Three Essays on Economics of Predictability of Stock Returns

by

Tymur Gabuniya

Submitted in Partial Fulfillment

of the

Requirements for the Degree of

Doctor of Philosophy

Supervised by

Professor Pedro Santa Clara

NOVA School of Business and Economics

Universidade Nova de Lisboa

Lisbon

2012

Curriculum Vitae

The author studied economics of entrepreneurship and finance at Kharkiv State University of Economics (Kharkiv, Ukraine) from 1997 to 2002, graduating with a diploma of a higher education in 2002. In 2002, he entered into a research program in Regional Management at Kharkiv National Karazin's University (Kharkiv, Ukraine). He was also working at different positions at PrivatBank (Kharkiv, Ukraine) from 2003 to 2004. In 2004 he received financial support in order to continue his graduate education at CERGE-EI (Prague, Czech Republic) from where he obtained a Master of Arts in Economics in 2006. He joined NOVA School of Business and Economics (Lisbon, Portugal) in 2006. He received a research grant from the Science and Technology Foundation of Portugal from 2006 to 2010. In 2008-2009 he was a visiting student at Toulouse School of Economics (Toulouse, France). He works in the area of financial econometrics under the supervision of Professor Pedro Santa-Clara from 2009.

Acknowledgements

I gratefully acknowledge the advice, encouragement and guidance of Pedro Santa Clara. I owe a special debt to Iliyan Georgiev for his advice and useful comments. I thank the Foundation for the Science and Technology and NOVA School of Business and Economics for financial support.

ABSTRACT

In the first essay of this dissertation I analyze predictability of returns generated by the long-run risks model of Bansal and Yaron (2004). I uncover some counterfactual features of the predictability and connect them with the specific features of the long-run risks processes. In the second essay, I analyze the effect of the aggregation on the predictability in the long-run risks model. I found that the aggregation implies that a part of expected dividend growth is observable and points at the nature of the additional to the dividend-price ratio variables which might help to predict returns. In the third essay, I use expected returns and expected dividend growth processes implied by the long-run risks and other models to analyze the out-of-sample performance of the predictive regression and some of its alternatives. My analysis suggests that the poor out-of-sample performance is due to the finite sample noise and a large unpredictable component in returns.

To my grandparents

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

After almost two decades of research the question of whether it is possible to predict stock returns is still open. On the one hand, the literature offers a wide list of the variables and methods for predicting stock returns (Fama and French (1988a); Fama and French (1989); Lettau and Ludvigson (2001); Lewellen (2004); Campbell and Yogo (2006); Rapach, Strauss and Zhou (2010)). On the other, the validity of the statistical tests for predictability is questioned. The problem is that many predictors are highly persistent and their innovations are contemporaneously correlated with returns. This leads to a finite sample bias in the coefficients and to the incorrect statistical inference (Nelson and Kim (1993), Stambaugh (1999)). Additional problem is data mining (Ferson, Sarkissian and Simin (2003)).

Recently more attention is paid to the out-of-sample predictability which is more relevant for the real-time forecasting of stock returns. Goyal and Welch (2008) find that popular predictors of returns perform poorly out-of-sample. There is a debate on the interpretation of this result. Goyal and Welch (2008) argue that the poor out-of-sample performance reflects instability of predictive relations, while Inoue and Kilian (2004), Campbell and Thompson (2008) and Cochrane (2008b) claim that the poor out-of-sample performance is related to the finite samples, high persistence of some of the predictors or to the low power of the out-of-sample tests.

A different branch of the literature addresses the question of what are the economic forces responsible for predictability of returns. Several asset pricing models capable of generating variation in expected returns were proposed. One of them–the long-run risks model by Bansal and Yaron (2004) emphases small and highly persistent components of consumption growth as driving variability of expected returns. Even though, this model is one of the most successful ones in explaining various asset-pricing puzzles, some studies have revealed some inconsistencies of the model with the financial data. Beeler and Campbell (2009) find counterfactual predictability patterns produced by the model, Constantinides and Ghosh (2009) reject the model while testing it against a cross-section of stock returns. Consequently, Bansal, Kiku and Yaron (2007a) and Bansal, Kiku and Yaron (2009) respond that the long-run risk model is a valid description of cross-sectional and time-series behavior of expected returns.

This work attempts to shed light on both, the correct interpretation of the poor out-of-sample performance of the conventional predictive regression–the one where the dividend-price ratio is used to predict returns and on the economic forces behind the predictability. The paper consists of three essays. In the first essay, I analyze the ability of the Bansal and Yaron (2004) model to generate predictability of returns consistently with the data. I add more evidence against the model by finding counterfactual patterns of the predictability of returns and riskless interest rate it produces. Furthermore, I connected specific features of the predictability generated by the model to the specific features of the long-run risks processes.

In the Bansal and Yaron (2004) model the decision interval of the representative agent is monthly while the calibration is done to match the annual data. Several existing papers stress the importance of the aggregation issue for the correct interpretation of the calibration parameters and for testing asset pricing models, see for example Hansen and Sargent (1983), Heaton (1995), Bansal, Kiku and Yaron (2007a). In the second essay I address the effect of the aggregation on the processes for expected returns and expected dividend growth and, consequently, on predictability of returns in the Bansal and Yaron (2004) model.

In the third essay, I study the out-of-sample forecasting performance of the conventional predictive regression and some alternatives to this method proposed in the literature. In order to do that, I set up a Monte Carlo simulation exercise based on different models for expected returns and expected dividend growth. The aim is to understand the source of the poor out-of-sample performance of the conventional predictive regression, compare it with the alternative methods and to analyze the effect of the differences in the existing models for expected returns and expected dividend growth on their performance.

2. **PREDICTABILITY IN THE LONG RUN RISKS MODEL**

Introduction

In their celebrated work Mehra and Prescott (1985) stress the difficulty faced by equilibrium asset pricing models to justify the magnitude of the equity premium. With a time-separable constant relative risk aversion (CRRA) utility function, an excessively high relative risk aversion (RRA) is needed to generate the 6% equity premium observed in the data. CRRA utility does not allow separation between the RRA and the elasticity of intertemporal substitution (EIS). As a result, a high level of RRA leads to a counterfactually high level of the riskless interest rate provided a reasonable bound on the subjective discount factor holds–the riskfree rate puzzle.

Bansal and Yaron (2004) (BY) offer a model capable of solving both the equity premium and riskfree rate puzzles. Their framework employs Epstein and Zin (1989) preferences which allow separation between RRA and EIS. Consequently, a high level of RRA needed to fit the equity premium does not lead to the risk free rate puzzle. The level of RRA needed to generate a sizable equity premium is further decreased to economically justifiable levels by the presence of two additional sources of risk: an economic growth risk which is caused by shocks to expected log-consumption growth and an economic uncertainty risk which stems from an unexpected change in the volatility of log-consumption growth.

The BY model is consistent with many asset pricing facts. It generates the observed level of the equity premium and produces time variation in its volatility. It successfully matches the means and standard deviations of market returns, dividend growth and the interest rate. In addition, BY claim that the long run risk model is capable of reproducing some of the predictability patterns observed in the data.

Some recent papers find empirical problems with the BY model. Using GMM, its validity for describing the cross-section of stock returns is questioned by Constantinides and Ghosh (2009). The persistence of consumption and dividend growth processes, some of the predictability implications as well as other dimensions of the model are challenged by Beeler and Campbell (2009). Moreover, Garcia, Meddahi and Tedongap (2010) set the BY model in a Markov switching context and claim that predictability of returns in the BY model is a finite sample phenomenon. Bansal, Kiku and Yaron (2007a), Bansal, Kiku and Yaron (2007b) and Bansal, Kiku and Yaron (2009) respond by providing more evidence in favor of high persistence of expected consumption growth, high persistence of the time-varying volatility and cross-sectional and predictability implications generated by the long-run risks model.

