	12
VRP Coeff	-1.127	-1.431	-1.411	-1.097	-0.870
t -stat	-1.543	-2.452	-2.964	-2.798	-2.556
Adj R^2 (%)	1.01	5.76	10.37	10.38	9.54
Panel B: Second Quintile					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.696	-0.781	-0.686	-0.486	-0.360
t -stat	-1.372	-2.024	-2.341	-2.069	-1.818
Adj R^2 (%)	0.61	2.95	4.66	3.99	3.25
Panel C: Third Quintile					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.567	-0.525	-0.433	-0.295	-0.241
t -stat	-1.222	-1.726	-1.862	-1.559	-1.493
Adj R^2 (%)	0.41	1.64	2.42	2.12	2.32
Panel D: Fourth Quintile					
Horizon (months)	1	3	6	9	12
VRP Coeff	-0.250	-0.295	-0.194	-0.147	-0.105
t -stat	-0.582	-1.021	-0.929	-0.894	-0.727
Adj R^2 (%)	0.04	0.58	0.48	0.62	0.63
Panel E: Fifth Quintile					
Horizon (months)	1	3	6	9	12
VRP Coeff	0.114	0.003	0.071	0.063	0.077
t -stat	0.314	0.011	0.347	0.351	0.508
Adj R^2 (%)	-0.14	-0.16	-0.03	0.15	0.61

Table 2.8 reports average estimation results in each quintile of β -values. In the first quintile (Panel A), the regression coefficient at any horizon is about 3 times larger than for the market index and the adjusted R^2 is higher at horizons of 6, 9, and 12 months. In the second quintile (Panel B), the values of the regression coefficient are close to those of the market but are significant only at horizons of 3, 6, and 9 months; in general, significance and R^2 decline rapidly as the value of beta increases (Panels C, D, and E). The regression coefficients switch sign in the fifth quintile but are not significantly different from zero. There seems to be a link between the value of the variance risk premium loading and stock capitalization, which is consistent with the model (see Section 2.3.3). In fact, the average capitalization of stocks in the first quintile is about 40% less than that of stocks in the fifth quintile. To investigate this link further, I run predictive regression (2.53) using CRSP cap-based portfolio returns and report the estimation results in Table 2.9. Decile 1 includes the largest stocks and decile 10 the smallest; results are reported for portfolios including deciles 1 and 2 (large-cap CRSP index), 3 to 5 (mid-cap CRSP index), and 6 to 10 (small-cap CRSP index). At the one-month horizon, there does not seem to be a clear pattern and overall significance is quite weak. At longer horizons, the predictive power of the market variance risk premium for excess returns is much stronger for small stocks, in line with the model and with the empirical evidence described previously. At the six-month horizon, for instance, adjusted R^2 for the small-cap portfolio is 10.17%—more than 50% larger than the 6.21% R^2 for the big-cap portfolio. The forecasting power with respect to small stocks is still impressive at the one-year horizon, with an R^2 of 9.67%. As a robustness check, I estimate the same regression on the 25 Fama and French portfolios that are sorted by the firm characteristics of size and book-to-market ratio (BM).⁴⁰ Figure 2.18 plots estimates of VRP loadings (left panel) and of adjusted R^2 in percentage (right panel) from regression (2.53), at the six-month horizon, for the 25 Fama and French portfolios. Lines connect portfolios of different book-to-market categories within each size category while focusing on the bottom and upper quintiles, which correspond to small and big stocks, respectively. On average, small stocks have larger (in absolute value) VRP loading and higher R^2 . This means that exposure to aggregate variance risk could partially explain the size premium.⁴¹ The predictive power of variance risk premium for future returns seems to be stronger also for growth stocks. Within the simulated model, however, no distinction can be drawn between size and value effects if assets have identically distributed cash flows.⁴²

As in the case of predictive regressions for the aggregate market, discussed in Section 2.4.2, the regression model in Equation (2.53) likewise assumes a monotone and linear structure (this is contrary to implications of the theoretical asset pricing model presented in Section 2.2). Therefore, also in the cross section I analyze the dependence of the linear regression coefficient β_i on the level of the variance risk premium and of the difference in beliefs (focusing on the six-month horizon); results are consistent with those reported in Section 2.4.2. Figure 2.19 shows the results of estimating regression (2.53), using CRSP cap-based portfolio returns, separately for different quantiles of the difference in beliefs (DB). Left panels display the distributions of regression coefficients and right panels display adjusted R^2 , where both are obtained by applying a block bootstrap procedure. Within each panel, the leftmost box plot corresponds to small values of the disagreement ($DB < q_{30\%}$), the rightmost to large values ($DB > q_{70\%}$),

⁴⁰Data are obtained from Kenneth French’s website. The portfolios are the intersections of five portfolios formed on size and five portfolios formed on the ratio of book equity to market equity. Breakpoints are the NYSE market equity and *BM* quintiles. *Size 1* corresponds to small stocks and *BM 1* to growth stocks.

⁴¹Bali and Zhou (2011) show that an asset pricing model in which both market risk and aggregate variance risk premium are priced can explain the premia for industry, size, and value.

⁴²Martin (2013) proposes different ways to generate variation along the value dimension that does not align perfectly with size in a general equilibrium model with multiple trees. He shows that the different alternatives generate very different patterns of alphas and betas across the size and value dimensions.

Table 2.9: Return predictive regressions for CRSP cap-based portfolios

Return predictability by variance risk premium for CRSP cap-based portfolios. Panel A include regression estimates at the one-month horizon, Panel B is for the six-month horizon and Panel C for the 12-month horizon.

Panel A: one-month horizon			
Portfolio	Big Cap	Mid Cap	Small Cap
VRP Coeff	-0.527	-0.597	-0.630
t -stat	-1.975	-2.075	-1.820
Adj R^2 (%)	1.69	1.45	1.12
Panel A: six-month horizon			
Portfolio	Big Cap	Mid Cap	Small Cap
VRP Coeff	-0.435	-0.573	-0.736
t -stat	-3.335	-3.844	-3.720
Adj R^2 (%)	6.21	8.19	10.17
Panel A: 12-month horizon			
Portfolio	Big Cap	Mid Cap	Small Cap
VRP Coeff	-0.229	-0.354	-0.450
t -stat	-2.310	-3.099	-3.237
Adj R^2 (%)	2.86	7.28	9.67

and the middle box plot to average values of DB . In accordance with the model and just as for the aggregate predictive regression (Figure 2.13), here predictive power increases with the level of the difference in beliefs. Table 2.10 summarizes these results. Differences among the DB quantiles seem to be stronger for the small-cap portfolio. Also, the difference between small- and big-cap beta is significant only in the state where there is a large difference in beliefs. In other states, the VRP loading and the adjusted R^2 are strongly similar for large and small stocks, and the big-cap portfolio beta is even higher (in absolute value) than the beta for small stocks when disagreement is low. The last panel of Table 2.10 reports results of regressing, on the aggregate variance risk premium, the return of a portfolio that is long the small-cap index and short the big-cap index. The regression coefficient is negative and strongly significant when disagreement is high, with an R^2 of more than 20%. Therefore, in these states a large (absolute) variance risk premium predicts a larger size premium, while the effect is not significant when difference in beliefs is low. This finding is related to the work of Lemmon and Portniaguina (2006), who show a negative relation between the size premium and consumer confidence. In fact, the size effect (of smaller firms having higher returns on average) seems to be concentrated in periods characterized by large disagreement, as shown in Table 2.11. It is natural for investors to require higher returns on assets that are more sensitive to systemic disaster risk. The economic intuition for this finding can be found in Jagannathan and Wang (1996), who argue that firms more likely to exhibit financial distress (e.g., small firms) have market betas that are more sensitive to changes in the business cycle. Investor sentiment, or disagreement, is thus related to time variation in the expected returns of those firms because these factors forecast future business conditions. However, this reasoning holds only when the perceived systemic jump premium is high. The model suggests that the systemic jump premium component could actually have a negative effect on the stock's excess returns if the consumption share of pessimists were low enough. Thus the size premium could go in opposite directions depending on which agent type dominates the market. This finding is consistent with some of

the later empirical research on the size effect, which suggests that the premium disappears in the 1980s.

Figure 2.18: Return predictive regressions for Fama and French portfolios

Standard OLS regression of excess returns of Fama and French portfolios at the six-month horizon on the lagged market variance risk premium. Left panel displays the regression coefficients and right panel the percentage R^2 . Lines connect portfolios of different book-to-market categories within each size category, focusing on the bottom and upper quintiles, which correspond to small and big stocks, respectively.

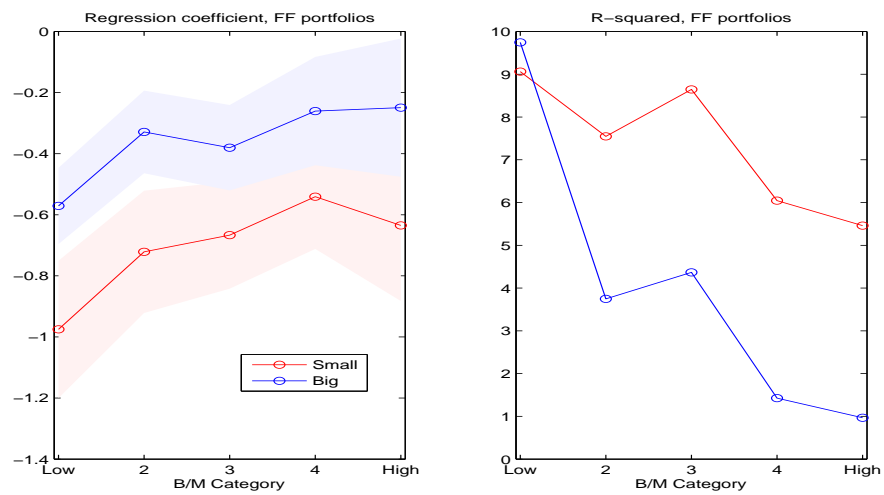


Figure 2.19: Return predictability of cap-based portfolios for different levels of difference in beliefs

Standard OLS regression of excess returns at the six-month horizon on the lagged market variance risk premium for the CRSP cap-based portfolios, for different levels of the difference in beliefs proxy, DB. Left panels display the distribution of regression coefficients and left panels of adjusted R^2 , both obtained applying a block bootstrap procedure.

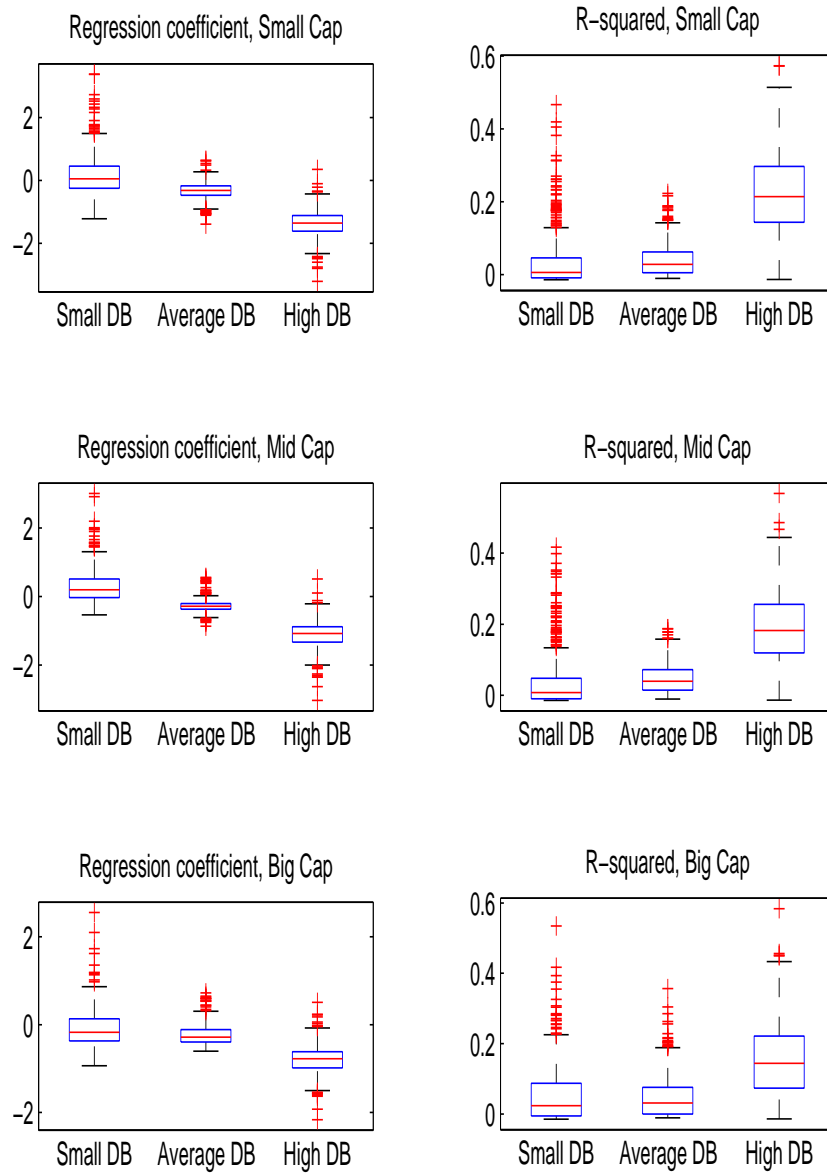


Table 2.10: Return predictability of cap-based portfolios for different levels of difference in beliefs

Return predictability by variance risk premium for small-, mid-, and big-cap portfolios from CRSP, at the 6-month horizon, for different levels of the difference in beliefs. For comparison, the last two panel report results of the same predictive regression for the *S&P500* index return and for small- minus big-cap portfolio return, respectively.

		Small DB	Average DB	Large DB
small-cap	VRP Coeff	0.06	-0.32	-1.37
	Adj R^2 (%)	-1.36	2.87	21.26
		Small DB	Average DB	Large DB
mid-cap	VRP Coeff	0.19	-0.29	-1.11
	Adj R^2 (%)	-0.41	3.82	17.79
		Small DB	Average DB	Large DB
big-cap	VRP Coeff	-0.18	-0.29	-0.82
	Adj R^2 (%)	-0.04	2.86	14.41
		Small DB	Average DB	Large DB
<i>S&P500</i>	VRP Coeff	-0.24	-0.31	-0.97
	Adj R^2 (%)	1.09	4.02	20.91
		Small DB	Average DB	Large DB
small-big	VRP Coeff	0.17	0.11	-0.60
	Adj R^2 (%)	-0.55	-0.35	21.42

Table 2.11: Size premium for different levels of difference in beliefs

Mean and standard deviation of returns (in annualized percentage) for small- and big-cap portfolios from CRSP, at the 6-month horizon, for different levels of the difference in beliefs.

		Small DB	Average DB	Large DB
small-cap	Mean (%)	5.40	18.17	11.37
	Standard Deviation (%)	21.54	22.90	45.48
		Small DB	Average DB	Large DB
big-cap	Mean (%)	7.94	14.53	2.04
	Standard Deviation (%)	15.00	19.62	32.61

2.5 Conclusion

This paper studies both theoretically and empirically how agent disagreement about the likelihood of systemic disasters affects the equity and variance risk premia and the relation between them, both for the market portfolio and in the cross section of stocks. The starting point is a general equilibrium model with multiple trees and disagreement about systemic rare event risk.

The main findings are the following. First, the equity (variance) risk premium of an individual stock tends to increase (decrease) with its dividend share and with the consumption share of the pessimistic agent. The variance risk premium can also switch sign, mainly for small stocks, and it is time varying; it alternates phases of small and smooth premia with periods in which the variance premium is larger (in absolute value) and more volatile. Second, the index variance risk premium is largely due to a covariance premium when assets are relatively evenly distributed or the number of stocks in the economy is large. The model-implied correlation risk premium, as the variance risk premium, is large (in absolute value) when pessimists hold a large fraction of the aggregate endowment. Third, rare event risk implies a tight link between the equity and the variance risk premia, both for the market and for the cross section of stock returns. This link however is state-dependent and varies with the asset's dividend share and the agents' consumption share. In particular, the relation is stronger when the consumption share of the pessimist is larger, i.e., in bad states of the economy, and for small stocks. Fourth, in the case of a large diversified economy only systemic risk is priced and the relation between equity and variance risk premia is conditionally linear. Moreover, infinitely small assets still earn a risk premium owing to the presence of systemic rare event risk.

I test empirically the main predictions and show that, as implied by the model, the relation between the equity premium and the index variance premium is time varying and systematically linked to the degree of risk sharing among disagreeing investors. In particular, the predictive power of variance premium for future excess returns is stronger in periods of large differences in investor beliefs. This relation holds especially for small stocks, whose returns are more dependent on the compensation for systemic rare event risk.

This work suggests several interesting lines of future research. On the theoretical side, the model's simplicity means that several extensions are possible. Examples include introducing a stochastic diffusion volatility of the dividend processes and allowing for learning based on an exogenous signal about the state of the economy. I could also introduce disagreement with regard to both the disaster intensity and the expected dividend growth in normal times. Given the link between volatility risk premia and option surfaces, the option pricing implications of an heterogeneous rare disaster model are also worth exploring.

On the empirical side, it would be natural to look for potential time variation and nonlinearity in the relation between stock returns and correlation risk premia as I find for the variance risk premium. It would also be worth investigating whether the same nonlinear relationships are present in other markets for which a link between disagreement or variance premia and excess returns has been documented, as for example the fixed income market (Mueller, Vedolin, and Yen (2011) and Buraschi and Whelan (2011)) and the foreign exchange market (Beber, Breedon, and Buraschi (2010)).

Dividend Growth Predictability and the Price-Dividend Ratio

Introduction

Are stock market returns and dividend growth predictable? [Campbell and Shiller \(1988\)](#) observation that the price-dividend ratio reflects information on both future expected returns and expected dividend growth has motivated a vast literature that studies predictability features based on predictive regressions of returns and cash flow growth on lagged dividend yields.

Predictive regression results typically imply an economically significant evidence of return predictability, even if the statistical significance is weaker in some subperiods, and an almost constant expected dividend growth. This evidence suggests that the price-dividend ratio varies mainly because of discount rate shocks; See [Campbell \(1991\)](#) and [Cochrane \(1992\)](#), among others.¹ In contrast, the Kalman filter estimation of a benchmark present-value model with hidden dividend and return expectations yields both a predictable return and a predictable dividend growth, indicating that the price-dividend ratio varies because of both dividend expectation and discount rate shocks; see, e.g., [Binsbergen and Koijen \(2010\)](#).

In this paper, we introduce a new method with reliable finite-sample accuracy and valid asymptotic properties, for testing general predictability hypotheses in present-value models. Our approach is based on a nonparametric Monte Carlo bootstrap, which avoids distributional assumptions in aggregating the information from the time series of price-dividend ratios and dividend growth. This approach allows us to study more sharply the diverging predictability implications emerging from present-value models and standard predictive regression settings.

We show that while conventional testing procedures can imply significant finite-sample biases that over-reject the null of no predictability, our testing method produces more reliable finite-sample inferences. Applying our testing methodology to benchmark present-value models, which are designed to parsimoniously aggregate dividend growth and price-dividend ratio information, we find a significant evidence of return predictability, but no evidence of dividend predictability in postwar US data, thus reconciling the diverging predictability conclusions in the literature.

Inference on return and dividend predictability in standard predictive regressions is difficult for a number of reasons. First, the large correlation between stock returns and predictive variables, combined with the high persistence of the latter, can create finite-sample biases and

¹Early predictive regression studies are [Rozeff \(1984\)](#), [Schiller \(1984\)](#), [Keim and Stambaugh \(1986\)](#), [Campbell and Shiller \(1988\)](#) and [Fama and French \(1988\)](#). The predictive regression findings of no dividend predictability also depend on the sample period used for estimation, as the conclusions are opposite for the prewar sample; see, e.g., [Chen \(2009\)](#).

a non-standard asymptotic behaviour for common tests of return or dividend predictability; see e.g., [Stambaugh \(1999\)](#) and [Torous, Valkanov, and Yan \(2004\)](#), among others. Second, as the dividend yield reflects expectations of both future stock returns and future cash flows, it is a noisy estimate of expected returns and expected dividend growth in univariate predictive regressions, thus creating a standard error-in-variable (EIV) problem; see, e.g., [Binsbergen and Koijen \(2010\)](#), among others. Third, powerful tests of predictability need to incorporate the fact that they test a joint null hypothesis on the return-dividend process. [Cochrane \(2008b\)](#) stresses the fact that return and dividend growth predictability have to be studied jointly, concluding that the weak evidence of return predictability in earlier univariate studies is stronger if one jointly considers the empirical evidence on dividend predictability.²

Standard predictive regressions are agnostic about the hidden economic link between price-dividend ratios, expected returns and expected cash flow growth, which is instead explicitly revealed within present-value models. These models offer a convenient framework to jointly estimate expected returns and dividend growth, while taking into account the joint no-arbitrage constraints on stock returns, cash flows and valuation ratios. Recent studies estimating market expectations for returns and dividends with different present-value models include [Menzly, Santos, and Veronesi \(2004\)](#), [Lettau and Ludvigson \(2005\)](#), [Ang and Bekaert \(2007\)](#), [Lettau and Van Nieuwerburgh \(2008\)](#), [Campbell and Thompson \(2008\)](#), [Pastor, Sinha, and Swaminathan \(2008\)](#), [Cochrane \(2008a,b\)](#), [Binsbergen and Koijen \(2010\)](#), and [Rytchkov \(2012\)](#), among others.

Consistent with the literature, the Kalman filter maximum likelihood estimation of a benchmark present-value model on postwar US stock market data, along the lines of, e.g., [Binsbergen and Koijen \(2010\)](#), yields different predictability implications than standard predictive regressions. In the sample period from January 1946 to December 2010, our point estimates imply predictable returns and dividend growth rates, a quite large fraction (18%) of future dividend variability explained by dividend expectations and a much lower fraction (9%) of future returns explained by return expectations. These findings are supported by the evidence produced in a standard likelihood ratio test of the null hypotheses of constant expected returns or expected dividend growth, which are both clearly rejected at significance levels below 0.5%. Expanding the predictive information set to include additional predictive variables for return and dividend growth expectations further strengthens this evidence; see [Yun \(2012\)](#). Simple univariate predictive regressions with the lagged price-dividend ratio as a predictive variable imply R^2 's of about 10% and 1% for market returns and aggregate dividend growth, respectively. Moreover, the return predictive regression test implies a significant predictability evidence with a p-value of 1.13%, while the null hypothesis of no dividend predictability is not rejected.

What drives the diverging predictability implications between benchmark present-value models and predictive regressions? A possible explanation is the EIV-bias inherent to predictability studies.³ This paper explores a different explanation, which is based on the finite-sample properties of estimators and tests in present-value models with latent return and dividend expectations. While the finite-sample properties of tests of predictability hypotheses in standard predictive regressions have been studied in detail,⁴ they have not yet been thoroughly

²Given the time variation in the price-dividend ratio, at least one between returns and dividend growth must be predictable. [Cochrane \(2008b\)](#) also derives upper bounds on price-dividend ratio autocorrelations, to deliver more powerful statistics in the joint testing of return and dividend growth predictability. In a present-value approach, such constraints are explicitly incorporated given the joint dynamics of the expectation processes.

³While the estimation results imply a small asymptotic EIV-bias in the return predictive regression, they are linked to a large negative EIV-bias in the slope parameter of the dividend predictive regression, with the limit of the slope coefficient being equal to 0.0197, much lower than the model-implied parameter value of 0.7038.

⁴[Stambaugh \(1999\)](#) derives an analytic expression for the bias in univariate predictive regressions. [Kothari and Shanken \(1997\)](#), [Amihud and Hurvich \(2004\)](#), [Lewellen \(2004\)](#), [Torous, Valkanov, and Yan \(2004\)](#), [Campbell and Yogo \(2006\)](#) and [Polk, Thompson, and Vuolteenaho \(2006\)](#) develop methods for hypothesis testing in univariate settings. [Amihud, Hurvich, and Whang \(2009\)](#) propose an analytic method for hypothesis testing in regression

studied in present-value models estimated with a latent-variables approach. Similar to predictive regressions, in state-space models inference about relevant hypotheses is made tractable by the existence of an asymptotic theory, which under appropriate conditions implies consistency and asymptotic normality of parameter and latent state estimates; see, e.g., [Liung and Caines \(1979\)](#) and [Spall and Wall \(1984\)](#). However, the short or moderate length of time-series data available in many predictability studies can make the use of asymptotic inference methods potentially suspect for latent variable approaches in present-value models as well.⁵

To improve over the conventional asymptotic inference, a useful nonparametric method, which does not rely on strong assumptions about the joint distribution of dividend growth and price-dividend ratios, is the nonparametric Monte Carlo bootstrap first suggested by [Efron \(1979\)](#). [Stoffer and Wall \(1991\)](#) prove that the bootstrap applied to the innovations of a time-invariant and stable state-space model yields asymptotically correct results. They also demonstrate, by Monte Carlo simulation and in a number of real-data applications, that a bootstrap approach can improve over the finite-sample inference of conventional asymptotics.

We start from these insights and propose a novel class of bootstrap likelihood ratio tests of predictability hypotheses. We show their asymptotic validity and demonstrate the improved finite-sample properties over the conventional asymptotics. Using the new testing methodology, we obtain novel findings and interpretations for the predictability evidence obtained by latent variable approaches within present-value models. The more detailed contributions to the literature are the following.

First, in order to study the finite-sample properties of tests of predictability hypotheses in present-value models, without assuming a particular error distribution, such as, e.g., a normal distribution, we introduce a simple nonparametric Monte Carlo simulation approach. Our simulations show that asymptotic likelihood ratio tests imply large finite-sample biases that often lead to an incorrect rejection of the null of no predictability. For instance, while according to the asymptotic chi-square distribution and a significance level $\alpha = 5\%$ the critical value of the asymptotic test of no dividend (return) predictability is 7.81 (9.49), the Monte-Carlo finite-sample critical value is 17.13 (15.99). Overall, the fraction of incorrect rejections of the null of no time-variation in dividend (return) expectations using the asymptotic test can be as large as 25.8% (60.5%). As a result, the evidence resulting from asymptotic tests needs to be taken with caution, because it could be generated by chance alone, using the sample sizes typically available in many predictability studies.

Second, the Monte Carlo evidence shows that large estimated R^2 's for dividends or returns can arise by chance alone, even under the null of constant expected dividend growth or expected return. These features are linked to the volatile point estimates for the persistence of dividend and return expectations, which is estimated unprecisely in finite samples and stays in a close relation to the estimated R^2 's. This evidence stresses the importance of combining a pure estimation approach with a reliable testing method, when quantifying the actual degree of dividend or return predictability.

Third, we propose a general nonparametric likelihood ratio test of predictability, by applying the bootstrap to the innovations from the latent state dynamics, generated under the relevant null hypothesis. We prove that our bootstrap likelihood ratio test implies an asymptotically valid inference, under standard conditions, and demonstrate by Monte Carlo simulation that

with multiple predictors, while [Lettau and Ludvigson \(2001\)](#) and [Ang and Bekaert \(2007\)](#), among others, use bootstrap methods in this setting.

⁵The close relation between present-value models and their (VAR) reduced-form predictive regression representations (see, e.g., [Cochrane \(2008b\)](#)) also suggests that if samples must be fairly large before asymptotic theory is applicable, then this should similarly hold both in predictive regressions and present-value models. See also Section A of the Supplemental Appendix, which is available at Piatti's webpage: <http://www.people.usi.ch/piattii>.

it improves in finite samples over the conventional asymptotic tests. Overall, these findings indicate that the bootstrap testing approach can better control the finite-sample probability of rejecting a null hypothesis because of chance alone, thus producing a more reliable predictability evidence in a number of applications.

Fourth, we apply our bootstrap test to US stock market data, using several specifications of the predictive information set. Overall, we find evidence in favour of time-varying expected returns, but no significant evidence against a constant expected dividend growth in post-war US data. These conclusions are robust with respect to different proxies of market cash flows and the inclusion of prewar dividend and return data. While these findings are different from those of conventional tests, they suggest that the postwar dividend predictability implied by present-value models aggregating dividend growth and price-dividend ratio information is similarly weak as for standard predictive regressions, thus reconciling the diverging conclusions in the recent literature.

Finally, we propose a modification of our bootstrap testing method that is useful to test the actual degree of out-of-sample predictability, while controlling the probability of detecting predictive relations by chance alone. Also in this context, our tests indicate that the larger estimated out-of-sample R-squared for dividends in the data can arise by chance alone, under the null of constant dividend growth expectations.

The paper proceeds as follows. Section 3.1 introduces the benchmark present-value model for aggregate dividends and market returns. It then briefly discusses the data and the estimation strategy, before reporting the standard estimation results. Section 3.2 studies the finite-sample biases of conventional asymptotic likelihood ratio tests, while Section 3.3 introduces the bootstrap likelihood ratio testing approach and studies the resulting improvements in finite-sample inference. Sections 3.4 and 3.5 analyse the robustness of our main predictability findings with respect to broader specifications of the predictive information set and to the inclusion of the prewar sample, respectively. Section 3.6 concludes.

3.1 Present-Value Approach

Borrowing from Binsbergen and Koijen (2010), we introduce the benchmark cash flow and discount rate dynamics. This model offers a tractable framework to estimate the expected return and expected dividend growth processes, by parsimoniously aggregating the time-series information from dividend growth and price-dividend ratios. Even though the benchmark model restricts the information set to be spanned by the history of dividend (or returns) and price-dividend ratios, it is flexible enough to capture the essential aspects related to the estimation and testing of predictive relations.⁶ Broader specifications of the predictive information set are studied in Section 3.4.

3.1.1 The benchmark model

Let

$$r_{t+1} \equiv \log \left(\frac{P_{t+1} + D_{t+1}}{P_t} \right) \quad (3.1)$$

⁶The same setting can result from a general equilibrium framework with multiple securities and time-varying risk aversions; see, e.g., Menzly, Santos, and Veronesi (2004)). Recent studies have investigated predictability in the context of the model considered in this paper, including Cochrane (2008a), Binsbergen and Koijen (2010), and Rytchkov (2012), among others. Model extensions and different special cases have also been considered in Lettau and Ludvigson (2005), Ang and Bekaert (2007), Lettau and Van Nieuwerburgh (2008), Campbell and Thompson (2008), Pastor, Sinha, and Swaminathan (2008), and Yun (2012).

be the cum-dividend log market return and denote by

$$\Delta d_{t+1} \equiv \log \left(\frac{D_{t+1}}{D_t} \right), \quad (3.2)$$

the aggregate log dividend growth. Expected dividend growth and return, conditional on the information at time t , are denoted by $g_t \equiv E_t[\Delta d_{t+1}]$ and $\mu_t \equiv E_t[r_{t+1}]$, respectively. They follow simple autoregressive processes:

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g, \quad (3.3)$$

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^\mu. \quad (3.4)$$

The dividend growth rate is the expected dividend growth plus an orthogonal shock:

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^d. \quad (3.5)$$

The vector of independent and identically distributed shocks $(\varepsilon_{t+1}^g, \varepsilon_{t+1}^\mu, \varepsilon_{t+1}^d)'$ has covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{\mu d} \\ \sigma_{gd} & \sigma_{\mu d} & \sigma_d^2 \end{bmatrix}. \quad (3.6)$$

The affine explicit expression for the log price-dividend ratio directly follows from a [Campbell and Shiller \(1988\)](#) log linearisation:

$$pd_t = A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0), \quad (3.7)$$

where A , B_1 and B_2 are simple functions of the model parameters such that, consistent with intuition, pd_t is decreasing in expected returns and increasing in expected dividend growth.⁷

3.1.2 Estimation results

We obtain the with- and without-dividend monthly returns on the value-weighted portfolio of all NYSE, Amex and Nasdaq stocks, in the period from January 1946 until December 2010, from the Center for Research in Security Prices (CRSP). We construct annual time series of aggregate dividends and prices, assuming that monthly dividends are cash-reinvested at the 30-day T-bill rate. Data on 30-day T-bill rates are also obtained from CRSP.

We estimate the model with a Kalman filter based on a Gaussian quasi likelihood function, from the observable time series of dividend growth Δd_{t+1} and price-dividend ratios pd_{t+1} . Due to the present-value relations, market return r_{t+1} is redundant with respect to Δd_{t+1} and pd_{t+1} .⁸

The parameter estimates are reported in [Table 3.1](#), with bootstrapped standard errors in parenthesis.⁹ We find an unconditional expected log return (dividend growth) of $\delta_0 = 8.3\%$ ($\gamma_0 = 5.7\%$). Both expectation processes feature some degree of persistence, with autoregressive roots γ_1 and δ_1 equal to 0.304 and 0.927, respectively, and expected returns are substantially more persistent than expected dividend growth. Finally, expected dividend growth is estimated as very volatile ($\sigma_g = 6.5\%$), while unexpected dividend growth variability is very low ($\sigma_d = 0.2\%$).

⁷See [Binsbergen and Koijen \(2010\)](#) and [Appendix B.1](#). [Campbell and Shiller \(1988\)](#) approximation in our sample holds almost exactly, for yearly data, if annual dividends and prices are constructed as described in [Section 3.1.2](#).

⁸Using (r_{t+1}, pd_{t+1}) as observable variables, the estimation results are almost identical and one can always recover the missing variable using [Campbell and Shiller \(1988\)](#) approximation; see also [Cochrane \(2008a\)](#), among others. Details on the estimation procedure are collected in [Appendix B.2](#).

⁹Parameter standard errors are obtained using the circular block-bootstrap of [Politis and Romano \(1992\)](#), in order to account for the potential serial correlation in the data.

Table 3.1: Estimation Results

Results of the estimation of the present-value model in Section 3.1. The model is estimated by maximum likelihood, using yearly data from 1946 to 2010 on log dividend growth rates and log price-dividend ratio. Panel A presents estimates of the coefficients of the underlying processes. Panel B reports resulting coefficients of the present-value decomposition $pd_t = A - B_1\hat{\mu}_t + B_2\hat{g}_t$. Bootstrapped standard errors are in parentheses.

Panel A: Maximum likelihood estimates				
γ_0	δ_0	γ_1	δ_1	
0.057	0.083	0.304	0.927	
(0.009)	(0.010)	(0.337)	(0.089)	
σ_g	σ_μ	σ_D	$\rho_{g,\mu}$	$\rho_{\mu,D}$
0.065	0.015	0.002	0.231	-0.972
(0.023)	(0.024)	(0.028)	(0.419)	(0.606)
Panel B: Implied present-value parameters				
ρ	A	B_1	B_2	
0.974	3.637	10.332	1.421	
(0.004)	(0.140)	(2.418)	(6.396)	

3.1.3 Dividend and return predictability

Let I_t denote the econometrician's information set at time t , generated by the history of dividends and price-dividend ratios. A nice feature of the Kalman filter is to provide filtered estimates of the unknown latent states μ_{t-1} and g_{t-1} , conditional on I_{t-1} . Thus, a standard measure of the degree of predictability in model (3.3)-(3.5) can be computed by the fraction of r_t and Δd_t variability explained by μ_{t-1} and g_{t-1} , respectively:

$$R_{Ret}^2 = 1 - \frac{\widehat{Var}(r_{t+1} - \mu_t)}{\widehat{Var}(r_{t+1})}, \quad (3.8)$$

$$R_{Div}^2 = 1 - \frac{\widehat{Var}(\Delta d_{t+1} - g_t)}{\widehat{Var}(\Delta d_{t+1})}, \quad (3.9)$$

where \widehat{Var} denotes sample variances.

We find that $R_{Ret}^2 = 8.82\%$ and $R_{Div}^2 = 17.58\%$, indicating that the degree of dividend predictability is about twice as large as the degree of return predictability. In contrast to these findings, simple regressions of returns and dividend growth on lagged price-dividend ratios yield R^2 of about 9.9% and 0.95%, respectively.

A possible interpretation for these diverging results is the noisiness of the price-dividend ratio (3.7) as a signal for expected returns and expected dividend growth, respectively, which creates a potential EIV problem in predictive regressions of returns and dividend growth on lagged price-dividend ratios. Indeed, the large persistence of return expectations is linked to a large sensitivity of price-dividend ratios to expected return shocks ($B_1 = 10.332$) and a smaller sensitivity to dividend expectation shocks ($B_2 = 1.421$). This feature obfuscates the predictive power of dividend expectations in dividend predictive regressions, leading to the low (biased) R^2 .

3.2 Testing Predictability Hypotheses

Based on the above estimation results, it is natural to ask whether the empirical evidence supports (i) the hypotheses that dividends and returns are predictable and (ii) the EIV problem interpretation for the different predictability findings with respect to standard predictive regressions.

These questions are best addressed using an appropriate hypothesis testing framework. As emphasized, e.g., in Cochrane (2008a), while a point estimate produces the most likely predictability structure according to the chosen statistical metric, an hypothesis testing framework is needed to simultaneously control for the probability that some estimated predictability features might be generated by chance alone.

3.2.1 Asymptotic tests

Most predictability hypotheses can be formulated by means of simple parametric constraints, which can be efficiently tested with a standard likelihood ratio (LR) test, using the statistic

$$LR_T = 2 \left(\max_{\Theta} \log \mathcal{L}(\theta, \{Y_t\}_{t=1}^T) - \max_{\Theta_0} \log \mathcal{L}(\theta, \{Y_t\}_{t=1}^T) \right), \quad (3.10)$$

where Θ_0 is the restricted set of parameters under the given null hypothesis H_0 and $\log \mathcal{L}$ is the log-likelihood of the model. Evidence against H_0 is collected when LR is sufficiently large:

$$\{LR_T > c_{1-\alpha}\}, \quad (3.11)$$

relative to a critical value $c_{1-\alpha}$ that is unlikely under H_0 . As $T \rightarrow \infty$, statistic LR_T follows a χ_r^2 distribution with r degrees of freedom, where r is the number of parameter constraints defining the constrained parameter set Θ_0 . Therefore, the choice $c_{1-\alpha} = \chi_{r,1-\alpha}^2$, where $\chi_{r,1-\alpha}^2$ is the $1 - \alpha$ quantile of the chi-square distribution, ensures asymptotically a small probability α of rejecting H_0 by chance alone:

$$\alpha = \lim_{T \rightarrow \infty} P_{H_0}(LR_T > \chi_{r,1-\alpha}^2). \quad (3.12)$$

3.2.2 Time-varying expectations

Testing return and cash flow predictability is equivalent to examining time-variation in expected returns and expected dividend growth, respectively. In terms of the model parameters, the null of constant return expectations is:

$$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0. \quad (3.13)$$

Similarly, the null of constant dividend expectations is:¹⁰

$$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0. \quad (3.14)$$

Using standard asymptotic critical values from a χ_r^2 distribution, where $r = 4$ and $r = 3$, respectively, Table 3.2 shows that both these null hypotheses are strongly rejected according to the asymptotic likelihood ratio test based on statistic (3.10), for significance levels α below 0.5%.