This paper contributes to this literature by focusing on predictability generated by the BY model. Two main elements distinguish my analysis from the others. First, I analyze predictability of returns and predictability of the riskless interest rate separately. Although, existing evidence shows similar predictability patterns in returns and excess returns over the riskless interest rate, implying a similar dynamics for expected returns and expected excess returns, I show that this does not hold in the BY model. Furthermore, in the model dividend-price ratio predicts both returns and dividend growth negatively and predictability of returns is excessively weak–the reverse to what is observed in the data. The negative sign of the slope coefficient from the predictive regression signals failure of the model to produce the discount rate effect while the stronger evidence in favor of predictability of dividend growth rather than returns indicates that most part of the variation in the dividend-price ratio stems from expected dividend growth. Additionally, I find that the BY model produces a highly predictable by the dividend-price ratio riskless interest rate which is inconsistent with the data. Lettau and Van Nieuwerburgh (2007) and Koijen and Van Nieuwerburgh (2010) provide evidence in favor of structural breaks in the mean of the dividend-price ratio. Since the model does not include this channel, it makes sense to test the model against the data adjusted for the structural breaks following the reasoning of Constantinides and Ghosh (2009). Using the dividend-price ratio adjusted for the structural breaks the model is rejected even strongly.

Second, my method is different from the methods used in the other papers in that I follow the logic behind the State Space Representation (SSR) for returns proposed by Cochrane (2008a). The SSR is useful, since it allows interpreting the results of any predictive VAR in terms of the variation in expected returns and the variation in expected dividend growth. Or, in other words, it provides a connection between the observed predictability patterns with the unobserved factors influencing expected returns and expected dividend growth.

The SSR obtained on the basis of the equations describing the model shows that expected returns and expected dividend growth are positively correlated and highly persistent. Moreover, the innovations to expected dividend growth are much more volatile than the innovations to expected returns. Therefore, it is expected dividend growth which accounts for the biggest part of the variation of the dividend-price ratio. Connecting expected returns and expected dividend growth with the processes for the long run risks, I find that the model fails to produce the predictability observed in the data since it attributes a large portion of the variation in both, prices and cash flows, to expected consumption growth.

Indeed, deriving the predictive relations from the SSR and substituting the parameters of the calibration of the model, I find that the performance of the model can be improved by decreasing the weight of expected consumption growth and increasing the weight of the time-varying volatility in the variation of prices. A further analysis, however, reveals the limitations of the model to generate the predictability patterns observed in the data. Those limitations are caused by the specification of the processes for the long run risks.

The paper proceeds as follows. Section 1 outlines the long run risks model. Section 2 tests it against the real data using a simulation exercise. Section 3 uses the SSR to explain the counterfactual predictability patterns and to indicate directions of a change in the calibration of the model to reproduce the predictability patterns from the data. Section 4 recalibrates the model. Section 5 concludes.

2.1 The long run risks model

Denoting Δc_{t+1} and Δd_{t+1} as log-consumption and log-dividend growth, BY specify the following processes:

$$\Delta c_{t+1} = m + x_t + \sigma_t \eta_{t+1}, \tag{2.1.1}$$

$$\Delta d_{t+1} = m_d + \phi x_t + \varphi_d \sigma_t u_{t+1}, \tag{2.1.2}$$

$$x_{t+1} = \rho x_t + \varphi_e \sigma_t e_{t+1}, \tag{2.1.3}$$

$$\sigma_{t+1}^2 = \bar{\sigma}^2(1-\nu_1) + \nu_1 \sigma_t^2 + \sigma_\omega \omega_{t+1}, \tag{2.1.4}$$

with all the shocks being mutually independent *i.i.d* Standard Normal. The values of the parameters assumed in BY are provided in Table 2.1.

Tab. 2.1: Calibration

This table presents the values of the parameters of the calibration of the processes describing the long run risks model. Panel A provides the calibration of the conditional means of the processes. Panel B, the calibration of the conditional variances.

Panel A: Conditional means						
Parameter:	m	m_d	$\bar{\sigma}$	ρ	ν_1	ϕ
Value:	0.0015	0.0015	0.0078	0.9790	0.9870	3.0000
Panel B: Conditional variances						
Parameter:	φ_e	φ_d	σ_ω			
Value:	0.0440	4.5000	2.3000×10^{-6}			

The conditional means and the conditional variances of dividend growth and consumption growth are time-varying. The dynamics of their moments is determined by highly persistent processes for x_t and σ_t^2, which are the sources of the long run risks in the BY model. From Eq.(2.1.1), x_t equals to the conditional mean and σ_t^2, to the conditional variance of consumption growth. In other words, x_t is a source of economic growth risk and σ_t^2 is a source of economic uncertainty risk. Table 2.1 shows that the model places a much larger weight on the variance of expected consumption growth than on the variance of the time-varying volatility. The innovations to the conditional mean are many times more volatile than the innovations to the conditional variance. The above numbers imply that the variance of the innovations to the conditional mean exceed the variance of innovations to the time-varying volatility by a factor of 20,000. High persistence of both processes then leads to the variance of expected consumption growth exceeding the variance of the time-varying volatility by a factor of 14000. Note from Table 2.1 that while dividend growth and consumption growth have equal unconditional means, dividend growth is more volatile than the consumption growth. Moreover, it has a leveraged exposure to expected consumption growth. The leverage is governed by the parameter ϕ.

BY use Epstein and Zin (1989) preferences that allow separation of the relative risk aversion (γ) and the intertemporal substitution (ψ). Denote by $r_{c,t+1}$ log-returns on the claim to the aggregate consumption at time $t+1$, by $r_{m,t+1}$ log-returns on the claim to the aggregate dividends (market returns) at time $t+1$ and β as the time discount factor. Epstein and Zin (1989) preferences lead to the following Euler equation for market returns:

$$E_t\left[exp\left(\theta log\beta - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta-1)r_{c,t+1} + r_{m,t+1}\right)\right] = 1, \tag{2.1.5}$$

with $\theta = \dfrac{1-\gamma}{1-\frac{1}{\psi}}$. BY assume that $\beta = 0.998$, $\gamma = 10$ and $\psi = 1.5$.

To obtain closed form solutions BY employ the Campbell and Shiller (1988) approximation, which in the case of returns on the claim to aggregate consumption is:

$$r_{c,t+1} = k_0 - k_1 cp_{t+1} + cp_t + \Delta c_{t+1}, \tag{2.1.6}$$

with cp_t being the consumption-price ratio at time t. The linearization constants: $k_1 = \dfrac{e^{-\bar{cp}}}{1+e^{-\bar{cp}}}$ and $k_0 = -(1-k_1)log(1-k_1) - k_1 log(k_1)$ are endogenous since they both, depend and determine the mean of the consumption-price ratio – $\bar{cp}$[1]. A similar to Eq.(2.1.6) expression also holds for market returns–$r_{m,t+1}$:

$$r_{m,t+1} = k_{0m} - k_{1m} dp_{t+1} + dp_t + \Delta d_{t+1}. \tag{2.1.7}$$

BY show that in the above specified economy, the price-consumption ratio can be written as the following function of expected consumption growth and the time-varying volatility:

[1]For details refer to Beeler and Campbell (2009) or Bansal, Kiku and Yaron (2007).

$$cp_t = -A_0 - A_1 x_t - A_2 \sigma_t^2 = -6.2704 - 14.5549 x_t + 470.2738 \sigma_t^2. \tag{2.1.8}$$

The solutions for A_0, A_1 and A_2 are provided in Beeler and Campbell (2009) or Constantinides and Ghosh (2009). The loadings A_1, A_2 as well as the variances of x_t and σ_t^2 determine correlations of price of a claim to the aggregate consumption with expected consumption growth and with the time-varying volatility. Additionally, A_0 determines the mean of the consumption-price ratio and, therefore, influences the linearization constants k_0 and k_1.