¹⁰Under the null (3.13) (the null (3.14)) all price-dividend ratio variation comes from variation in expected dividend growth (returns) and the present-value model collapses to a standard linear regression of dividend growth rates (returns) on the lagged price-dividend ratio. Note that $\rho_{gd} = 0$ is imposed also in the unconstrained model for identification purposes, as for instance in Binsbergen and Kojen (2010), see Appendix B.2.

Table 3.2: Standard Likelihood Ratio tests

Constrained ML estimates of the present-value model and LR statistics for the tests of constant expected returns ($H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$), constant expected dividend growth ($H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$) and equal autoregressive parameters ($H_0 : \delta_1 = \gamma_1$). The first column reports the results of the unconstrained estimation, from Table 3.1. *LogL* denotes the pseudo log-likelihood obtained, *LR* is the value of the Likelihood Ratio statistic computed using (3.10), *p-value* denotes percentage p-values of the tests, and the last row reports finite-sample sizes of the tests, in percentage.

	Unconstrained	$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$	$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$	$H_0 : \delta_1 = \gamma_1$
γ_0	0.057	0.072	0.055	0.054
δ_0	0.083	0.079	0.082	0.081
γ_1	0.304	0.996	0	0.926
δ_1	0.927	0	0.903	0.926
σ_g	0.065	0.002	0	0.004
σ_μ	0.015	0	0.021	0.021
σ_d	0.002	0.069	0.068	0.068
$\rho_{g\mu}$	0.231	0	0	0.950
$\rho_{\mu d}$	-0.972	0	0.357	0.312
<i>LogL</i>	230.84	215.91	224.33	224.79
<i>LR</i>		29.87	13.02	12.11
<i>p-value (%)</i>		0.00	0.46	0.05
<i>empirical size (%)</i>		60.5	25.8	26.40

3.2.3 Expectation persistence and EIV-problem

Consistent with the literature, e.g., Campbell and Shiller (1988), Campbell (1991) and Cochrane (1992), the largest estimated fraction of price-dividend ratio variation is generated by expected return shocks. The loadings B_1 and B_2 of expected dividend growth and expected returns in the price-dividend ratio (3.7) are in a close relation to their persistence features. Therefore, the relative persistence of dividend and return expectations is a key parameter for quantifying the potential EIV problem in the present-value model.

Figure 3.1 reproduces graphically this link, by plotting the asymptotic EIV-induced bias for dividend and return predictive regressions, as a function of different hypotheses about the relative persistence, $\delta_1 - \gamma_1$, of dividend and return expectations.¹¹ While the bias for the return predictive regression coefficient (top panel) is moderate and less than 20% across all values of $\delta_1 - \gamma_1$, the one for the dividend predictive regression coefficient (bottom panel) is very sensitive to differences in the persistence of the two expectations.

The predictability features implied by a present-value model are also strongly dependent on the relative persistence of the unobservable expected returns and expected dividend growth processes. Figure 3.2 shows the R-squared of returns and dividend growth implied by the present-value model described in Section 3.1 as a function of the difference between the autoregressive coefficients in the expected returns and cash flow growth dynamics. For low values of

¹¹The asymptotic EIV-induced bias for dividend and return predictive regressions in the benchmark present-value model can be computed explicitly, as a function of the model parameters; see Appendix B.3.

Figure 3.1: Model-implied EIV bias

Model-implied EIV bias for standard predictive regressions for returns (upper panel) and dividend growth (lower panel) as a function of the difference between the autoregressive coefficients in the dynamics of expected returns (δ_1) and expected dividend growth (γ_1). Solid blue lines denote the true model-implied value of the regression coefficients, $b_r = -1/B_1$ and $b_d = 1/B_2$, while dashed red lines denote the limit of the OLS estimator of b_r and b_d .

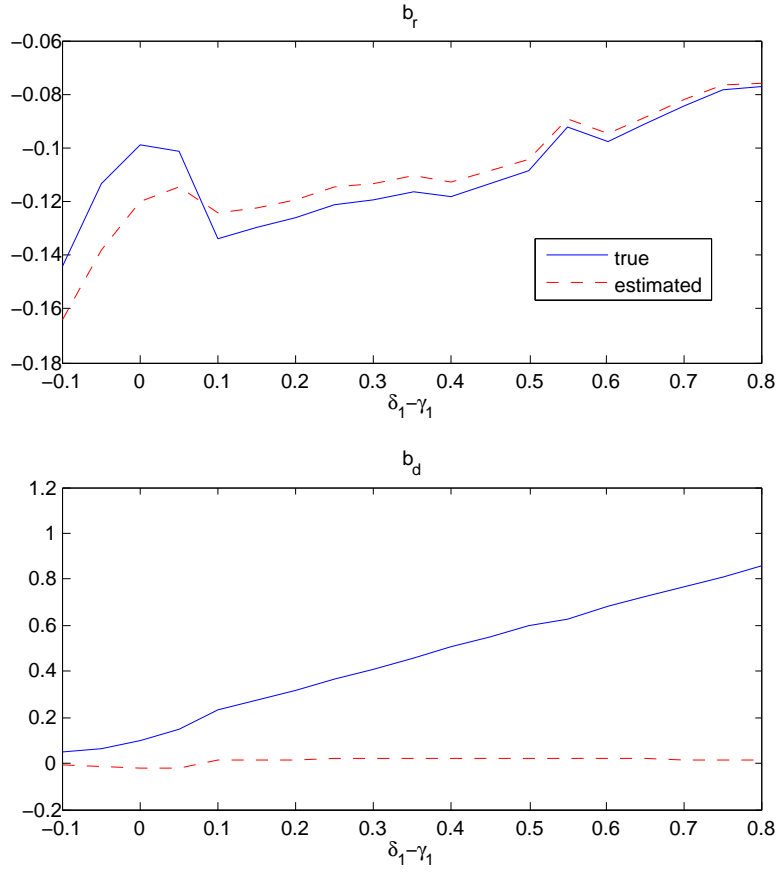
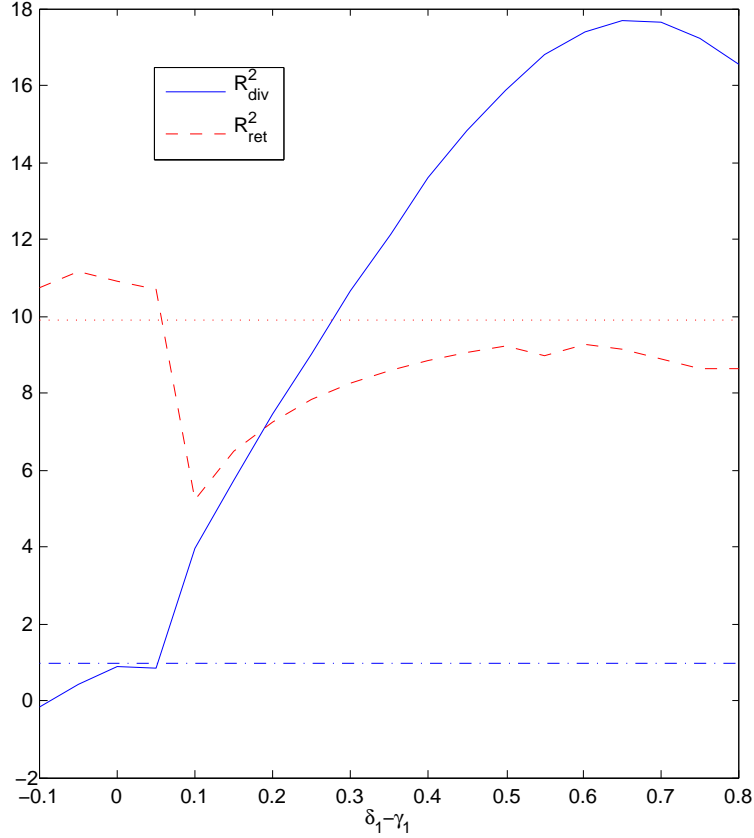


Figure 3.2: Model-implied R-squared as a function of relative persistence

Percentage R-squared of returns (dashed red line) and dividend growth (solid blue line), implied by a simple present-value model as a function of the difference between the autoregressive coefficients in the dynamics of expected returns (δ_1) and expected dividend growth (γ_1). Horizontal lines denote the R-squared of standard predictive regressions of returns (dotted red line) and dividend growth (dash-dotted blue line) on lagged price-dividend ratio for the same sample period, i.e. 1946 to 2010.



the difference between the persistence parameters, the predictability of growth rates implied by the model is almost zero, contrary to what we obtain in Section 3.1.3. Note that under the null hypothesis of equal expectation persistences:¹²

$$H_0 : \gamma_1 = \delta_1 , \quad (3.15)$$

the model implies an identical sensitivity of price-dividend ratios to return and dividend expectation shocks. Under this constraint, the EIV-induced asymptotic bias for both dividend and return predictive regressions is very small.¹³

Based on asymptotic critical values from a χ_r^2 distribution ($r = 1$), we find in Table 3.2 that null hypothesis (3.15) is also clearly rejected, at a significance level α well below 1%. In

¹²At the estimated parameters, the difference between the autoregressive parameters in the expected returns and dividend growth dynamics is equal to 0.623. The parameter constraint (3.15) was also imposed, e.g., for the present-value models in Cochrane (2008a) and Lettau and Van Nieuwerburgh (2008).

¹³Under null hypothesis (3.15), pd_t follows a standard AR(1) process; see also Stambaugh (1999) and Lewellen (2004), among others, and Section A of the Supplemental Appendix.

summary, the evidence from asymptotic tests supports the following predictability features:

- (i) A time-variation in expected returns and expected dividends;
- (ii) A large (a negligible) EIV problem in standard predictive regressions for dividends (returns);
- (iii) A larger degree of predictability, in terms of estimated R^2 s, for dividends than for returns.

3.2.4 Finite-sample reliability of asymptotic tests

How reliable is the asymptotic approximation (3.12) for the probability of rejecting H_0 by chance alone? Since the conventional asymptotics might provide inaccurate results, we investigate the quality of this approximation for tests of predictability in the benchmark present-value model.

3.2.4.1 Nonparametric Monte Carlo bootstrap

Under null hypothesis H_0 , we can consistently estimate the quantiles of statistic LR_T , using a simulation approach that does not rely on distributional assumptions on observed dividend and price-dividend ratio shocks. We borrow from [Stoffer and Wall \(1991\)](#) and apply a nonparametric Monte Carlo bootstrap to the fitted innovations in the present-value model. Details and formal justification for this approach are provided in Section 3.3.¹⁴

We impose the null hypothesis H_0 using the constrained Maximum Likelihood estimator $\hat{\theta}_0$ and we simulate $B = 1000$ time series of dividend growth and price-dividend ratios. Given the actual probability $\alpha_T := P_{H_0}(LR_T \geq \chi_{r,1-\alpha}^2)$ of rejecting H_0 by chance alone in the asymptotic tests, we consistently estimate α_T with the bootstrap estimator:

$$\hat{\alpha}_T := \frac{1}{B} \sum_{b=1}^B \mathbb{I}(LR_{T,b}^* > \chi_{r,1-\alpha}^2), \quad (3.16)$$

where $LR_{T,b}^*$ is the value of the likelihood ratio statistic in simulated bootstrap sample $b = 1, \dots, B$ and $\mathbb{I}(A)$ denotes the indicator function of event A .¹⁵

3.2.4.2 Constant return or dividend expectations

For null hypothesis (3.13) (null hypothesis (3.14)), the first (second) Panel of Figure 3.3 displays the estimated quantiles of the empirical distribution of likelihood ratio statistic (3.10) under H_0 , against the quantiles of the asymptotic χ_4^2 (χ_3^2) distribution.

Apparently, the finite-sample distributions of the test statistics deviate substantially from their asymptotic limit. For instance, while for a significance level $\alpha = 5\%$ the asymptotic critical value for the hypothesis of constant expected dividend growth is $\chi_{3,0.95}^2 = 7.81$, the finite-sample critical value is more than two times larger (17.13).

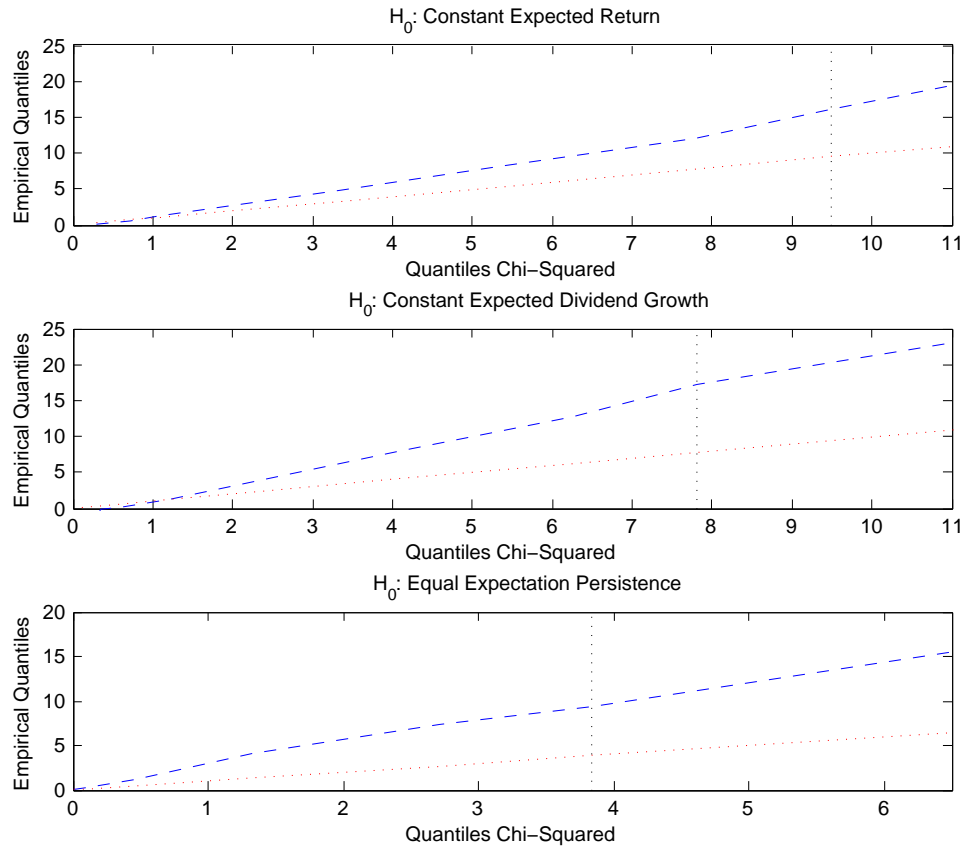
Overall, we find that both asymptotic tests tend to reject H_0 too often. For instance, the estimated probability of rejecting null hypothesis (3.14) (null hypothesis (3.13)) by chance alone in a test of asymptotic significance level $\alpha = 5\%$ is as large as 25.8% (60.5%); see the last row of Table 3.2.

¹⁴[Rytchkov \(2012\)](#) also recognizes that inference based on standard asymptotics may be incorrect and applies a parametric Monte Carlo method with Gaussian shocks to estimate the finite-sample distribution of the LR statistic for the test of no return predictability. We prefer to avoid a parametric Monte Carlo simulation with jointly normal dividend and price-dividend ratios, because for several null hypotheses relevant to our analysis we have found the fitted model residuals under H_0 to deviate quite significantly from normality; see Section 3.2.4.4.

¹⁵ $\mathbb{I}(A)(\omega) = 1$ ($\mathbb{I}(A)(\omega) = 0$) if and only if $\omega \in A$ ($\omega \notin A$).

Figure 3.3: Quantiles of the LR statistics

First (second) panel displays the quantiles of the empirical distribution of the LR statistics for the tests of constant expected returns (dividend growth), while third panel shows the quantiles of the empirical distribution of the LR statistics for the test of equal persistence parameters, all obtained through a nonparametric bootstrap simulation procedure, against the quantiles of the asymptotic chi-squared distribution of the statistics (dotted red line). The vertical dotted line denotes the 95% quantile of this distribution.



3.2.4.3 Equal expectation persistence

For null hypothesis (3.15), the third panel of Figure 3.3 plots the quantiles of the empirical distribution of likelihood ratio statistic (3.10) under H_0 , against those of the asymptotic χ_1^2 distribution. The finite-sample quantiles under H_0 are quite different from their asymptotic limit: The asymptotic test is again excessively liberal. For instance, for a significance level $\alpha = 5\%$ a test of null hypothesis $\delta_1 = \gamma_1$ has an asymptotic critical value $\chi_{0.95,1}^2 = 3.84$, which is less than half the finite-sample critical value of 9.43 estimated with the bootstrap approach. This difference implies a finite-sample probability of rejecting the null by chance alone as large as 26.4%, according to the bootstrap estimate (3.16); see again the last row of Table 3.2.

3.2.4.4 Why a nonparametric bootstrap?

The simplest way to simulate random samples from the present-value model, under the relevant null hypothesis, is by means of a parametric Monte Carlo simulation, e.g., under a normality assumption for the innovations in dividends and returns. In contrast, our nonparametric bootstrap approach renders the inference not dependent on such distributional assumptions. This is an important property, because the empirical error distribution in the present-value model can substantially deviate from normality for some of the relevant null hypotheses.¹⁶

To quantify the empirical deviations from normality, we can test the normality of dividends and price-dividend ratios in the innovation form representation of Section 4.1 for the fitted present-value model. Jarque and Bera (1987) Lagrange multiplier test is among the most common tests of normality. It tests the joint null hypothesis that sample skewness and sample kurtosis equal 0 and 3, respectively. The null hypothesis is rejected whenever the statistic

$$JB = \frac{T}{6} \left(S^2 + \frac{1}{4}(K - 3)^2 \right) \stackrel{a}{\sim} \chi^2(2), \quad (3.17)$$

exceeds the test critical value, where S and K are sample skewness and sample kurtosis, respectively, and T is the sample size.¹⁷

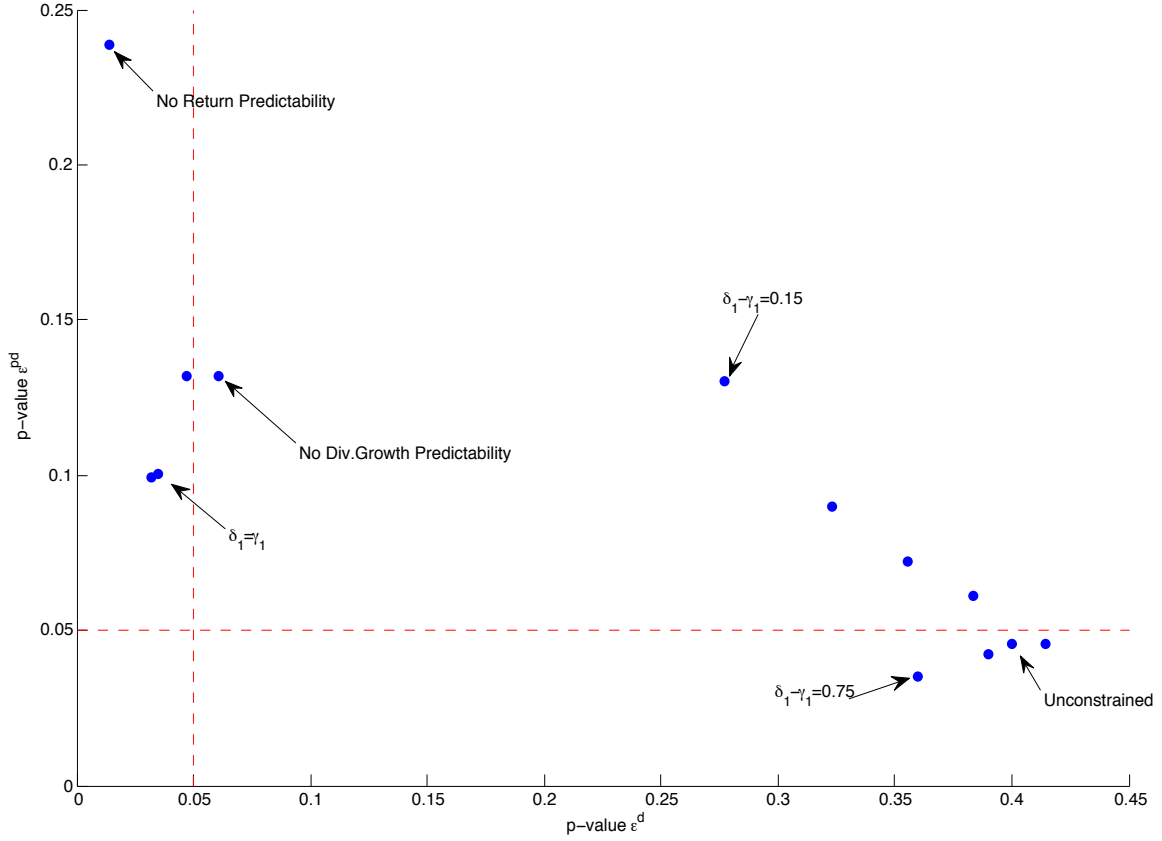
Figure 3.4 plots the p-values of the JB test applied to the standardized dividend and price-dividend innovations in the present-value model, for the unconstrained model and under several null hypotheses. We find that the p-values for the dividend and price-dividend ratio shocks in the unconstrained estimation are 40.01% and 4.54%, respectively. When the parameters are estimated under the null of constant expected returns, constant expected dividend growth and equal persistence, the p-value of the test for the dividend innovation is much lower, with values of 1.43%, 6.09% and 3.51%. In contrast, we never reject the null of normality of the price-dividend ratio innovations in these cases. Figure 3.4 reports also the p-values of tests of different hypotheses about the relative persistence, $\delta_1 - \gamma_1$, of dividend and return expectations. In most cases, we obtain significant evidence of a deviation from normality at the 5% level, for at least one of the shocks in the present-value model. More importantly, we find that the form

¹⁶Unreported empirical evidence also shows that mean, standard deviation and autocorrelation of dividend growth, price-dividend ratio and returns implied by our nonparametric bootstrap tend to be closer to the empirical sample moments than those obtained using a parametric Monte Carlo simulation.

¹⁷While the asymptotic critical values follow from a chi-squared distribution with two degrees of freedom, it has been noted (e.g., in Deb and Sefton (1996)) that the small-sample quantiles of the test statistic are quite different from their asymptotic counterparts. Therefore, we interpolate p-values using critical values computed by Monte Carlo simulation, as provided by the Matlab function *jbtest*.

Figure 3.4: Normality test

Test for normality of the filtered innovations for the unconstrained model and under different null hypotheses (No return predictability, no dividend growth predictability, equal autoregressive coefficients and $\delta_1 - \gamma_1 = 0.05, 0.15, \dots, 0.95$). The two axes show the p-value of the Jarque-Bera test applied to the filtered dividend and price-dividend shocks, respectively.



of the distribution of the fitted innovations in the present-value model is very sensitive to the particular null hypothesis being tested.¹⁸

Overall, we find that a nonparametric bootstrap approach, which does not rely on distributional assumptions about dividend and price dividend ratio shocks, is more appropriate for testing predictability hypotheses in the benchmark present-value model. In such a setting, the Kalman filter remains valid for error distributions different from the normal and it produces the best linear filter, even if global optimality is lost. For estimation purposes, consistency is preserved whenever the first two conditional moments implied by the filter are correctly specified, with other distributional assumptions beyond this being immaterial.

¹⁸In order to understand better the need for a nonparametric procedure, Figures IV and V of the Supplemental Appendix show the bootstrapped distribution of skewness and kurtosis of the filtered innovations, under the unconstrained estimation and for different null hypotheses. The standardized residuals in all cases clearly show either asymmetry or fat tails, or both.

3.3 Bootstrap Tests in the Present-Value Model

A powerful approach to obtain asymptotically valid tests that are less susceptible to finite-sample distortions or specific distributional assumptions, can rely on nonparametric Monte Carlo methods, such as the bootstrap.¹⁹ We first introduce our bootstrap tests of predictability hypotheses in present-value models. We then show their asymptotic validity and quantify by Monte Carlo simulation the improvements over conventional asymptotic tests. Finally, we revisit the conclusions about return and dividend predictability in the benchmark present-value model.

3.3.1 State-space representation

For observed variables $Y_t := (\Delta d_t, pd_t)'$ and expanded state vector $X_t := (\hat{g}_{t-1}, \varepsilon_t^g, \varepsilon_t^\mu, \varepsilon_t^d)'$, where $\hat{g}_t := g_t - \gamma_0$, the present-value model can be written in state-space form (see Appendix B.2):

$$X_{t+1} = FX_t + \Gamma \varepsilon_{t+1}^X, \quad (3.18)$$

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t, \quad (3.19)$$

with matrices F , Γ , M_0 , M_1 , M_2 that are functions of parameter vector

$$\theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{\mu d}, \rho_{gd})'.$$

Let $X_{t,t-1}$ be the best linear prediction of X_t based on observable data $\{Y_s\}_{s=1}^{t-1}$, obtained via the Kalman filter, and $\eta_t = Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}$ the corresponding prediction error. The innovations form representation of the present-value model follows from the Kalman filter as:

$$X_{t+1,t} = FX_{t,t-1} + FK_t \eta_t, \quad (3.20)$$

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_{t,t-1} + \eta_t, \quad (3.21)$$

where the Kalman gain K_t and the conditional covariance matrix S_t of innovation η_t are given explicitly in Appendix B.2.

The advantage of representation (3.20)-(3.21) for an efficient nonparametric bootstrap procedure, is that it allows to easily simulate forward the dynamics of observable variables $\{Y_1, \dots, Y_T\}$, given initial conditions Y_0 , $X_{0,0}$ and random innovations $\{\eta_1, \dots, \eta_T\}$.²⁰

3.3.2 Nonparametric Monte Carlo bootstrap

Let $\hat{\theta}$ and $\hat{\theta}_0$ be the unconstrained and the constrained estimators of the model parameters, obtained by maximizing the likelihood function (B.10) in Appendix B.2 over the full and the H_0 -constrained parameter set, $\hat{\Theta}$ and $\hat{\Theta}_0$, respectively. The observed value of the likelihood ratio statistic LR_T then follows from definition (3.10).²¹

¹⁹As shown in Hall and Horowitz (1996) and Andrews (2002), among others, a desirable property of the bootstrap is that it may provide more accurate finite-sample approximations of the sampling distribution of standard t -test statistics for testing the null of no predictability in predictive regression models. Ang and Bekaert (2007) use bootstrap methods to quantify the bias of standard estimators of regression in predictive regressions of future returns on the lagged price-dividend ratio and interest rate. Amihud, Hurvich, and Whang (2009) compare the performance of bootstrap tests to bias-corrected procedures in multi-predictor regressions and find the two to provide similar finite-sample accuracy.

²⁰In practice, we first apply a nonparametric bootstrap to efficiently simulate the joint distribution of innovations $\{\eta_1, \dots, \eta_T\}$. In a second step, we simulate the joint distribution of $\{Y_1, \dots, Y_T\}$ using the forward dynamics (3.20)-(3.21).

²¹Bootstrap inference is always conditional on the observed sample of data. With a slight abuse of notation, in the sequel we denote by LR_T the sample value of the likelihood ratio statistics.

We apply a nonparametric Monte Carlo bootstrap to the (standardized) innovations $\{\hat{e}_t := S_t^{-1/2}(\hat{\theta})\eta_t(\hat{\theta})\}_{t=1}^T$, in order to obtain the standardized bootstrap residuals $\{\hat{e}_t^*\}_{t=1}^T$. The bootstrap residuals are used to compute a bootstrap distribution of maximum likelihood estimators $\hat{\theta}^*$:

$$\hat{\theta}^* = \arg \max_{\theta} \log \mathcal{L}(\theta, \{Y_t^*\}_{t=1}^T) , \quad (3.22)$$

where the Monte Carlo sequence $\{Y_t^*\}_{t=1}^T$ is simulated with the dynamics (3.20)-(3.21) applied to the unstandardized bootstrap residuals $\{\hat{\eta}_t^* := S_t^{1/2}(\hat{\theta})e_t^*\}_{t=1}^T$. [Stoffer and Wall \(1991\)](#) prove that this approach gives rise to a valid bootstrap distribution for $\sqrt{T}(\hat{\theta}^* - \hat{\theta})$, which is equivalent in large samples to the distribution of $\sqrt{T}(\hat{\theta} - \theta^*)$, where θ^* is the true unknown parameter value. We start from this result, to construct a valid nonparametric bootstrap likelihood ratio test of null hypothesis H_0 in the present-value model.²²

3.3.3 Nonparametric Monte Carlo bootstrap likelihood ratio test

Our bootstrap likelihood ratio test for state-space model (3.18)-(3.19) is based on the following six-steps algorithm.

- 1) Using the estimated parameter vector under null hypothesis H_0 , construct the (constrained) time series of standardized innovations $\{\hat{e}_{0t}\}_{t=1}^T$, by setting:

$$\hat{e}_{0t} = S_t^{-1/2}(\hat{\theta}_0)\eta_t(\hat{\theta}_0) , \quad (3.23)$$

where $S_t^{-1/2}$ is the inverse of the unique square root of S_t .

- 2) Applying a nonparametric bootstrap procedure (such as, e.g., the circular block-bootstrap in [Politis and Romano \(1992\)](#)) to time series $\{\hat{e}_{0t}\}_{t=1}^T$, compute a bootstrap sample $\{\hat{e}_{0t}^*\}_{t=1}^T$ of standardized innovations.
- 3) Using the innovation form representation (3.20)-(3.21), construct a bootstrap sample $\{Y_t^*\}_{t=1}^T$ as follows:

$$X_{t+1,t}^* = F X_{t,t-1}^* + F \mathcal{K}_t S_t^{1/2} \hat{e}_{0t}^* , \quad (3.24)$$

$$Y_t^* = M_0 + M_1 Y_{t-1}^* + M_2 X_{t,t-1}^* + S_t^{1/2} \hat{e}_{0t}^* , \quad (3.25)$$

where matrices F , \mathcal{K}_t , S_t , M_0 , M_1 , M_2 are all evaluated in $\hat{\theta}_0$ and the initial conditions are $Y_0^* = Y_0$, $X_{0,-1}^* = X_{0,0}$.

- 4) Using bootstrap sample $\{Y_t^*\}_{t=1}^T$, compute constrained and unconstrained maximum likelihood point estimates $\hat{\theta}_0^*$ and $\hat{\theta}^*$, respectively, by maximizing the log likelihood function $\log \mathcal{L}(\theta, \{Y_t^*\}_{t=1}^T)$, while imposing and not imposing null hypothesis H_0 , respectively.
- 5) Following definition (3.10), compute the value LR_T^* of the likelihood ratio statistic in the bootstrap sample, defined by:

$$LR_T^* = 2 \left(\log \mathcal{L}(\hat{\theta}^*, \{Y_t^*\}_{t=1}^T) - \log \mathcal{L}(\hat{\theta}_0^*, \{Y_t^*\}_{t=1}^T) \right) . \quad (3.26)$$

²²In a robustness check, [Rytchkov \(2012\)](#) applies a version of a nonparametric bootstrap method, which simulates the full state-space dynamics, in order to test the null hypothesis of constant expected returns in a present-value setting. In contrast, we develop a bootstrap method for the innovation form representation of the state-space model. In addition to producing a less computationally demanding procedure, this approach yields a more transparent bootstrap simulation scheme that allows us to prove the formal asymptotic validity of our approach.

- 6) Repeat steps 2)-5) a large number of times, B , to obtain a collection of bootstrap values of the likelihood ratio statistics, $\{LR_{T,b}^*, 1 \leq b \leq B\}$. The empirical distribution of these values provides an approximation of the distribution of the likelihood ratio statistic under the null hypothesis H_0 .

Remark 1 (i) In step 2) of the algorithm, several bootstrap procedures are applicable to the standardized innovations $\{\hat{e}_{0t}\}_{t=1}^T$. We recommend a time-series bootstrap, such as the circular block-bootstrap, in order to robustify the test against a potentially left time series dependence, not captured by the estimated conditional moment dynamics. ii) In some cases, it may help to exclude the random sampling of the innovations for the first 2-3 data points in step 2) of the algorithm, e.g., by setting $\hat{e}_{0t}^* = \hat{e}_{0t}$ for $t = 1, 2, 3$. This is useful to avoid start-up problems of the algorithm, when the Kalman filter might have an initially transient behavior, e.g., with large values of the Kalman gain K_t .

An important question is whether the proposed bootstrap test delivers correct results in large samples, i.e., whether the bootstrap likelihood ratio statistic LR_T^* follows the same asymptotic distribution as LR_T under H_0 . The next theorem justifies our bootstrap likelihood ratio test of null hypothesis H_0 .

Theorem 1 Under regularity conditions detailed in Appendix B.4, it follows as $B, T \rightarrow \infty$: $LR_T^* \rightarrow \chi_r^2$, in distribution.

According to Theorem 1, the bootstrap statistic LR_T^* has an asymptotically equivalent distribution to LR_T under H_0 .²³ Therefore, it gives rise to bootstrap tests with the correct significance level asymptotically.

The most convenient way to define a bootstrap likelihood ratio test of H_0 is by means of the so-called bootstrap p -value:

$$p^*(LR_T) := P^*(LR_T^* > LR_T) = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(LR_{T,b}^* > LR_T) , \quad (3.27)$$

where P^* denotes the bootstrap probability measure. Using bootstrap p -values, the bootstrap test rejects H_0 whenever:

$$p^*(LR_T) < \alpha . \quad (3.28)$$

From Theorem 1, this test implies the correct asymptotic size α . The interesting question then is whether the bootstrap test delivers more reliable results in finite samples.

A useful property in this respect is that the inference based on bootstrap procedures applied to asymptotically pivotal statistics, such as the likelihood ratio statistic, is generally more accurate than the inference of conventional asymptotics, in the sense that the errors made are of lower order in the sample size T ;²⁴ see Beran (1988), Davidson and MacKinnon (1999b), Hall and Horowitz (1996) and Andrews (2002), among others. As a consequence, we can hope that bootstrap likelihood ratio tests will improve over the conventional asymptotic inference in realistic applications.

²³A similar result can be proven with respect to sequences of shrinking local alternative hypotheses $H_{A,T}$, by applying the above algorithm to innovations defined by $\hat{e}_{At} = S_t^{-1/2}(\hat{\theta}_{A,T})\eta_t(\hat{\theta}_{A,T})$ in step 1), where $\hat{\theta}_{A,T}$ is the constrained maximum likelihood estimator computed under the local alternative $H_{A,T}$.

²⁴A pivotal statistic is a statistic with sampling distribution independent of nuisance parameters.

3.3.4 Finite-sample reliability of bootstrap likelihood ratio tests

In this section, we investigate by Monte Carlo simulation the finite-sample properties of our bootstrap tests, in the context of the benchmark present-value model.

We first impose the null hypothesis H_0 , using the constrained ML estimator $\hat{\theta}_0$, and simulate S time series of dividend growth and price dividend ratios, using our nonparametric bootstrap procedure applied to the fitted innovations in the present-value model. In this way, we can simulate the empirical distribution of observed data, under the null hypothesis H_0 , without making strong parametric assumptions on the joint distribution of dividend and price-dividend ratio shocks. For each simulated time series $s = 1, \dots, S$, we compute the corresponding value of the likelihood ratio statistic, denoted by $LR_{T,s}$.

We then apply our bootstrap testing method to the simulated data. For each simulated time series $s = 1, \dots, S$, we compute a bootstrap distribution of likelihood ratio statistics $LR_{T,b,s}^*$, $b = 1, \dots, B$, and we compute the resulting bootstrap p-value:

$$p^*(LR_{T,s}) = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(LR_{T,b,s}^* > LR_{T,s}) ,$$

following the algorithm in Section 3.3.3. For a significance level $\alpha = 5\%$, we finally compute the frequency of rejections of H_0 in the bootstrap test, i.e., the fraction of time series $s = 1, \dots, S$ in which $p^*(LR_{T,s}) < \alpha$, and compare it to the frequency of rejections obtained by following the asymptotic testing approach. We denote these rejection frequencies by α_T^* and α_T , respectively. Overall, our Monte Carlo simulation is based on a double-bootstrap simulation scheme with $2S(B+1)$ estimations of the parameters in the present-value model, which is a computationally demanding procedure. We present our Monte Carlo results for the parameter choices $S = 200$, $B = 99$ and an optimal bootstrap block size of 2.²⁵ Other parameter choices produce similar results.

For null hypothesis (3.13), we obtain $\alpha_T^* = 5\%$ for the bootstrap test, which is exactly equal to the nominal level ($\alpha = 5\%$), while for null hypothesis (3.14), the empirical size of the bootstrap test is $\alpha_T^* = 8\%$, which is clearly closer to the given nominal level ($\alpha = 5\%$) than the rejection frequency $\alpha_T = 25.8\%$ implied by the asymptotic test. Even though the bootstrap test is slightly too liberal in the Monte Carlo simulation, it does not reject the null of constant dividend expectations in our data. For null hypothesis (3.15), we obtain an empirical rejection frequency $\alpha_T^* = 7.5\%$ for the bootstrap test, which is again clearly lower than the rejection frequency $\alpha_T = 26.4\%$ of the asymptotic test. Also in this case, the bootstrap test corrects the asymptotic critical values in the correct direction, even though it is again slightly too liberal in the Monte Carlo simulation.

3.3.5 The empirical evidence revisited

We make use of the bootstrap likelihood ratio test in Section 3.3.3 and compare the results with those of conventional asymptotic tests in Table 3.2. Based on a bootstrap size $B = 1000$ and an optimal block size of 2, Table 3.3 shows that the null hypothesis of a constant expected return is rejected at a significance level $\alpha = 1\%$ by the bootstrap test, but the null hypothesis of constant expected dividend growth is not rejected, with a bootstrap p-value of 9.5%.

The p-value of the null hypothesis of equal autoregressive coefficients in Table 3.2 is 2.4% (see Table 3.3) for the bootstrap test, compared to the p-value of about 0.05% implied by the

²⁵We apply a data driven calibration method for the selection of the block size, similar to the one introduced in Romano and Wolf (2001) and Camponovo, Scaillet, and Trojani (2009). We choose the block size that minimizes the difference between empirical and nominal size of the bootstrap test for equal expectation persistences.

Table 3.3: p-values of the bootstrap tests

Results of the bootstrap LR tests of constant expected returns ($H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$), constant expected dividend growth ($H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$) and equal autoregressive parameters ($H_0 : \delta_1 = \gamma_1$). *p-value* denotes percentage p-values of the tests.

	$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$	$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$	$H_0 : \delta_1 = \gamma_1$
<i>p-value</i> (%)	0.5	9.5	2.4

standard asymptotic results.²⁶ Thus, null hypothesis (3.15) cannot be rejected at a 1% significance level, but it is significantly rejected at the 5% level by our bootstrap testing procedure. Overall, when considering also the slightly too liberal behaviour of bootstrap tests in our Monte Carlo simulations, the evidence against a similar persistence of expected returns and expected dividend growth is more ambiguous than under the conventional asymptotic tests.