Since the time-varying volatility process is assumed to be homoscedastic, the variance of its innovation affects A_0. Moreover, the coefficient A_1 summarizes dependence of expected consumption growth and expected returns on a claim to the aggregated consumption on x_t. BY assume that $\psi > 1$ which implies that A_1 is positive or that a rise in expected cash flows increases prices. The parameters entering A_2 reflect the time-heteroskedasticity of processes in Eq.(2.1.1)–Eq.(2.1.4). Note that since in the BY model expected consumption growth itself is assumed to be a heteroskedastic process, its variance affects A_2. The calibration of the model implies that A_2 is negative or an increase in the uncertainty decreases prices. The variance of the time-varying volatility is small. Therefore, a large in absolute value A_2 is needed to magnify its effect on prices.

An expression relating the dividend-price ratio to expected consumption and to the time-varying volatility is:

$$dp_t = -A_{0m} - A_{1m} x_t - A_{2m} \sigma_t^2 = -5.6343 - 93.2178 x_t + 2.3978 \times 10^3 \sigma_t^2. \tag{2.1.9}$$

A_{0m}, A_{1m} and A_{2m} have a similar interpretation to the coefficients in Eq.(2.1.8). A_{1m} summarizes dependence of expected market returns and expected dividend growth on expected consumption growth. The assumption that $\phi > \dfrac{1}{\psi}$, ensures that A_{1m} is positive or that an increase in expected cash flows increases prices. Moreover, A_{2m} is negative and high in absolute value implying a large negative effect of an increase in economic uncertainty on a price of a claim to the aggregate dividends. Returns on the aggregate consumption claim enter the pricing kernel and, therefore, affect returns on any asset. This implies a dependence of A_{0m}, A_{1m} and A_{2m} on A_0, A_1 and A_2.

2.2 Predictability of stock returns in the long run risks model

Beeler and Campbell (2009) report abnormal patterns in predictability of excess returns and dividend growth by the dividend-price ratio in the BY model. In this section, I extend their analysis by studying predictability of returns and of the risk free rate separately.

2.2.1 Data

The proxy for the market portfolio is the value-weighted portfolio of stocks in NYSE, AMEX and NASDAQ obtained from CRSP. Following the common procedure, yearly log-returns are constructed by summing up twelve monthly log-returns. The annual dividend growth and the dividend-price ratio are obtained from the aggregated value weighted total returns and the aggregated value weighted returns excluding cash flows. The proxy for the riskless interest rate is the return on the one-month Treasury Bill from Ibbotson Associates Inc. downloaded from Prof. Kenneth French's Data Library. All the nominal quantities are adjusted by the CPI. Finally, the data set covers 80 years from 1930 to 2009.

Using the equations describing the BY economy, I generate 1,000 paths of 960 monthly data points and aggregate them to the annual frequency following the procedure described in Beeler and Campbell (2009). Annual dividend growth series is constructed using sample averages of the monthly data and the annualized dividend-price ratio employs sample averages of dividend growth together with the price of the last month of a year.

Table 2.2 provides the means and the standard deviations of the data on returns, dividend growth and the dividend-price ratio generated from the BY model as well as the data from CRSP. For the model I report the medians of the corresponding statistics taken across simulation paths [2]. Constantinides and Ghosh (2009) argues that the BY model misses structural breaks observed in the data. Indeed, Lettau and Van Nieuwerburgh

[2]Since the mean values are close to the median ones they are not reported.

(2007) and Koijen and Van Nieuwerburgh (2010) provide evidence in favor of structural breaks in the dividend-price ratio. I repeat the analysis of the Koijen and Van Nieuwerburgh (2010) and Lettau and Van Nieuwerburgh (2007) and demean the dividend-price ratio to account for the structural breaks in 1954 and 1994. In Table 2.2 the adjusted dividend-price ratio is denoted as dp^{sb}[3]

Tab. 2.2: Descriptive Statistics

This table reports the descriptive statistics. Columns marked *Model* report the descriptive statistics for the data generated by the model. Columns marked *Data* report the descriptive statistics of the annual stock market data adjusted by the CPI for a period 1930-2009. All the data is in logs. dp^s is the dividend-price ratio adjusted for the structural breaks in 1954 and 1994 by demeaning. For the model, I report the medians of the descriptive statistics taken across simulation paths. Standard deviations are in %

Statistics	dp	dp^{sb}	r_m	r_f	$r_m - r_f$	Δd
Panel A: Model						
Mean	−3.0116	–	0.0674	0.0258	0.0413	0.0178
Standard Deviation	18.38	–	16.52	1.22	16.65	11.05
Panel B: Data						
Mean	−3.3273	–	0.0571	0.0043	0.0528	0.0136
Standard Deviation	43.19	20.16	20.03	3.93	20.36	14.03

The model captures the mean of the market return, dividend-price ratio and dividend growth well (see Bansal and Yaron (2004), Beeler and Campbell (2009)). The values of the standard deviations for most of the variables are also close to the ones in the data. The biggest discrepancy is observed for the standard deviation of the dividend-price ratio, which in the model is more then twice lower than in the data. [4] Note, however, that the model matches very closely the standard deviation of the adjusted dividend-price ratio. Moreover, the riskless interest rate generated by the model has a higher mean and a lower standard deviation than the one observed in the data. This is attributed to inflation surprises and the 1951 Accord affecting returns on the Treasury Bill rate which are not present in the model.

[3] I also used a data set which starts in 1952 to control for the 1951 Accord. The mean of the return on the Treasury Bill in this subsample is somewhat higher while its standard deviation is lower. Using this smaller data set to analyze predictability of the riskless interest rate and of the excess returns does not change the results reported in this paper.

[4] The fact that the model produces a less volatile than in the data dividend price ratio has already been mentioned in Beeler and Campbell (2009).

2.2.2 Regression analysis

Following Cochrane (2008b) I analyze predictability of returns jointly with predictability of dividend growth. More specifically, in Table 2.3 I estimate the following predictive VAR[5]:

$$dp_{t+1} = \alpha^{dp} + \beta^{dp} dp_t + \epsilon^{dp}_{t+1}, \tag{2.2.1}$$

$$r_{m,t+1} = \alpha^{r} + \beta^{r} dp_t + \epsilon^{r}_{t+1}, \tag{2.2.2}$$

$$\Delta d_{t+1} = \alpha^{d} + \beta^{d} dp_t + \epsilon^{d}_{t+1}. \tag{2.2.3}$$

To account for heteroskedasticity, I adjust the standard errors by GMM using the codes downloaded from Prof. John H. Cochrane web page. Additionally to the usual regression output, I report the absolute values of the long-run slope coefficients of Cochrane (2008b):

$$b^{r,lr} = \frac{b^r}{1 - k^{'}_{1m}\beta^{dp}}, \tag{2.2.4}$$

$$b^{d,lr} = \frac{b^d}{1 - k^{'}_{1m}\beta^{dp}}. \tag{2.2.5}$$

$b^{r,lr}$ is interpreted as a fraction of the variation in the dividend-price ratio attributed to expected returns and $b^{d,lr}$ as the fraction attributed to expected dividend growth. To obtain the long-run coefficients in the model, in each simulation I calculate $k^{'}_{1m}$ as:

$$k^{'}_{1m} = \frac{e^{-\bar{dp}}}{1 + e^{-\bar{dp}}}$$

using the sample mean of the dividend price ratio $\bar{dp}$.