In summary, the non-rejection of null hypotheses (3.14) and (3.15) based on bootstrap tests suggests the following different predictability features, when compared to the findings in Section 3.2 for the asymptotic tests:

- (i) A time-variation in expected returns, but no apparent evidence of a time-varying expected dividend growth;
- (ii) A moderate EIV-problem in dividends and return predictive regressions;
- (iii) Returns and dividend growth predictability features roughly consistent with those of standard predictive regressions.

Section D of the Supplemental Appendix shows that these findings are robust to the choice of the cash-flow proxy, as the estimation and test results are unchanged when using total payouts (dividend plus repurchases) instead of cash dividends.

3.3.6 How much predictability?

The weak evidence of dividend growth predictability produced by bootstrap likelihood-ratio tests raises the question of the interpretation of the large R-squared ($R_{div}^2 = 17.58\%$) estimated in Section 3.1.3 for future dividends.

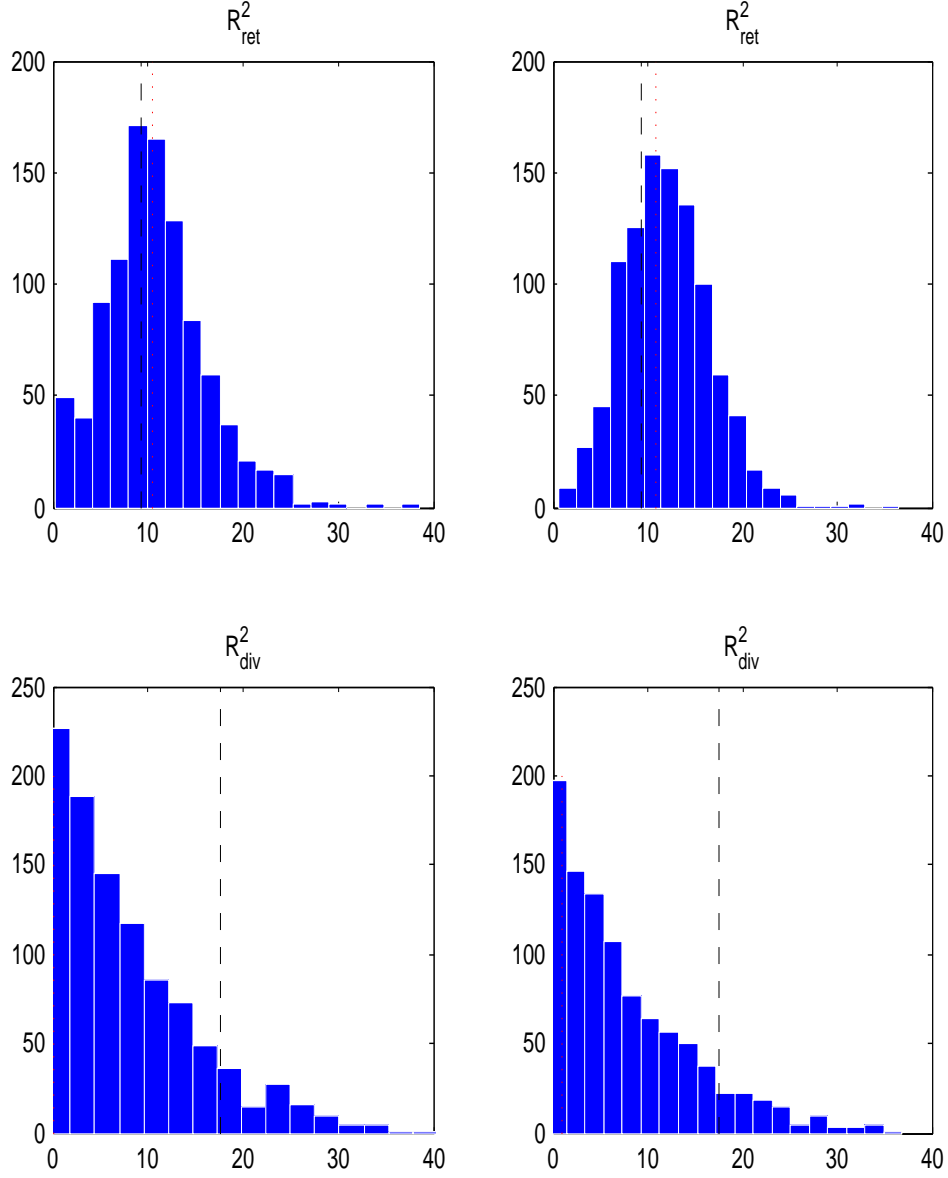
Differently from standard predictive regressions, the asymptotic distribution of estimated R-squares in the present-value model is not known in closed-form. Therefore, the conventional asymptotic approach cannot be used, e.g., to quantify the probability of estimating large R-squares because of chance alone. In contrast, our bootstrap methodology can be applied with no major modification to consistently estimate such probability, under the assumption that an asymptotic distribution for the estimated R-squares exists.

Using steps 1)-3) of the algorithm in Section 3.3.3, we can compute bootstrap estimates of parameter θ in the present-value model and obtain the bootstrap distribution of estimated R-squared statistics under a given null hypothesis H_0 . Figure 3.5 displays the histogram of the bootstrap distribution of estimated R-squares for future returns and future dividend growth, under the null hypotheses of a constant expected dividend growth and an equal persistence of dividend and return expectations, respectively.

²⁶Consistent with the standard asymptotic results, a parametric bootstrap likelihood ratio test with Gaussian errors yields p-values of 0%, 0.3% and 0.2% for the null hypotheses of constant expected return, constant expected dividend growth and equal autoregressive coefficients, respectively.

Figure 3.5: Distribution of R-squares under H_0

Bootstrapped distribution of the R-squared of returns (upper panels) and dividend growth (lower panels), starting from the estimates under the constraint of constant expected dividend growth ($\gamma_1 = \sigma_g = \rho_{g\mu} = 0$, left panels) and of equal persistence parameters ($\delta_1 = \gamma_1$, right panels). Vertical red lines and dashed black lines denote R-squared from constrained and unconstrained estimations on real data, respectively. Distributions are based on 1000 bootstrap samples.



Apparently, the bootstrap distribution of estimated R-squares under the two null hypotheses is similar. Moreover, even though the model-implied R-squared for dividend growth under H_0 is 0% and 0.9%, respectively, we find that the variability of estimated R-squares is quite large. For instance, the median estimated R_{Div}^2 -value is 6.02% under the null of constant expected dividend growth (5.34% under the null of equal persistence parameters) and the most frequently estimated R-squared value is 0% in both cases, but the probability of estimating a dividend R-squared of at least 17.58%, as in the data, is 11.3% (10.5%).

Overall, these findings highlight that finite-sample variability is important for appropriately interpreting the finite-sample distribution of estimated R-squares, as large R-squares as in the data can arise by chance alone, in a present-value model where dividend predictability is absent or weak.

3.3.7 Out-of-sample predictability

All R^2 values reported in the previous sections are estimated using in-sample data. From the perspective of real-time predictability, out-of-sample prediction is an additional important aspect. For instance, Goyal and Welch (2008) study the out-of-sample predictive power of a large set of variables for market returns and find that most of them perform worse than the historical mean.

Following Campbell and Thompson (2008) and Goyal and Welch (2008), the incremental out-of-sample predictive power for returns and dividend growth in the present-value model of Section 3.1 can be estimated using the metrics:

$$R_{Ret,OS}^2 = 1 - \frac{\sum_{t=0}^T (r_{t+1} - \tilde{\mu}_t)^2}{\sum_{t=0}^T (r_{t+1} - \bar{r}_t)^2}, \quad (3.29)$$

$$R_{Div,OS}^2 = 1 - \frac{\sum_{t=0}^T (\Delta d_{t+1} - \tilde{g}_t)^2}{\sum_{t=0}^T (\Delta d_{t+1} - \bar{\Delta d}_t)^2}, \quad (3.30)$$

where $\tilde{\mu}_t$ and \tilde{g}_t are the estimated expected return and expected dividend growth in the present-value model, using observations up to time t , while \bar{r}_t and $\bar{\Delta d}_t$ are the sample means of returns and dividend growth using data up to time t .

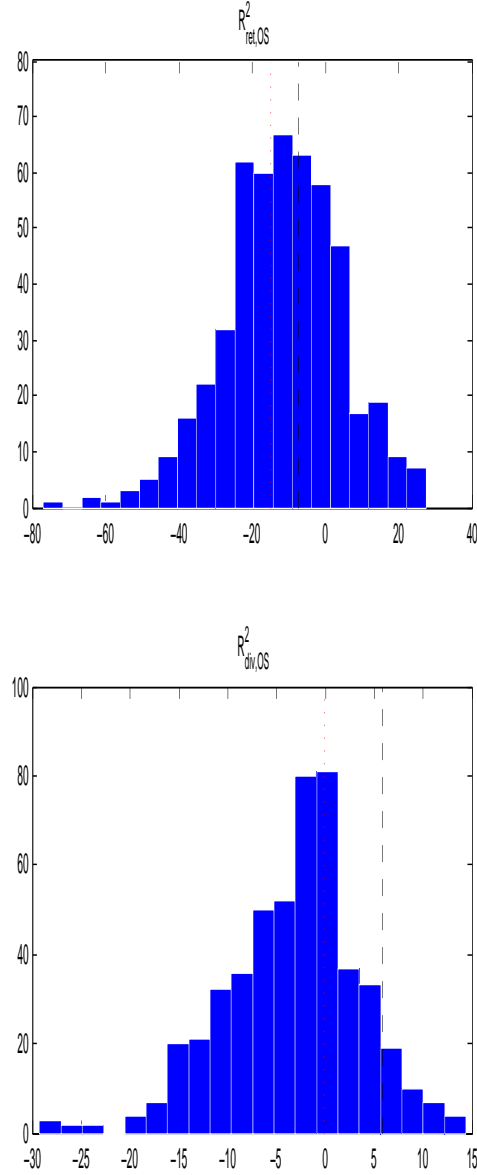
We estimate the degree of out-of-sample predictability according to measures (3.29) and (3.30), using an out-of-sample period starting in 1985. Standard predictive regressions of returns and dividend growth on the lagged price-dividend ratio yield $R_{Ret,OS}^2 = -12.32\%$ and $R_{Div,OS}^2 = -4.38\%$, while we obtain $R_{Ret,OS}^2 = -7.31\%$ and $R_{Div,OS}^2 = 5.88\%$ for the present-value model.²⁷ Thus, the point estimates for the benchmark present-value model might indicate an incremental degree of out-of-sample predictability for dividend growth with respect to the sample mean forecast.

Using a slight modification of our bootstrap method, we can estimate the distribution of out-of-sample R-squares (3.29) and (3.30) under the null of no return or dividend predictability; details of the procedure are given in Appendix B.5. This approach is useful, e.g., to better quantify the probability of estimating large out-of-sample R^2 values as in the data by chance alone.

²⁷Precisely, we use data between 1946 and 1985 to estimate the parameters of the model and compute expected return and expected dividend growth for 1986, which are compared to the realized return and dividend growth in the same year. We then use data between 1946 and 1986 to compute predictions for 1987 and proceed in this way until the end of the sample. Using data from 1946 to 2007 and starting the out-of-sample computations in 1972, Binsbergen and Koijen (2010) find $R_{Ret,OS}^2 = 1.06\%$ and $R_{Div,OS}^2 = 5.76\%$.

Figure 3.6: Distribution of out-of-sample R-squares under H_0

Bootstrapped distribution of the out-of-sample R-squared of returns (upper panel) and dividend growth (lower panel), starting from the estimates under the constraint of constant expected dividend growth ($\gamma_1 = \sigma_g = \rho_{g\mu} = 0$). Vertical red lines and dashed black lines denote out-of-sample R-squared from constrained and unconstrained estimations on real data, respectively. Distributions are based on 1000 bootstrap samples.



Given the variability of estimated in-sample R_{Div}^2 values highlighted in Section 3.3.6, it is plausible that the out-of-sample R-squared distribution might inherit similar features. Figure 3.6 illustrates the properties of the bootstrap distributions of out-of-sample R-squares (3.29) and (3.30), generated under the null hypothesis of constant expected cash flow growth in the present-value model. Both distributions imply a large variability of estimated out-of-sample measures of predictability for returns (upper panel) and dividend growth (lower panel). Even though under H_0 expected returns are time-varying, the estimated $R_{Ret,OS}^2$ distribution puts a large mass in regions where no evidence of incremental predictability is estimated. Moreover, despite the absence of dividend predictability under the null, the distribution of estimated $R_{Div,OS}^2$'s puts a significant mass of about 15% in regions of positive $R_{Div,OS}^2$ values, with a probability of estimating an out-of-sample R-squared for dividends at least as large as in the data that is almost 10%.

Overall, these findings show that the conclusions produced by estimated common measures of out-of-sample predictability in present-value models have to be taken with caution and put in relation to the finite-sample variability of these quantities under the null of no predictability. On the one side, the limited amount of data information available can lead to a difficulty in detecting predictive relations for returns when they are there. On the other side, high out-of-sample R-squares for dividends can arise by chance alone, in a setting with constant expected dividend growth. In this respect, our nonparametric bootstrap approach provides a useful tool to better interpret also the information provided by estimated out-of-sample measures of predictability.

3.4 Broader Specifications of the Predictive Information Set

While the benchmark present-value model in Section 3.1 is useful for highlighting the main issues of tests of predictability hypotheses, it might not provide the most accurate description for the dynamics of dividend-return expectations and their link to price-dividend ratios. Richer specifications might improve the evidence of predictability and it is useful to study the robustness of our previous results, with respect to an enlarged specification of the predictive information set.

Several potential predictors have been considered in the literature, to improve the statistical evidence of univariate predictive regressions with the lagged price-dividend ratio.²⁸ Such predictive variables can naturally extend the benchmark present-value model, in order to parsimoniously aggregate the joint information generated by the time series of dividend growth, price-dividend ratios and additional predictors, following the present-value approach proposed in Yun (2012).

Using the conventional asymptotic approach, variables such as the book-to-market ratio (*bm*), the stock market variance (*svar*), the consumption-wealth-income ratio (*cay*) and the BAA-rated corporate bond yield (*BAA*) significantly improve the forecasts of future returns and future dividend growth in the present-value model.²⁹ Using our general bootstrap tests of Section 3.3, we study the robustness of our findings on dividend and return predictability, with respect to the choice of the predictive information set.

²⁸Goyal and Welch (2008) and Kojien and Van Nieuwerburgh (2011) give an excellent review of this literature. Even though less studies have focused on dividend growth predictability, Lettau and Ludvigson (2005) and Favero, Gozluklu, and Tamoni (2011), among others, provide evidence that predictive variables like *cay* and proxies of demographics help forecasting cash flow growth.

²⁹ The benchmark present-value model assumes a constant return volatility. Piatti and Trojani (2012b) develop a present-value approach with time-varying return and dividend growth risks to predictive regression.

3.4.1 The present-value model with additional predictive variables: estimation results

Expected dividend growth, expected return and an additional predictive variable, z_t , follow the following first-order vector autoregression:

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \gamma_2(z_t - \xi_0) + \varepsilon_{t+1}^g, \quad (3.31)$$

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \delta_2(z_t - \xi_0) + \varepsilon_{t+1}^\mu, \quad (3.32)$$

$$z_{t+1} = \xi_0 + \xi_1(z_t - \xi_0) + \varepsilon_{t+1}^z. \quad (3.33)$$

In contrast to the benchmark dynamics (3.3)-(??), the additional predictive variable z_t can help to better explain expected returns or expected dividend growth. As such, it appears in the price-dividend ratio implied by a standard Campbell and Shiller (1988) log linearization:

$$pd_t = A - (B_1\hat{\mu}_t + B_3\hat{z}_t) + (B_2\hat{g}_t + B_4\hat{z}_t), \quad (3.34)$$

where $B_3 = \frac{\delta_2}{\delta_1 - \xi_1} \left(\frac{1}{1 - \rho\delta_1} - \frac{1}{1 - \rho\xi_1} \right)$, $B_4 = \frac{\gamma_2}{\gamma_1 - \xi_1} \left(\frac{1}{1 - \rho\gamma_1} - \frac{1}{1 - \rho\xi_1} \right)$ and $\hat{z}_t = z_t - \xi_0$ is the demeaned additional predictive variable at time t ; see, e.g., Yun (2012).

The model is again estimated in state-space form with a Kalman filter.³⁰ For brevity, we report results only for additional predictive variables that significantly predict returns and dividend growth using standard asymptotic tests. These include the book-to-market ratio (*bm*), the stock market variance (*svar*), the consumption-wealth-income ratio (*cay*) and the corporate bond yield on BAA-rated bonds (*BAA*). The description of the variables is provided by Goyal and Welch (2008) and their updated time series through 2010 are available at Goyal's website.³¹

Estimated present-value model parameters and R-squares for returns and dividend growth are collected in Table 3.4, together with the R-squared estimated from standard predictive regressions with the additional predictive variable z_t . In each present-value model, the predictive information set enlarged by the additional predictor z_t increases the estimated R-squares for dividends and returns, relative to the findings for the benchmark model in Section 3.1.3. While estimated R-squares for returns are similar to those obtained from the standard predictive regressions in Panel C of Table 3.4, the estimated R-squared values for dividend growth are much higher, consistently with the findings of Section 3.1.3 for the benchmark present-value model.

3.4.2 Tests of constant dividend and return expectations

Cash flow predictability is again tested by testing the null hypothesis of constant expected dividend growth. In the extended present-value model, this hypothesis is equivalent to the following constraints, which are tested using a standard *LR* statistic that is asymptotically χ_5^2 distributed:

$$H_0 : \gamma_1 = \gamma_2 = \sigma_g = \rho_{g\mu} = \rho_{gz} = 0. \quad (3.35)$$

We test this null hypothesis for $z_t = bm$ and $z_t = svar$, which are the variables that seem to increase more model-implied dividend growth predictability, measured in terms of R-squared, compared to the benchmark model (see again Panel B of Table 3.4). Panel B of Table 3.5 shows that the asymptotic likelihood ratio test rejects null hypothesis (3.35) for both choices of predictive variable z_t , with a p-value below 0.5%.

³⁰For completeness, Appendix B.2 also describes the state-space representation and the Kalman filter estimation procedure for the present-value model with the additional predictor z_t .

³¹See the web page <http://www.hec.unil.ch/agoyal/>.

Table 3.4: Estimation results for the extended model

Panel A reports estimation results of the present-value model in Section 3.4.1, using as predictor variables the book-to-market ratio (BM), stock variance ($SVAR$), CAY and the BAA corporate bond yield (BAA), respectively. The models are estimated using annual data from 1946 to 2010. Panel B reports the model-implied R-squared values for return and dividend growth, in percentage, computed as in (3.8)-(3.9), while Panel C reports R-squared from standard OLS predictive regressions of returns and dividend growth on lagged price-dividend ratio and each predictive variable z_t

	BM	$SVAR$	CAY	BAA
Panel A: Maximum-likelihood estimates				
γ_0	0.051	0.056	0.050	0.057
δ_0	0.071	0.129	0.077	0.090
γ_1	0.234	0.475	0.338	0.296
δ_1	0.878	0.993	0.926	0.920
σ_g	0.064	0.078	0.066	0.065
σ_μ	0.016	0.018	0.031	0.018
σ_d	0.013	0.018	0.015	0.008
$\rho_{g\mu}$	0.220	-0.454	-0.308	0.177
$\rho_{\mu d}$	-0.144	-0.758	-0.596	-0.167
ξ_0	0.487	0.020	0	0.082
ξ_1	0.913	0.418	0.733	0.939
ρ_{gz}	-0.298	-0.764	-0.528	-0.250
$\rho_{\mu z}$	0.567	0.804	0.882	0.485
ρ_{dz}	0.597	-0.594	-0.775	0.702
σ_z	0.102	0.021	0.014	0.010
δ_2	0.019	-0.285	-0.396	-0.025
γ_2	0.067	1.652	0.837	-0.014
Panel B: Model-implied R-squared				
R^2_{ret}	10.13	9.70	17.58	9.65
R^2_{div}	22.32	25.71	19.81	18.29
Panel C: Predictive regression R-squared				
R^2_{ret}	10.34	12.24	15.40	13.81
R^2_{div}	4.26	8.02	0.95	3.03

Table 3.5: Dividend growth predictability test for the extended model

Test of no dividend growth predictability in the context of the present-value model in Section 3.4.1. Panel A reports constrained estimation results, using as predictor variables the book-to-market ratio (BM) and stock variance ($SVAR$), respectively. The models are estimated using annual data from 1946 to 2010. Panel B reports the p-values of the test, using the asymptotic distribution of the LR statistic, and the effective size of the asymptotic test, for a nominal size $\alpha = 5\%$, while Panel C reports the p-values of the bootstrap test. Finite sample size computations and bootstrap tests are based on 1000 bootstrap samples.

	BM	$SVAR$
Panel A: Constrained Maximum-likelihood estimates		
γ_0	0.056	0.055
δ_0	0.078	0.094
γ_1	0	0
δ_1	0.887	0.960
σ_g	0	0
σ_μ	0.019	0.031
σ_d	0.068	0.068
$\rho_{g\mu}$	0	0
$\rho_{\mu d}$	0.382	-0.029
ξ_0	0.504	0.021
ξ_1	0.904	0.363
ρ_{gz}	0	0
$\rho_{\mu z}$	0.681	0.927
ρ_{dz}	0.136	-0.216
σ_z	0.101	0.022
δ_2	0.006	-0.701
γ_2	0	0
Panel B: Asymptotic test		
$p - value$ (%)	0.22	0.19
$empirical\ size$ (%)	22.70	28.60
Panel C: Bootstrap test		
$p - value$ (%)	7.40	10.20

To apply our bootstrap testing approach, we introduce the extended vectors of observed variables $Y_t := (\Delta d_t, pd_t, z_t)'$ and state variables $X_t := (\hat{g}_{t-1}, \epsilon_t^g, \epsilon_t^\mu, \epsilon_t^d, \epsilon_t^z)'$, in order to write the present-value model (3.31)-(3.33) in state-space form (see Appendix B.2):

$$X_{t+1} = FX_t + Bu_{t+1} + \Gamma \epsilon_{t+1}^X, \quad (3.36)$$

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t, \quad (3.37)$$

with parameter-dependent matrices $F, B, \Gamma, M_0, M_1, M_2$ and variable $u_t := z_{t-1} - \xi_0$.

Given $X_{t,t-1}$ the best linear prediction of X_t based on data $\{Y_s\}_{s=1}^{t-1}$ and $\eta_t = Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}$, the innovations form representation of model (3.31)-(3.33) follows from the Kalman filter:

$$X_{t+1,t} = FX_{t,t-1} + Bu_{t+1} + FK_t \eta_t, \quad (3.38)$$

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_{t,t-1} + \eta_t, \quad (3.39)$$

where the Kalman gain K_t is given explicitly in Appendix B.2. From this dynamics, the bootstrap likelihood ratio test in the extended present-value model is performed with the algorithm presented in Section 3.3.3.³²

Panel C of Table 3.5 shows that the bootstrap likelihood ratio test produces different conclusions from the asymptotic test. The bootstrap test p-values are always bigger than the asymptotic p-values and we can never reject null hypothesis (3.35) at the 5% significance level, indicating that the evidence of dividend growth predictability is similarly weak in the extended present-value models, as it was in Section 3.3.5 for the benchmark model.

The null hypothesis of no return predictability in the extended present-value model is equivalent to the following parametric constraints:

$$H_0 : \delta_1 = \delta_2 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = \rho_{\mu z} = 0. \quad (3.40)$$

For brevity, we test again this null hypothesis using the two predictive variables that mostly increase the return predictability evidence, as measured by the model-implied R_{Ret}^2 , namely $z_t = bm$ and $z_t = cay$; see again Panel B of Table 3.4. Panel B of Table 3.6 shows that the asymptotic likelihood ratio test rejects null hypothesis (3.40) for all choices of the predictive variable z_t , with a p-value below 0.05%. The p-values for the bootstrap test are reported in Panel C of Table 3.6. Consistently with the asymptotic test results and the bootstrap test results of in Section 3.3.5 for the benchmark model, null hypothesis (3.40) is again clearly rejected, with p-values of about 0.5%.

3.4.3 Variability of estimated R-squared values

To explain the weak evidence of dividend growth predictability and the large estimated dividend R-squares in the extended present-value models, Figure VI of the Supplemental Appendix plots the bootstrap distribution of estimated R-squares for returns and future dividend growth, simulated under null hypothesis (3.35), for two different choices $z_t = bm$ (left panels) and $z_t = svar$ (right panels) of the additional predictive variable.

The bootstrap distribution of estimated R-squares under the null of constant expected dividend growth is similar to the one estimated in the benchmark model, with a large variability

³²To run the bootstrap algorithm in the extended present-value model, we replace equation (3.20) in step 3) of the algorithm in Section 3.3 by the following bootstrap simulation scheme:

$$X_{t+1,t}^* = FX_{t,t-1}^* + Bu_{t+1} + FK_t S_t^{1/2} \hat{e}_{0t}^*,$$

using parameter matrices detailed in Appendix B.2.

Table 3.6: Return predictability test for the extended model

Test of no return predictability in the context of the present-value model in Section 3.4.1. Panel A reports constrained estimation results, using as predictor variables the book-to-market ratio (BM) and CAY , respectively. The models are estimated using annual data from 1946 to 2010. Panel B reports the p-values of the test, using the asymptotic distribution of the LR statistic, and the effective size of the asymptotic test, for a nominal size $\alpha = 5\%$, while Panel C reports the p-values of the bootstrap test. Finite sample size computations and bootstrap tests are based on 1000 bootstrap samples.

	BM	CAY
Panel A: Constrained Maximum-likelihood estimates		
γ_0	0.070	0.072
δ_0	0.075	0.080
γ_1	0.960	0.996
δ_1	0	0
σ_g	0.008	0.003
σ_μ	0	0
σ_d	0.067	0.069
ξ_0	0.234	-0.006
ξ_1	0.997	0.857
ρ_{gz}	0.811	-0.796
$\rho_{\mu z}$	0	0
ρ_{dz}	0.215	-0.253
σ_z	0.104	0.014
δ_2	0	0
γ_2	-0.001	0.015
Panel B: Asymptotic test		
$p - value$ (%)	0.01	0.00
$empirical\ size$ (%)	14.10	28.10
Panel C: Bootstrap test		
$p - value$ (%)	0.50	0.30

of estimated R-squares. The increased predictive information generated by z_t tends to rise the probability of correctly estimating an R-squared of 0% for dividend growth under the given null hypothesis for $z_t = bm$, while the distribution of R_{Div}^2 displays more variability for $z_t = svar$. To illustrate, while the median estimated R_{Div}^2 -value is 6.23% for $z_t = bm$ (9.41% for $z_t = svar$), the most frequently estimated R-squared value is 0%, but the probability of estimating a dividend R-squared of at least 22.32% (25.71%), as in the data, is still as large as 8.20% (10.60%). In summary, finite-sample variability again produces large estimated R-squares by chance alone, within a present-value model where dividend predictability is absent.

3.5 A Tale of Two Periods

The time series of US aggregate dividend growth for the prewar and the postwar periods exhibit substantially different properties, suggesting a potential structural break in the dividend process between these two sample periods. While tests based on standard predictive regressions for the postwar sample find no evidence of dividend predictability, the evidence is reversed for the prewar sample; see [Chen \(2009\)](#), among others. Therefore, it is natural to test whether our bootstrap tests of predictability in present-value models can produce consistent results for both the prewar and postwar samples.

We can parsimoniously account for the structural break in the parameters of the present-value model, between the prewar and the postwar samples, by allowing the persistence and the variability of expected returns and dividend growth, parameterized by δ_1 , γ_1 , σ_g , σ_μ and σ_d , respectively, to differ before and after 1946. The parameter estimates and p-values for the *LR* tests of predictability in the prewar and postwar samples are collected in [Table 3.7](#).

The estimation results support the evidence of a regime shift in the parameters of the dividend process, approximately in 1946, since expected dividend growth is estimated as much more volatile in the prewar sample. Similarly, expected returns are estimated as much less persistent before 1946. The asymptotic LR test clearly rejects the null of no dividend predictability in the prewar and the postwar samples, with a p-value of 0% and 0.76%, respectively. The asymptotic test also rejects the null of no return predictability for the prewar and the postwar samples, with a p-value of 0.51% and 0%, respectively.

The results of the bootstrap test again indicate that asymptotic tests in present-value models tend to overreject the null of no predictability, since all bootstrap p-values are larger than the p-values of the corresponding asymptotic test. However, while the null of no dividend predictability cannot be rejected in the postwar sample at the 10% significance level, it is clearly rejected with a bootstrap p-value of 2% in the prewar sample. In contrast, the null of no return predictability is clearly rejected by the bootstrap test, both for the prewar and the postwar samples, with a p-value of 2% and 0%, respectively.

In summary, when accounting for a regime shift in the parameters of the dividend and return processes, our bootstrap test reconciles the conclusions produced by standard predictive regressions and present-value models, producing dividend predictability findings consistent with the evidence in [Chen \(2009\)](#), among others. The findings for the prewar sample also indicate that our bootstrap test has power to detect dividend and return predictability structures based on a quite limited data information, as it rejects the null of no predictability in the prewar sample, both for dividends and returns, using the information provided by only about 20 yearly observations.

Table 3.7: Predictability tests assuming two regimes

Unconstrained and constrained ML estimates of the present-value model using long sample (1927-2010) and assuming a regime shift in 1946 for the persistence and volatility parameters. The table also shows LR statistics for the tests of constant expected returns (all sample, only prewar or postwar) and constant expected dividend growth (all sample, only prewar and postwar). *LogL* denotes the pseudo log-likelihood obtained and *LR* is the value of the Likelihood Ratio statistic. *As - pval* and *Boot - pval* denote percentage p-values of the asymptotic and bootstrap *LR* tests, respectively.

	Unconstr	No Ret Pred - all	No Ret Pred - pre	No Ret Pred - post	No Div Pred - all	No Div Pred - pre	No Div Pred - post
γ_0	0.056	0.053	0.045	0.048	0.055	0.059	0.058
δ_0	0.102	0.091	0.091	0.091	0.083	0.109	0.102
γ_{1p}	0.255	0.556	0.420	0.156	0	0	0.173
γ_{1d}	0.380	0.998	0.182	0.997	0	0.233	0
δ_{1p}	0.614	0	0	0.623	0.988	0.608	0.624
δ_{1d}	0.973	0	0.973	0	0.905	0.980	0.955
σ_{gp}	0.220	0.141	0.164	0.230	0	0	0.219
σ_{gd}	0.065	0.007	0.055	0.008	0	0.054	0
$\sigma_{\mu p}$	0.049	0	0	0.049	0.013	0.125	0.047
$\sigma_{\mu d}$	0.011	0	0.011	0	0.021	0.011	0.015
σ_{dp}	0.001	0.138	0.155	0.005	0.227	0.235	0.027
σ_{ddp}	0.011	0.072	0.035	0.074	0.069	0.038	0.068
$\rho_{g\mu}$	0.300	0	0	0	0	0	0
$\rho_{\mu d}$	-0.364	0	0	0	0.404	0.315	0.350
<i>LogL</i>	279.08	257.47	271.67	261.40	257.82	261.41	273.10
<i>LR</i>		43.22	14.81	35.36	42.52	35.34	11.95
<i>As - pval (%)</i>		0.00	0.51	0.00	0.00	0.00	0.76
<i>Boot - pval (%)</i>		0.00	2.00	0.00	0.00	0.00	13.00

3.6 Conclusion

The Campbell and Shiller (1988) present-value logic, implying that price-dividend ratios vary because of shocks to expected returns or expected dividend growth, has motivated a vast literature studying the predictability of market returns and aggregate dividend growth. Univariate predictive regressions of future returns and dividend growth on predictive variables including the lagged price-dividend ratio have produced no apparent evidence of dividend predictability in the postwar sample, suggesting that price-dividend ratios have mostly varied because of discount rate shocks in that period. In contrast, latent variable approaches within present-value models, which parsimoniously incorporate information from the joint time-series of dividends and returns, have found a stronger evidence of a time-varying expected dividend growth.

A natural explanation for these contrasting conclusions is the error-in-variable (EIV) problem inherent to predictability studies, which can be explicitly modelled using the present-value relations that connect return and dividend growth dynamics to the price-dividend ratio. This paper provides sharp evidence for a different explanation, linked to the so far unexplored finite-sample properties of conventional tests of predictability in models with latent return and dividend expectations. Using a nonparametric Monte Carlo simulation approach, which avoids restrictive distributional assumption, such as, e.g., a normal distribution for dividend and return shocks, we show that the conventional tests have similar finite-sample drawbacks as many tests of predictability in predictive regressions with lagged persistent predictors and correlated innovations.

First, we show that conventional tests frequently reject the null of no dividend predictability because of chance alone. Moreover, we find that large estimated R-squares for dividends can arise by chance alone, even under the null of a constant expected dividend growth. These findings stress the importance of combining a pure estimation approach with a reliable testing method, when testing and quantifying the actual degree of predictability within present-value models.

Second, in order to introduce a general and more reliable testing approach, we propose a class of nonparametric bootstrap tests of predictability hypotheses in present-value models, by applying the bootstrap to the innovations from the latent state dynamics, generated under the relevant null hypothesis. We prove that the bootstrap tests imply a valid asymptotic inference and demonstrate their improved properties in finite samples. Precisely, we find that the bootstrap test can better control the finite-sample probability of rejecting a null hypothesis by chance alone, thus producing a more reliable predictability evidence in a number of applications.

Third, we apply our bootstrap tests to US stock market data, based on a variety of specifications of the predictive information set. In contrast to the results implied by standard asymptotic tests, we find a significant evidence in favour of time-varying expected returns, both in the prewar and the postwar samples, no evidence of time-varying dividend expectations in the postwar sample and a strong dividend predictability in the prewar sample. This evidence is consistent with the one implied by standard predictive regression, thus reconciling the diverging conclusions in the literature, and it indicates that the major source of price-dividend ratio variation in postwar US data are discount rate shocks.

We finally propose a slight modification of our bootstrap testing method, which can be used also to test the presence of out-of-sample predictability, while controlling the probability of detecting predictive relations by chance alone. We find that the conclusions produced by estimated common measures of out-of-sample predictability in present-value models have to be taken with caution and need to be set in relation to the finite-sample variability of these quantities under the null of no predictability. In this respect, our nonparametric bootstrap testing method provides a useful tool for more comprehensively interpreting also the information

provided by estimated measures of out-of-sample predictability.

From a broader methodological perspective, our bootstrap testing approach and our results have implication for a number of potentially more general aspects. First, while our bootstrap tests can help to control more systematically the probability of rejecting a null hypothesis by chance alone, our results also indicate that the information generated by the joint time series of stock market returns and dividends might be insufficient to reliably identify time-variations in dividend expectations, i.e., tests of dividend predictability in such settings may have a low power.

A low power might arise because of the short time series available for many predictability studies or because market price-dividend ratios aggregate into a single observable signal the expectations of future dividends for different horizons, which are potentially difficult to identify separately. As shown in [Binsbergen, Brandt, and Koijen \(2012\)](#) and [Binsbergen, Hueskes, Koijen, and Vrugt \(2013\)](#), a more direct identification of dividend expectations at distinct horizons can rely on the equity yield of dividend strips, which are dividend claims for single maturities. Annual dividend growth is strongly predictable in the period from October 2002 to April 2011, with univariate predictive regression R^2 s between 48% for the 5 year yield and 76% for the 1 year yield. This evidence suggests that dividend strip information can potentially improve the power of tests of dividend predictability more generally. Unfortunately, quotes of liquid dividend claims are available only since recently. Another possibility is to replicate synthetically the prices of dividend strips from quoted index option or futures prices, in which case data are available starting approximately in 1986; see, e.g., [Binsbergen, Brandt, and Koijen \(2012\)](#).

The study of reliable inference methods in present-value models estimating the joint dynamics of dividend growth, stock returns and dividend strip returns is an interesting direction for future research. Also in this domain, our bootstrap testing methods can prove useful in order to better control the probability of rejecting a null of no predictability by chance alone.

Finally, our bootstrap testing method is also applicable more generally, in order to more reliably test the relevant null hypotheses in models estimated by a latent variable approach using their state-space form. Concrete but not exhaustive examples of possible applications include the testing of the expectation and similar hypotheses in the context of affine factor models for the yield curve (see, e.g., [Piazzesi \(2010\)](#) for a review) or tests of predictability hypotheses in present-value models with time-varying return and dividend risks (see, e.g., [Piatti and Trojani \(2012b\)](#)).

Predictable Risks and Predictive Regression in Present-Value Models

Introduction

Within a tractable present-value model with time-varying cash flow and discount rate risks, we propose a latent variable approach to estimate the joint predictability features of aggregate dividends, market returns and their conditional risks. Given latent exogenous time series processes for expected returns, expected dividend growth and the covariance matrix of cash flow and discount rate shocks, we derive the implied price-dividend ratio dynamics following Campbell and Shiller (1988). We then use a Kalman filter to estimate the model parameters by Quasi Maximum Likelihood (QML). This approach allows us to aggregate information from the history of dividend growth, price-dividend ratios and the covariance matrix between returns and dividends, in order to uncover expected returns and dividend growth rate dynamics that are coherent with their conditional risk features.

We first find that expected dividend growth and expected returns are both time-varying, and explain an economically relevant fraction of actual dividend growth and future returns, with average R^2 values of about 2.6% and 9.4%, respectively. However, the evidence for dividend growth predictability does not seem to be statistically significant. The estimated expected return is more persistent than expected dividend growth and gives rise to a large price-dividend ratio component, which masks the predictive power of valuation ratios for future dividend growth in standard predictive regressions. Second, we study the time-varying risk features and we uncover useful additional model implications for the estimated dynamics of market Sharpe ratios and the term structure of expectations and risks: We find quite volatile market Sharpe ratios, which are more countercyclical than under the assumption of constant risks, and highly time-varying term structures of both expectations and volatilities of long-horizon dividend growth and returns, which have the flexibility to take different shapes. Third, using a nonparametric Monte Carlo bootstrap approach, we show that our model findings are consistent with a number of well-known predictability features estimated using standard OLS predictive regression approaches, including e.g. the very weak post-war evidence of dividend growth predictability at yearly horizons and the low real-time predictability of stock returns.

Our approach builds on the recent literature advocating the use of present-value models to jointly uncover market expectations for returns and dividends, including Menzly, Santos, and Veronesi (2004), Lettau and Ludvigson (2005), Ang and Bekaert (2007), Lettau and Van Nieuwerburgh (2008), Campbell and Thompson (2008), Pastor, Sinha, and Swaminathan

(2008), Rytchkov (2012), Cochrane (2008a,b), Ferreira and Santa-Clara (2010) and Binsbergen and Koijen (2010), among others. We add to this literature by introducing a tractable present-value model incorporating the latent time-varying features of return and dividend risks, in which we study the implications for the identification of potentially persistent components, the detection of predictive relations and the estimation of time-varying risk features. Notably, we include time-varying risks maintaining the same level of tractability of present-value models with constant risks.

We show that this inclusion helps us to better reconcile under a single common framework a number of predictability findings in the literature. First, we show that our model is well consistent with the very low dividend predictability estimated in standard post-war predictive regressions and with Goyal and Welch (2008) observation that aggregate price-dividend ratios have no additional out-of-sample predictive power for market returns, relative to a straightforward sample mean forecast. Second, our model allows for a negative correlation between expected and realized dividend growth, which has been shown (see Lettau and Wachter (2007)) to play a crucial role in explaining the value premium and the decreasing term structure of zero-coupon equity volatility documented by Binsbergen, Brandt, and Koijen (2012). In our setting, the estimated correlation between dividend growth and expected cash flow growth is highly time-varying and it can also switch sign: it becomes positive in periods in which the conditional covariance between returns and cash flow growth is large. These periods can be often linked to financial turmoil such as the 1989 saving and loans crisis or the beginning of the recent financial crisis in 2007. Interestingly, a countercyclical covariance between returns and dividend growth could help explain the documented countercyclical variation in expected returns within an equilibrium model in which the stochastic discount factor prices dividend growth shocks, as in Lettau and Wachter (2007). Third, conditional Sharpe ratios estimated by our model are often countercyclical, consistently with the empirical evidence, and quite volatile, with a standard deviation of about 0.65, which is a useful implication with respect to the “Sharpe ratio volatility puzzle” highlighted in Lettau and Ludvigson (2010), among others. Fourth, our model is consistent with Binsbergen, Hueskes, Koijen, and Vrugt (2013) observation of time variation in the slope of the term structures of expected dividend growth and with the empirical evidence reported in Binsbergen, Brandt, and Koijen (2012), which suggests a decreasing term structure of volatilities on dividend strips (i.e. claims to dividends paid over some specified future time interval), but has also the flexibility to let this term structure change over time, for example with the business cycle. Finally, we provide independent evidence on the role of time-varying risk features to uncover joint predictive relations within present-value models.¹

The paper proceeds as follows. Section 4.1 introduces our present-value model with time-varying return and dividend risks. In Section 4.2, we discuss our data set and the estimation strategy, while Section 4.3 presents estimation results, analyzes the model implications and shows that they are consistent with a number of predictive regression findings in the literature. Section 4.4 discusses additional implications of the model, including time-varying risk features and the term structure of market risks, and Section 4.5 concludes.

¹Using a particle filter, Johannes, Korteweg, and Polson (2011) estimate a set of Bayesian predictive regressions of market returns on aggregate payout yields and show that return predictability coupled with time-varying risk features can produce economic value, from the perspective of a Constant Relative Risk Aversion investor maximizing the predictive utility of her terminal wealth. In contrast, models with constant risks imply no substantial economic gain in incorporating predictability features. Ang and Liu (2007) study the link between Stochastic volatility and predictability in the context of a univariate, continuous-time asset pricing model.

4.1 Present-Value Model

As shown in [Cochrane \(2008b\)](#), among others, dividend growth and returns are better studied jointly in order to understand their predictability features. Following [Campbell and Shiller \(1988\)](#), this section introduces a tractable present-value model with time-varying risks for the joint dynamics of aggregate dividends and market returns. We denote by

$$r_{t+1} \equiv \log \left(\frac{P_{t+1} + D_{t+1}}{P_t} \right), \quad (4.1)$$

the cum-dividend log market return, and by

$$\Delta d_{t+1} \equiv \log \left(\frac{D_{t+1}}{D_t} \right), \quad (4.2)$$

the aggregate log dividend growth. Expected return and dividend growth, conditional on investors' information set at time t , are denoted by $\mu_t \equiv E_t[r_{t+1}]$ and $g_t \equiv E_t[\Delta d_{t+1}]$, respectively, while the conditional variance-covariance matrix of returns and dividend growth is denoted by Σ_t .

μ_t, g_t and Σ_t follow exogenous latent processes that model the time-varying second-order structure of returns and dividends:

$$\begin{pmatrix} \Delta d_{t+1} \\ r_{t+1} \end{pmatrix} = \begin{pmatrix} g_t \\ \mu_t \end{pmatrix} + \Sigma_t^{1/2} \begin{pmatrix} \varepsilon_{t+1}^D \\ \varepsilon_{t+1}^r \end{pmatrix}, \quad (4.3)$$

where $(\varepsilon_{t+1}^D, \varepsilon_{t+1}^r)'$ is a bivariate iid process. Expected returns and expected dividends follow simple linear autoregressive processes:

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g, \quad (4.4)$$

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^\mu, \quad (4.5)$$

with real valued parameters $\gamma_0, \gamma_1, \delta_0, \delta_1$.

Shocks $(\varepsilon_{t+1}^g, \varepsilon_{t+1}^\mu)'$ have zero conditional means, but they feature a potentially time-varying risk structure, which has to be consistent with the present-value constraints imposed on the dynamics of dividends, returns and price-dividend ratios, discussed in detail below. The conditional mean of (g_{t+1}, μ_{t+1}) has a simple linear autoregressive structure. However, process (g_{t+1}, μ_{t+1}) does not follow a standard linear autoregressive process with constant risk, as for instance the one studied in [Binsbergen and Kojen \(2010\)](#), because shocks $(\varepsilon_{t+1}^g, \varepsilon_{t+1}^\mu)'$ feature a degree of heteroskedasticity, induced by present-value constraints when Σ_t is time-varying.

We specify the dynamics of Σ_t by a simple autoregressive process that implies a degree of tractability comparable to the case of constant risks. Precisely, we assume that Σ_t follows a Wishart Autoregressive process of order one, denoted $WAR(1)$ (see [Gourieroux, Jasiak, and Sufana \(2009\)](#) and [Gourieroux \(2006\)](#)), which is completely characterized by the (affine) Laplace transform:

$$\Psi_t(\Gamma) = E_t[\exp \text{Tr}(\Gamma \Sigma_{t+1})] = \frac{\exp \text{Tr} [M' \Gamma (I_2 - 2V\Gamma)^{-1} M \Sigma_t]}{[\det(I_2 - 2V\Gamma)]^{k/2}}, \quad (4.6)$$

where $k > 1$ is the scalar degree of freedom, M is a 2×2 matrix of autoregressive parameters and V is a 2×2 symmetric and positive-definite volatility of volatility matrix. Process Σ_t admits the following autoregressive representation:

$$\Sigma_{t+1} = M \Sigma_t M' + kV + \nu_{t+1}, \quad (4.7)$$

and it takes positive semi-definite values, making dynamics (4.7) a naturally suited model for multivariate time-varying risks.²

The unconditional mean of stationary variance-covariance process Σ_t , μ^Σ , is the unique solution of the (implicit) steady state equation:

$$\mu^\Sigma = kV + M\mu^\Sigma M'. \quad (4.8)$$

The Wishart process allows for a good degree of flexibility of the volatility dynamics, while preserving tractability. Useful properties of this model include, e.g., a potentially negative dynamic dependence between variances (diagonal elements of Σ_t) and a covariance (out-of-diagonal element) which is unrestricted in sign (Gourieroux (2006)).

4.1.1 Price-dividend ratio

Let $pd_t \equiv \log \frac{P_t}{D_t}$ denote the log price-dividend ratio. To derive the expression for the price-dividend ratio implied by our model, we follow Campbell and Shiller (1988) log linearization approach:³

$$r_{t+1} \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t, \quad (4.9)$$

where $\overline{pd} = E[pd_t]$, $\kappa = \log(1 + \exp(\overline{pd})) - \rho \overline{pd}$ and $\rho = \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})}$. By iterating this equation using dynamics (4.4)-(4.7), we obtain a log price-dividend ratio that is an affine function of μ_t and g_t . For convenience of interpretations and in order to obtain pd_t expressions that are easily manageable in our Kalman filter estimation, we directly express pd_t as an affine function of a demeaned expected return and dividend growth ($\hat{\mu}_t = \mu_t - \delta_0$ and $\hat{g}_t = g_t - \gamma_0$).

Proposition 4 (Price-dividend ratio) *Under model (4.3)-(4.7), the log price-dividend ratio takes the affine form:*

$$pd_t = A - B_1 \hat{\mu}_t + B_2 \hat{g}_t, \quad (4.10)$$

with

$$A = \frac{\kappa + \gamma_0 - \delta_0}{1 - \rho}, \quad (4.11)$$

$$B_1 = \frac{1}{1 - \rho \delta_1}, \quad (4.12)$$

$$B_2 = \frac{1}{1 - \rho \gamma_1}. \quad (4.13)$$

The proof is given in Appendix C.1.2.

Price-dividend ratio pd_t is an affine function of expected returns and expected dividend growth. According to intuition, pd_t is decreasing in expected returns and increasing in expected dividend growth. Note that the dependence of price-dividend ratio pd_t on $\hat{\mu}_t$ and \hat{g}_t in Proposition 4 is identical to the dependence obtained in the model with constant dividend and return risks. Thus, this setting allows us to obtain simple interpretations for the additional effect of time-varying risks on dividend and return predictability features.

² Σ_t is positive definite if $k > n$, where n is the dimension of Σ_t .

³Expression (4.9) is obtained from a first order Taylor expansion of (4.1) around the unconditional mean of pd . The approximation error is related to the variance of the price-dividend ratio (see, e.g., Engsted, Pedersen, and Tanggaard (2010)), which is time-varying in our model. In the data the identity is virtually exact.

4.1.2 Time-varying risks in the present-value model

For Quasi Maximum Likelihood estimation with a Kalman Filter, we assume independence between shocks to returns and dividends $(\varepsilon_{t+1}^D, \varepsilon_{t+1}^r)'$ and shocks to time-varying risk ν_{t+1} , in equations (4.3) and (4.7), respectively, where $(\varepsilon_{t+1}^D, \varepsilon_{t+1}^r)'$ follows a bivariate standard normal distribution.

Time-varying risks in dynamics (4.3) and (4.7) have implications for the conditional risk features of expected returns and expected dividend growth in equations (4.4) and (4.5). Let

$$\tilde{\varepsilon}_{t+1}^D = e_1' \Sigma_t^{1/2} \begin{pmatrix} \varepsilon_{t+1}^D \\ \varepsilon_{t+1}^r \end{pmatrix} \quad (4.14)$$

and

$$\tilde{\varepsilon}_{t+1}^r = e_2' \Sigma_t^{1/2} \begin{pmatrix} \varepsilon_{t+1}^D \\ \varepsilon_{t+1}^r \end{pmatrix} \quad (4.15)$$

be the total shocks to dividends and returns in dynamics (4.3), where e_i denotes the i -th unit vector in \mathbb{R}^2 . Campbell and Shiller (1988) approximation (4.9) implies, together with the explicit price-dividend ratio expression (4.10):

$$\tilde{\varepsilon}_{t+1}^r = \tilde{\varepsilon}_{t+1}^D + \rho \varepsilon_{t+1}^{pd}, \quad (4.16)$$

and

$$\varepsilon_{t+1}^{pd} = B_2 \varepsilon_{t+1}^g - B_1 \varepsilon_{t+1}^\mu, \quad (4.17)$$

so that

$$\frac{1}{\rho} (\tilde{\varepsilon}_{t+1}^r - \tilde{\varepsilon}_{t+1}^D) = B_2 \varepsilon_{t+1}^g - B_1 \varepsilon_{t+1}^\mu. \quad (4.18)$$

The redundancy of return shocks in equation (4.16) implies that the state dynamics of our present-value model can be fully described by the joint dynamics of state vector $(\Delta d_{t+1}, pd_{t+1}, \hat{\Sigma}_t, \hat{g}_t, \hat{\mu}_t)$. Moreover, equation (4.18) implies that the distribution of the shocks in expected returns and expected dividends is constrained and an identification assumption has to be imposed. We introduce a parameter p that controls the attribution of price-dividend ratio shocks to expected return and dividend growth shocks:

$$\varepsilon_{t+1}^g = \frac{p}{\rho B_2} (\tilde{\varepsilon}_{t+1}^r - \tilde{\varepsilon}_{t+1}^D), \quad (4.19)$$

$$\varepsilon_{t+1}^\mu = \frac{p-1}{\rho B_1} (\tilde{\varepsilon}_{t+1}^r - \tilde{\varepsilon}_{t+1}^D), \quad (4.20)$$

Under this assumption, the conditional variance of return and dividend growth expectation shocks are both time-varying and given by

$$Var_t(\varepsilon_{t+1}^g) = \frac{p^2}{(\rho B_2)^2} (\Sigma_{11,t} + \Sigma_{22,t} - 2\Sigma_{12,t}), \quad (4.21)$$

and

$$Var_t(\varepsilon_{t+1}^\mu) = \frac{(p-1)^2}{(\rho B_1)^2} (\Sigma_{11,t} + \Sigma_{22,t} - 2\Sigma_{12,t}), \quad (4.22)$$

respectively, while their conditional covariance is the following:

$$Cov_t(\varepsilon_{t+1}^g, \varepsilon_{t+1}^\mu) = \frac{p(p-1)}{\rho^2 B_1 B_2} (\Sigma_{11,t} + \Sigma_{22,t} - 2\Sigma_{12,t}), \quad (4.23)$$

where $\Sigma_{ij,t}$ denotes the ij -component of the variance covariance matrix Σ_t .

4.2 Data and Estimation Strategy

This section describes our data set and introduces our estimation strategy based on a Quasi Maximum Likelihood estimation with a Kalman filter.

4.2.1 Data

We obtain the with-dividend and without dividend monthly returns on the value-weighted portfolio of all NYSE, Amex and Nasdaq stocks from January 1946 until December 2010 from the Center for Research in Security Prices (CRSP). We use this data to construct annual series of aggregate dividends and prices. We assume that monthly dividends are reinvested in 30-day T-bills and obtain annual series for cash-reinvested log dividend growth. Data on 30-day T-bill rates are also obtained from CRSP.

In order to produce useful information to identify latent time-varying risk components in our present-value model, we consider proxies for the yearly realized volatility of market returns and dividend growth and for their yearly realized covariance, which can be measured with a moderate estimation error. We compute a proxy for the yearly realized return variance as the sum of squared monthly market returns over the corresponding year:

$$RV_t^r = \sum_{i=1}^{12} r_{i,t}^2, \quad (4.24)$$

where $r_{i,t}$ is the demeaned market return on month i of year t . We do not correct for autocorrelation effects in daily returns (see French, Schwert, and Stambaugh (1987)), since we found the impact of this adjustment to be negligible. Analogously, the yearly realized variance of dividend growth and the realized covariance between returns and dividend growth are computed as follows:

$$RV_t^{\Delta d} = \sum_{i=1}^{12} \Delta d_{i,t}^2, \quad (4.25)$$

$$RC_t = \sum_{i=1}^{12} r_{i,t} \Delta d_{i,t}, \quad (4.26)$$

$$(4.27)$$

where $\Delta d_{i,t}$ is the demeaned and deseasonalized monthly dividend growth on month i of year t .⁴ Let us define the observed variance-covariance matrix of dividend growth and returns as:

$$RV_t \equiv \begin{bmatrix} RV_t^{\Delta d} & RC_t \\ RC_t & RV_t^r \end{bmatrix}.$$

Figure 4.1 represents our realized measures of return and dividend growth volatility and of the correlation between return and dividend growth, and summary statistics are provided in Table 4.1. As expected, the volatility of dividend growth is much lower than the market return volatility, however, both are unquestionably time-varying. The correlation between returns and dividend growth is close to zero on average, but it is also highly time-varying and it often switches sign.

⁴Deseasonalized monthly dividend growth is defined as the average of the last 12 log dividend growth, as e.g. in Goyal and Welch (2008).

Figure 4.1: Realized volatilities and correlation of returns and dividends

Realized volatility of yearly dividend growth ($\sqrt{RV^{\Delta d}}$, first panel), of returns ($\sqrt{RV^r}$, third panel), and their realized correlation ($RC/\sqrt{RV^{\Delta d}RV^r}$, second panel), where $RV^{\Delta d}$, RV^r and RC are given in equations (4.25), (4.24) and (4.26), respectively.

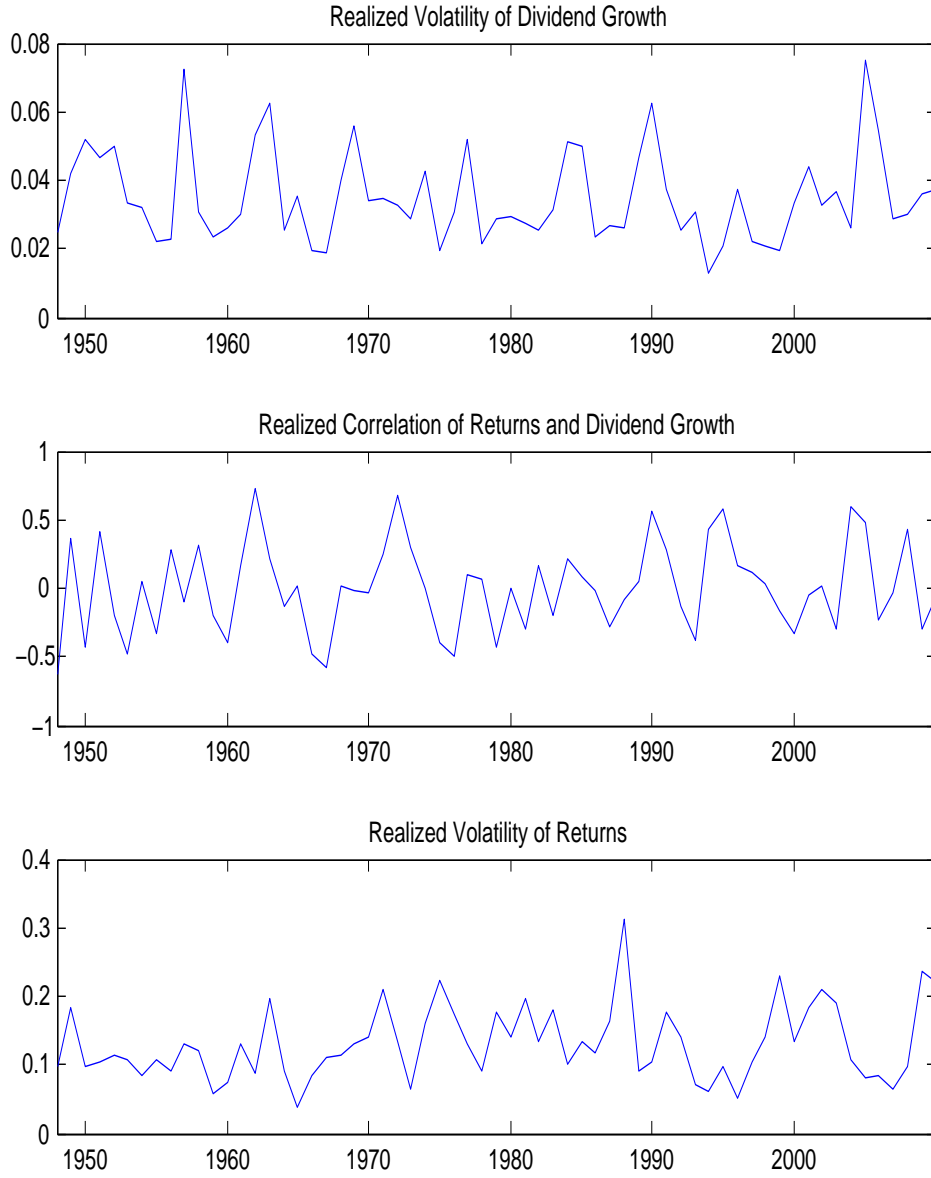


Table 4.1: Summary statistics

Summary statistics of the observed time series of returns (Ret), dividend growth (Div), log price-dividend ratio (pd), realized volatilities of returns and dividend growth ($RVol^r$ and $RVol^d$) and their realized correlation ($RCorr$). *mean* and *stdev* denote sample mean and standard deviation, while *auto* denotes autocorrelation. Time series are annual, from 1946 to 2010.

	Ret	Div	pd	$RVol^r$	$RVol^d$	$RCorr$
mean	0.0998	0.0564	3.4565	0.1320	0.0355	0.0035
stdev	0.1670	0.0686	0.4311	0.0539	0.0135	0.3211
auto	-0.0669	0.3048	0.9146	0.2735	0.1607	0.0539

4.2.2 State space representation

The relevant state variables in model (4.3)-(4.7) are the expected return and dividend growth μ_t , g_t and variance-covariance matrix Σ_t , but for simplicity we assume that Σ_t is observable, equal to the realized variance-covariance matrix of dividend growth and returns RV_t .⁵

We propose a Kalman filter to estimate the model parameters together with the values of these latent states μ_t and g_t . To this end, we cast the model in state space form, using demeaned state variables $\hat{\mu}_t$ and \hat{g}_t defined in Section 4.1.1. In this way, we obtain the following linear transition dynamics with heteroskedastic error terms for present-value model (4.3)-(4.7):

$$\begin{aligned}\hat{g}_{t+1} &= \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g, \\ \hat{\mu}_{t+1} &= \delta_1 \hat{\mu}_t + \varepsilon_{t+1}^\mu.\end{aligned}$$

Observable variables in our model are dividend growth Δd_t , the price-dividend ratio pd_t and the realized variance-covariance of returns and dividend growth RV_t . Note that while the market return r_{t+1} produces redundant information, relative to linear combinations of Δd_{t+1} and pd_{t+1} , the realized covariance of market returns and dividends produces useful information to identify time-varying risk structures, summarized by state $\hat{\Sigma}_t$. This is a sharp difference of our setting, relative to present-value models with constant risks, in which dividend growth and price-dividend ratio provide sufficient information to identify the latent state dynamics.

Measurement equations for Δd_t , pd_t , RV_t are derived from the model-implied expressions for dividend growth, price-dividend ratio and the conditional variance-covariance of returns and dividends. The measurement equation for dividend growth follows from the first row of dynamics (4.3):

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \tilde{\varepsilon}_{t+1}^D. \quad (4.28)$$

Thanks to our assumption of observability of Σ_t ($RV_t = \Sigma_t$), the measurement equation for the realized variance-covariance of returns and dividend growth, half-vectorized, can be written as:

$$vech(RV_{t+1}) = (I_2 - S)vech(\mu^\Sigma) + S \cdot vech(RV_t) + \varepsilon_{t+1}^\Sigma, \quad (4.29)$$

⁵This last assumption is consistent with the literature on variance risk premia and makes our Kalman filter estimation consistent even in the presence of time-varying risks. Moreover, it could be difficult to identify a latent variance-covariance state given the limited amount of information available. We explored several alternatives, including the introduction of a measurement error in the realized covariance or exponential smoothing of the monthly series and the main features of the estimated model do not seem to be affected. Details are available upon request.

where S is a function only of parameter M , specified explicitly in Appendix C.1.1.

The measurement equation for the log price-dividend ratio in equation (4.10) contains no error term. As shown by Binsbergen and Koijen (2010), this feature can be exploited to reduce the number of transition equations in the model. By substituting the equation for pd_t in the measurement equation for dividend growth, we arrive at a final system with one transition equation,

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \varepsilon_{t+1}^\mu, \quad (4.30)$$

$$(4.31)$$

and three measurement equations:

$$\Delta d_{t+1} = \gamma_0 + \frac{1}{B_2} (pd_t - A + B_1 \hat{\mu}_t) + \tilde{\varepsilon}_{t+1}^D, \quad (4.32)$$

$$pd_{t+1} = (1 - \gamma_1)A + B_1(\gamma_1 - \delta_1)\hat{\mu}_t + \gamma_1 pd_t + B_2 \varepsilon_{t+1}^g - B_1 \varepsilon_{t+1}^\mu, \quad (4.33)$$

$$vech(RV_{t+1}) = (I_2 - S)vech(\mu^\Sigma) + S \cdot vech(RV_t) + \varepsilon_{t+1}^\Sigma. \quad (4.34)$$

We use the Kalman filter to derive the likelihood of the model and we estimate it using QML. The parameters to be estimated are the following:

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, M, k, V, p).$$

For identification purposes, we impose some parameter constraints. M is assumed lower triangular, with positive diagonal elements less than one. V is assumed diagonal with positive components and $k \geq 2$ is integer. Parameters δ_1 and γ_1 are bounded to be less than one in absolute value. Overall, the most general version of our present-value model contains 11 parameters. Details on the estimation procedure are presented in Appendix C.2.

4.3 Predictability Features

In this section we discuss the estimation results, focusing on the structural quantification of the predictability implications of present-value models with time-varying risks, i.e., the characterization of the dynamic features of processes μ_t and g_t for expected returns and expected dividend growth. First, we quantify the estimated degree of predictability for returns and dividend growth and we analyse the implications of the estimated price-dividend ratio decomposition for the predictability features of returns and dividends by aggregate valuation ratios. Second, we perform hypothesis tests to establish the statistical significance of the predictability results and finally we evaluate the consistency of the model implications with a number of well-known predictive regression findings in the literature.

4.3.1 Estimation results

Table 4.2, Panel A, presents our QML estimation results for present-value model (4.30)-(4.34). The value of the quasi log-likelihood is 1094.1.⁶ The unconditional expected log return is $\delta_0 = 8.9\%$, while the unconditional expected growth rate of dividends is $\gamma_0 = 5.7\%$. Expected return features an high autoregressive root, $\delta_1 = 0.864$, which is an indication of a highly persistent process, having an half-life of about 5 years. Expected dividend growth are persistent,

⁶Parameter standard errors are obtained using the circular block-bootstrap of Politis and Romano (1992), in order to account for the potential serial correlation in the data. We use eight years blocks. Results are unchanged using the stationary bootstrap in Politis and Romano (1994).

Table 4.2: Estimation results

We present results of the estimation of the present-value model in equations (4.3)-(4.7). The model is estimated by quasi maximum-likelihood using yearly data from 1946 to 2010 on log dividend growth rates, log price-dividend ratio and realized variance-covariance of returns. Panel A presents estimates of the coefficients of the underlying processes. Panel B reports resulting coefficients of the present-value model in equation (4.10). Bootstrapped standard errors are in parentheses.

Panel A: Quasi maximum-likelihood estimates					
γ_0	δ_0	γ_1	δ_1	M_{11}	M_{21}
0.057	0.089	0.121	0.864	0.748	0.006
(0.014)	(0.011)	(0.405)	(0.134)	(0.186)	(0.338)
M_{22}	k	V_{11}	V_{22}	p	
0.926	2	0.0004	0.0039	0.054	
(0.116)	(3.429)	(0.0001)	(0.0013)	(0.068)	
Panel B: Implied present-value parameters					
ρ	A	B_1	B_2		
0.969	3.426	6.127	1.133		
(0.012)	(0.299)	(1.764)	(7.957)		

but much less persistent than expected return, with an autoregressive root $\gamma_1 = 0.121$ and an half-life of about 0.8 years.⁷ For comparison, the estimated persistence of expected returns and expected dividend growth in a model with constant risks is slightly larger, with an estimated root $\delta_1 = 0.927$ and $\gamma_1 = 0.304$, respectively and half-lives of 9.5 and 1 years.⁸ Estimation results also indicate persistent dividend and return risks. The autoregressive matrix M in the risk dynamics (4.7) features two quite persistent components, with estimated eigenvalues $M_{11} = 0.748$ and $M_{22} = 0.926$, respectively, and a slightly positive out-of-diagonal element $M_{21} = 0.006$. The low estimated degrees of freedom parameter $k = 2$ indicates a slightly fat tailed distribution for the components of Σ_t .

4.3.2 Dividend and return predictability

In order to quantify the degree of predictability implied by present-value model (4.3)-(4.7), we can measure the fraction of variability in r_t , Δd_t explained by μ_{t-1} and g_{t-1} , respectively.⁹ We present in Figure 4.2 the estimated expected return and expected dividend growth implied by our present-value model. In each panel, we also plot the fitted values of an OLS regression of r_t

⁷The first order autoregressive coefficient is equivalent to $1 - \lambda\Delta t$, where λ is the mean reversion speed and Δt is one year in our setting. The half-life is defined as $\frac{\ln 2}{\lambda}$.

⁸To derive the implications for the model with constant risks, we estimate the model in Binsbergen and Koijen (2010) for the case of cash-reinvested dividends, using data for the sample period 1946-2010. Our parameter estimates are very similar to their ones, which are based on the sample period 1946-2007. Detailed estimation results are given in Table II of the Supplemental Appendix, which is available at Piatti's webpage: <http://www.people.usi.ch/piattii>.

⁹Let I_t denote the econometrician's information set at time t , generated by the history of dividends, price-dividend ratios and realized variances and covariances up to time t . Given estimated parameter $\hat{\Theta}$, the Kalman filter provides expressions to compute filtered estimates of the unknown latent states μ_{t-1} and g_{t-1} , conditional on I_{t-1} .

and Δd_t on the lagged log price-dividend ratio, as well as the actual value of these variables.¹⁰ The expected returns estimated by the present-value model and the one implied by the predictive regression are quite smooth and close to each other, while there are larger differences between the expected dividend growth estimated by our present-value model and those of a standard predictive regression: The model-implied expected dividend growth varies more over time and it is less persistent. These findings are consistent with the autoregressive coefficients of expected return and dividend growth estimated by the present-value model with time-varying risks.

Table 4.3: R-squares of returns and dividends

Sample R-squared values of returns and dividend growth, computed using equations (4.35)-(4.36). In the first row, R^2 are computed from our present-value model, estimated using yearly data from 1946 to 2010, while the second row gives results for a standard OLS predictive regression of observed returns and dividend growth on price-dividend ratio.

	R-squared values (%)	
	R^2_{Ret}	R^2_{Div}
Present-value model	9.36	2.62
OLS	9.90	0.95

We can quantify the degree of predictability in returns and dividend growth within our present-value model and a standard predictive regression, by the following sample R^2 goodness-of-fit measures:

$$R^2_{Ret} = 1 - \frac{\widehat{Var}(r_{t+1} - \mu_t)}{\widehat{Var}(r_{t+1})}, \quad (4.35)$$

$$R^2_{Div} = 1 - \frac{\widehat{Var}(\Delta d_{t+1} - g_t)}{\widehat{Var}(\Delta d_{t+1})}, \quad (4.36)$$

where \widehat{Var} denotes sample variances and μ_t , g_t , are, with a slight abuse of notation, the estimated expected return and expected dividend growth in the present-value model and the standard predictive regression model, respectively.

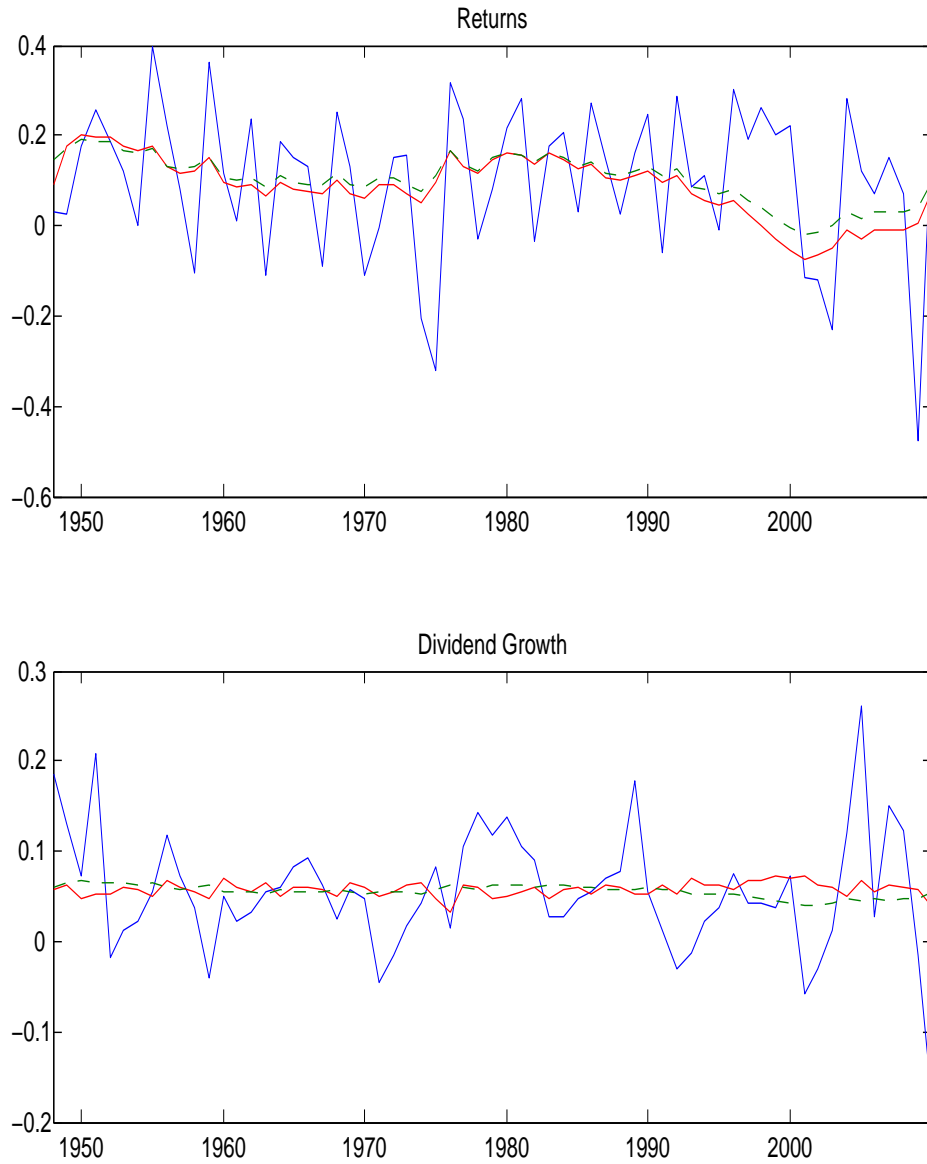
The results in Table 4.3 show that the estimated R^2 for returns in the present-value model is about 9.36% and the estimated R^2 for dividends is about 2.62%. Therefore, expected returns seem to explain a relatively large fraction of actual returns, while dividend growth predictability is quite low, even if it is higher than the one implied by standard predictive regressions.¹¹ The predictability results of standard predictive regressions are consistent with the evidence

¹⁰The predictive regression for returns takes the form $r_{t+1} = a_r + b_r pd_t + \varepsilon_{t+1}^r$, while the predictive regression for dividend growth is $\Delta d_{t+1} = a_d + b_d pd_t + \varepsilon_{t+1}^d$.

¹¹The basic intuition for the potentially different degrees of predictability implied by standard predictive regressions, relative to the latent expected return and dividend growth processes in our model, is provided in Cochrane (2008a), who derives the relation between state-space models and their observable VAR counterparts in settings with constant risks. Using the Kalman filter in Appendix C.2, we borrow from Binsbergen and Kojen (2010) to derive approximate expressions for the observable model-implied VAR representation with respect to the econometrician's information set. Such VAR contains several lag polynomials of returns, dividend growth rates and return realized variances.

Figure 4.2: Expected vs Realized yearly returns and dividend growth

These graphs show the model-implied (filtered) series (red lines) of expected returns μ_t (first panel) and expected dividend growth g_t (second panel), as well as the realized (blue lines) return, r_{t+1} and log dividend growth, Δd_{t+1} , respectively. The two panels also show the fitted values (dashed green lines) of an OLS regression of realized quantities (r_{t+1} and Δd_{t+1}) on the lagged log price-dividend ratio.



in the literature. While the R^2 for returns is about 10%, the one for dividends is about 1%. In summary, while the model-implied return predictability is close to the one implied by a standard predictive regression of returns on lagged price-dividend ratio, the dividend growth predictability uncovered by standard predictive regressions is lower than the model-implied dividend growth predictability. The estimated structure of the price-dividend ratio decomposition in our model offers an intuition for this finding: Since price-dividend ratios are only noisy signals of expected dividend growth, which are contaminated by an expected return component, these predictive regressions are affected by the well-known EIV problem;¹² see also Binsbergen and Koijen (2009). According to the estimated parameters in Panel B of Table 4.2, the expected return (expected dividend growth) loads negatively (positively) on price-dividend ratios, with an estimated coefficient $-B_1 = -6.127$ ($B_2 = 1.133$). Therefore, the persistent expected return component has a large loading on the model-implied price-dividend ratio. This large loading is associated with a large fraction of the price-dividend ratio that is driven by expected return shocks. Therefore, the large and persistent expected return component in price-dividend ratios likely obfuscates the predictive power of expected dividend growth for actual dividend growth. Since the expected return component is difficult to estimate from actual returns, due to a low signal-to-noise ratio, isolating it from aggregate price-dividend ratios in a model-free way is a potentially difficult task.

The conditional variance decomposition of the price-dividend ratio is given by:

$$\begin{aligned} Var_t(pd_{t+1}) &= B_1^2 Var_t(\varepsilon_{t+1}^\mu) + B_2^2 Var_t(\varepsilon_{t+1}^g) - 2B_1B_2 Cov_t(\varepsilon_{t+1}^\mu, \varepsilon_{t+1}^g) \\ &= \frac{1}{\rho^2} (\Sigma_{11,t} + \Sigma_{22,t} - 2\Sigma_{12,t}) [(p-1)^2 + p^2 - 2p(p-1)]. \end{aligned} \quad (4.37)$$

Standardizing all terms by the left-hand side of (4.37) we find that the variation due to discount rates is $(p-1)^2$, the variation due to expected dividend growth is p^2 , and the variation due to their covariance is $-2p(1-p)$. Thus, the conditional variance decomposition of the price-dividend ratio is constant over time and depends only on p . At the estimated parameters, these percentage variations are 89.49%, 0.29% and 10.22%, respectively. Therefore, the price-dividend ratio varies mainly because of expected return shocks, consistent with Campbell (1991) and Cochrane (1992), among others. The same conclusion holds if we look at the unconditional structure of the price-dividend ratio variation, with percentages of 80.45%, 0.07% and 19.48%, respectively. These results are summarized in Panel A of Table 4.4. For comparison, Panel B reports the same variance decompositions for the constant risks present-value model: Results are similar, but the contribution of expected dividend growth to the overall variation is slightly larger and the covariance has a negative impact.

4.3.3 Testing predictability hypotheses

Our estimates suggest that both expected returns and expected dividend growth vary over time, which is equivalent to saying that both returns and cash flow growth are predictable. In this section, we perform formal hypothesis tests to establish the statistical significance of these results. In our setting, most predictability hypotheses can be formulated by means of simple parametric constraints, which, given a likelihood-based estimation approach, can be efficiently

¹²Appendix C.1.3 provides analytic expressions for the model-implied asymptotic bias in standard predictive regression coefficients. These expressions allow us to evaluate the relative importance of EIV and small sample bias (Stambaugh (1999)), using simulations from the model. We find that, for the dividend regression, the EIV and small sample biases go in opposite directions and the EIV bias dominates. Therefore, the typical small sample bias correction (Stambaugh (1999)) produces even more biased point estimates in dividend growth predictive regressions.