Fama and French (1988a) and Pastor and Stambaugh (1999) provide evidence for the discount rate effect–a negative contemporaneous covariance between the innovations to expected and realized returns. To check for the presence of the discount rate effect in the model, in the last row of Table 2.3 I add an estimate of this covariance given by

[5] Note that I can use any two regression equations from Eq.(2.2.1)-Eq.(2.2.3) together with the Campbell and Shiller (1988) identity to identify the coefficients and the standard error of the other regression equation, see Cochrane (2008a) and Cochrane (2008b).

$\beta^r cov(\epsilon^r_{t+1}, \epsilon^{dp}_{t+1})$. All the values reported for the model are the medians of the corresponding statistics taken across simulation paths. For the data, I report the estimation results using the conventional dividend-price ratio in the column *Data(dp)* and the dividend-price ratio adjusted for structural breaks in the column *Data(dp^s)*. Finally, in the columns named %, I report a percentile of the distribution of a given statistics in the model which corresponds to its value in the data.

Tab. 2.3: Predictability of returns

In this table I report the results of estimating a predictive VAR where only the dividend-price ratio is used to predict returns and dividend growth. In the column *Model* I report the medians of the corresponding statistics taken across simulation paths. In Data(dp) I report the values obtained in the data. In Data(dp^s) I report the values obtained in the data using the dividend-price ratio adjusted for structural breaks in 1954 and 1994 by demeaning. In %– I report a percentile of the distibution the statistics obtained in the model which corresponds to the value obtained in the data. For each regression, I report an estimated slope coefficient, its t-statistics and the R-square statistics. $\left|\beta^{d,lr}\right|$ and $\left|\beta^{r,lr}\right|$ are the absolute values of the long-run slope coefficients of Cochrane (2008b), defined in Eq.(2.2.4) and Eq.(2.2.5) in the text. Standard errors are GMM adjusted to account for heteroskedasticity. Values of R-square statistics are in %

Statistics	Model	Data(dp)	%	Data(dp^s)	%
Panel A: AR(1) process for the dividend-price ratio					
β^{dp}	0.6723	0.9399	100	0.6651	48
t-stat	8.2917	21.9996	–	6.7951–	
R^2	45.08	88.07	100	41.47	39
Panel B: Predictive regression for returns					
β^r	−0.0168	0.0950	84	0.3343	99
t-stat	−0.1830	1.9184	–	3.4916	–
R^2	00.65	4.31	94	11.14	100
$100 \times \left|\beta^{r,lr}\right|$	4.77	102.48	100	90.95	99
$\beta^r cov(\epsilon^r_{t+1}, \epsilon^{dp}_{t+1})$	0.0003	−0.0019	16	−0.0047	2
Panel C: Predictive regression for dividend growth					
β^d	−0.3739	0.0048	100	0.0477	100
t-stat	−7.1629	0.1153	–	0.5506	–
R^2	38.88	00.02	0	00.45	0
$100 \times \left|\beta^{d,lr}\right|$	104.77	2.48	0	9.05	0

Table 2.3 shows that in the data the dividend-price ratio is a highly persistent process. There is stronger evidence in favor of predictability of returns rather than dividend growth. For returns, the slope coefficient is positive and marginally significant. The value of the R^2 indicates that the dividend-price ratio captures a bit more than 4 % of the variation in future returns. The value of the long horizon coefficient implies that over 100 % of the variation in the dividend-price ratio is due to expected returns. A positive slope coefficient in the returns regression signals the presence of the discount rate

effect–the estimated contemporaneous covariance between the innovations to expected and realized returns is negative. For dividend growth, the slope coefficient is also positive, but small. Moreover, dividend-price ratio captures almost no variation in the future dividend growth. The value of the long run coefficient indicates that virtually none of the variation in the dividend-price ratio stems from expected dividend growth. Finally, adjusting dividend-price ratio for structural breaks decreases its persistence and reinforces predictability evidence for returns. The slope coefficient in the returns regression with the adjusted dividend-price ratio is highly statistically significant and R^2 is over 11%.

Dividend-price ratio in the model is characterized by a low persistence. The median AR(1) coefficient is 0.6723 against the value of 0.9399 in the data. Judging from the value of % the model is clearly rejected. The model, however, matches very well dynamics of the adjusted dividend price ratio generating the AR(1) coefficient and the R-square statistics very close to the one observed in the data.

More importantly, Panel B and Panel C reveal that predictability patterns in the model are strikingly different from those in the data. In the model, the slope coefficients are negative for both returns and dividend growth regressions while they are positive for both regressions in the data. R^2 is excessively small for returns and excessively large for dividend-growth. The value of the estimated covariance between the innovations to expected and realized returns is small and positive indicating a failure of the model to produce the discount rate effect. Moreover, judging from the long-run coefficients, in the model most part of the variance of the dividend-price ratio stems from expected dividend growth-an absolutely reverse pattern to the data. While from Panel A the dividend-price ratio generated by the model matches the dynamics of the adjusted dividend-price ratio well, there is even stronger evidence against the model in this case. The values of % for both, returns and dividend growth regressions indicate a complete rejection of the model.

In Table 2.4 I analyze predictability of excess returns and of the interest rate. More specifically, I estimate:

$$r_{m,t+1} - r_{f,t+1} = \alpha^{er} + \beta^{er} dp_t + \epsilon^{er}_{t+1}, \tag{2.2.6}$$

$$r_{f,t+1} = \alpha^{rf} + \beta^{rf} dp_t + \epsilon^{rf}_{t+1}. \tag{2.2.7}$$

Tab. 2.4: Predictability of excess returns and riskless interest rate

In this table I report the results of predicting the excess returns and the riskless interest rate with the dividend-price ratio. In the column *Model* I report the medians of the corresponding statistics taken across simulation paths. In Data(dp) I report the values obtained in the data. In Data(dp^s) I report the values obtained in the data using the dividend-price ratio adjusted for structural breaks in 1954 and 1994 by demeaning. In %– I report a percentile of the distibution of the statistics obtained in the model which corresponds to the value obtained in the data. For each regression, I report the estimated slope coefficient, its t-statistics and the R-square statistics. Standard errors are GMM adjusted to account for heteroskedasticity. Values of R-square statistics are in %

Statistics	Model	Data(dp)	%	Data(dp^s)	%
Panel A: Predictive regression for the excess returns					
β^{er}	0.0419	0.1080	74	0.3215	99
t-stat	0.4253	2.1716	–	3.3416	–
R^2	0.63	5.56	96	10.30	99
Panel B: Predictive regression for the interest rate					
β^{rf}	−0.0557	−0.0130	100	0.0128	100
t-stat	−14.3086	−1.4490	–	0.5264	–
R^2	71.06	2.17	0	0.44	0

I start with the predictability patterns observed in the data. Comparing the values from Table 2.4 with the corresponding ones from Table 2.3, excess returns are more predictable than returns. The slope coefficient, the t-statistics and the R-square statistics increase. The riskless interest rate is not well forecasted by the dividend-price ratio. The t-statistics and the R-square statistics are low. Adjusting dividend-price ratio for structural breaks reinforces the evidence for predictability of excess returns. The slope coefficient is highly statistically significant and R^2 reaches 10 %. There is even less evidence for predictability of the riskless rate, however. Comparing with the unadjusted dividend-price ratio case, both the t-statistics and the R-square statistics decrease.