Table 4.4: Price-dividend ratio decomposition

Conditional and unconditional variance decomposition of the price-dividend ratio for our time-varying risks model (Panel A), using the estimated parameters in Table 4.2, and in the constant risks model (Panel B) of van Binsbergen and Koijen (2010). Parameters used in the simulations are estimated using yearly price-dividend ratio and dividend growth from 1946 to 2010.

	Discount Rates	Div.Growth	Covariance
Panel A: Time-varying risks model			
Conditional	89.49%	0.29%	10.22%
Unconditional	80.45%	0.07%	19.48%
Panel A: Constant risks model			
Conditional	93.61%	32.89%	−25.50%
Unconditional	99.89%	5.50%	−5.39%

tested with a standard likelihood ratio (LR) test, using the statistic

$$LR_T = 2 \left(\max_{\Theta} \log \mathcal{L}(\theta, \{Y_t\}_{t=1}^T) - \max_{\Theta_0} \log \mathcal{L}(\theta, \{Y_t\}_{t=1}^T) \right), \quad (4.38)$$

where Θ_0 is the restricted set of parameters under the given null hypothesis H_0 and $\log \mathcal{L}$ is the log-likelihood of the model. Evidence against H_0 is collected when LR is sufficiently large:

$$\{LR_T > c_{1-\alpha}\}, \quad (4.39)$$

relative to a critical value $c_{1-\alpha}$ that is unlikely under H_0 . As $T \rightarrow \infty$, statistic LR_T follows a χ_r^2 distribution with r degrees of freedom, where r is the number of parameter constraints defining the constrained parameter set Θ_0 . However, given the limited length of the available data sample, the asymptotic theory is likely to provide a bad approximation of the true finite-sample distribution of the LR statistics. Therefore, we apply the nonparametric Monte Carlo bootstrap likelihood ratio test proposed by Piatti and Trojani (2012a).¹³

First, we test the hypothesis of constant return expectation, i.e. no return predictability. In terms of the model parameters, this null is given by:

$$H_0 : \delta_1 = 0 \quad \text{and} \quad p = 1. \quad (4.40)$$

Under this H_0 all price-dividend ratio variation is due to variation in expected cash flow growth. The value of the likelihood ratio statistic for this null is equal to $LR_T = 41.1$, which corresponds to a p-value of 3.4% of the bootstrap LR test. Therefore null hypothesis (4.40) can be rejected at a 5% confidence level.

Second, we test for constant expected dividend growth. The corresponding null hypothesis is the following:

$$H_0 : \gamma_1 = 0 \quad \text{and} \quad p = 0. \quad (4.41)$$

Under null hypothesis (4.41), dividend growth is unpredictable, and all variation in the log price-dividend ratio comes from variation in expected returns. The value of the likelihood ratio

¹³Piatti and Trojani (2012a) show that standard asymptotic tests of present-value models with constant risks tend to over-reject the null of no predictability, and propose a nonparametric Monte Carlo testing method with more reliable finite-sample properties. With slight modifications we can apply their testing method to our time-varying risks framework.

statistic for this null is equal to $LR_T = 10.6$, which corresponds to a p-value of 10.6% of the bootstrap LR test. Therefore null hypothesis (4.41) cannot be rejected at standard confidence levels.¹⁴

As suggested by Piatti and Trojani (2012a), the lack of statistical significance of null hypothesis (4.41) does not necessarily mean that dividend growth is not predictable, but it indicates that the information set might be insufficient to reliably identify time variations in dividend expectations, i.e., tests of dividend predictability may have a low power, probably due to the short time series available. Therefore, in the sequel we study the implications of the estimated unconstrained model, which allow for time variation in both return and dividend growth expectations, even if the predictability evidence is weak.

4.3.4 Consistency with Predictive Regression Results

A good specification of the time-varying risk and return structure of present-value models has to produce predictability features consistent with the empirical evidence of standard OLS predictive regressions for dividends and returns. This feature is essential in order to avoid a model misspecification along some potentially important predictability dimension, which would weaken the interpretation of predictability and time-varying risks structures estimated by a latent variables approach.

In order to assess the main implications of our estimated model with respect to (i) the empirical features of standard predictive regressions with aggregate price-dividend ratios (ii) the long-horizon predictability properties and (iii) the real-time predictability patterns, we generate 10000 paths of length 65 years for all state variables and observable variables in our model, following a nonparametric Monte Carlo bootstrap approach (see Stoffer and Wall (1991) and Piatti and Trojani (2012a)).

4.3.4.1 Joint dividend-return predictability features

The predictive regression results in the data indicate the presence of return predictability (with an R^2 of about 9.90%) and a weak dividend predictability (with an R^2 of about 0.95%) by aggregate price-dividend ratios. As emphasized in Cochrane (2008b), this joint evidence implies sharp restrictions that are useful to validate or test the ability of a model in generating appropriate predictability properties. We follow this insight and compute by Monte Carlo simulation the model-implied joint distribution of estimated R^2 's for dividend and return predictive regressions with lagged log price-dividend ratios. Table 4.5 (columns *SV*) reports confidence intervals for the degree of predictability in OLS predictive regressions, under the assumptions of our estimated present-value model. Our model implies OLS predictive regression results in line with the empirical evidence. For instance, the median OLS R^2 's for return and dividend predictive regressions are about 12.36% and 1.43%, respectively, and are similar to the 9.9% and 0.95% OLS R^2 's estimated on real data. Overall, real data OLS R^2 's for return and dividend growth predictive regressions are all well inside the 80% confidence interval of estimated OLS R^2 's simulated from our present-value model. It is useful to compare the predictability implications of the model with time-varying return and dividend risks with those of present-value models with constant risks. Columns *CV* in Table 4.5 summarize the results of the same simulation exercise for the present-value model with constant risks studied in Binsbergen and Koijen (2010). The median R^2 implied by OLS predictive regressions for returns (dividends) is about 11.82% (1.15%), which is similar to what we find from our time-varying risks present-value

¹⁴Using the asymptotic χ^2_2 distribution of the LR statistics, we would reject both null hypotheses, (4.40) and (4.41) with a p-values of about 0 and 0.5%, respectively.

model. Therefore, the introduction of time-varying risks do not undermine model's consistency with standard predictive regression results for returns and dividend growth.¹⁵

4.3.4.2 Long-horizon predictive regressions

Cochrane (2008b) shows how to derive regression coefficients of long-horizon returns and dividend growth on price-dividend ratio, implied by yearly predictive regressions. By applying recursively the following regressions,

$$\begin{aligned} r_{t+1} &= a_r + b_r pd_t + \varepsilon_{t+1}^r \\ \Delta d_{t+1} &= a_d + b_d pd_t + \varepsilon_{t+1}^d \\ pd_{t+1} &= a_{pd} + \phi pd_t + \varepsilon_{t+1}^{pd} \end{aligned}$$

the regression coefficient of long-run returns, $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$, on pd_t is

$$b_r^{lr} = \frac{b_r}{1 - \rho\phi}.$$

Similarly, the regression coefficient of long-run dividend growth is

$$b_d^{lr} = \frac{b_d}{1 - \rho\phi}.$$

We compute, as in Section 4.3.4.1, these regression coefficients from the data and by simulation.¹⁶ We find that the model with time-varying risks produces with a frequency of about 36% estimated long-run coefficients within one standard deviation of the observed sample values (jointly).

4.3.4.3 Out-of-sample predictability

From the perspective of real-time forecasting, out-of-sample prediction is more relevant than in-sample prediction. Goyal and Welch (2008) study the out-of-sample explanatory power of a large set of predictive variables for market returns, finding that most of them perform worse than the historical mean in forecasting future returns. As explained in Cochrane (2008b), among others, a weak out-of-sample forecasting power does not imply a rejection of the null of predictability itself, but it rather raises important doubts about the practical usefulness of such return forecasts in forming real-time portfolios, given the persistence of forecasting variables and the short span of available data.

Given the degree of return and dividend growth predictability implied by the estimation results in Table 4.2, a useful reality check for our present-value model is the absence of excessive incremental out-of-sample forecasting power, relative to a simple mean forecast, when using simple predictive regressions based on aggregate price-dividend ratios. Following Goyal and Welch (2008), we quantify incremental out-of-sample predictive power by the metric:

$$R_{i,OS}^2 = 1 - \frac{MSE_{i,A}}{MSE_{i,M}}, \quad (4.42)$$

¹⁵If we generate samples from the constant risks model using standard MC simulation instead of bootstrapping residuals we find too low return predictability (median $R_{Ret}^2 = 7.25\%$) and an excessive dividend growth predictability (median $R_{Div}^2 = 3.81\%$), suggesting that the model shocks may be far from normal. The model with time-varying risks suffers less from this misspecification of the innovation's distribution, with a median simulated R_{Div}^2 of 1.9%.

¹⁶Using yearly data from 1946 to 2009 we find $\phi = 0.9199$, $b_r = -0.1224$ and $b_d = -0.0156$, so that $b_r^{lr} = -1.1310$ (with a standard deviation of 0.6796) and $b_d^{lr} = -0.1443$ (with a standard deviation of 0.1984). Standard deviations are obtained by the delta method from the standard deviations of b_r , b_d and ϕ . Note that b_r^{lr} and b_d^{lr} satisfy the approximate identity $b_d^{lr} - b_r^{lr} = 1$; see Cochrane (2008b).

Table 4.5: Simulated R-squares of returns and dividends

10%-, 50%- and 90%-quantile of the empirical distribution of R-squared values of returns and dividend growth, computed using (4.35)-(4.36), from OLS regressions of simulated returns and dividend growth on lagged log price-dividend ratio. Distributions are based on 10000 bootstrap samples of length 65 years of the state variables and observables in our time-varying risks model (*SV*), using the estimated parameters in Table 4.2, and in the constant risks model (*CV*) of van Binsbergen and Koijen (2010). Parameters used in the simulations are estimated using yearly price-dividend ratio and dividend growth from 1946 to 2010.

	R_{Ret}^2		R_{Div}^2	
	<i>SV</i>	<i>CV</i>	<i>SV</i>	<i>CV</i>
10%	7.27	7.09	0.05	0.04
50%	12.36	11.82	1.43	1.15
90%	20.55	18.25	7.77	6.72

where $MSE_{i,A}$ ($MSE_{i,M}$) is the out-of-sample mean squared forecast error of the predictive regression model (historical mean) for returns ($i = r$) and dividend growth ($i = d$), respectively. We simulate 10000 paths of observables and state variables from our estimated present-value models and compute the joint Monte Carlo distribution of $(R_{d,OS}^2, R_{r,OS}^2)$ realizations. We find that it is unlikely that predictive regressions for returns or dividends can produce a significantly larger out-of-sample predictive power than the historical mean: The estimated probability of the event $\{R_{d,OS}^2 \leq 0, R_{r,OS}^2 \leq 0\}$ is about 81%, while the estimated probability of the event $\{R_{d,OS}^2 > 0, R_{r,OS}^2 > 0\}$ is less than 0.8%.

4.4 Time-varying Risk Features

In this section, we study additional implications of the estimated present-value model with time-varying risks, by focusing on (i) the model-implied correlation dynamics, (ii) the conditional Sharpe ratio (iii) the term structure of long-horizon expectations and risks.

4.4.1 Correlation between expected and realized returns and dividends

Our time-varying risks assumption and our specification of the shocks in the expectation processes (see Section 4.1.2) allows both the volatility of expected return and of expected cash flow growth to be time-varying (see Equations (4.21) and (4.22)). More importantly, we allow for a time-varying correlation between innovations in the expected and unexpected part of both returns and dividend growth:

$$corr_t(\varepsilon_{t+1}^\mu, \tilde{\varepsilon}_{t+1}^r) = sign(p-1) \frac{\Sigma_{22,t} - \Sigma_{12,t}}{\sqrt{\Sigma_{22,t}(\Sigma_{22,t} + \Sigma_{11,t} - 2\Sigma_{12,t})}}, \quad (4.43)$$

$$corr_t(\varepsilon_{t+1}^g, \tilde{\varepsilon}_{t+1}^D) = sign(p) \frac{\Sigma_{12,t} - \Sigma_{11,t}}{\sqrt{\Sigma_{11,t}(\Sigma_{22,t} + \Sigma_{11,t} - 2\Sigma_{12,t})}}. \quad (4.44)$$

Figure 4.3 reproduces the time series of correlations (4.43) and (4.44) in our model, using estimated parameters in Table 4.2 and the corresponding filtered states in our Kalman filter. We find that the estimated correlation (4.43) is strongly negative (with a mean of about -0.94),

as expected, but it varies substantially over time, with a maximum of about -0.56 in 2004. Similarly, the conditional correlation between expected and realized dividend growth is negative on average, with a mean of about -0.27 . The standard assumption in constant risks present-value models is to put the correlation between expected and realized dividend growth to zero (see e.g. Binsbergen and Koijen (2010), Rytchkov (2012) and Yun (2012)). However, Lettau and Wachter (2007) show that a negative value of this correlation plays a crucial role in explaining the value premium¹⁷ and the decreasing term structure of zero-coupon equity volatility documented by Binsbergen, Brandt, and Koijen (2012), and they calibrate it to a constant value of -0.83 , using the consumption-dividend ratio as a proxy for expected cash flow growth as in Lettau and Ludvigson (2005). In our setting, the estimated correlation between dividend growth and expected cash flow growth is highly time-varying and it can also switch sign: it becomes positive in periods in which the conditional covariance between returns and cash flow growth, Σ_{12} , is positive and bigger than the conditional variance of dividend growth (see equation (4.44)). These periods can be often linked to financial turmoil such as the 1989 saving and loans crisis or the beginning of the recent financial crisis in 2007. Interestingly, a countercyclical covariance between returns and dividend growth could help explain the documented countercyclical variation in expected returns within an equilibrium model in which the stochastic discount factor prices dividend growth shocks, as in Lettau and Wachter (2007). Moreover, our model is consistent with Binsbergen, Hueskes, Koijen, and Vrugt (2013) observation of time variation in the slope of the term structures of expected dividend growth and dividend growth volatility, as we will discuss in detail in sections 4.4.3 and 4.4.4.

Within our model, correlations (4.43) and (4.44) are directly related to the conditional correlations of returns and dividend growth with the price-dividend ratio:

$$\text{corr}_t(\varepsilon_{t+1}^{pd}, \tilde{\varepsilon}_{t+1}^r) = \text{sign}(p-1) \text{corr}_t(\varepsilon_{t+1}^\mu, \tilde{\varepsilon}_{t+1}^r), \quad (4.45)$$

$$\text{corr}_t(\varepsilon_{t+1}^{pd}, \tilde{\varepsilon}_{t+1}^D) = \text{sign}(p) \text{corr}_t(\varepsilon_{t+1}^g, \tilde{\varepsilon}_{t+1}^D). \quad (4.46)$$

Therefore, the correlation between price-dividend ratio and returns (dividend growth) is on average positive (negative) and highly time-varying. This link is due to our specification of the expectation shocks in (4.19) and (4.20), which are proportional to the price-dividend ratio innovation, where the proportion is controlled by parameter p .

4.4.2 Conditional Sharpe ratio dynamics

Using the filtered expected return $\hat{\mu}_t$ and our estimate of the conditional variance of returns $\hat{\Sigma}_{22,t}$, we find a relatively large degree of variability in expected market returns and market risk. The average correlation between expected returns and return volatilities is slightly negative, at about -0.12 , even if in some subperiods these variables tend to move in the same direction (see Figure 4.4). Therefore, we cannot draw a unique conclusion on the sign of the relation between the conditional mean and volatility of returns, which appears instead to be time-varying.

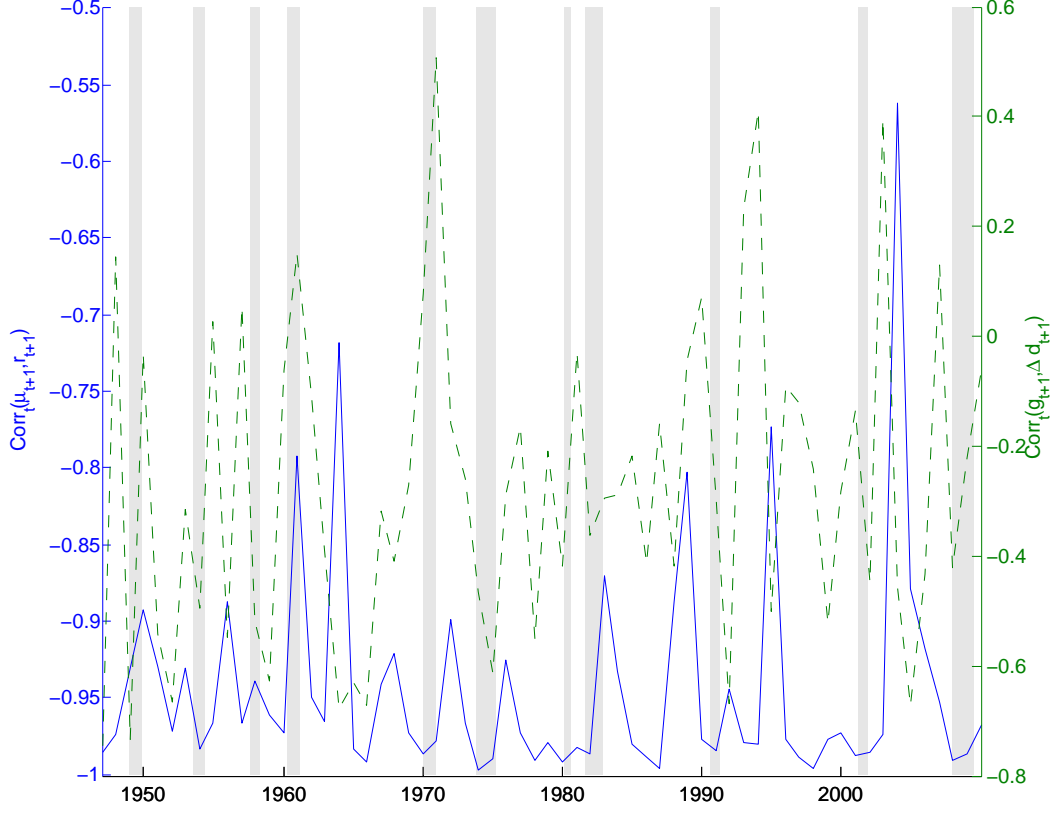
Conditional Sharpe ratios are defined as the ratio of conditional excess expected returns and conditional volatility, which requires assumptions on the riskless interest rate r_t^f :

$$SR_t = \frac{E_t(r_{t+1}) - r_t^f}{\sqrt{\text{Var}_t(r_{t+1})}} = \frac{\mu_t - r_t^f}{\sqrt{\Sigma_{22,t}}}.$$

¹⁷Lettau and Wachter (2007) specify the stochastic discount factor so that shocks to aggregate dividends are priced. The negative correlation between expected dividend growth and dividend growth leads to lower risk premia for growth stocks, since shocks to expected dividend growth act as a hedge.

Figure 4.3: Correlation between expected and realized returns and dividends

Conditional correlation between shocks in expected and unexpected returns (solid blue line, left axis), $\text{corr}_t(\varepsilon_{t+1}^\mu, \varepsilon_{t+1}^r)$, and conditional correlation between shocks in expected and unexpected dividend growth (dashed green line, right axis), $\text{corr}_t(\varepsilon_{t+1}^g, \varepsilon_{t+1}^D)$.

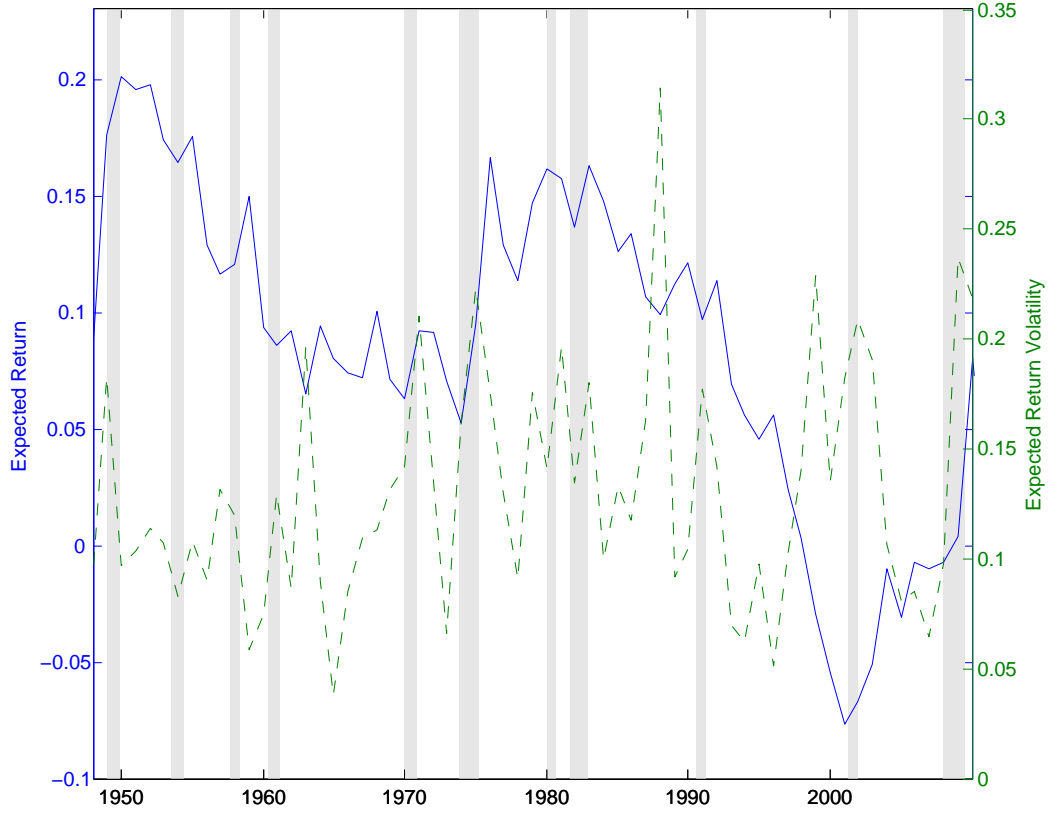


To compute our proxy for SR_t , we fix r_t^f as the annualized 30-day T-Bill rate at time t . Figure 4.5 shows that conditional Sharpe ratios estimated by our model are often countercyclical, consistently with the empirical evidence, and quite volatile, with a standard deviation of about 0.65, which is a useful implication with respect to the “Sharpe ratio volatility puzzle” highlighted in Lettau and Ludvigson (2010), among others. In contrast, we find that the conditional Sharpe ratio implied by a model with constant risks is both less countercyclical and not as volatile, with a standard deviation of 0.30. For comparison, the conditional Sharpe ratio constructed using estimates of fitted return mean and variance obtained using a vector of standard predictor variables, as in Lettau and Ludvigson (2010), has a volatility ranging from about 0.45 to about

0.7, according to the variables used.¹⁸

Figure 4.4: Risk-Return tradeoff

Filtered values of conditional expected returns, μ_t (blue line, left axis) against conditional volatility of returns, $\sqrt{\Sigma_{22,t}}$ (dashed green line, right axis). Shaded areas corresponds to NBER recessions.



¹⁸Fitted mean and variance are constructed using regressions of the form:

$$r_{t+1} = \beta_r' Z_t + \varepsilon_{t+1}^r$$

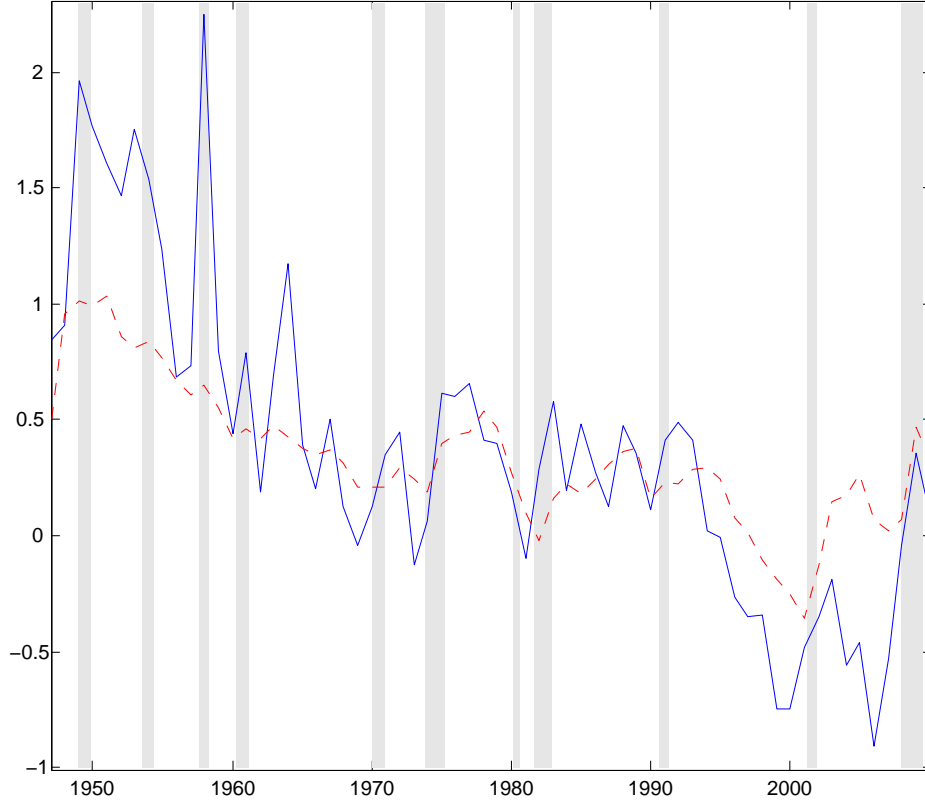
and

$$RV_{t+1}^r = \beta_v' Z_t + \varepsilon_{t+1}^v,$$

respectively, where Z_t is a vector of predetermined conditioning variables. In Lettau and Ludvigson (2010) for example Z_t contains cay_t and two lags of the realized variance to construct fitted volatility, while it contains cay_t and the risk free rate when computing fitted mean.

Figure 4.5: Conditional Sharpe ratio

The blue line shows the conditional Sharpe ratio implied by our model, obtained from filtered values of conditional expected returns and conditional volatility of returns, using as risk-free rate the annualized 30-day T-Bill rate at each time t . The red line is obtained in the same way, but for a version of the model with constant risks. Shaded areas corresponds to NBER recessions.



4.4.3 Term structure of long-horizon expectations

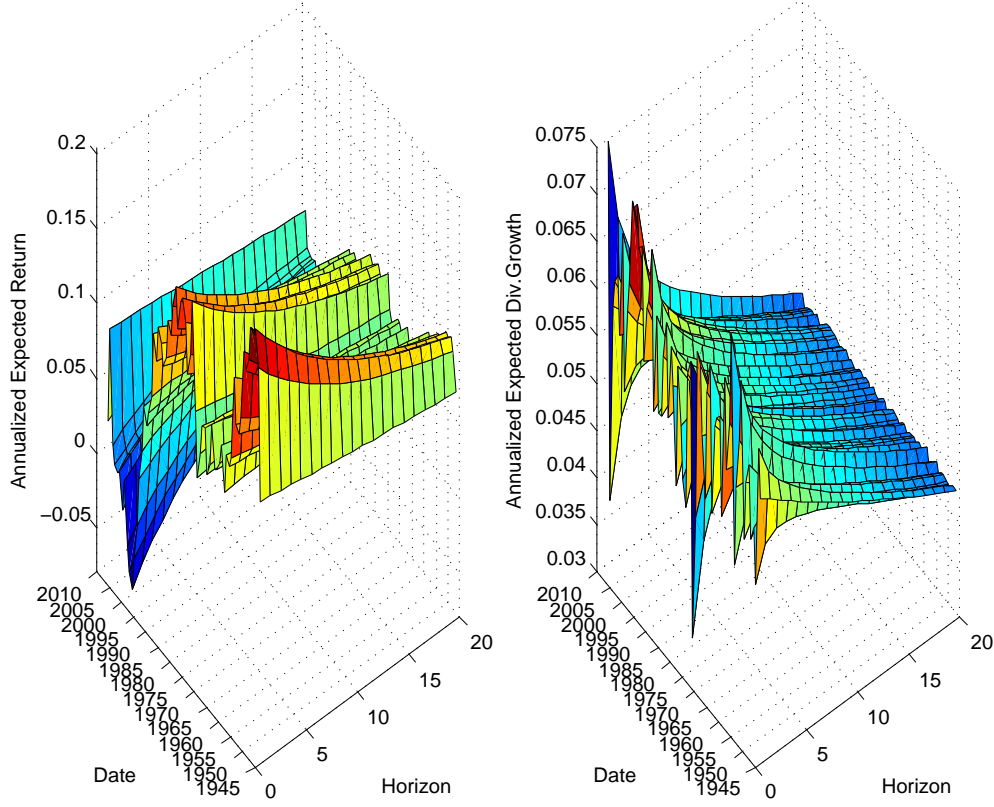
By applying recursively equations (4.3)-(4.5), we obtain the following explicit expressions for the model-implied n -year return and dividend growth:

$$\sum_{j=1}^n \rho^{j-1} r_{t+j} = \frac{1 - \rho^n}{1 - \rho} \delta_0 + \frac{1 - (\rho \delta_1)^n}{1 - \rho \delta_1} \hat{\mu}_t + \sum_{j=1}^{n-1} \rho^j \frac{1 - (\rho \delta_1)^{n-j}}{1 - \rho \delta_1} \varepsilon_{t+j}^\mu + \sum_{j=1}^n \rho^{j-1} \tilde{\varepsilon}_{t+j}^r, \quad (4.47)$$

$$\sum_{j=1}^n \rho^{j-1} \Delta d_{t+j} = \frac{1 - \rho^n}{1 - \rho} \gamma_0 + \frac{1 - (\rho \gamma_1)^n}{1 - \rho \gamma_1} \hat{g}_t + \sum_{j=1}^{n-1} \rho^j \frac{1 - (\rho \gamma_1)^{n-j}}{1 - \rho \gamma_1} \varepsilon_{t+j}^g + \sum_{j=1}^n \rho^{j-1} \tilde{\varepsilon}_{t+j}^D. \quad (4.48)$$

Figure 4.6: Term Structure of conditional expected long-horizon returns and dividends

Dynamics of the term structure of the conditional per-period expected long-horizon return (left panel) and dividend growth (right panel), from equations (4.49) and (4.50), respectively, computed using estimated parameters and filtered state. We consider horizons of 1 to 20 years.



The model-implied expected n -year return and dividend growth follow as:

$$E_t \left[\sum_{j=1}^n \rho^{j-1} r_{t+j} \right] = \frac{1 - \rho^n}{1 - \rho} \delta_0 + \frac{1 - (\rho \delta_1)^n}{1 - \rho \delta_1} \hat{\mu}_t, \quad (4.49)$$

$$E_t \left[\sum_{j=1}^n \rho^{j-1} \Delta d_{t+j} \right] = \frac{1 - \rho^n}{1 - \rho} \gamma_0 + \frac{1 - (\rho \gamma_1)^n}{1 - \rho \gamma_1} \hat{g}_t. \quad (4.50)$$

Left panel in Figure 4.6 plots the estimated term structure of return predictability implied by formula (4.49). The term structure is quite time-varying at short horizons and can be increasing or decreasing depending on the level of short-horizon expectations, but stabilizes with the horizon, around a long-term expected market return of approximately 6%. The term structure of dividend growth expectations (right panel) is also time-varying and usually decreasing, but it is hump-shaped in periods when the yearly expected dividend growth is particularly low (crisis periods), with a peak at an horizon of about 5 years. This finding is consistent with term structures of expectations obtained by more direct approaches: Using a new data set of dividend derivatives with maturities up to 10 years, [Binsbergen, Hueskes, Koijen, and Vrugt](#)

(2013) find that the slope of the term structure of growth is countercyclical. In particular, the 5-year expected dividend growth rate is higher than the 2-year expected growth rate during recessions, and lower during expansions. Level and slope of the term structure of dividend expectations implied by our estimated model are represented in Figure 4.7.

4.4.4 Term structure of risks

Siegel (2008) reports that unconditional (sample) variances realized over long investment horizons are lower than short-horizon variances on a per-year basis. Based on an estimated VAR model for returns and predictors, Campbell and Viceira (2005) conclude that also the term structure of conditional variances is decreasing with the investment horizon. Taking a slightly different view, Pastor and Stambaugh (2012) show that from the perspective of an investor subject to parameter uncertainty and imperfect predictors stocks can be more risky over longer horizons. Using Bayesian Model Averaging to account for model uncertainty, Diris (2011) finds that stocks are at least as risky in the long-run as in the short-run.

The model-implied conditional variance of a n -year return in the setting with time-varying risks is derived from equation (4.47) as follows:

$$\begin{aligned} Var_t \left[\sum_{j=1}^n \rho^{j-1} r_{t+j} \right] &= \sum_{j=1}^{n-1} \rho^{2j} \left(\frac{1 - (\rho\delta_1)^{n-j}}{1 - \rho\delta_1} \right)^2 Var_t(\varepsilon_{t+j}^\mu) + \sum_{j=1}^n \rho^{2(j-1)} Var_t(\tilde{\varepsilon}_{t+j}^r) \\ &+ 2 \sum_{j=1}^{n-1} \rho^{2j-1} \frac{1 - (\rho\delta_1)^{n-j}}{1 - \rho\delta_1} Cov_t(\varepsilon_{t+j}^\mu, \tilde{\varepsilon}_{t+j}^r), \end{aligned} \quad (4.51)$$

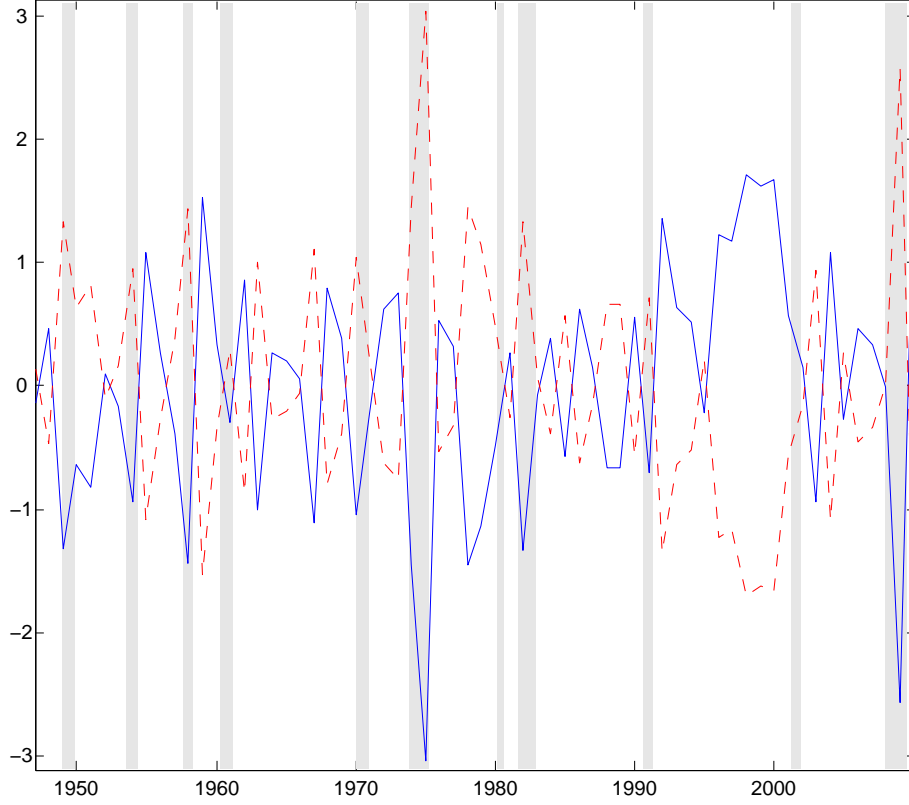
where $Var_t(\varepsilon_{t+j}^\mu)$, $Var_t(\tilde{\varepsilon}_{t+j}^r)$ and $Cov_t(\varepsilon_{t+j}^\mu, \tilde{\varepsilon}_{t+j}^r)$ are affine functions of the variance-covariance state Σ_t , given explicitly in Appendix C.1.4. The model-implied term structure of market risk, which we define, as it is typical in the literature, as the annualized volatility of cumulative returns versus the investment horizon,¹⁹ is thus time-varying. Left panel of Figure 4.8 plots its estimated dynamics. We find a time-varying term-structure of market risk, which can be decreasing or hump-shaped with a peak at around 5 years maturity and flattens for long maturities. In our model this potentially increasing pattern of market risk between short and medium maturities arises directly from the uncertainty of future expected returns, even in absence of imperfect predictors (Pastor and Stambaugh (2012)) or an explicit concern for model uncertainty (Diris (2011)).

To understand these findings, it is useful to split conditional variance (4.51) in its three components: A first term reflecting uncertainty about future expected returns, a second term capturing the risk of future return shocks and a third part reflecting the mean reversion of returns, due to the negative correlation between realized and expected return shocks. Figure 4.9 plots the estimated term structure of market risk and its three components at three different points in time, characterized by different levels of short term market volatility. Consistent with the intuition that return mean reversion tends to produce a decreasing term structure of market risk, we find that the mean reversion component has a strongly negative term structure effect, which is however partly offset by the impact of the other two components. The uncertainty about future expected returns has a large positive and increasing effect, which, as highlighted by

¹⁹Note that this definition differs from the one used by Binsbergen, Brandt, and Koijen (2012): they consider the volatility of single future cash flows using dividend strips, while we evaluate the annualized volatility of the stream of all future cash flows.

Figure 4.7: Level and slope of the term structure of the dividend growth expectation

The level (blue line) is measured as the 2-year expected growth rate, standardized, while the slope is measured by the difference between the 5- and 2-year expected dividend growth, also standardized (dashed red). Shaded areas corresponds to NBER recessions.



Pastor and Stambaugh (2012), is often underestimated or neglected and its relative contribution is positively linked to the degree of predictability in returns, which is large for long horizons. The term structure effect of return shock risk is positive and typically hump-shaped or increasing with the horizon, leading to the hump in the overall return variance.

The bottom right panel of Figure 4.9 reports for comparison the decomposition of the (constant) term structure of market risks in the model with constant risks.²⁰ In this case, we find that the effect of future expected return risk is not large enough to offset the impact of the other two term structure components, leading to a downward sloping term structure of risk.

Finally, it is interesting to note that even if the time-varying risks model suggests a term structure of market risks that can take different shapes, it might be difficult to identify this feature without appropriate assumptions about the latent risk dynamics. To illustrate this feature, Figure 4.10 presents sample variance ratios for horizons from two to 20 years, computed from our 65-year sample of observed annual log returns, together with the 10%, 50% and 90%-quantile of the variance ratio's Monte Carlo distribution (obtained from 10000 samples of returns), obtained using a nonparametric bootstrap procedure. The model is consistent

²⁰The expression for the conditional variance in the constant risks case is analogous to expression (4.51), but with $Var_t(\varepsilon_{t+j}^\mu)$, $Var_t(\tilde{\varepsilon}_{t+j}^r)$ and $Cov_t(\varepsilon_{t+j}^\mu, \tilde{\varepsilon}_{t+j}^r)$ that are constant functions only of the model parameters.

with Siegel (2008) observation that variance ratios decrease with the horizon and observed sample values are inside the 80% confidence interval of the Monte Carlo simulation.²¹ Similar unreported Monte Carlo results show that the term structure of market risk estimated by iterated VAR forecasts under the assumptions of our present-value model is monotonically decreasing, as in the data (see Campbell and Viceira (2005)).