Note that in the model, while still being statistically insignificant, the slope coefficient changes its sign from negative in the case of returns to positive in the case of the excess returns. The rejection of the model becomes weaker since the value of % deccreases. Using the adjusted dividend-price ratio, however, still leads to a strong rejection of the model. The difference between the model and the data is especially strong in the case of the predictive regression for the riskless rate. While in the data the riskless rate is not well forecasted by the dividend-price ratio, in the model lagged dividend-price ratio predicts riskless rate with a negative and highly statistically significant coefficient and explains around 70 % of its variation. The value of % for both the conventional dividend-price ratio and the adjusted one indicate an absolute rejection of the model.

2.3 Expected returns and expected dividend growth in the long run risks model

Cochrane (2008a), Cochrane (2008b), Binsbergen and Koijen (2010) and Koijen and Van Nieuwerburgh (2010) show the usefulness of the state-space representation for returns (SSR) for interpreting predictive relations between returns, dividend growth and the dividend-price ratio. In this paper, I analyze predictability patterns in the BY economy through an SSR. Let:

$$\mu_t^r = E_t\left[r_{t+1}\right],$$
$$\mu_t^d = E_t\left[\Delta d_{t+1}\right]$$

In the Appendix, I show that the above economy implies the following SSR:

$$\mu_{t+1}^r = \alpha^{\mu r} + \delta^{\mu r}\mu_t^r + \tau\mu_t^d + \epsilon_{t+1}^{\mu r}, \tag{2.3.1}$$

$$\mu_{t+1}^d = \alpha^{\mu d} + \delta^{\mu d}\mu_t^d + \epsilon_{t+1}^{\mu d}, \tag{2.3.2}$$

$$\Delta d_{t+1} = \mu_t^d + \epsilon_{t+1}^d, \tag{2.3.3}$$

I provide the expressions for the coefficients and errors of Eq.(2.3.1)–Eq.(2.3.3) in Table 2.5.

From Table 2.5, expected returns inherit persistence of the time-varying volatility while expected dividend growth that of expected consumption growth. The Euler equation implies a dependence of expected returns on any asset on expected consumption growth. A perfect correlation between expected dividend growth and expected consumption growth assumed in the BY model then leads to the lagged expected dividend growth entering the equation for expected returns. The value of persistence of the time-varying volatility

Tab. 2.5: State-space representation for returns

This table provides the expressions for the coefficients and errors of the state-space representation for returns in terms of parameters of the long run risks model.

Variable	Expression in terms of the model parameters	Value
$\alpha^{\mu r}$	$(1-\nu_1)\left[k_{0m}-(1-k_{1m})A_{0m}+k_{1m}A_{2m}\bar{\sigma}^2(1-\nu_1)+m_d\right]-$ $-A_{2m}\bar{\sigma}^2(1-k_{1m}\nu_1)(1-\nu_1)-\dfrac{(\rho-\nu_1)m_d}{\phi\psi}$	7.5804×10^{-5}
$\alpha^{\mu d}$	$(1-\rho)m_d$	3.1500×10^{-5}
$\delta^{\mu r}$	ν_1	0.9870
$\delta^{\mu d}$	ρ	0.9790
τ	$\dfrac{(\rho-\nu_1)}{\phi\psi}$	−0.0018
$\epsilon^{\mu r}_{t+1}$	$\dfrac{1}{\psi}\varphi_e\sigma_t e_{t+1}-(1-\nu_1 k_{1m})A_{2m}\sigma_\omega\omega_{t+1}$	–
$\epsilon^{\mu d}_{t+1}$	$\phi\varphi_e\sigma_t e_{t+1}$	–
ϵ^{d}_{t+1}	$\varphi_d\sigma_t u_{t+1}$	–

is close to the value of persistence of expected consumption growth. Additionally, the dividend leverage on expected consumption growth and elasticity of intertemporal substitution are more than one. This implies a small absolute value of τ and, consequently, a small contribution of the lagged expected dividend growth to the volatility of expected returns. The errors to expected dividend growth are completely determined by the risk of poor economic growth. A $\phi > 1$ implies an even higher variance of the innovations to expected dividend growth than of the innovations to expected consumption growth. The errors to expected returns, on the other hand, are determined by both risks. The effect of the innovations to expected consumption growth on expected returns is decreased by a value of elasticity of intertemporal substitution larger than one. The effect of the innovations to the time-varying volatility on expected returns is magnified by a large $(1-\nu_1 k_{1m})A_{2m}$. Note that whenever expected consumption growth shocks contribute the most to the volatility of the innovations to expected returns, expected returns are correlated with expected dividend growth. In an extreme case, when the innovations to expected consumption growth determine all the variation in the innovations to expected returns, which happens if $(1-\nu_1 k_{1m})A_{2m}$ is small, expected returns will be perfectly correlated with but less volatile than expected dividend growth.

Using Eq.(2.3.1)–Eq.(2.3.3) in the Appendix I derive an identity linking the dividend-price ratio to expected returns and expected dividend growth. In Panel A of Table 2.6 I report the standard deviations on the diagonal and the cross-correlations on the off-diagonal of the dividend-price ratio, expected returns and expected dividend growth and in the Panel B I decompose the standard deviation of the dividend-price ratio into the parts related to the variation in expected returns and to the variation in expected dividend growth. The values reported are calculated using the original calibration of the BY model.

Tab. 2.6: Implications of state-space representation for returns

Panel A of this table presents the values of the standard deviations on the diagonal and the correlations on the off-diagonal of the dividend-price ratio, expected returns and expected dividend growth. Panel B provides the values of fractions of the standard deviations of expected returns and expected dividend growth in the standard deviation of the dividend-price ratio. The calculations are based on the state-space representation for returns derived using equations describing the long run risks model. All the numbers are in %.

Panel A: Correlations and standard deviations			
Components	dp	μ^r	μ^d
dp	16.11	-74.50	-96.75
μ^r	-74.50	00.13	85.25
μ^d	-96.75	85.25	0.51

Panel B: Decomposition of the volatility of the dividend-price ratio		
	y	
Share	$\frac{\mu^r}{1-k_{1m}\delta^{\mu r}}$	$\frac{(1-(\tau+\delta^{\mu r})k_{1m})\mu^d}{(1-k_{1m}\delta^{\mu d})(1-k_{1m}\delta^{\mu r})}$
$std(y)/std(dp)$	47.29	139.76

In line with the evidence on the negative slope coefficient in the predictive regression for returns and dividend growth reported in Section 2, dividend-price ratio is negatively correlated with both expected returns and expected dividend growth. Moreover, expected returns and expected dividend growth are positively correlated and the variance of expected dividend growth is higher than the variance of expected returns. Expected dividend growth is very volatile. From Table 2.6, the standard deviation of expected dividend growth exceeds the standard deviation of expected returns by a factor of five. Given that both processes are highly persistent, it should not come as a surprise that expected dividend growth accounts for 140 % of the standard deviation of the dividend price ratio against slightly more than 47 % of the standard deviation stemming from expected returns as reported in the Panel B. Since:

$$corr(dp, \mu^r) = \left(\frac{1}{1 - k_{1m}\delta^{\mu r}}\right)\frac{\sigma(\mu^r)}{\sigma(dp)} - \left(\frac{(1 - (\tau + \delta^{\mu r})k_{1m})}{(1 - k_{1m}\delta^{\mu d})(1 - k_{1m}\delta^{\mu r})}\right)\frac{\sigma(\mu^d)}{\sigma(dp)}corr(\mu^r, \mu^d); \tag{2.3.4}$$

and

$$corr(dp, \mu^d) = \left(\frac{1}{1 - k_{1m}\delta^{\mu r}}\right)\frac{\sigma(\mu^r)}{\sigma(dp)}corr(\mu^r, \mu^d) - \left(\frac{(1 - (\tau + \delta^{\mu r})k_{1m})}{(1 - k_{1m}\delta^{\mu d})(1 - k_{1m}\delta^{\mu r})}\right)\frac{\sigma(\mu^d)}{\sigma(dp)}, \tag{2.3.5}$$

a positive correlation between expected returns and expected dividend growth together with an excessive volatility of expected dividend growth leads to a negative correlation of the dividend-price ratio with both, expected returns and expected dividend growth.