In the same way we can analyse the model-implied conditional variance of a n -year dividend growth, which is derived from equation (4.48) as follows:

$$\begin{aligned} Var_t \left[\sum_{j=1}^n \rho^{j-1} \Delta d_{t+j} \right] &= \sum_{j=1}^{n-1} \rho^{2j} \left(\frac{1 - (\rho\gamma_1)^{n-j}}{1 - \rho\gamma_1} \right)^2 Var_t(\varepsilon_{t+j}^g) + \sum_{j=1}^n \rho^{2(j-1)} Var_t(\tilde{\varepsilon}_{t+j}^D) \\ &+ 2 \sum_{j=1}^{n-1} \rho^{2j-1} \frac{1 - (\rho\gamma_1)^{n-j}}{1 - \rho\gamma_1} Cov_t(\varepsilon_{t+j}^g, \tilde{\varepsilon}_{t+j}^D), \end{aligned} \quad (4.52)$$

where $Var_t(\varepsilon_{t+j}^g)$, $Var_t(\tilde{\varepsilon}_{t+j}^D)$ and $Cov_t(\varepsilon_{t+j}^g, \tilde{\varepsilon}_{t+j}^D)$ are affine functions of the variance-covariance state Σ_t , given explicitly in Appendix C.1.4. Analogously to the model-implied term structure of market risk, we define the term structure of dividend growth risk as the annualized volatility of cumulative cash flow growth versus the investment horizon. Right panel of Figure 4.8 shows that this term structure is highly time-varying and decreasing on average, but it can also be increasing or hump-shaped and reverts to a long-term value as maturity increases.

Figure 4.11 plots the estimated term structure of dividend risk and its three components at three different points in time. The first term (blue line) reflects uncertainty about future expected dividend growth and is generally positive and slightly increasing with maturity, but always close to zero. The second term (red line) is related to the risk of future dividend growth shocks and captures the main part of the overall dividend growth variance. This term can be increasing, decreasing or hump shaped and for large maturities converges to long-term value of about 0.001 due to the mean reversion properties of the conditional dividend volatility dynamics. The third variance component reflects the correlation between realized and expected dividend growth shocks, which is time-varying and can also switch sign. However, this explains only a small part of the total variance. Within the constant risk model (bottom right panel of Figure 4.11) this third component is put to zero by assumption, the effect of the dividend growth shock is also negligible and all the term structure of dividend growth risk is driven by the uncertainty about future expected dividend growth, which strongly increases with the horizon, at least until a maturity of about 15 years and then slightly decreases.

Our model is thus consistent with the empirical evidence reported in Binsbergen, Brandt, and Koijen (2012) and Binsbergen, Hueskes, Koijen, and Vrugt (2013), which suggests a decreasing term structure of volatilities on dividend strips (i.e. claims to dividends paid over some specified future time interval), but has also the flexibility to let this term structure change over time, for example with the business cycle.

²¹The width of these confidence intervals increases rapidly with the horizon, due to the decreasing number of long-horizon returns. The variance ratio at horizon n is defined as the sample variance of n -year returns, divided by n times the sample variance of 1-year returns. Calculations are based on overlapping returns and unbiased variance estimates, as for instance in equation (2.4.37) of Campbell, Lo, and MacKinlay (1997).

Figure 4.8: Term Structure of conditional return and dividend growth volatility

Dynamics of the term structure of the conditional per-period long-horizon return (left panel) and dividend growth (right panel) volatility, from equations (4.51) and (4.52), respectively, computed using estimated parameters and filtered state. We consider horizons of 1 to 20 years.

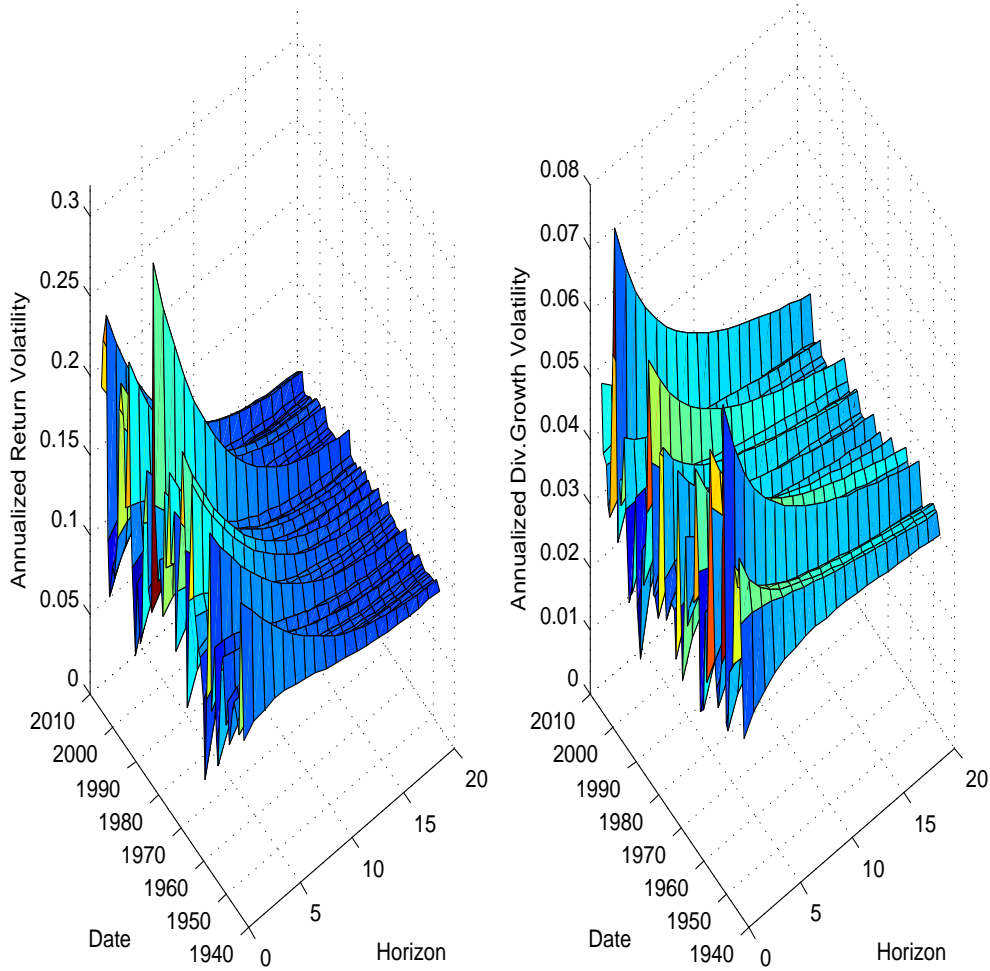


Figure 4.9: Decomposition of the term structure of conditional variance

Decomposition of the term structure of the conditional per-period variance of long-horizon returns, computed using estimated parameters and filtered states. The blue line denotes the component of the variance that is due to uncertainty about future expected returns, the red line denotes the component due to future return shocks, while the green line denotes the mean reversion component. The black dashed line denotes the total conditional variance, for horizons of 1 to 20 years. The first three panels show the decomposition implied by our model at different points in time, while the last (bottom right) panel considers the term structure estimated for the constant risks model.

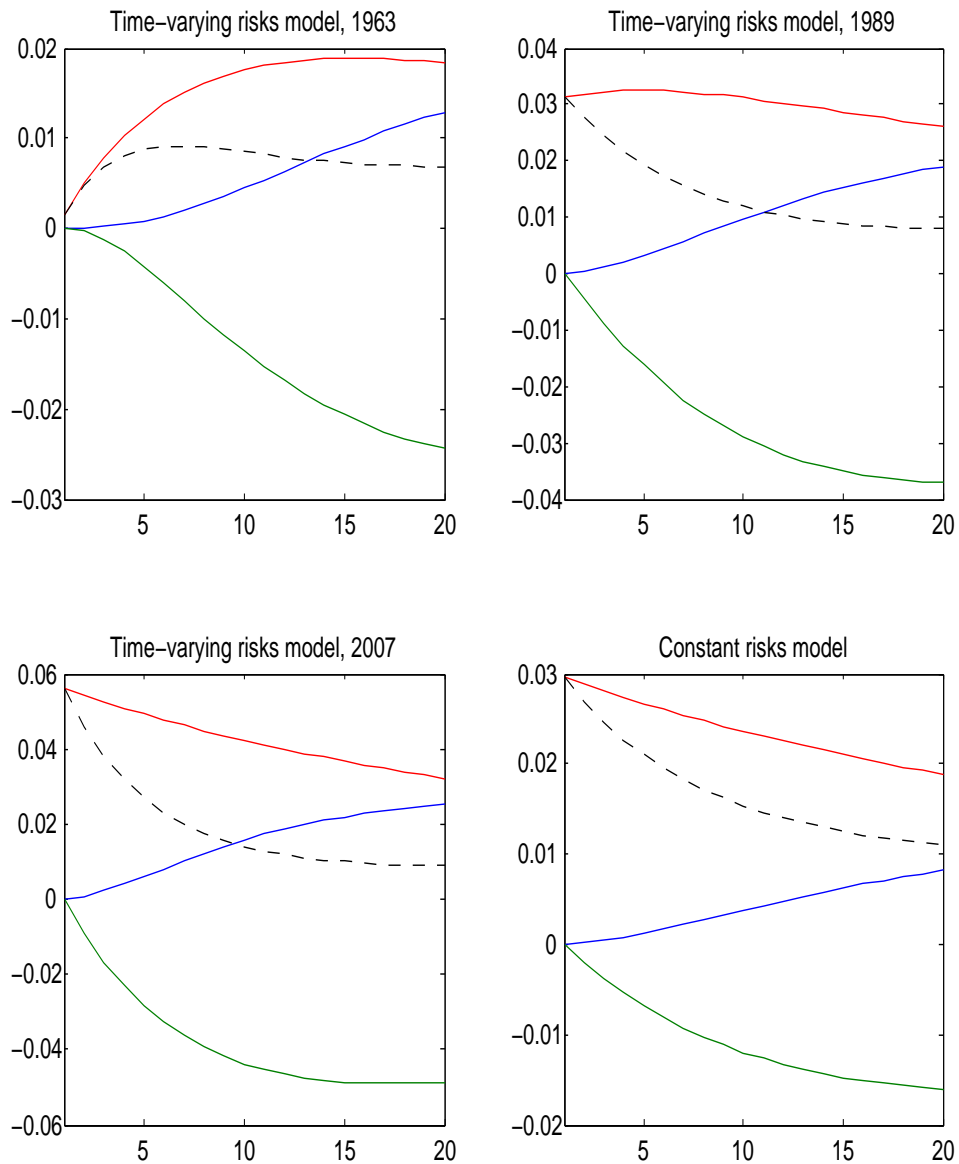


Figure 4.10: Observed vs simulated variance ratios

Sample variance ratios for horizons of 2 to 20 years, computed from the 65-year sample of annual log stock market returns (red line) and from 10000 samples of returns simulated from the model (blue lines) using a nonparametric bootstrap procedure. Solid blue line denotes the median variance ratios of the 10000 simulations, while dashed lines represent 10%- and 90%-quantiles.

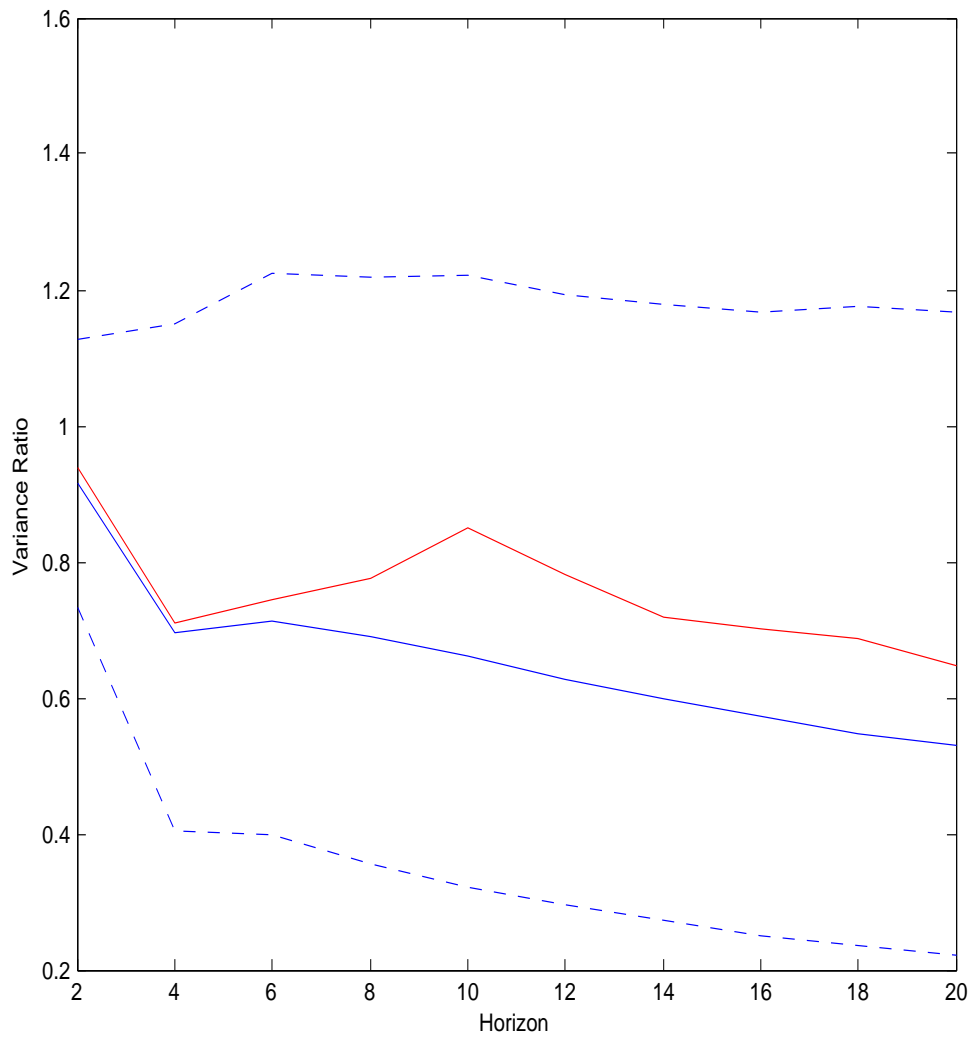
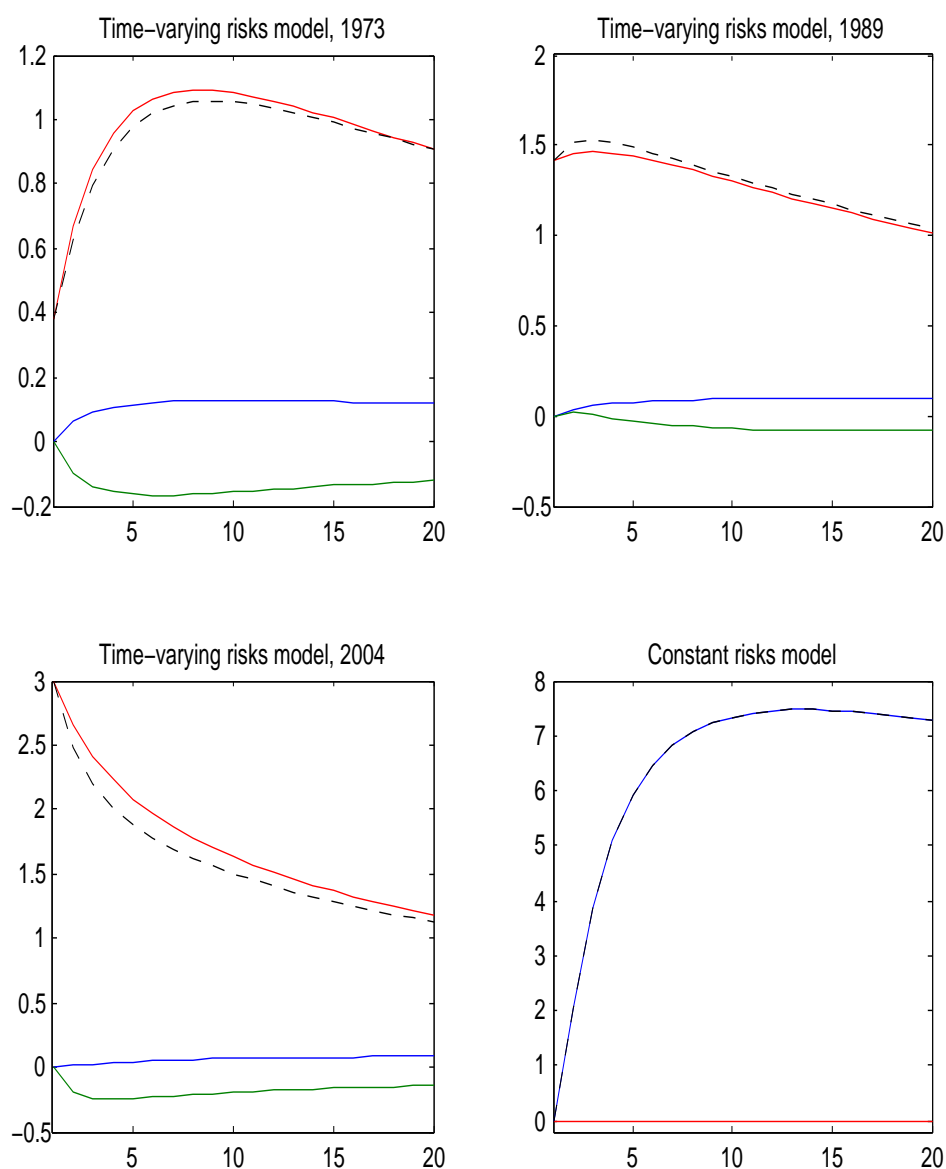


Figure 4.11: Decomposition of the term structure of conditional variance of dividend growth

Decomposition of the term structure of the conditional per-period variance of long-horizon dividend growth, computed using estimated parameters and filtered states. The blue line denotes the component of the variance that is due to uncertainty about future expected cash flow growth, the red line denotes the component due to future dividend growth shocks, while the green line denotes the mean reversion component. The black dashed line denotes the total conditional variance, for horizons of 1 to 20 years. The first three panels show the decomposition implied by our model at different points in time, while the last (bottom right) panel considers the term structure estimated for the constant risks model. In all panels, values are multiplied by 10^3 for readability.



4.5 Conclusion

In a tractable Campbell and Shiller (1988) present-value model with predictable risks, we propose a latent variable approach to estimate the joint predictability features of aggregate dividends, market returns and their time-varying risks. We specify exogenous latent processes for expected returns, expected dividend growth and the conditional variance-covariance matrix of dividends and returns and use filtering methods to uncover their joint time series process. We then use the model solutions to study in a single common framework the implications for (i) the dividend and return predictability features at short and long horizons, (ii) the dynamics of market Sharpe ratios and (iii) the time-series and cross-sectional properties of the term structure of market risks.

First, we find that expected dividend growth and expected return are both time-varying and explain about 2.6% and 9.4% of actual return and dividend growth variability, respectively, at an annual frequency. The estimated expected return process is more persistent and gives rise to an economically relevant price-dividend ratio component that masks the predictive power of valuation ratios in standard predictive regressions for dividend growth. Second, we estimate a quite volatile market Sharpe ratio featuring a more pronounced countercyclicality than in models assuming constant risks. Third, we obtain highly time-varying term structures of market expectations and risks of both dividend growth and returns, which have the flexibility to take different shapes. Through these mechanics, our model implies a variety of predictive features that are consistent with a number of findings in the literature. These include e.g. the very weak post-war evidence of dividend growth predictability at yearly horizons and the low real-time predictability of stock returns..

Therefore, these findings show more generally the usefulness of jointly controlling for predictable risks and persistent dividend and return forecasts when studying predictive relations in present-value models.

The features that we document could also be a guideline to build new equilibrium asset pricing models. Binsbergen, Brandt, and Koijen (2012) analyse the ability of leading asset pricing models²² to replicate the observed properties of dividend strips and find e.g. that none of them is fully able to match the decreasing pattern of the term structure of dividend strips volatility. In the same spirit, we could study which kind of asset pricing model is able to generate the predictability and time-varying risk features that we document, such as the time-varying and often hump-shaped term structure of market risk.

A practically relevant question concerns the extent to which estimated predictability features can be exploited in real time, e.g., in order to build successful portfolio strategies. Given the economically relevant degree of persistence of expected market returns estimated by our model, a key aspect is likely the identification of relevant predictive variables, spanning the information set available to investors for building their return expectation μ_t . A first interesting approach can make use of cross-sectional information on individual stocks, in order to better span investors' information set. Kelly and Pruitt (2013) propose and estimate a predictive factor model with constant risks, in which cross-sectional information from individual price-dividend ratios is aggregated to forecast market returns and dividends. In a similar spirit, Brennan and Taylor (2010) extract aggregate discount rate news from returns of equity portfolios based on individual stock characteristics like size and book-to-market. A second useful way of enlarging the predictive information set can make use of more direct proxies of cash-flow forecasts derived, e.g., from either synthetic prices of dividend strips, which can be synthesized from options and

²²In particular, they consider external habit formation model (Campbell and Cochrane (1999)), long-run risk model (Bansal and Yaron (2004)), variable rare disasters model (Gabaix (2012)) and the model by Lettau and Wachter (2007).

futures data, or quotes for swaps, futures or options on dividends, which have been recently introduced in several exchanges; see, for instance, [Binsbergen, Brandt, and Koijen \(2012\)](#) and [Binsbergen, Hueskes, Koijen, and Vrugt \(2013\)](#). [Golez \(2011\)](#) corrects the price-dividend ratio in standard predictive regressions using information on dividend growth implied by S&P 500 options and futures and finds evidence of improved and more stable predictive relations. Such enriched predictive information set can prove useful also for a more accurate identification of the dynamics of the term structure of market risks in the context of present-value models with time-varying risks.

Appendix A

A.1 Proofs

A.1.1 Proof of Proposition 1

Given multiplier ξ for the constraint in representative agent problem (2.17),

$$U^A(C^A(t))' = \xi = \phi(t)U^B(C^B(t))' \Rightarrow \phi(t) = \frac{U^A(C^A(t))'}{U^B(C^B(t))'} = \frac{y^A\eta^A(t)}{y^B\eta^B(t)}, \quad (\text{A.1})$$

where the last equality follows from individual agents optimality. Equation (A.1) and the optimality condition of agent A imply:

$$U(C(t), \phi(t))' = U^A(C^A(t))' \frac{\partial C^A}{\partial C} + \phi U^B(C^B(t))' \frac{\partial C^B}{\partial C} = U^A(C^A(t))' = y^A e^{\delta t} \eta^A(t). \quad (\text{A.2})$$

Therefore, individual agents optimal consumptions can be written as

$$C^A(t) = (y^A e^{\delta t} \eta^A(t))^{-1/\gamma} = U'(C(t), \phi(t))^{-1/\gamma}, \quad (\text{A.3})$$

$$C^B(t) = (y^B e^{\delta t} \eta^B(t))^{-1/\gamma} = \left(\frac{y^A e^{\delta t} \eta^A(t)}{\phi(t)} \right)^{-1/\gamma} = \left(\frac{U'(C(t), \phi(t))}{\phi(t)} \right)^{-1/\gamma}. \quad (\text{A.4})$$

Using the market clearing condition $C(t) = C^A(t) + C^B(t)$,

$$C(t) = U'(C(t), \phi(t))^{-1/\gamma} + \left(\frac{U'(C(t), \phi(t))}{\phi(t)} \right)^{-1/\gamma},$$

which can be solved for the marginal utility of the representative agent, leading to

$$U'(C(t), \phi(t)) = \frac{(1 + \phi(t)^{1/\gamma})^\gamma}{C(t)^\gamma}. \quad (\text{A.5})$$

Inserting (A.5) in (A.3) and (A.4) lead to the equilibrium consumption allocations in equation (2.19) and the investors' state price densities (2.20).

A.1.2 Proof of Proposition 2

The state price density of agent A is given in Proposition 1:

$$\eta^A(t) = e^{-\delta t} \frac{(1 + \phi(t)^{1/\gamma})^\gamma}{y^A C(t)^\gamma}$$

Applying Ito's lemma and using the dynamics of $\phi(t)$ and $C(t)$ in equations (2.4) and (2.11), respectively, the dynamics of η^A is given by:

$$\begin{aligned}
d\eta^A(t) = & -\delta\eta^A(t)dt + \eta^A(t)\frac{\phi(t)^{1/\gamma}}{1+\phi(t)^{1/\gamma}}(\beta^A - \beta^B)X(t)dt - \gamma\eta^A(t)(\mu dt + \sigma \sum_{j=1}^N s_j dW_{jt}) + \\
& + \frac{1}{2}\gamma(\gamma+1)\eta^A(t)\sigma^2 \sum_{j=1}^N s_j^2 dt + \eta^A(t) \sum_{j=1}^N [(s_j k + 1)^{-\gamma} - 1] dN_{jt} + \\
& + \eta^A(t) \left[\left(\frac{1 + \left(\phi(t) \frac{\beta^B}{\beta^A} \right)^{1/\gamma}}{1 + \phi(t)^{1/\gamma}} \right)^\gamma (k+1)^{-\gamma} - 1 \right] dN_{ct}. \tag{A.6}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{d\eta^A(t)}{\eta^A(t)} = & \left[-\delta + \frac{\phi(t)^{1/\gamma}}{1 + \phi(t)^{1/\gamma}}(\beta^A - \beta^B)X(t) - \gamma\mu + \frac{1}{2}\gamma(\gamma+1)\sigma^2 \sum_{j=1}^N s_j^2 \right] dt - \gamma\sigma \sum_{j=1}^N s_j dW_{jt} + \\
& + \sum_{j=1}^N [(s_j k + 1)^{-\gamma} - 1] dN_{jt} + \left[\left(\frac{1 + \left(\phi(t) \frac{\beta^B}{\beta^A} \right)^{1/\gamma}}{1 + \phi(t)^{1/\gamma}} \right)^\gamma (k+1)^{-\gamma} - 1 \right] dN_{ct}. \tag{A.7}
\end{aligned}$$

Comparing the drift, diffusion and jump terms of this expression with those in equation (2.16) directly leads to the solution for θ_j , λ_j^Q , λ_c^Q and r in Proposition 2.

A.1.3 Proof of Proposition 3

For simplicity, I first provide the derivation of the stock price expression for the case of a two-trees economy, to then extend it to the case of N stocks.

A.1.3.1 Stock prices in an economy with $N = 2$ stocks

For notational convenience, I compute the price of stock 1, the price of stock 2 follows an analogous expression with obvious modifications. Let me define $y_{it} \equiv \ln D_i(t)$ and $\tilde{y}_{i\tau} \equiv y_{i\tau} - y_{it}$. The price of stock 1 is given by the discounted value of all its future dividends:

$$S_1(t) = E_t^A \left[\int_t^T \frac{\eta^A(s)}{\eta^A(t)} D_1(s) ds \right]. \tag{A.8}$$

This can also be viewed as a portfolio of zero coupon dividend claims:

$$S_1(t) = \int_t^T S_1^\tau(t) d\tau$$

with

$$\begin{aligned}
S_1^\tau(t) &= E_t^A \left[\frac{\eta^A(\tau)}{\eta^A(t)} D_1(\tau) \right] \\
&= e^{-\delta(\tau-t)} \frac{C(t)^\gamma}{(1 + \phi(t)^{1/\gamma})^\gamma} E_t^A \left[\frac{(1 + \phi(\tau)^{1/\gamma})^\gamma}{C(\tau)^\gamma} D_1(\tau) \right] \\
&= e^{-\delta(\tau-t)} \frac{C(t)^\gamma}{(1 + \phi(t)^{1/\gamma})^\gamma} E_t^A \left[\sum_{k=0}^{\gamma} \binom{\gamma}{k} \phi(\tau)^{k/\gamma} \frac{e^{y_{1t} + \tilde{y}_{1\tau}}}{(e^{y_{1t} + \tilde{y}_{1\tau}} + e^{y_{2t} + \tilde{y}_{2\tau}})^\gamma} \right] \\
&= e^{-\delta(\tau-t)} \frac{C(t)^\gamma}{(1 + \phi(t)^{1/\gamma})^\gamma} E_t^A \left[\sum_{k=0}^{\gamma} \binom{\gamma}{k} \phi(\tau)^{k/\gamma} \frac{e^{(1-\gamma/2)y_{1t} - \gamma/2 y_{2t} + (1-\gamma/2)\tilde{y}_{1\tau} - \gamma/2 \tilde{y}_{1\tau}}}{\left(2 \cosh \left(\frac{y_{2t} - y_{1t} + \tilde{y}_{1\tau} - \tilde{y}_{2\tau}}{2} \right)\right)^\gamma} \right] \\
&= e^{-\delta(\tau-t)} \frac{C(t)^\gamma}{(1 + \phi(t)^{1/\gamma})^\gamma} e^{(1-\gamma/2)y_{1t} - \gamma/2 y_{2t}} \sum_{k=0}^{\gamma} \binom{\gamma}{k} E_t^A \left[\frac{e^{k/\gamma \ln \phi(\tau) + (1-\gamma/2)\tilde{y}_{1\tau} - \gamma/2 \tilde{y}_{1\tau}}}{\left(2 \cosh \left(\frac{y_{2t} - y_{1t} + \tilde{y}_{1\tau} - \tilde{y}_{2\tau}}{2} \right)\right)^\gamma} \right] \quad (A.9)
\end{aligned}$$

assuming an integer coefficient of relative risk aversion γ . Then I use the fact that

$$\frac{1}{[2 \cosh(u/2)]^\gamma} = \int_{-\infty}^{\infty} e^{iuz} \mathcal{F}_\gamma(z) dz,$$

where the Fourier transform $\mathcal{F}_\gamma(z)$ is given by (see [Martin \(2013\)](#)):

$$\mathcal{F}_\gamma(z) \equiv \frac{\Gamma(\gamma/2 + iz)\Gamma(\gamma/2 - iz)}{2\pi\Gamma(\gamma)}. \quad (A.10)$$

The conditional expectation in Equation (A.9) can thus be written as

$$\begin{aligned}
E_t^A \left[\frac{e^{k/\gamma \ln \phi(\tau) + (1-\gamma/2)\tilde{y}_{1\tau} - \gamma/2 \tilde{y}_{1\tau}}}{\left(2 \cosh \left(\frac{y_{2t} - y_{1t} + \tilde{y}_{1\tau} - \tilde{y}_{2\tau}}{2} \right)\right)^\gamma} \right] &= \int_{-\infty}^{\infty} \mathcal{F}_\gamma(z) e^{iz(y_{2t} - y_{1t})} E_t^A \left[e^{k/\gamma \ln \phi(\tau) + (1-\gamma/2 - iz)\tilde{y}_{1\tau} + (-\gamma/2 + iz)\tilde{y}_{2\tau}} \right] dz \\
&= \int_{-\infty}^{\infty} \mathcal{F}_\gamma(z) e^{iz(y_{2t} - y_{1t})} D_1(t)^{-(1-\gamma/2 - iz)} D_2(t)^{-(iz - \gamma/2)} \\
&\quad \cdot e^{(\tau-t)[(1-\gamma)(\mu - 1/2\sigma^2) + 1/2\sigma^2((1-\gamma/2 - iz)^2 + (-\gamma/2 - iz)^2) + (1-\gamma/2 - iz)y_1^c + (-\gamma/2 - iz)y_2^c]} \\
&\quad \cdot E_t^A \left[e^{k/\gamma \ln \phi(\tau) + (1-\gamma/2 - iz)y_1^d(\tau) + (-\gamma/2 + iz)y_1^d(\tau)} \right] dz, \quad (A.11)
\end{aligned}$$

where $y_i^c(t)$ and $y_i^d(t)$, for $i = 1, 2$ are the diffusion and jump components of log dividends and their dynamics are given by:

$$dy_i^c(t) = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_{it}, \quad (A.12)$$

$$dy_i^d(t) = \ln(k+1) (dN_{it} + dN_{ct}), \quad (A.13)$$

and

$$d \ln \phi(t) = (\beta^A - \beta^B) X(t) dt + \ln \left(\frac{\beta^B}{\beta^A} \right) dN_{ct}, \quad (A.14)$$

from equation (2.4).

Thus, denoting $x(t) \equiv \frac{k}{\gamma} \ln \phi(\tau) + (1 - \gamma/2 - iz)y_1^d(\tau) + (-\gamma/2 + iz)y_1^d(\tau)$, its dynamics follows:

$$\begin{aligned}
dx(t) &= \frac{k}{\gamma} (\beta^A - \beta^B) X(t) dt + \underbrace{(1 - \gamma/2 - iz) \ln(k+1)}_{k_x^1} dN_{1t} + \underbrace{(iz - \gamma/2) \ln(k+1)}_{k_x^2} dN_{2t} \\
&\quad + \underbrace{\left[(1 - \gamma) \ln(k+1) + \frac{k}{\gamma} \ln \left(\frac{\beta^B}{\beta^A} \right) \right]}_{k_x^c} dN_{ct}, \quad (A.15)
\end{aligned}$$

and also depends on the state variable $X(t)$. Let me define $Y = \begin{bmatrix} x \\ X \end{bmatrix}$, whose dynamics can be written as

$$dY = \left(\begin{bmatrix} 0 \\ \varphi \end{bmatrix} + \begin{bmatrix} 0 & \frac{k}{\gamma}(\beta^A - \beta^B) \\ 0 & -\varphi \end{bmatrix} Y \right) dt + \begin{bmatrix} 0 \\ \sigma_X \sqrt{X} \end{bmatrix} dW_t^X + \begin{bmatrix} k_x^1 \\ 0 \end{bmatrix} dN_{1t} + \begin{bmatrix} k_x^2 \\ 0 \end{bmatrix} dN_{2t} + \begin{bmatrix} k_x^c \\ 0 \end{bmatrix} dN_{ct}, \quad (\text{A.16})$$

with jump intensities $\lambda_Y^1 = \lambda_Y^2 = \lambda$ and $\lambda_Y^c = [0 \quad \beta^A]Y$.

The conditional expectation in equation (A.11) can thus be written as

$$f(Y, t) = E_t^A [e^{wY}],$$

with $w = [1 \quad 0]$ and since Y follows an affine jump diffusion we know (see Duffie, Pan, and Singleton (2000)) that $f(Y, t)$ is of the form $f(Y, t) = e^{\alpha_{0,k}(s) + \alpha_{1,k}(s)x(t) + \alpha_{2,k}(s)X(t)}$, where $s = \tau - t$ and $\alpha_{0,k}(s)$, $\alpha_{1,k}(s)$ and $\alpha_{2,k}(s)$ follow the system of Riccati equations:

$$\alpha'_{0,k}(s) = \alpha_{2,k}(s)\varphi + \lambda \left(e^{k_x^1} - 1 + e^{k_x^2} - 1 \right), \quad (\text{A.17})$$

$$\alpha'_{1,k}(s) = 0, \quad (\text{A.18})$$

$$\alpha'_{2,k}(s) = \frac{k}{\gamma}(\beta^A - \beta^B)\alpha_{1,k}(s) - \varphi\alpha_{2,k}(s) + \frac{1}{2}\sigma_X^2\alpha_{2,k}(s)^2 + \beta^A \left(e^{k_x^c} - 1 \right), \quad (\text{A.19})$$

with initial conditions, $\alpha_{0,k}(0) = \alpha_{2,k}(0) = 0$ and $\alpha_{1,k}(0) = 1$.

From (A.18) we find $\alpha_{1,k}(s) = \alpha_{1,k}(0) = 1$, while $\alpha_{0,k}(s)$ and $\alpha_{2,k}(0)$ are easily solved numerically.¹

Therefore, $f(Y, t) = e^{\alpha_{0,k}(s) + x(t) + \alpha_{2,k}(s)X(t)}$, and the price of a zero coupon dividend claim on the is

$$\begin{aligned} S_1^\tau(t) &= e^{-\delta(\tau-t)} \frac{C(t)^\gamma}{(1 + \phi(t)^{1/\gamma})^\gamma} \frac{D_1(t)}{(D_1(t)D_2(t))^{\gamma/2}} \sum_{k=0}^{\gamma} \binom{\gamma}{k} \phi(t)^{k/\gamma} \int_{-\infty}^{\infty} \mathcal{F}_\gamma(z) e^{iz(y_{2t} - y_{1t})} \\ &\quad \cdot e^{(\tau-t)[(1-\gamma)(\mu-1/2\sigma^2)+1/2\sigma^2((1-\gamma/2-iz)^2+(-\gamma/2-iz)^2)] + \alpha_{0,k}(\tau-t) + \alpha_{2,k}(\tau-t)X(t)} dz \\ &= D_1(t) \left(\frac{C(t)}{\sqrt{D_1(t)D_2(t)}} \right)^\gamma \sum_{k=0}^{\gamma} \binom{\gamma}{k} \frac{\phi(t)^{k/\gamma}}{(1 + \phi(t)^{1/\gamma})^\gamma} \int_{-\infty}^{\infty} \mathcal{F}_\gamma(z) e^{iz u_t} \\ &\quad \cdot e^{(\tau-t)[- \delta + (1-\gamma)(\mu-1/2\sigma^2)+1/2\sigma^2((1-\gamma/2-iz)^2+(-\gamma/2-iz)^2)] + \alpha_{0,k}(\tau-t) + \alpha_{2,k}(\tau-t)X(t)} dz \quad (\text{A.20}) \end{aligned}$$

Therefore the price of the aggregate consumption claim is

$$\begin{aligned} S_1(t) &= \int_t^T S_1^\tau(t) d\tau \\ &= D_1(t) \left[2 \cosh\left(\frac{u_t}{2}\right) \right]^\gamma \sum_{k=0}^{\gamma} a_k(\phi) \int_{-\infty}^{\infty} \mathcal{F}_\gamma(z) e^{iz u_t} b_k(X, t, z) dz, \quad (\text{A.21}) \end{aligned}$$

where

$$a_k(\phi) = \binom{\gamma}{k} \frac{\phi(t)^{k/\gamma}}{(1 + \phi(t)^{1/\gamma})^\gamma}, \quad (\text{A.22})$$

$$b_k(X, t) = \int_t^T e^{(\tau-t)[- \delta + (1-\gamma)(\mu-1/2\sigma^2)+1/2\sigma^2((1-\gamma/2-iz)^2+(-\gamma/2-iz)^2)] + \alpha_{0,k}(\tau-t) + \alpha_{2,k}(\tau-t)X(t)} d\tau, \quad (\text{A.23})$$

¹In my simulations I use a Runge-Kutta 4th order method.

and the price-dividend ratio of stock 1 is

$$g_1(\phi, X, u, t) \equiv \frac{S_1(t)}{D_1(t)} = \left[2 \cosh \left(\frac{u_t}{2} \right) \right]^\gamma \sum_{k=0}^{\gamma} a_k(\phi) \int_{-\infty}^{\infty} \mathcal{F}_\gamma(z) e^{izut} b_k(X, t, z) dz.$$

In the same way it is possible to obtain the closed form expression for the price of stock 2, $S_2(t)$.

A.1.3.2 General stock price expressions in an economy with N stocks

The basic approach is the same with $N > 2$ assets. The main technical difficulty lies in generalizing $\mathcal{F}_\gamma(z)$ to the N -asset case, but this problem is solved by [Martin \(2013\)](#), who defines

$$\mathcal{F}_\gamma^N(\mathbf{z}) \equiv \frac{\Gamma(\gamma/N + iz_1 + iz_2 + \dots + iz_{N-1})}{(2\pi)^{N-1} \Gamma(\gamma)} \prod_{k=1}^{N-1} \Gamma(\gamma/N - iz_k). \quad (\text{A.24})$$

The price-dividend ratio on an asset j is thus:

$$g_j(\phi, X, \mathbf{u}, t) = e^{-\gamma \sum_{j=2}^N u_j/N} (1 + e^{u_2} + \dots + e^{u_N})^\gamma \sum_{k=0}^{\gamma} a_k(\phi) \int \mathcal{F}_\gamma^N(\mathbf{z}) e^{i\mathbf{u}'\mathbf{z}} b_{jk}(X, t, \mathbf{z}) d\mathbf{z}, \quad (\text{A.25})$$

where the integral is evaluated on \mathbb{R}^{N-1} , $a_k(\phi)$ is given in Equation (A.22) and $b_{jk}(X, t, \mathbf{z})$ generalizes Equation (A.23) as follows:

$$b_{jk}(X, t, \mathbf{z}) = \int_t^T e^{(\tau-t)[-\delta + (\mu - \frac{1}{2}\sigma^2)\mathbf{1}'_N(\mathbf{e}_j - \gamma/N + i\mathbf{U}'\mathbf{z}) + \frac{1}{2}\sigma^2(\mathbf{e}_j - \gamma/N + i\mathbf{U}'\mathbf{z})'(\mathbf{e}_j - \gamma/N + i\mathbf{U}'\mathbf{z})] + \alpha_{0,k}^N(\tau-t) + \alpha_{2,k}^N(\tau-t)X(t)} d\tau,$$

where \mathbf{e}_j is the N -vector with a 1 at the j -th entry and zeros elsewhere, $\mathbf{1}_N$ is a N -dimensional vector of ones and $\alpha_{0,k}^N(\tau)$ and $\alpha_{2,k}^N(\tau)$ satisfy the following system of Riccati equations:

$$\begin{aligned} \alpha_{0,k}^N(s)' &= \alpha_{2,k}^N(s)\varphi + \lambda \sum_{i=1}^N \left(e^{k_x^i} - 1 \right), \\ \alpha_{2,k}^N(s)' &= \frac{k}{\gamma}(\beta^A - \beta^B) - \varphi \alpha_{2,k}^N(s) + \frac{1}{2}\sigma_X^2 \alpha_{2,k}^N(s)^2 + \beta^A \left(e^{k_x^c} - 1 \right), \end{aligned}$$

with initial conditions, $\alpha_{0,k}^N(0) = \alpha_{2,k}^N(0) = 0$.

A.1.3.3 The case of a large economy: $N \rightarrow \infty$

In the case of a large diversified economy considered in Section 2.3.4, i.e. N large and $s_j = 1/N$, price expressions simplify since $u_j = 0 \quad \forall j$. The price-dividend ratio of any stock j is given by:

$$g_j(\phi, X, \mathbf{u}, t) = N^\gamma \sum_{k=0}^{\gamma} a_k(\phi) \int \mathcal{F}_\gamma^N(\mathbf{z}) b_{jk}(X, t, \mathbf{z}) d\mathbf{z}. \quad (\text{A.26})$$

A.2 Variance risk premium due to large jumps

[Bollerslev and Todorov \(2011\)](#) develop a nonparametric method to isolate the fraction of the observed variance risk premium due to large jumps. Since the model-implied variance risk premium only includes compensation for jump risk, it is useful, as a robustness check, to run the

predictive regressions in Section 2.4 using Bollerslev and Todorov (2011)'s time series of market variance risk premium only due to large jumps (VRP^j), which is available from February 1996 through July 2007. Figure A.1 compares this measure of the variance premium only due to jumps with the variance risk premium measure used in the main text. The correlation between the two series is about 73%. I first consider the predictive regression

$$r_{t+6}^e = \alpha + \beta VRP_t^j + \varepsilon_{t+6}, \quad (\text{A.27})$$

where r_{t+6}^e is the excess return of the *S&P500* index at a 6 months horizon. Figure A.2 shows regression coefficient estimates with 95% confidence bounds (upper panel) and adjusted R^2 in percentage (upper panel) estimated on a rolling window of 50 months. Consistent with the results in the main text, predictive power is stronger in phases of large disagreement, such as in the early 2000 and at the onset of the recent financial crisis. Then, analogously to Figure 2.14, Figure A.3 shows the distributions of regression coefficients (upper panel) and R^2 (lower panel), for small, average, and large values of the difference in beliefs, obtained applying a bock bootstrap procedure. Both the regression coefficient and the adjusted R^2 increase (in absolute value) with the level of DB. Results are a bit less strong than what I find using the aggregate variance risk premium in the man text, but they have to be taken with caution since the number of observations in every bin is small.²

Then I estimate regressions of the form:

$$r_{i,t+6}^e = \alpha_i + \beta_i VRP_t^j + \varepsilon_{t+6}, \quad (\text{A.28})$$

for $i = S, M, B$, where $r_{i,t+h}^e$ is the monthly excess returns on small-, mid-, and big-cap portfolios, respectively, at the 6-month horizon. Consistent with the results in the main text, predictive power is stronger for small stocks, with an adjusted R^2 of 2.1% against an adjusted R^2 of -0.4% for large stocks.

²Since VRP^j is available only from February 1996 through July 2007, there are less than 50 monthly observations for each DB quantile.

Figure A.1: Variance risk premium due to jumps

Time series of variance risk premium only due to jumps, from Bollerslev and Todorov (2011), versus the total variance risk premium measure described in Section 2.4.1, for the overlapping sample, that goes from February 1996 through July 2007. Both measures are in monthly squared percentage.

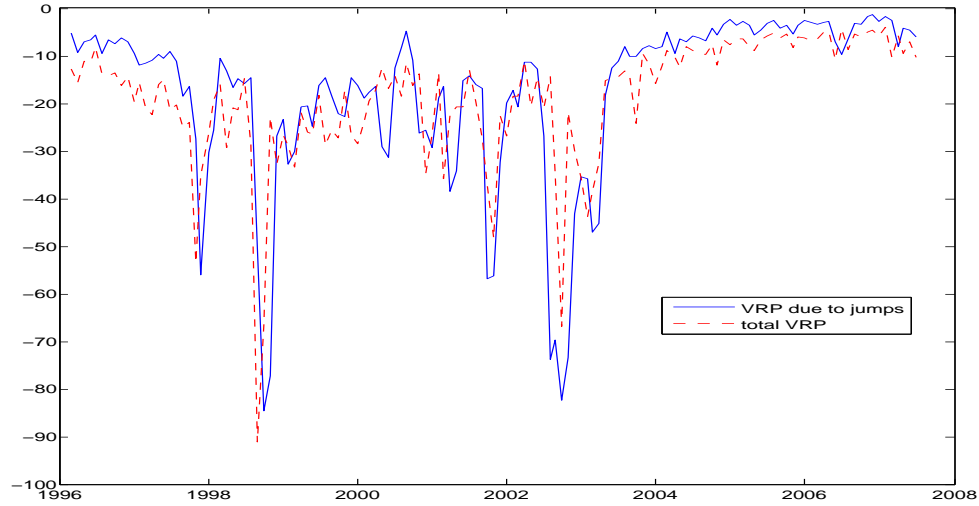


Figure A.2: Market return predictive regressions using jumps variance risk premium

Predictive regressions of market 6-month excess returns on lagged variance risk premium only due to jumps, from Bollerslev and Todorov (2011), on a rolling window of 50 months. Upper panel shows regression coefficient estimates with 95% confidence bounds, while lower panel reports adjusted R^2 in percentage.

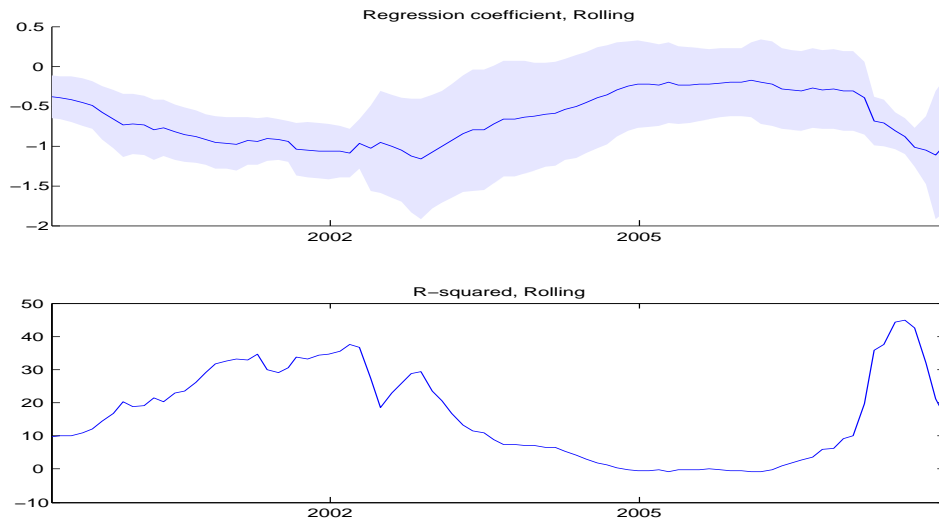
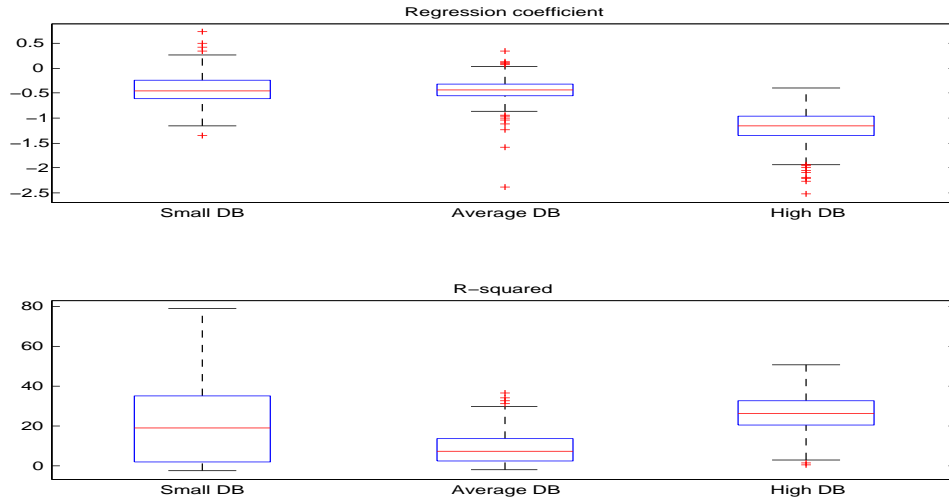


Figure A.3: Market return predictive regressions using jumps variance risk premium for different values of the difference in beliefs

Standard OLS regression of excess market returns at the six-month horizon on the lagged variance risk premium only due to jumps, from Bollerslev and Todorov (2011), for different levels of the difference in beliefs (DB). The first box plot corresponds to small values of disagreement ($DB < q_{30\%}$), the last to large values ($DB > q_{70\%}$) and the middle box plot to average values of DB. Upper panel display the distribution of regression coefficients and lower panel of percentage R^2 , both obtained applying a block bootstrap procedure.



Appendix B

B.1 Price-dividend ratio

In this section we present the detailed derivation of equation (3.7) in the text. From Campbell and Shiller (1988) we have

$$pd_t \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1}. \quad (\text{B.1})$$

By iterating this equation we find:

$$\begin{aligned} pd_t &\simeq \kappa + \rho(\kappa + \rho pd_{t+2} + \Delta d_{t+2} - r_{t+2}) + \Delta d_{t+1} - r_{t+1} \\ &= \sum_{j=0}^{\infty} \rho^j \kappa + \rho^\infty pd_\infty + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \\ &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}), \end{aligned} \quad (\text{B.2})$$

assuming that $\rho^\infty pd_\infty = \lim_{j \rightarrow \infty} \rho^j pd_{t+j} = 0$, at least in expectation. Then, we take expectation conditional to time t :

$$\begin{aligned} pd_t &\simeq \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[\Delta d_{t+j} - r_{t+j}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[g_{t+j-1} - \mu_{t+j-1}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j E_t[g_{t+j} - \mu_{t+j}]. \end{aligned} \quad (\text{B.3})$$

Iterating the dynamics of $\hat{\mu}_{t+1}$ and \hat{g}_{t+1} and taking conditional expectation we find

$$E_t[\hat{\mu}_{t+j}] = \delta_1^j \hat{\mu}_t$$

and

$$E_t[\hat{g}_{t+j}] = \gamma_1^j \hat{g}_t.$$

Therefore,

$$\begin{aligned}
pd_t &\simeq \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1^j \hat{g}_t - \delta_0 - \delta_1^j \hat{\mu}_t] \\
&= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \frac{\hat{g}_t}{1-\rho\gamma_1} - \frac{\hat{\mu}_t}{1-\rho\delta_1} \\
&= A + B_2 \hat{g}_t - B_1 \hat{\mu}_t.
\end{aligned} \tag{B.4}$$

The explicit expressions for the present-value coefficients A , B_1 and B_2 are the following:

$$\begin{aligned}
A &= \frac{\kappa + \gamma_0 - \delta_0}{1-\rho}, \\
B_1 &= \frac{1}{1-\rho\delta_1}, \\
B_2 &= \frac{1}{1-\rho\gamma_1}.
\end{aligned}$$

B.2 Estimation Methodology

This appendix describes in detail the estimation procedure, first for the benchmark model in Section 3.1 and then for the extended model in Section 3.4.

B.2.1 Benchmark model

For estimation purposes, we cast the model in state-space form, using demeaned state variables $\hat{\mu}_t \equiv \mu_t - \delta_0$ and $\hat{g}_t \equiv g_t - \gamma_0$. We obtain the following linear transition dynamics:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g, \tag{B.5}$$

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \varepsilon_{t+1}^\mu. \tag{B.6}$$

The observable variables are dividend growth Δd_{t+1} and the price-dividend ratio pd_{t+1} . Measurement equations for Δd_{t+1} and pd_{t+1} are derived from the model-implied expressions for dividend growth and price-dividend ratio. The measurement equation for dividend growth is given by (??) while log price-dividend ratio is given by (3.7). Note however that Equation (3.7) contains no error term, and as shown by Binsbergen and Koijen (2010), this feature can be exploited to reduce the number of transition equations in the model. By substituting the equation for pd_t in the measurement equation for dividend growth, we arrive at a final system with one transition equation, (B.5), and two measurement equations:

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^d. \tag{B.7}$$

$$pd_{t+1} = (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + \delta_1 pd_t - B_1 \varepsilon_{t+1}^\mu + B_2 \varepsilon_{t+1}^g. \tag{B.8}$$

We use the Kalman filter to derive the likelihood of the model and we estimate it using ML. The parameters to be estimated are the following:

$$\theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{\mu d}, \rho_{gd}).$$

We assume that expectation processes are stationary, therefore parameters δ_1 and γ_1 are bounded to be less than one in absolute value. The covariance matrix of the shocks, (3.6), has to be positive definite, thus σ_g , σ_μ and σ_d are constrained to be positive, while the correlation parameters are between -1 and 1 .¹ Rytchkov (2012) shows that it is impossible to

¹Moreover, the condition $\rho_{g\mu}^2 + \rho_{\mu d}^2 + \rho_{gd}^2 < 1$ has to hold for Σ to be positive definite

identify the whole covariance structure of shocks even when an infinitely long history of returns and dividends is given, but only one element of Σ must be fixed to identify the whole matrix. Thus, for identification purposes, we impose the constraint $\rho_{gd} = 0$, as in [Binsbergen and Koijen \(2010\)](#). Overall the model implies 9 free parameters to estimate. The estimation procedure is the following: We first define an expanded 4-dimensional state vector by the concatenation of the original state variable \hat{g} and the process and observation noise random variables:

$$X_t = \begin{pmatrix} \hat{g}_{t-1} \\ \varepsilon_t^g \\ \varepsilon_t^\mu \\ \varepsilon_t^d \end{pmatrix},$$

which satisfies:

$$X_{t+1} = F X_t + \Gamma \varepsilon_{t+1}^X,$$

where

$$\varepsilon_{t+1}^X = \begin{pmatrix} \varepsilon_{t+1}^g \\ \varepsilon_{t+1}^\mu \\ \varepsilon_{t+1}^d \end{pmatrix},$$

with conditional variance Σ , given in (3.6). Moreover,

$$F = \begin{bmatrix} \gamma_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad \Gamma = \begin{bmatrix} 0_{1 \times 3} \\ I_3 \end{bmatrix},$$

The measurement equation,

$$Y_t = \begin{pmatrix} \Delta d_t \\ p d_t \end{pmatrix},$$

is of the form

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t,$$

where

$$M_0 = \begin{bmatrix} \gamma_0 \\ (1 - \delta_1)A \end{bmatrix}, \quad M_1 = \begin{bmatrix} 0 & 0 \\ 0 & \delta_1 \end{bmatrix},$$

and

$$M_2 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ B_2(\gamma_1 - \delta_1) & B_2 & -B_1 & 0 \end{bmatrix}.$$

The steps of the filter algorithm are the following:

- Initialize with the unconditional mean and covariance of the expanded state:

$$\begin{aligned} X_{0,0} &= 0_{4 \times 1}, \\ P_{0,0} &= E(X_t X_t'). \end{aligned}$$

- The time-update equations are

$$\begin{aligned} X_{t,t-1} &= F X_{t-1,t-1}, \\ P_{t,t-1} &= F P_{t-1,t-1} F' + \Gamma \Sigma \Gamma', \end{aligned}$$

- The prediction error η_t and the variance-covariance matrix of the measurement equations are then:

$$\begin{aligned}\eta_t &= Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}, \\ S_t &= M_2 P_{t,t-1} M_2',\end{aligned}\tag{B.9}$$

where Y_t is the observed value of the measurement equation at time t .

- Update filtering:

$$\begin{aligned}\mathcal{K}_t &= P_{t,t-1} M_2' S_t^{-1}, \\ X_{t,t} &= X_{t,t-1} + \mathcal{K}_t \eta_t, \\ P_{t,t} &= (I - \mathcal{K}_t M_2) P_{t,t-1},\end{aligned}$$

where \mathcal{K}_t is the *Kalman gain*.

To estimate model parameters, θ , we define the log-likelihood for each time t , assuming normally distributed observation errors, as

$$l_t(\theta) = -\frac{1}{2} \log |S_t| - \frac{1}{2} \eta_t' S_t^{-1} \eta_t,$$

where η_t and S_t denote prediction error of the measurement series and the covariance of the measurement series, respectively, obtained from the KF. Model parameters are chosen to maximize the log-likelihood of the data series:

$$\hat{\theta} \equiv \arg \max_{\theta} \mathcal{L}(\theta, \{Y_t\}_{t=1}^T),\tag{B.10}$$

with

$$\mathcal{L}(\theta, \{Y_t\}_{t=1}^T) = \sum_{t=1}^T l_t(\theta),$$

where T denotes the number of time periods in the sample of estimation.²

B.2.2 Extended Model

In the case of the extended model in Section 3.4.1, the transition dynamics are the following:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \gamma_2 \hat{z}_t + \varepsilon_{t+1}^g,\tag{B.11}$$

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \delta_2 \hat{z}_t + \varepsilon_{t+1}^\mu.\tag{B.12}$$

The observable variables are dividend growth Δd_{t+1} , the price-dividend ratio pd_{t+1} and an additional observable predictor variable, z_t . Since Equation (3.34) contains no error term, as for the benchmark model we can reduce the number of transition equations and we arrive at a final system with one transition equation, (B.11), and three measurement equations:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^d.\tag{B.13}$$

$$\begin{aligned}pd_{t+1} &= (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + [\gamma_2 B_2 + (\xi_1 - \delta_1)(B_4 - B_3) - \delta_2 B_1]\hat{z}_t + \\ &\quad + \delta_1 pd_t - B_1 \varepsilon_{t+1}^\mu + B_2 \varepsilon_{t+1}^g + (B_4 - B_3)\varepsilon_{t+1}^z.\end{aligned}\tag{B.14}$$

²For yearly data, as in our application, T is the number of years in the sample.

We use the Kalman filter to derive the likelihood of the model and we estimate it using ML. The parameters to be estimated are the following:

$$\theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{\mu d}, \rho_{gd}, \xi_0, \xi_1, \rho_{gz}, \rho_{\mu z}, \rho_{dz}, \sigma_z, \delta_2, \gamma_2).$$

For identification purposes, we impose the constraint $\rho_{gd} = 0$, as in [Binsbergen and Koijen \(2010\)](#) and [Yun \(2012\)](#). Overall the model implies 17 free parameters to estimate. The estimation procedure is the following: We first define an expanded 5-dimensional state vector by the concatenation of the original state variable \hat{g} and the process and observation noise random variables:

$$X_t = \begin{pmatrix} \hat{g}_{t-1} \\ \varepsilon_t^g \\ \varepsilon_t^\mu \\ \varepsilon_t^d \\ \varepsilon_t^z \end{pmatrix},$$

which satisfies:

$$X_{t+1} = FX_t + Bu_{t+1} + \Gamma \varepsilon_{t+1}^X,$$

where $u_t = z_{t-1} - \xi_0$ and

$$\varepsilon_{t+1}^X = \begin{pmatrix} \varepsilon_{t+1}^g \\ \varepsilon_{t+1}^\mu \\ \varepsilon_{t+1}^d \\ \varepsilon_{t+1}^z \end{pmatrix},$$

with conditional variance

$$\Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} & \sigma_{gz} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{\mu d} & \sigma_{\mu z} \\ \sigma_{gd} & \sigma_{\mu d} & \sigma_d^2 & \sigma_{dz} \\ \sigma_{gz} & \sigma_{\mu z} & \sigma_{dz} & \sigma_z^2 \end{bmatrix}. \quad (\text{B.15})$$

Moreover,

$$F = \begin{bmatrix} \gamma_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad B = [\gamma_2 \quad 0_{1 \times 4}]' \quad \text{and} \quad \Gamma = \begin{bmatrix} 0_{1 \times 4} \\ I_4 \end{bmatrix},$$

The measurement equation,

$$Y_t = \begin{pmatrix} \Delta d_t \\ pd_t \\ z_t \end{pmatrix},$$

is of the form

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t,$$

where

$$M_0 = \begin{bmatrix} \gamma_0 \\ (1 - \delta_1)A \\ \xi_0(1 - \xi_1) \end{bmatrix}, \quad M_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \delta_1 & \xi_2 \\ 0 & 0 & \xi_1 \end{bmatrix},$$

$$\xi_2 = \gamma_2 B_2 + (\xi_1 - \delta_1)(B_4 - B_3) - \delta_2 B_1,$$

and

$$M_2 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ B_2(\gamma_1 - \delta_1) & B_2 & -B_1 & 0 & B_4 - B_3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The steps of the filter algorithm are exactly as for the benchmark model (see previous subsection) apart from a slight change in the time-update equation for the state, which becomes:

$$X_{t,t-1} = FX_{t-1,t-1} + Bu_t.$$

B.3 Asymptotic EIV bias in standard predictive regressions

Standard predictive regressions of either returns or dividend growth rates on the lagged log price-dividend ratio suffer from an error-in-variables (EIV) problem, which does not disappear as the sample size increases. Indeed, the true model for aggregate stock returns is:

$$r_{t+1} = \delta_0 + \hat{\mu}_t + \varepsilon_{t+1}^r,$$

but we wrongly assume the following model to hold:

$$r_{t+1} = a_r + b_r pd_t + \tilde{\varepsilon}_{t+1}^r, \quad (\text{B.16})$$

where $pd_t = A - B_1 \hat{\mu}_t + B_2 \hat{g}_t$, and we try to estimate the true parameter $b_r = -1/B_1$ from (B.16). The p-limit of the OLS slope coefficient is the following:³

$$\hat{b}_r \longrightarrow \frac{\text{Cov}(pd_t, r_{t+1})}{\text{Var}(pd_t)},$$

where

$$\begin{aligned} \text{Cov}(pd_t, r_{t+1}) &= \text{Cov}(A - B_1 \hat{\mu}_t + B_2 \hat{g}_t, \delta_0 + \hat{\mu}_t + \tilde{\varepsilon}_{t+1}^r) \\ &= -B_1 \text{Var}(\hat{\mu}_t) + B_2 \text{Cov}(\hat{g}_t, \hat{\mu}_t) \\ \text{Var}(pd_t) &= B_1^2 \text{Var}(\hat{\mu}_t) + B_2^2 \text{Var}(\hat{g}_t) - 2B_1 B_2 \text{Cov}(\hat{g}_t, \hat{\mu}_t) \end{aligned}$$

so that

$$\hat{b}_r \longrightarrow \frac{1}{-B_1 + \frac{B_2^2 \text{Var}(\hat{g}_t) - B_1 B_2 \text{Cov}(\hat{g}_t, \hat{\mu}_t)}{B_2 \text{Cov}(\hat{g}_t, \hat{\mu}_t) - B_1 \text{Var}(\hat{\mu}_t)}},$$

and the unconditional variances and covariance of demeaned expected return and dividend growth are the following:

$$\begin{aligned} \text{Var}(\hat{\mu}_t) &= \frac{\sigma_\mu^2}{1 - \delta_1^2}, \\ \text{Var}(\hat{g}_t) &= \frac{\sigma_g^2}{1 - \gamma_1^2}, \\ \text{Cov}(\hat{g}_t, \hat{\mu}_t) &= \frac{\sigma_{g\mu}}{1 - \gamma_1 \delta_1}. \end{aligned}$$

Thus, the OLS slope coefficient in the regression of returns on lagged price-dividend ratio is biased. However, at the estimated parameters the bias is small due to the relative persistence of expected dividend growth and returns.

The model for aggregate log dividend growth is:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^d,$$

³Note that here we denote with \hat{b}_r the OLS estimate of the slope coefficient b_r in (B.16).

while the wrong model is:

$$\Delta d_{t+1} = a_d + b_d p d_t + \tilde{\varepsilon}_{t+1}^d, \quad (\text{B.17})$$

and we try to estimate the true parameter $b_d = 1/B_2$ from (B.17). The p-limit of the OLS slope is the following:

$$\hat{b}_d \longrightarrow \frac{\text{Cov}(p d_t, \Delta d_{t+1})}{\text{Var}(p d_t)},$$

where

$$\begin{aligned} \text{Cov}(p d_t, \Delta d_{t+1}) &= \text{Cov}(A - B_1 \hat{\mu}_t + B_2 \hat{g}_t, \gamma_0 + \hat{g}_t + \tilde{\varepsilon}_{t+1}^d) \\ &= B_2 \text{Var}(\hat{g}_t) - B_1 \text{Cov}(\hat{g}_t, \hat{\mu}_t) \end{aligned}$$

so that

$$\hat{b}_d \longrightarrow \frac{1}{B_2 + \frac{B_1^2 \text{Var}(\hat{\mu}_t) - B_1 B_2 \text{Cov}(\hat{g}_t, \hat{\mu}_t)}{B_2 \text{Var}(\hat{g}_t) - B_1 \text{Cov}(\hat{g}_t, \hat{\mu}_t)}}.$$

Therefore, the OLS slope coefficient in the regression of dividend growth on lagged price-dividend ratio is also biased. This bias is negative and, at the estimated parameters, much more significant than the one for standard return regressions.

B.4 Asymptotic Validity of the Bootstrap Likelihood Ratio Test

In this appendix we prove the validity of our nonparametric bootstrap likelihood ratio testing procedure, i.e., the equivalence in distribution of LR_T and LR_T^* in equations (3.10) and (3.26), respectively, when $B, T \rightarrow \infty$, under the null hypothesis H_0 . It is well known that if H_0 holds, as $T \rightarrow \infty$, LR_T follows a χ_r^2 distribution with r degrees of freedom, where r is the number of parameter constraints defining the null hypothesis H_0 . Therefore, we only need to show that also LR_T^* is asymptotically χ_r^2 distributed.

Without loss of generality, let us consider for brevity the case in which the null hypothesis to be tested is formed by zero restrictions, i.e., some of the model parameters are equal to zero. In such cases, the r restrictions can be written as $\theta_2 = 0_{r \times 1}$, where the parameter vector θ is partitioned as $\theta = [\theta_1' \ \theta_2']'$, possibly after some reordering of the elements, where θ_1 is $(k - r) \times 1$ and θ_2 is $r \times 1$ -dimensional.

Let $\hat{\theta}$ be the unconstrained ML estimator of θ , while the pseudo-true value of θ in the population under H_0 is denoted by $\theta^* = [\theta_1^{*'} \ 0_{1 \times r}]'$, where θ_1^* is the *pseudo-true value* of θ_1 , i.e., the maximum of the population expected log likelihood function with respect to θ_1 under the (potentially) incorrect assumption $H_0 : \theta_2 = 0_{r \times 1}$.

Stoffer and Wall (1991) show that nonparametric Monte Carlo bootstrap applied to the (standardized) innovations $\{\hat{e}_t := S_t^{-1/2}(\hat{\theta})\eta_t(\hat{\theta})\}_{t=1}^T$ yields a distribution of bootstrap residuals $\{\hat{e}_t^*\}_{t=1}^T$, which can be used to compute a bootstrap distribution of ML estimators $\hat{\theta}^*$:

$$\hat{\theta}^* = \arg \max_{\theta} \log \mathcal{L}(\theta, \{Y_t^*\}_{t=1}^T), \quad (\text{B.18})$$

where the Monte Carlo sequence $\{Y_t^*\}_{t=1}^T$ is obtained by simulating dynamics (3.20)-(3.21) based on bootstrap residuals $\{\hat{e}_t^*\}_{t=1}^T$ (see steps 1)-3) in Section 3.3.3). Stoffer and Wall (1991) also provide an asymptotic justification of this procedure, showing, under general conditions, the equivalence in distribution of $\sqrt{T}(\hat{\theta}^* - \hat{\theta})$ and $\sqrt{T}(\hat{\theta} - \theta^*)$ as $B, T \rightarrow \infty$, and assuming for simplicity $B = T$. For simplicity of notation we assume that the ML setting holds, but all results hold true with obvious modifications in a PML setting, using sandwich variance-covariance matrix estimators, see Stoffer and Wall (1991):

$$\sqrt{T}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, \mathcal{I}(\theta^*)^{-1}), \quad (\text{B.19})$$

where $\mathcal{I}(\theta) = \text{plim}_{T \rightarrow \infty} \frac{1}{T} E[-\partial^2 \log \mathcal{L}(\theta) / \partial \theta \partial \theta']$ is the asymptotic information matrix, and

$$\sqrt{T}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{d} N(0, \mathcal{I}(\theta^*)^{-1}). \quad (\text{B.20})$$

The constrained ML estimator $\hat{\theta}_0$ can then be expressed as $\hat{\theta}_0 = \begin{bmatrix} \hat{\theta}_1' & 0_{1 \times r} \end{bmatrix}'$, and the asymptotic distribution of $\hat{\theta}_1$ is given by:⁴

$$\sqrt{T}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N(0, \mathcal{I}_{11}(\theta_1^*)^{-1}), \quad (\text{B.21})$$

where $\mathcal{I}_{11}(\cdot)$ is the $(k-r) \times (k-r)$ top left block of the asymptotic information matrix $\mathcal{I}(\cdot)$ of the unrestricted model. Analogously, the constrained bootstrap Maximum Likelihood estimator $\hat{\theta}_0^*$ can be partitioned as $\hat{\theta}_0^* = \begin{bmatrix} \hat{\theta}_1^{*'} & 0_{1 \times r} \end{bmatrix}'$ and its asymptotic distribution is given by:

$$\sqrt{T}(\hat{\theta}_1^* - \hat{\theta}_1) \xrightarrow{d} N(0, \mathcal{I}_{11}(\theta_1^*)^{-1}). \quad (\text{B.22})$$

For ease of notation, let us denote by $l(\theta, y)$ the log-likelihood of the model, i.e. $l(\theta, y) \equiv \log \mathcal{L}(\theta, \{Y_t\}_{t=1}^T)$. Using a second order Taylor expansion around $\hat{\theta}^*$, the bootstrap log-likelihood $l(\hat{\theta}_0^*, y^*)$ can be written as

$$l(\hat{\theta}_0^*, y^*) = l(\hat{\theta}^*, y^*) - \frac{1}{2}(\hat{\theta}_0^* - \hat{\theta}^*)' H(\bar{\theta})(\hat{\theta}_0^* - \hat{\theta}^*). \quad (\text{B.23})$$

where $H(\cdot)$ is the Hessian matrix,⁵ and $\bar{\theta} \in (\hat{\theta}_0^*, \hat{\theta}^*)$. Using (B.23), the bootstrap likelihood ratio statistics LR_T^* in (3.26) becomes

$$\begin{aligned} LR_T^* &= 2 \left(l(\hat{\theta}^*, y^*) - l(\hat{\theta}_0^*, y^*) \right) \\ &= -(\hat{\theta}_0^* - \hat{\theta}^*)' H(\bar{\theta})(\hat{\theta}_0^* - \hat{\theta}^*). \end{aligned}$$

Consistency of $\hat{\theta}^*$ implies consistency of $\bar{\theta}$, and using information matrix inequality⁶ we get:

$$LR_T^* \stackrel{a}{=} T(\hat{\theta}_0^* - \hat{\theta}^*)' \mathcal{I}(\theta^*)(\hat{\theta}_0^* - \hat{\theta}^*). \quad (\text{B.24})$$

Let us now define the score vector $g(\theta, y)$ of first derivatives of $l(\theta, y)$ with respect to the elements of θ ,⁷ and the asymptotic score vector $s \equiv \text{plim} T^{-1/2} g(\theta^*, y)$. From a Taylor expansion of the likelihood equation $g(\hat{\theta}^*, y^*) = 0$ we obtain the following asymptotic equalities:

$$\begin{aligned} T^{1/2}(\hat{\theta}^* - \hat{\theta}) &\stackrel{a}{=} \mathcal{I}^{-1} T^{-1/2} g(\theta^*) \\ T^{1/2}(\hat{\theta}_1^* - \hat{\theta}_1) &\stackrel{a}{=} \mathcal{I}_{11}^{-1} T^{-1/2} g_1(\theta^*), \end{aligned}$$

which can be used to eliminate the estimators in (B.24) when we take the limit, obtaining an expression that involves only asymptotic information matrix and asymptotic score vector, as follows:

$$\begin{aligned} \text{plim} T^{1/2}(\hat{\theta}^* - \hat{\theta}_0^*) &= \text{plim} T^{1/2}(\hat{\theta}^* - \hat{\theta}) - \text{plim} T^{1/2}(\hat{\theta}_0^* - \hat{\theta}) \\ &= \mathcal{I}^{-1} s - \mathcal{I}_{11}^{-1} s_1 \\ &= \mathbf{J} s, \end{aligned} \quad (\text{B.25})$$

⁴See e.g. Davidson and MacKinnon (1999a), chapter 10.

⁵The $k \times k$ matrix of second derivatives of the log-likelihood with respect to θ .

⁶Let the asymptotic Hessian matrix be defined as $\mathcal{H}(\theta) \equiv \text{plim} \frac{1}{T} H(\theta)$. The information matrix equality, which assumes correct specification of the model, implies that $\mathcal{I}(\theta) = -\mathcal{H}(\theta)$.

⁷In the same way, $g_1(\theta, y)$ is the subvector of first derivatives of $l(\theta, y)$ with respect to the elements of θ_1

where s_1 is the subvector of s that corresponds to θ_1 , and

$$\mathbf{J} \equiv \mathcal{I}^{-1} - \begin{bmatrix} \mathcal{I}_{11}^{-1} & 0_{(k-r) \times r} \\ 0_{r \times (k-r)} & 0_{r \times r} \end{bmatrix}. \quad (\text{B.26})$$

Using (B.25), the probability limit of LR_T^* for $T \rightarrow \infty$ becomes:

$$\text{plim } LR_T^* = s' \mathbf{J} \mathcal{I} \mathbf{J} s. \quad (\text{B.27})$$

Moreover, from (B.26), we have that

$$\mathcal{I} \mathbf{J} = I_k - \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix} \begin{bmatrix} \mathcal{I}_{11}^{-1} & 0_{(k-r) \times r} \\ 0_{r \times (k-r)} & 0_{r \times r} \end{bmatrix} = \begin{bmatrix} 0_{(k-r) \times (k-r)} & 0_{(k-r) \times r} \\ -\mathcal{I}_{21} \mathcal{I}_{11}^{-1} & I_r \end{bmatrix} \equiv \mathbf{Q}, \quad (\text{B.28})$$

which implies $\mathcal{I}^{-1} \mathbf{Q} = \mathbf{J}$, $\mathbf{J} \mathbf{Q} = \mathbf{J}$ and $\mathbf{J} \mathcal{I} \mathbf{J} = \mathbf{J}$, from which we conclude that (B.27) can be written as

$$\text{plim } LR_T^* = s' \mathbf{J} s. \quad (\text{B.29})$$

Now, notice that s is asymptotically $N(0, \mathcal{I})$, thus $s = \mathcal{I}^{1/2} \tilde{s}$, where \tilde{s} is asymptotically standard normal. Therefore, (B.29) can be written as

$$\text{plim } LR_T^* = \tilde{s}' \mathcal{I}^{1/2} \mathbf{J} \mathcal{I}^{1/2} \tilde{s}, \quad (\text{B.30})$$

which is χ^2 distributed with degrees of freedom equal to the rank of matrix $\mathcal{I}^{1/2} \mathbf{J} \mathcal{I}^{1/2}$:

$$r(\mathcal{I}^{1/2} \mathbf{J} \mathcal{I}^{1/2}) = r(\mathcal{I}^{-1/2} \mathbf{Q} \mathcal{I}^{1/2}) = r,$$

using the fact that \mathcal{I} has full rank and that the rank of \mathbf{Q} is r since its first $k - r$ rows are zero. Therefore, we can conclude that $LR_T^* \xrightarrow{d} \chi_r^2$, as we wanted to show.