To elaborate further, in Table 2.7 I connect expected returns, expected dividend growth, the expected excess returns, the interest rate and the dividend-price ratio to the sources of the long run risks-the processes for expected consumption growth and the time-varying volatility. In Panel A of Table 2.7 I present the shares of the volatility attributed to expected consumption growth and to the time-varying volatility. In the Panel B I provide the values of the correlations of each of the variables with x_t and σ_t^2. All the values are computed using the calibration the BY model.

Table 2.7 offers another explanation to the counterfactual features generated by the model. Almost all the variance of the dividend-price ratio, expected returns and the riskless interest rate and all the variance of expected dividend growth are attributed to expected consumption growth. An increase in expected consumption leads to an increase in prices in the BY economy. Consequently, the dividend-price ratio is negatively correlated with expected consumption growth. The riskless interest rate, expected returns and expected dividend growth, however, are positively correlated with it. This explains a negative sign in the predictive regressions for returns, dividend growth and riskless interest rate. Correlations of over 95 % between expected consumption growth and the dividend-price ratio, expected dividend growth and the riskless interest rate explain

Tab. 2.7: Variation in expected consumption growth and time-varying volatility

Panel A of this table presents the fractions of the variance of the dividend-price ratio, expected dividend growth, expected returns, the riskless interest rate and the excess returns attributed to expected consumption growth and to the time-varying volatility. Panel B, provides the correlations of each of the variables with x_t and σ_t^2. All the numbers are reported in %. All the values are computed using the calibration of the long run risks model.

Series	x_t	σ_t^2
Panel A: The fractions of the variance		
dp_t	95.44	4.56
μ_t^d	100.00	0.00
μ_t^r	78.61	21.39
$r_{f,t+1}$	97.03	2.97
$\mu_t^r - r_{f,t+1}$	0.00	100.00
Panel B: The correlations		
dp_t	−98.06	21.23
μ_t^d	100.00	0.00
μ_t^r	87.18	45.04
$r_{f,t+1}$	94.44	−16.36
$\mu_t^r - r_{f,t+1}$	0.00	100.00

a highly significant predictive relations between the dividend-price ratio and dividend growth and riskless interest rate.

Note that while expected consumption growth determines the biggest fraction of the variation in expected returns it contributes nothing to the variation in the expected excess returns, implying a very different economic meaning of the predictability of these processes. An increase in economic uncertainty leads to a decrease in prices and to an increase in the dividend-price ratio. Increase in economic uncertainty also increases expected excess returns. This explains a positive sign of the slope coefficient in the excess returns regression. These features of the model point at the need in giving more weight to the variance of the time-varying volatility in the variation of prices.

Finally, the contemporaneous covariance between the innovations to expected and realized returns is related to the innovations to expected consumption growth and to the time-varying volatility through:

$$cov\left(\epsilon_{t+1}^r, \epsilon_{t+1}^{\mu r}\right) = 0.1198\sigma_t^2 - 5.1686 \times 10^{-7} = k_{1m}\frac{1}{\psi}A_{1m} - k_{1m}(1 - k_{1m}\nu_1)(A_{2m}\sigma_\omega)^2 = \\ = k_{1m}\frac{1}{\psi}\frac{\phi - \frac{1}{\psi}}{1 - k_{1m}\rho}\varphi_e^2\sigma_t^2 - k_{1m}(1 - k_{1m}\nu_1)(A_{2m}\sigma_\omega)^2. \tag{2.3.6}$$

The discount rate effect states that the covariance between the innovations to the realized and expected returns is negative. The unconditional mean of σ^2 is 6.084×10^{-5} implying that, on average, the above covariance is positive rather than negative. In the second equality of Eq.(2.3.6), I express the discount rate effect in terms of the parameters which govern the shares of the variance of the dividend-price ratio attributed to expected consumption growth and to the time-varying volatility–A_{1m} and A_{2m}. In line with the above conclusions, then the second equality shows that the model fails to produce the discount rate effect because it assigns too much weight to the poor economic growth risk and too little weight to the high economic uncertainty risk in the variation of prices.

To get more insights, in the third equality I substitute for A_{1m}. Going back to Table 2.5, $\frac{1}{\psi}\varphi_e\sigma_t$ measures the variance of the component of the innovations to expected returns attributed to the risk of the poor economic growth and $(1-k_{1m}\nu_1)A_{2m}\sigma_\omega$ of the component attributed to the risk of a high economic uncertainty. Similarly, $\phi\varphi_e\sigma_t$ governs the variance of the innovation to expected dividend growth, while ρ measures its persistence. Thus, the third equality in Eq.(2.3.6) indicates the need in a decrease in the volatility and persistence of expected dividend growth and in an increase in the volatility of expected returns. Additionally, it provides guidance on the specific parameters which should be affected. Specifically, the third equality in Eq.(2.3.6) indicates a need in a decrease in ϕ, φ_e and ρ and in an increase in σ_ω and A_{2m}[6]

Exactly the same conclusions are made based on the implications of the SRR to the predictive regressions. Given the SSR described by Eq.(2.3.1)-Eq.(2.3.3), in the Appendix I derive the following predictive relations implied by the model:

$$\begin{aligned} dp_{t+1} &= \\ &= -0.0717 + 0.987 dp_t + 0.2486\mu_t^d + \epsilon_{t+1}^{dp} \\ &= -0.0713 + 0.987 dp_t + 0.7458 x_t + \epsilon_{t+1}^{dp} \\ &= \alpha^{dp} + \nu_1 dp_t + (\nu_1-\rho)\frac{\phi-\frac{1}{\psi}}{1-k_{1m}\rho}x_t - \frac{\phi-\frac{1}{\psi}}{1-k_{1m}\rho}\varphi_e\sigma_t^2 e_{t+1} - A_{2m}\sigma_\omega\omega_{t+1}, \end{aligned} \tag{2.3.7}$$

[6] An increase in the volatility of expected returns can also be achieved through a decrease in the elasticity of intertemporal substitution. The long run risk model, however, relies on the value of this parameter larger than one. See for example Beeler and Campbell (2009) and Bonomo, Garcia, Meddahi and Tedongap (2010).

with $\alpha^{dp} = (1-\nu_1)\left[-A_{0m} - A_{2m}\bar{\sigma}^2\right]$;

$$
\begin{aligned}
r_{t+1} &= \\
&= 0.0982 + 0.0171 dp_t + 0.7524\mu_t^d + \epsilon_{t+1}^r \\
&= -0.0372 - 0.0072 dp_t + 58.0643\sigma_t^2 + \epsilon_{t+1}^r \\
&= \alpha^r + (1-k_{1m}\rho)\left[\frac{-\frac{1}{\psi}}{\phi - \frac{1}{\psi}}\right] dp_t + \left((1-k_{1m}\rho)\left[\frac{-\phi}{\phi-\frac{1}{\psi}}\right] + k_{1m}(\nu_1-\rho)A_{2m}\right)\sigma_t^2 + \epsilon_{t+1}^d - k_1\epsilon_{t+1}^{dp},
\end{aligned}
\tag{2.3.8}
$$

with $\alpha^r = k_{0m} - k_{1m}\alpha^{dp} + \alpha^d + k_{1m}(\nu_1 - \rho)A_{0m}$;

$$
\begin{aligned}
\Delta d_{t+1} &= \\
&= -0.1304 - 0.0227 dp_t + 1.329\mu_t^r + \epsilon_{t+1}^d \\
&= -0.1798 - 0.0322 dp_t + 77.1677\sigma_t^2 + \epsilon_{t+1}^d \\
&= \alpha^d + (1-k_{1m}\rho)\frac{-\phi}{\phi-\frac{1}{\psi}} dp_t + (1-k_{1m}\rho)\frac{-\phi}{\phi-\frac{1}{\psi}} A_{2m}\sigma_t^2 + \varphi_d\sigma_t u_{t+1},
\end{aligned}
\tag{2.3.9}
$$

with $\alpha^d = m_d + (1-k_{1m}\rho)\frac{-\phi}{\phi-\frac{1}{\psi}}A_{0m}$.

In Eq.(2.3.7)-Eq.(2.3.9) the first equality follows directly form the SSR. In the second equality I substitute out expected returns/expected dividend with the time-varying volatility or expected consumption growth. Finally, in the third equality I present predictive relations in terms of the calibration parameters.

The first equality in Eq.(2.3.7)-Eq.(2.3.9) clearly shows that it is a negative correlation of the dividend-price ratio with expected returns and expected dividend growth that leads to all the counterfactual predictability features generated by the model. According to Eq.(2.3.7), regressing an uncorrelated with expected dividend growth dividend-price ratio should deliver a high AR(1) coefficient estimate. The dividend-price ratio, however, is highly negatively correlated with expected dividend growth implying a large negative bias in the AR(1) coefficient estimate.

Similarly, Eq.(2.3.8) shows that the dividend price ratio uncorrelated or positively correlated with expected dividend growth should predict returns positively and that a high in absolute value negative correlation between the dividend-price ratio and expected dividend growth leads to a negative coefficient in the predictive regression for returns. Moreover, Eq.(2.3.9) shows that a negative correlation between the dividend-price ratio and expected returns decreases further the slope coefficient in the predictive regression for dividend growth, reinforcing its predictability.

The second equality in Eq.(2.3.7)-Eq.(2.3.9) points at a need in a smaller correlation of the dividend-price ratio with expected consumption growth and in its larger correlation with the time-varying volatility. Eq.(2.3.7) shows that regressing the dividend-price ratio uncorrelated with expected consumption growth should deliver a high AR(1) coefficient estimate. Eq.(2.3.8) and Eq.(2.3.9) imply that it may be possible to obtain a positive slope coefficient in predictive regressions for returns and dividend growth rate by increasing the correlation between the dividend-price ratio and the time-varying volatility. Finally, the third equality indicates a need in a smaller φ_e, ρ and ϕ and a larger A_{2m} and σ_ω. I explore the possibility of adjusting these parameters in the next section.

2.4 Long run risks processes

In this section I conduct a comparative statics analysis to show that it is the specification of the long run risks processes which limits the ability of the model to match the predictability patterns. I start with the equations describing the model and then support the analysis by simulations. The BY model relies of the following process governing expected consumption growth:

$$x_t = \rho x_{t-1} + \varphi_e \sigma_{t-1} e_t = 0.979 x_{t-1} + 0.044 \sigma_{t-1} e_t.$$

Expected consumption growth is heteroskedastic. This heteroskedasticity leads to the dependence of A_{2m} on the parameters determining the share of the expected consumption growth in the volatility of prices. To elaborate, the solution to A_{2m} is:

$$A_{2m} = \frac{(1-\theta)A_2(1-k_1\nu_1) + 0.5\left[\gamma^2 + \varphi_d^2 + ((\theta-1)k_1A_1 + k_{1m}A_{1m})^2\varphi_e^2\right]}{1-k_{1m}\nu_1}, \qquad (2.4.1)$$

Recall that A_{2m} is negative, implying that higher uncertainty decreases prices. Additionally, in order a small variance of the time-varying volatility to have an effect on prices, A_{2m} should be large in absolute value. It is visible from Eq.2.4.1, that negativity and the value of A_{2m} is crucially dependent on A_2. Since:

$$A_2 = 0.5\left[\frac{\left(-\frac{\theta}{\psi}+\theta\right)^2}{\theta\,(1-k_1\nu_1)} + \frac{(\theta k_1 A_1 \varphi_e)^2}{\theta\,(1-k_1\nu_1)}\right] = \frac{0.5\theta}{1-k_1\nu_1}\left[\left(1-\frac{1}{\psi}\right)^2 + \left(k_1\varphi_e\frac{1-\frac{1}{\psi}}{1-k_1\rho}\right)^2\right] =$$
$$= -100.5537 - 369.7201, \qquad (2.4.2)$$

the value of A_2 in turn is largely determined by a value the variance of expected consumption growth, that is by φ_e and by ρ. Decrease in the value of these parameters leads to an increase in A_{2m}. Note that the parameter ϕ which governs the variance of expected dividend growth enters A_{2m} through A_{1m}. $(\theta-1)k_1A_1$ is negative and $k_{1m}A_{1m}$ is positive. This implies that a decrease in ϕ also leads to an increase in A_{2m}. In other words, decreasing the share of expected dividend growth in the variance of the dividend-price ratio leads to an increase in A_{2m} or, equivalently, to a decrease in the share of expected returns. Moreover, a large decrease in φ_e, ϕ and ρ may lead to a counterintuitively positive A_{2m} which implies a positive effect of the uncertainty on prices.

On the other hand, the time-varying volatility in the BY model is a homoscedastic process:

$$\sigma_t^2 = \bar{\sigma}^2(1-\nu_1) + \nu_1\sigma_{t-1}^2 + \sigma_\omega\omega_t = 7.909\times 10^{-7} + 0.987\sigma_{t-1}^2 + 2.300\times 10^{-6}\omega_t.$$

Heteroskedasticity of expected consumption growth and homoscedasticity of the time-varying volatility together imply that the dynamics of the excess returns depends on the variance of the expected consumption growth. Specifically, it can be shown that:

$$\mu_t^r - r_{f,t+1} = \alpha^{\mu r} - \alpha^{r_f} - 0.5\left[(k_{1m}A_{1m}\varphi_e)^2 - 2(1-\theta)k_{1m}k_1A_{1m}A_1(\varphi_e)^2 + \varphi_d^2\right]\sigma_t^2 = \\ = A_{0r^e} + A_{1r^e}\sigma_t^2, \tag{2.4.3}$$

with $\alpha^{\mu r}$ defined in Table 2.5 and $\alpha^{r_f} = -\theta log\beta + \frac{\theta}{\psi}m - (\theta-1)(k_0 + (k_1-1)A_0 + k_1A_2(1-\nu_1)\sigma^2 + m) - 0.5(\theta-1)^2(k_1A_2\sigma_\omega)^2$. Loadings of the consumption-price ratio and the dividend-price ratio on expected consumption growth–A_1 and A_{1m} and the parameter which determines the variance of the innovations to expected consumption growth $-\varphi_e$ affect the sign and the value of the response of the excess returns to the time-varying volatility. Consequently, a decrease in the share of expected dividend growth in the variance of the dividend-price ratio may lead to a decrease in the effect of the time-varying volatility on the excess returns and weaken the predictive relation between the excess returns and the dividend-price ratio.