B.5 Bootstrap Distribution of out-of-sample R-squares

The distribution of the out-of-sample R-squares of returns and dividend growth under the null hypothesis H_0 is computed based on the following algorithm:

- 1) Using the estimated model parameters obtained using the first T years of data, under the null hypothesis H_0 , denoted $\hat{\theta}_{T,0}$, construct the (constrained) time series of standardized innovations $\{\hat{e}_{0t}\}_{t=1}^T$, and a bootstrap sample of observations, $\{Y_t^*\}_{t=1}^T$ as in steps 1)-3) in Section 3.3.3.
- 2) Using bootstrap sample $\{Y_t^*\}_{t=1}^T$, compute unconstrained maximum likelihood point estimates $\hat{\theta}_T^*$, by maximizing the log likelihood function $\log \mathcal{L}(\theta, \{Y_t^*\}_{t=1}^T)$, without imposing null hypothesis H_0 .
- 3) Based on estimated parameters $\hat{\theta}_T^*$ and filtered state using data until time T , compute the expected return and dividend growth for year $T + 1$, $\tilde{\mu}_T$ and \tilde{g}_T , respectively.
- 4) Repeat steps 1)-3) for $T = T_{in}, \dots, T_{max} - 1$, where T_{in} is the minimum length of the in-sample period and T_{max} is the length of the full sample of data.⁸

⁸We start our out-of-sample computations in 1985, which means that the first estimation is done using $T_{in} = 40$ years of data, and the length of the full sample in our case is $T_{max} = 65$ years.

5) The out-of-sample R^2 statistics for returns and dividend growth are computed as

$$R_{Ret,OS}^2 = 1 - \frac{\sum_{T=T_{in}}^{T_{max}-1} (r_{T+1} - \tilde{\mu}_T)^2}{\sum_{T=T_{in}}^{T_{max}-1} (r_{T+1} - \bar{r}_T)^2},$$

$$R_{Div,OS}^2 = 1 - \frac{\sum_{T=T_{in}}^{T_{max}-1} (\Delta d_{T+1} - \tilde{g}_T)^2}{\sum_{t=T_{in}}^{T_{max}-1} (\Delta d_{T+1} - \overline{\Delta d}_T)^2},$$

where \bar{r}_T and $\overline{\Delta d}_T$ are historical means of returns and dividend growth up until time T .

6) Repeat steps 1)-5) a large number of times, B , to obtain a collection of bootstrap values of the out-of-sample R^2 statistics. The empirical distribution of these values provides an approximation of the distribution of the $R_{Ret,O}^2$ and $R_{Div,OS}^2$ statistics under the null hypothesis H_0 .

This procedure borrows from [Rodriguez and Ruiz \(2009\)](#), who show how to compute nonparametric bootstrap prediction intervals in state space models, while taking into account the uncertainty linked to parameter estimation and not resorting to parametric assumptions for the shock distribution in the model.

Appendix C

C.1 Present-value model

C.1.1 Main notation

The state variables of the model are:

$$\begin{aligned}\hat{\mu}_t &= \mu_t - \delta_0, \\ \hat{g}_t &= g_t - \gamma_0, \\ \hat{\Sigma}_t &= vech(\Sigma_t - \mu^\Sigma),\end{aligned}$$

where μ^Σ is the solution of (4.8), which is such that

$$vech(\mu^\Sigma) = [I_3 - L_2(M \otimes M)D_2]^{-1}kL_2vec(V),$$

where I_2 is the identity matrix of dimension two, D_2 and L_2 are 2-dimensional duplication and elimination matrices, respectively, i.e for a symmetric 2×2 matrix A :

$$D_2vech(A) = vec(A), \quad L_2vec(A) = vech(A),$$

where vec denotes vectorization and $vech$ half-vectorization.

The dynamics of the state variables are obtained from (4.4)-(4.7) as follows:

$$\begin{aligned}\hat{g}_{t+1} &= \gamma_1\hat{g}_t + \varepsilon_{t+1}^g, \\ \hat{\mu}_{t+1} &= \delta_1\hat{\mu}_t + \varepsilon_{t+1}^\mu, \\ \hat{\Sigma}_{t+1} &= S\hat{\Sigma}_t + \varepsilon_{t+1}^\Sigma,\end{aligned}$$

where $S = L_2(M \otimes M)D_2$.

In terms of these demeaned states, the dynamics of realized returns and dividend growth in equation (4.3) is the following:

$$\begin{aligned}\Delta d_{t+1} &= \gamma_0 + \hat{g}_t + \tilde{\varepsilon}_{t+1}^D \\ r_{t+1} &= \delta_0 + \hat{\mu}_t + \tilde{\varepsilon}_{t+1}^r\end{aligned}$$

where

$$\tilde{\varepsilon}_{t+1}^D = e_1'\Sigma_t^{1/2} \begin{pmatrix} \varepsilon_{t+1}^D \\ \varepsilon_{t+1}^r \end{pmatrix},$$

and

$$\tilde{\varepsilon}_{t+1}^r = e_2'\Sigma_t^{1/2} \begin{pmatrix} \varepsilon_{t+1}^D \\ \varepsilon_{t+1}^r \end{pmatrix}.$$

Since ε_{t+1}^μ is a linear combination of the other shocks (see equation (4.20)), to complete the specification of the model we only need to specify the conditional covariance matrix of

$$\begin{pmatrix} \tilde{\varepsilon}_{t+1}^D \\ \tilde{\varepsilon}_{t+1}^r \\ \varepsilon_{t+1}^\Sigma \end{pmatrix},$$

which is given by:

$$Q_t = \begin{bmatrix} \Sigma_t & 0_{2 \times 3} \\ 0_{3 \times 2} & Var_t(\varepsilon_{t+1}^\Sigma) \end{bmatrix}, \quad (C.1)$$

where $Var_t(\varepsilon_{t+1}^\Sigma)$ is given by:

$$Var_t(\varepsilon_{t+1}^\Sigma) = L_2(I_4 + K_{2,2})[M\Sigma_t M' \otimes V + k(V \otimes V) + V \otimes M\Sigma_t M']L_2',$$

with $K_{2,2}$ being the commutation matrix of order two, i.e. the 4×4 matrix such that, for any 2×2 matrix A , $K_{2,2}vec(A) = vec(A')$.

C.1.2 Price-dividend ratio

In this section we present the detailed derivation of equation (4.10) in the text. From Campbell-Shiller approximation (4.9) we have

$$pd_t \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1}.$$

By iterating this equation we find:

$$\begin{aligned} pd_t &\simeq \kappa + \rho(\kappa + \rho pd_{t+2} + \Delta d_{t+2} - r_{t+2}) + \Delta d_{t+1} - r_{t+1} \\ &= \sum_{j=0}^{\infty} \rho^j \kappa + \rho^\infty pd_\infty + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \\ &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}), \end{aligned} \quad (C.2)$$

assuming that $\rho^\infty pd_\infty = \lim_{j \rightarrow \infty} \rho^j pd_{t+j} = 0$, at least in expectation. Then, we take expectation conditional to time t :

$$\begin{aligned} pd_t &\simeq \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[\Delta d_{t+j} - r_{t+j}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[g_{t+j-1} - \mu_{t+j-1}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j E_t[g_{t+j} - \mu_{t+j}]. \end{aligned} \quad (C.3)$$

Iterating the dynamics of $\hat{\mu}_{t+1}$ and \hat{g}_{t+1} and taking conditional expectation we find

$$E_t[\hat{\mu}_{t+j}] = \delta_1^j \hat{\mu}_t$$

and

$$E_t[\hat{g}_{t+j}] = \gamma_1^j \hat{g}_t.$$

Therefore,

$$\begin{aligned} pd_t &\simeq \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1^j \hat{g}_t - \delta_0 - \delta_1^j \hat{\mu}_t] \\ &= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \frac{\hat{g}_t}{1-\rho\gamma_1} - \frac{\hat{\mu}_t}{1-\rho\delta_1} \\ &= A + B_2 \hat{g}_t - B_1 \hat{\mu}_t. \end{aligned} \tag{C.4}$$

The explicit expressions for the present-value coefficients A , B_1 and B_2 are the following:

$$\begin{aligned} A &= \frac{\kappa + \gamma_0 - \delta_0}{1-\rho}, \\ B_1 &= \frac{1}{1-\rho\delta_1}, \\ B_2 &= \frac{1}{1-\rho\gamma_1}, \end{aligned}$$

C.1.3 Asymptotic bias in standard predictive regressions

We have shown in Section 4.3.2 that standard predictive regressions of either returns or dividend growth rates on the lagged log price-dividend ratio suffer from an error-in-variables (EIV) problem, which does not disappear as the sample size increases. Indeed, the true model for aggregate stock returns is:

$$r_{t+1} = \delta_0 + \hat{\mu}_t + \tilde{\varepsilon}_{t+1}^r,$$

but we wrongly assume the following model to hold:

$$r_{t+1} = a_r + b_r pd_t + \varepsilon_{t+1}^r, \tag{C.5}$$

where $pd_t = A - B_1 \hat{\mu}_t + B_2 \hat{g}_t$,¹ and we try to estimate the true parameter $b_r = -1/B_1$ from (C.5). The p-limit of the OLS slope coefficient is the following:²

$$\hat{b}_r \longrightarrow \frac{Cov(pd_t, r_{t+1})}{Var(pd_t)},$$

where

$$\begin{aligned} Cov(pd_t, r_{t+1}) &= Cov(A - B_1 \hat{\mu}_t + B_2 \hat{g}_t, \delta_0 + \hat{\mu}_t + \tilde{\varepsilon}_{t+1}^r) \\ &= -B_1 Var(\hat{\mu}_t) + B_2 Cov(\hat{g}_t, \hat{\mu}_t) \\ Var(pd_t) &= B_1^2 Var(\hat{\mu}_t) + B_2^2 Var(\hat{g}_t) - 2B_1 B_2 Cov(\hat{g}_t, \hat{\mu}_t) \end{aligned}$$

so that

$$\hat{b}_r \longrightarrow \frac{1}{-B_1 + \frac{B_2^2 Var(\hat{g}_t) - B_1 B_2 Cov(\hat{g}_t, \hat{\mu}_t)}{B_2 Cov(\hat{g}_t, \hat{\mu}_t) - B_1 Var(\hat{\mu}_t)}},$$

¹Remind that, as in Sections 4 and 5, we consider the case in which $B_3 = 0_{1 \times 3}$.

²Note that here we denote with \hat{b}_r the OLS estimate of the slope coefficient b_r in (C.5).

and the unconditional variances and covariance of demeaned expected return and dividend growth are the following:

$$\begin{aligned} Var(\hat{\mu}_t) &= (p-1)^2 \frac{(1 \quad -2 \quad 1)vech(\mu^\Sigma)}{\rho^2 B_1^2 (1 - \delta_1^2)}, \\ Var(\hat{g}_t) &= p^2 \frac{(1 \quad -2 \quad 1)vech(\mu^\Sigma)}{\rho^2 B_2^2 (1 - \gamma_1^2)}, \\ Cov(\hat{g}_t, \hat{\mu}_t) &= p(p-1) \frac{(1 \quad -2 \quad 1)vech(\mu^\Sigma)}{\rho^2 B_1 B_2 (1 - \delta_1)(1 - \gamma_1)}. \end{aligned}$$

Thus, the OLS slope coefficient in the regression of returns on lagged price-dividend ratio is biased and converges to a value that, at the estimated parameters, is larger than the true one in absolute value, resulting in more evidence for return predictability, but at the estimated parameters the bias is very small due to the relative persistence of expected dividend growth and returns.

The model for aggregate log dividend growth is:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^D,$$

while the wrong model is:

$$\Delta d_{t+1} = a_D + b_D p d_t + \varepsilon_{t+1}^D, \quad (\text{C.6})$$

and we try to estimate the true parameter $b_D = 1/B_2$ from (C.6). The p-limit of the OLS slope is the following:

$$\hat{b}_D \longrightarrow \frac{Cov(p d_t, \Delta d_{t+1})}{Var(p d_t)},$$

where

$$\begin{aligned} Cov(p d_t, \Delta d_{t+1}) &= Cov(A - B_1 \hat{\mu}_t + B_2 \hat{g}_t, \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^D) \\ &= B_2 Var(\hat{g}_t) - B_1 Cov(\hat{g}_t, \hat{\mu}_t) \end{aligned}$$

so that

$$\hat{b}_D \longrightarrow \frac{1}{B_2 + \frac{B_1^2 Var(\hat{\mu}_t) - B_1 B_2 Cov(\hat{g}_t, \hat{\mu}_t)}{B_2 Var(\hat{g}_t) - B_1 Cov(\hat{g}_t, \hat{\mu}_t)}},$$

Therefore, the OLS slope coefficient in the regression of dividend growth on lagged price-dividend ratio is also biased. This bias is negative and, at the estimated parameters, much more significant than the one for standard return regressions.

C.1.4 Term structure of conditional variances

The conditional variance of model-implied n -year returns and dividend growth in equations (4.47) and (4.48), are the following:

$$\begin{aligned} Var_t \left[\sum_{j=1}^n \rho^{j-1} r_{t+j} \right] &= \sum_{j=1}^{n-1} \rho^{2j} \left(\frac{1 - (\rho \delta_1)^{n-j}}{1 - \rho \delta_1} \right)^2 Var_t(\varepsilon_{t+j}^\mu) + \sum_{j=1}^n \rho^{2(j-1)} Var_t(\varepsilon_{t+j}^r) \\ &\quad + 2 \sum_{j=1}^{n-1} \rho^{2j-1} \frac{1 - (\rho \delta_1)^{n-j}}{1 - \rho \delta_1} Cov_t(\varepsilon_{t+j}^\mu, \varepsilon_{t+j}^r), \\ Var_t \left[\sum_{j=1}^n \rho^{j-1} \Delta d_{t+j} \right] &= \sum_{j=1}^{n-1} \rho^{2j} \left(\frac{1 - (\rho \gamma_1)^{n-j}}{1 - \rho \gamma_1} \right)^2 Var_t(\varepsilon_{t+j}^g) + \sum_{j=1}^n \rho^{2(j-1)} Var_t(\varepsilon_{t+j}^D) \\ &\quad + 2 \sum_{j=1}^{n-1} \rho^{2j-1} \frac{1 - (\rho \gamma_1)^{n-j}}{1 - \rho \gamma_1} Cov_t(\varepsilon_{t+j}^g, \varepsilon_{t+j}^D), \end{aligned}$$

where

$$\begin{aligned}
Var_t(\varepsilon_{t+j}^\mu) &= \frac{(p-1)^2}{\rho^2 B_1^2} (1 \quad -2 \quad 1) vech E_t(\Sigma_{t+j-1}), \\
Var_t(\tilde{\varepsilon}_{t+j}^r) &= (0 \quad 0 \quad 1) vech E_t(\Sigma_{t+j-1}), \\
Cov_t(\varepsilon_{t+j}^\mu, \tilde{\varepsilon}_{t+j}^r) &= \frac{p-1}{\rho B_1} (0 \quad 1 \quad -1) vech E_t(\Sigma_{t+j-1}), \\
Var_t(\varepsilon_{t+j}^g) &= \frac{p^2}{\rho^2 B_2^2} (1 \quad -2 \quad 1) vech E_t(\Sigma_{t+j-1}), \\
Var_t(\tilde{\varepsilon}_{t+j}^D) &= (1 \quad 0 \quad 0) vech E_t(\Sigma_{t+j-1}), \\
Cov_t(\varepsilon_{t+j}^g, \tilde{\varepsilon}_{t+j}^D) &= \frac{p}{\rho B_2} (-1 \quad 1 \quad 0) vech E_t(\Sigma_{t+j-1}),
\end{aligned}$$

and

$$\begin{aligned}
E_t(\Sigma_{t+j-1}) &= M^j \Sigma_t (M^j)' + kV(j), \\
V(j) &= V + MV M' + \dots + M^{j-1} V (M^{j-1})',
\end{aligned}$$

Note that non-contemporaneous correlation between return and expected return shocks are equal to zero and that the conditional variance of long-run returns is an affine functions of the variance-covariance state Σ_t .

C.2 Kalman Filter

In this section we describe the estimation procedure of the model in Section 4.1.

We first define an expanded 6-dimensional state vector by the concatenation of the original state variables and the process and observation noise random variables:

$$X_t = \begin{pmatrix} \hat{\mu}_{t-1} \\ \tilde{\varepsilon}_t^D \\ \tilde{\varepsilon}_t^r \\ \varepsilon_t^\Sigma \end{pmatrix},$$

which satisfies:

$$X_{t+1} = F X_t + \Gamma \varepsilon_{t+1}^X,$$

where

$$\varepsilon_{t+1}^X = \begin{pmatrix} \tilde{\varepsilon}_{t+1}^D \\ \tilde{\varepsilon}_{t+1}^r \\ \varepsilon_{t+1}^\Sigma \end{pmatrix},$$

with conditional variance Q_t given in (C.1). Moreover,

$$F = \begin{bmatrix} \delta_1 & -\frac{1}{\rho B_1}(p-1) & \frac{1}{\rho B_1}(p-1) & 0_{1 \times 3} \\ & & 0_{5 \times 6} & \end{bmatrix}, \quad \text{and} \quad \Gamma = \begin{bmatrix} 0_{1 \times 4} \\ I_5 \end{bmatrix},$$

The measurement equation,

$$Y_t = \begin{pmatrix} \Delta d_t \\ pd_t \\ vech(RV_t) \end{pmatrix},$$

is of the form

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t,$$

where

$$M_0 = \begin{bmatrix} \gamma_0 - \frac{A}{B_2} \\ (1 - \gamma_1)A \\ (I_3 - S)vech(\mu^\Sigma) \end{bmatrix}, \quad M_1 = \begin{bmatrix} 0 & \frac{1}{B_2} & 0_{1 \times 3} \\ 0 & \gamma_1 & 0_{1 \times 3} \\ 0_{3 \times 1} & 0_{3 \times 1} & S \end{bmatrix},$$

and

$$M_2 = \begin{bmatrix} \frac{B_1}{B_2} & 1 & 0 & 0_{1 \times 3} \\ -B_1(\delta_1 - \gamma_1) & -\frac{1}{\rho} & \frac{1}{\rho} & 0_{1 \times 3} \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & I_3 \end{bmatrix}.$$

The steps of the filter algorithm are the following:

- Initialize with the unconditional mean and covariance of the expanded state:

$$\begin{aligned} X_{0,0} &= 0_{6 \times 1}, \\ P_{0,0} &= E(X_t X_t'). \end{aligned}$$

- The time-update equations are

$$\begin{aligned} X_{t,t-1} &= F X_{t-1,t-1}, \\ P_{t,t-1} &= F P_{t-1,t-1} F' + \Gamma \tilde{Q}_t \Gamma'. \end{aligned}$$

- The prediction error η_t and the variance-covariance matrix of the measurement equations are then:

$$\begin{aligned} \eta_t &= Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t,t-1}, \\ S_t &= M_2 P_{t,t-1} M_2', \end{aligned}$$

where Y_t is the observed value of the measurement equation at time t .

- Update filtering:

$$\begin{aligned} K_t &= P_{t,t-1} M_2' S_t^{-1}, \\ X_{t,t} &= X_{t,t-1} + K_t \eta_t, \\ P_{t,t} &= (I - K_t M_2) P_{t,t-1}, \end{aligned}$$

where K_t is called *Kalman gain*.

To estimate model parameters, Θ , we define the log-likelihood for each time t , assuming normally distributed observation errors, as

$$l_t(\Theta) = -\frac{1}{2} \log |S_t| - \frac{1}{2} \eta_t' S_t^{-1} \eta_t,$$

where η_t and S_t denote prediction error of the measurement series and the covariance of the measurement series, respectively, obtained from the KF. Model parameters are chosen to maximize the log-likelihood of the data series:

$$\Theta \equiv \arg \max_{\Theta} \mathcal{L}(\Theta, \{Y_t\}_{t=1}^T),$$

with

$$\mathcal{L}(\Theta, \{Y_t\}_{t=1}^T) = \sum_{t=1}^T l_t(\Theta),$$

where T denotes the number of time periods in the sample of estimation.³

³For yearly data, as in our application, T is the number of years in the sample.

Bibliography

- Ait-Sahalia, Y., M. Karaman, and L. Mancini (2012). The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis. Working Paper. Princeton University and Swiss Finance Institute.
- Amihud, Y. and C. Hurvich (2004). Predictive Regression: a Reduced Bias Estimation Method. *Journal of Financial and Quantitative Analysis* 39, 813–841.
- Amihud, Y., C. Hurvich, and Y. Whang (2009). Multiple-Predictor Regressions: Hypothesis Testing. *Review of Financial Studies* 22, 413–434.
- Anderson, E. W., E. Ghysels, and J. L. Juergens (2009). The Impact of Risk and Uncertainty on Expected Returns. *Journal of Financial Economics* 94, 233–263.
- Andrews, D. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* 70(6), 119–162.
- Ang, A. and G. Bekaert (2007). Stock Return Predictability: Is it There? *Review of Financial Studies* 20(3), 651–708.
- Ang, A. and J. Liu (2007). Risk, return and dividends. *Journal of Financial Economics* 85(1), 1–38.
- Baker, M. and J. Wurgler (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance* 61, 1645–1680.
- Bali, G. and H. Zhou (2011). Risk, Uncertainty, and Expected Returns. Finance and Economics Discussion Series, Federal Reserve Board, Washington, D. C.
- Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59(4), 1481–1509.
- Barro, R. J. (2006). Rare Disasters and Asset Markets in the Twentieth Century. *Quarterly Journal of Economics* 121, 823–66.
- Basak, S. and D. Cuoco (1998). An Equilibrium Model With Restricted Stock Market Participation. *Review of Financial Studies* 11, 309–341.
- Beber, A., F. Breedon, and A. Buraschi (2010). Differences in beliefs and currency risk premiums. *Journal of Financial Economics* 98, 415–438.
- Bekaert, G. and E. Engstrom (2010). Asset Return Dynamics under Bad Environment-Good Environment Fundamentals. Working Paper, Columbia University and NBER.
- Bekaert, G. and M. Hoerova (2014). The VIX, the Variance Premium and Stock Market Volatility. *Journal of Econometrics* forthcoming.

- Benzoni, L., P. Collin-Dufresne, R. S. Goldstein, and J. Helwege (2012). Modeling Credit Contagion via the Updating of Fragile Beliefs. Working Paper, Federal Reserve Bank of Chicago.
- Beran, R. (1988). Prepivoting Test Statistics. *Journal of the American Statistical Association* 83, 687–697.
- Berkman, H., B. Jacobsen, and J. B. Lee (2011). Time-Varying Rare Disaster Risk and Stock Returns. *Journal of Financial Economics* 101, 313–332.
- Berrada, T. (2009). Bounded Rationality and Asset Pricing with Intermediate Consumption. *Review of Finance* 13(4), 693–725.
- Binsbergen, J., M. Brandt, and R. S. Koijen (2012). On the timing and Pricing of Dividends. *American Economic Review* 102(4), 1596–1618.
- Binsbergen, J., W. H. Hueskes, R. S. Koijen, and E. B. Vrugt (2013). Equity Yields. *Journal of Financial Economics* 110(3), 503–519.
- Binsbergen, J. and R. S. Koijen (2009). Likelihood-based estimation of exactly-solved present-value models. Working Paper. University of Chicago, Booth School of Business.
- Binsbergen, J. and R. S. Koijen (2010). Predictive Regressions: A Present-Value Approach. *Journal of Finance* 65(4), 1439–1471.
- Bollerslev, T., G. Tauchen, and H. Zhou (2009). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22(11), 4463–4491.
- Bollerslev, T. and V. Todorov (2011). Tails, Fears, and Risk Premia. *Journal of Finance* 66(6), 2165–2211.
- Borovicka, J. (2012). Survival and long-run dynamics with heterogeneous beliefs under recursive preferences. Working Paper. New York University.
- Bowman, A. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis : The Kernel Approach with S-Plus Illustrations*. Oxford University Press.
- Brandt, M. and L. Wang (2010). Measuring the Time-Varying Risk-Return Relation from the Cross-Section of Equity Returns. Working Paper. Fuqua School of Business.
- Brennan, M. J. and A. P. Taylor (2010). Predicting the Market using Information from Equity Portfolio Returns. Working Paper. UCLA and Manchester University.
- Buraschi, A. and A. Jiltsov (2006). Model Uncertainty and Option Markets with Heterogeneous Agents. *Journal of Finance* 61, 2841–2897.
- Buraschi, A., F. Trojani, and A. Vedolin (2013). Economic Uncertainty, Disagreement, and Credit Markets. *Management Science* forthcoming.
- Buraschi, A., F. Trojani, and A. Vedolin (2014). When Uncertainty Blows in the Orchard: Comovement and Equilibrium Volatility Risk Premia. *Journal of Finance* 69(1), 101–137.
- Buraschi, A. and P. Whelan (2011). Macroeconomic Uncertainty, Difference in Beliefs, and Bond Risk Premia. Working paper. Imperial College London.
- Campbell, J. Y. (1991). A variance decomposition for stock returns. *Economic Journal* 101, 157–179.
- Campbell, J. Y. (2003). Consumption-Based Asset Pricing. *Handbook of Economics and Finance*. G. Constantinides, M. Harris, and R. Stulz (eds.), Amsterdam: North Holland.
- Campbell, J. Y. and J. H. Cochrane (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107, 205–251.

- Campbell, J. Y., A. W. Lo, and C. A. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Campbell, J. Y. and R. J. Shiller (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies* 1(3), 195–228.
- Campbell, J. Y. and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21(4), 1509–1531.
- Campbell, J. Y. and L. M. Viceira (2005). The Term Structure of the Risk-Return Tradeoff. *Financial Analysts Journal* 61(1), 34–44.
- Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* 81(1), 27–60.
- Camponovo, L., O. Scaillet, and F. Trojani (2009). Robust resampling methods for time series. Swiss Finance Institute Research Paper No. 09-38.
- Carr, P. and L. Wu (2009). Variance risk premia. *Review of Financial Studies* 22(3), 1311–1341.
- Chabakauri, G. (2013). Dynamic Equilibrium with Two Stocks, Heterogeneous Investors, and Portfolio Constraints. *Review of Financial Studies* 26, 3104–3141.
- Charoenrook, A. (2002). Change in Consumer Sentiment Index and Aggregate Stock Returns. Working paper. Vanderbilt University.
- Chen, H., S. Joslin, and N. Tran (2010). Affine Disagreement and Asset Pricing. *American Economic Review, Papers, and Proceedings* 100, 52226.
- Chen, H., S. Joslin, and N. Tran (2012). Rare Disasters and Risk Sharing with Heterogeneous Beliefs. *Review of Financial Studies* 25(7), 2189–2224.
- Chen, L. (2009). On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics* 92(1), 128–151.
- Cochrane, J. H. (1992). Explaining the variance of price-dividend ratios. *Review of Financial Studies* 5(2), 243–280.
- Cochrane, J. H. (2008a). State-space vs. var models for stock returns. Working Paper. University of Chicago.
- Cochrane, J. H. (2008b). The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies* 21(4), 1533–1576.
- Cochrane, J. H., F. A. Longstaff, and P. Santa-Clara (2008). Two Trees. *Review of Financial Studies* 21, 347–385.
- Corsi, F. (2009). A Simple Approximate Long Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7, 174–196.
- Cox, J. and C. F. Huang (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory* 49, 33–83.
- Crain, M. A. (2011). A Literature Review of the Size Effect. Working paper, Manchester Business School and Florida Atlantic University.
- Cuoco, D. and H. He (1994). Dynamic Equilibrium in Infinite-Dimensional Economies with Incomplete Financial Markets. Working paper, University of Pennsylvania.
- Davidson, R. and J. MacKinnon (1999a). *Econometric Theory and Methods*. Oxford University Press.

- Davidson, R. and J. MacKinnon (1999b). The Size Distortion of Bootstrap Tests. *Econometric Theory* 15, 361–376.
- Deb, P. and M. Sefton (1996). The distribution of a lagrange multiplier test of normality. *Economics Letters* 51, 123–130.
- Dieckmann, S. (2011). Rare Event Risk and Heterogeneous Beliefs: The Case of Incomplete Markets. *Journal of Financial and Quantitative Analysis* 46, 459–488.
- Dieckmann, S. and M. Gallmeyer (2005). The Equilibrium Allocation of Diffusive and Jump Risks with Heterogeneous Agents. *Journal of Economic Dynamics and Control* 29, 1547–76.
- Diris, B. (2011). Model Uncertainty for Long-Term Investors. Working Paper. Erasmus University Rotterdam and Netspar.
- Drechsler, I. (2013). Uncertainty, Time-Varying Fear, and Asset Prices. *Journal of Finance* 68(5), 1843–1889.
- Drechsler, I. and A. Yaron (2011). What’s Vol Got to Do With It. *Review of Financial Studies* 24(1), 1–45.
- Drissen, J., P. Maenhout, and G. Vilkov (2012). Option-Implied Correlations and the Price of Correlation Risk. Working paper. Tilburg University, INSEAD and Goethe University Frankfurt.
- Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asst Pricing for Affine Jump-Diffusions. *Econometrica* 68(6), 1343–1376.
- Dumas, B., A. Kurshev, and R. Uppal (2009). Equilibrium Portfolio Strategies in the Presence of Sentiment Risk and Excess Volatility. *Journal of Finance* 64(2), 579–629.
- Edelen, R. M., A. J. Marcus, and H. Tehranian (2010). Relative Sentiment and Stock Returns. *Financial Analysts Journal* 66(4), 1–13.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7, 1–26.
- Engsted, T., T. Q. Pedersen, and C. Tanggaard (2010). The Log-Linear Return Approximation, Bubbles, and Predictability. CREATES Research Papers 2010-37, School of Economics and Management, University of Aarhus.
- Fama, E. F. and K. R. French (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 22(3), 3–25.
- Favero, C. A., A. E. Gozluklu, and A. Tamoni (2011). Demographic Trends, the Dividend-Price Ratio, and the Predictability of Long-Run Stock Market Returns. *Journal of Financial and Quantitative Analysis* 46(5), 1493–1520.
- Ferreira, M. A. and P. Santa-Clara (2010). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, forthcoming.
- Fisher, K. and M. Statman (2003). Consumer Confidence and Stock Returns. *Journal of Portfolio Management* 30(1), 115–127.
- French, A., G. W. Schwert, and R. F. Stambaugh (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics* 19(1), 3–29.
- Fusari, N. and M. T. Gonzalez-Perez (2012). Volatility Dynamics and the Term Structure of the Variance Risk Premium. Working paper. Kellogg School of Management and Colegio Universitario de Estudios Financieros.

- Gabaix, X. (2012). Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance. *Quarterly Journal of Economics* 127(2), 645–700.
- Garleanu, N. and S. Panageas (2014). Young, Old, Conservative, and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing. *Journal of Political Economy* Forthcoming.
- Golez, B. (2011). Expected Returns and Dividend Growth Rates Implied in Derivative Markets. Universitat Pompeu Fabra.
- Gourieroux, C. (2006). Continuous time wishart process for stochastic risk. *Econometric Reviews* 25, 177–217.
- Gourieroux, C., J. Jasiak, and R. Sufana (2009). The Wishart Autoregressive Process of Multivariate Stochastic Volatility. *Journal of Econometrics* 150, 167–181.
- Goyal, A. and I. Welch (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21(4), 1455–1508.
- Hall, P. and J. Horowitz (1996). Bootstrap critical values for tests based on Generalized-Method-of-Moment estimators. *Econometrica* 64(4), 891–916.
- Han, B. and Y. Zhou (2011). Variance Risk Premium and Cross-Section of Stock Returns. Working Paper, University of Texas at Austin and University of Oklahoma.
- Hansen, P. R. and A. Lunde (2006). Realized Variance and Market Microstructure Noise. *Journal of Business and Economic Statistics* 24, 127–161.
- Henkel, S., J. Martin, and F. Nardari (2011). Time-Varying Short-Horizon Predictability. *Journal of Financial Economics* 99, 560–580.
- Hong, H. and D. Sraer (2012). Speculative Betas. Working Paper, Princeton University.
- Jagannathan, R. and Z. Wang (1996). The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance* 51(1), 3–53.
- Jarque, C. M. and A. K. Bera (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review* 55(2), 163–172.
- Jiang, G. J. and Y. S. Tian (2007). Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index. *Journal of Derivatives* 14(1), 1–26.
- Jin, J. (2013). Jump-Diffusion Long-Run Risks Models, Variance Risk Premium and Volatility Dynamics. Bank of Canada Working Paper 2013-12.
- Johannes, M., A. Korteweg, and N. Polson (2011). Sequential Learning, Predictive Regressions, and Optimal Portfolio Returns. Columbia University, Stanford University and University of Chicago.
- Keim, D. B. and R. F. Stambaugh (1986). Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics* 17(2), 357–390.
- Kelly, B. and S. Pruitt (2013). Market Expectations in the Cross Section of Present Values. *Journal of Finance* 68(5), 1729–1756.
- Kim, S. (2013). Asset Prices in Turbulent Markets with Rare Disasters. Working Paper. Kellogg School of Management.
- Kogan, L., S. A. Ross, J. Wang, and M. M. Westerfield (2006). The Price Impact and Survival of Irrational Traders. *Journal of Finance* 61(1), 195–229.
- Koijen, R. S. and S. Van Nieuwerburgh (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, forthcoming.

- Kothari, S. P. and J. Shanken (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time Series Analysis. *Journal of Financial Economics* 44(2), 169–203.
- Koulovatianos, C. and V. Wieland (2011). Asset Pricing under Rational Learning about Rare Disasters. Working paper. University of Nottingham and Goethe University of Frankfurt.
- Lemmon, M. and E. Portniaguina (2006). Consumer Confidence and Asset Prices: Some Empirical Evidence. *Review of Financial Studies* 19(4), 1499–1529.
- Lettau, M. and S. C. Ludvigson (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance* 3, 815–849.
- Lettau, M. and S. C. Ludvigson (2005). Expected Returns and Expected Dividend Growth. *Journal of Financial Economics* 76(3), 583–626.
- Lettau, M. and S. C. Ludvigson (2010). *Handbook of Financial Econometrics*, Volume 1, Chapter Measuring and Modeling Variation in the Risk-Return Trade-off, pp. 617–690. North-Holland, Amsterdam.
- Lettau, M. and S. Van Nieuwerburgh (2008). Reconciling the Return Predictability Evidence. *Review of Financial Studies* 21(4), 1607–1652.
- Lettau, M. and J. A. Wachter (2007). Why is long-horizon equity less risky? a duration-based explanation of the value premium. *Journal of Finance* 62(1), 55–92.
- Lewellen, J. (2004). Predicting Returns with Financial Ratios. *Journal of Financial Economics* 74(2), 209–235.
- Liung, L. and P. Caines (1979). Asymptotic Normality of Prediction Error Estimators for Approximate System Models. *Stochastics* 3, 29–46.
- Londono, J. M. (2011). The Variance Risk Premium Around the World. Board of Governors of the Federal Reserve System, IFDP Working Paper.
- Lucas, R. (1978). Asset Prices in an Exchange Economy. *Econometrica* 46, 1429–1445.
- Ludvigson, S. C. (2004). Consumer Confidence and Consumer Spending. *Journal of Economic Perspectives* 18, 29–50.
- Martin, I. (2013). The Lucas Orchard. *Econometrica* 81(1), 55–111.
- Menzly, L., T. Santos, and P. Veronesi (2004). Understanding predictability. *Journal of Political Economy* 112(1), 1–47.
- Mueller, P., A. Vedolin, and Y. Yen (2011). Bond Variance Risk Premia. Working Paper. London School of Economics.
- Pastor, L., M. Sinha, and B. Swaminathan (2008). Estimating the intertemporal risk-return tradeoff using the implied cost of capital. *Journal of Finance* 63(6), 2859–2897.
- Pastor, L. and R. Stambaugh (2012). Are Stocks Really Less Volatile in the Long Run? *Journal of Finance* 67(2), 431–478.
- Patton, A. J. and A. Timmermann (2010). Why do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion. *Journal of Monetary Economics* 57, 803–820.
- Piatti, I. and F. Trojani (2012a). Dividend growth predictability and the price-dividend ratio. *SSRN eLibrary*.
- Piatti, I. and F. Trojani (2012b). Predictable Risks and Predictive Regression in Present-Value Models. *SSRN eLibrary*.

- Piazzesi, M. (2010). *Handbook of Financial Econometrics*, Volume 1, Chapter Affine Term Structure Models, pp. 691–766. North-Holland, Amsterdam.
- Politis, D. N. and J. P. Romano (1992). *Exploring the Limits of Bootstrap*, Chapter A circular block-resampling procedure for stationary data, pp. 263–270. Wiley, New York.
- Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Polk, C., S. Thompson, and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics* 81, 101–141.
- Rietz, T. A. (1988). The Equity Risk Premium: A Solution. *Journal of Monetary Economics* 22, 117–131.
- Rodriguez, A. and E. Ruiz (2009). Bootstrap Prediction Intervals in State-space Models. *Journal of Time Series Analysis* 30(2), 167–178.
- Rozeff, M. S. (1984). Dividend Yields are Equity Risk Premiums. *Journal of Portfolio Management* 11(1), 68–75.
- Rytchkov, O. (2012). Filtering Out Expected Dividends and Expected Returns. *Quarterly Journal of Finance* 2(3), 1250–12.
- Schiller, R. (1984). Stock Prices and Social Dynamics. *Brookings Papers on Economic Activity* 2, 457–498.
- Siegel, J. J. (2008). *Stocks for the Long Run* (4th ed.). McGraw Hill.
- Spall, J. and K. Wall (1984). Asymptotic Distribution Theory for the Kalman Filter State Estimator. *Communications in Statistics-Theory and Methods* 13, 1981–2003.
- Stambaugh, R. (1999). Predictive Regressions. *Journal of Financial Economics* 54(3), 375–421.
- Stoffer, D. and K. Wall (1991). Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter. *Journal of the American Statistical Association* 86, 1024–1033.
- Torous, W., R. Valkanov, and S. Yan (2004). On Predicting Stock Returns with Nearly Integrated Explanatory Variables. *Journal of Business* 77, 937–966.
- Tsai, J. and J. A. Wachter (2013). Rare booms and disasters in a multi-sector endowment economy. Working paper. University of Pennsylvania.
- Wachter, J. A. (2013). Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility? *Journal of Finance* 68, 987–1035.
- Wu, L. (2012). What Does Variance Risk Premium Tell Us? The Fourth Annual Volatility Institute Conference.
- Yan, H. (2008). Natural Selection in Financial Markets: Does It Work? *Management Science* 54(11), 1935–1950.
- Yu, J. and Y. Yuan (2011). Investor Sentiment and the Mean-Variance Relation. *Journal of Financial Economics* 100, 367–381.
- Yun, J. (2012). Evaluating Predictors within a Present-Value Framework. Working Paper. University of Chicago Booth School of Business.
- Zhou, G. and Y. Zhu (2010). A Long-Run Risk Model with Long- and Short-Run Volatilities: Explaining Predictability and Volatility Risk Premium. Working Paper. Washington University.

Zhou, H. (2010). Variance Risk Prema, Asset Predictability Puzzles, and Macroeconomic Uncertainty. Working Paper. Federal Reserve Board.