Following Eq.(2.4.1) and Eq.(2.4.3), heteroskedasticity in expected consumption growth and homoscedasticity in the time varying volatility imply that any decrease in the share of expected dividend growth in the variance of the dividend-price ratio should be compensated with an increase in the share of expected returns through parameters ν_1 and σ_ω. Moreover, below it will be shown that an increase in these parameters will lead to a large number of negative realizations of σ_t^2 and, therefore, requires an increase in the mean of the time-varying volatility measured by $\bar{\sigma}^2$. A large increase in $\bar{\sigma}^2$, however, leads to a counterfactually high variance of consumption growth.

The assumption of homoscedasticity of the time-varying volatility process leads to another difficultly. It implies the dependence of A_{0m} on the parameters determining the share of expected returns in the variance of the dividend-price ratio. More specifically, A_{0m} is related to A_{2m} and σ_ω through:

$$A_{0m} = \bar{A_{0m}} + 0.5\frac{\left[(\theta-1)k_1A_2 + k_{1m}A_{2m}\right]^2\sigma_\omega^2}{1-k_{1m}}. \tag{2.4.4}$$

$\bar{A_{0m}}$ is a part of A_{0m} which depends on the unconditional means of dividend growth, consumption growth and of the time-varying volatility, on the linearization constants k_0 and k_{0m} and on the time discount factor β. Since A_{0m} defines the mean of the dividend-price ratio, a change in A_{2m} or in σ_ω then may lead to excessively high or low values of the mean of the dividend-price ratio and to an undefined solution to the endogenous linearization constant k_{1m}.

To quantify the analysis, in Figure 2.1 I plot the loading of the excess returns on the time-varying volatility and loadings of the dividend-price ratio on the long run risk processes, decreasing the variance of the innovation to both, expected consumption growth and expected dividend growth, through a decrease in φ_e. While changing the value of φ_e I adjust $\bar{\sigma}$ to keep the variance of the realized consumption growth constant. Figure 2.1 presents, from the right to the left, in the first row the loading of the excess returns on the time-varying volatility, A_{1r^e}, followed by the mean of the dividend-price ratio, A_{0m}. In the second row I plot the loading of the dividend-price ratio on expected consumption growth, A_{1m}, followed by the loading of the dividend-price ratio on the time-varying volatility A_{2m}. The values of the loadings which correspond to changing the value of φ_e from the original 0.044 to 0.004 are represented by a solid line while the values under the original calibration are indicated by a line in stars.

Figure 2.1 shows that, keeping a negative effect of the economic uncertainty on prices, a decrease in the variance of expected consumption growth leads to a decrease in the predictability of both returns and excess returns. A decrease in φ_e leads to a fast decrease in A_{1r^e}. Reducing φ_e by half leaves A_{1r^e} with less than one-fifth of its initial value. A decrease in φ_e also leads to a rapid increase in A_{2m} implying a decrease in the share of expected returns in the variance of the dividend-price ratio. A decrease in φ_e from the original 0.044 to 0.036 leads to an increase in A_{2m} from the original -2409.308 to around -1207.294. A further decrease in φ_e leads to a positive A_{2m} which implies a counter intuitively positive effect of the uncertainty on prices.

In Figure 2.2 I plot the loading of the excess returns together with the loadings of dividend-price ratio decreasing the value of ϕ, that is decreasing the variance of expected dividend growth and leaving the variance of expected consumption growth fixed. While

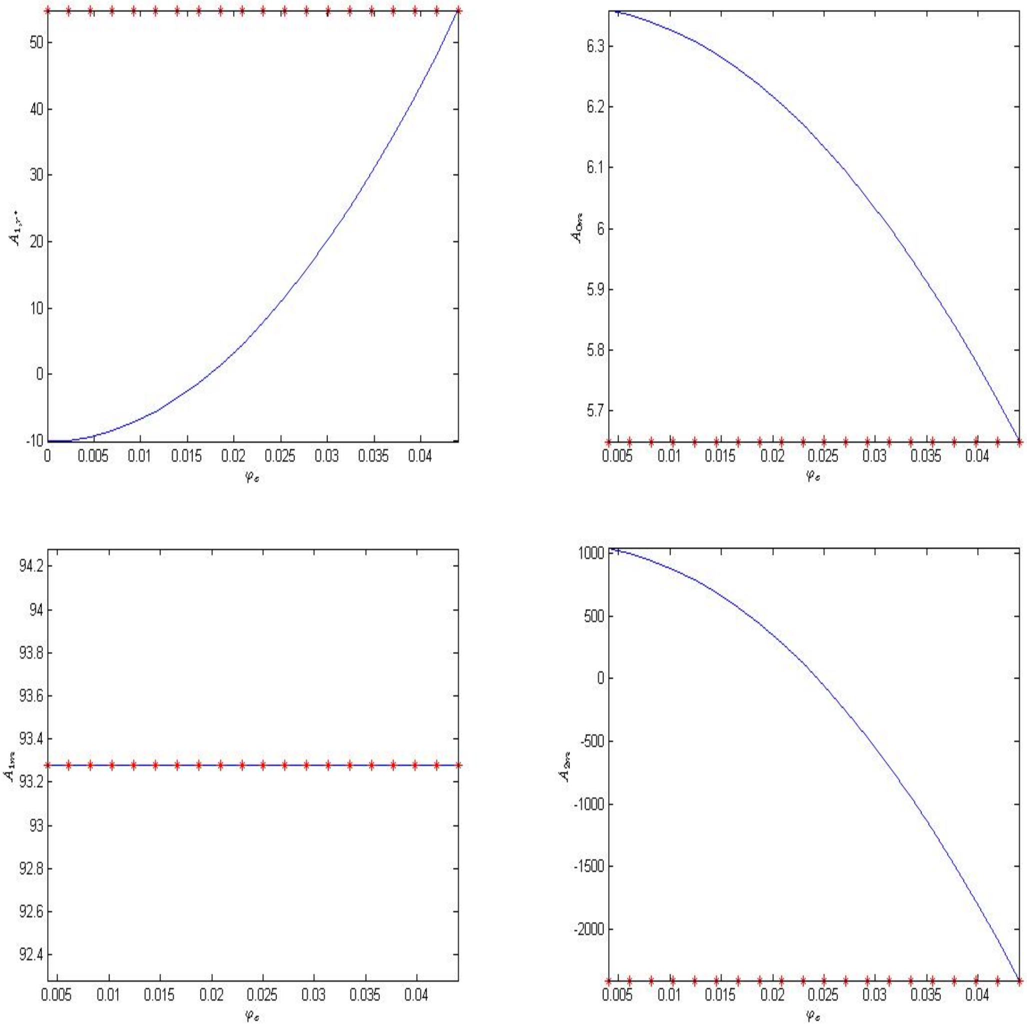

Fig. 2.1: Elasticity of loadings of expected excess returns and dividend-price ratio to variance of expected consumption growth

In this figure I plot the values of loadings of the expected excess returns and of the dividend-price ratio on the long run risks against the variance of the innovation to expected consumption growth. The values of the loadings which correspond to changing the value of φ_e from the original 0.044 to 0.004 are represented by a solid line while the values under the original calibration are indicated by a line in stars.

changing ϕ, I adjust φ_d to keep the unconditional variance of the dividend growth constant[7]. Similarly to Figure 2.1, the values of the loadings which correspond to changing a value of ϕ are depicted using a solid line while the values under the original calibration, by a line in stars.

Starting with the loading of the excess returns on the time-varying volatility in the upper-right corner, decreasing the variance of expected dividend growth leads to a rapid decrease in the predictability of excess returns. Figure 2.2 shows that decreasing ϕ by a

[7]Note that for values of $\phi < 1$, expected dividend growth is less volatile than expected consumption growth.

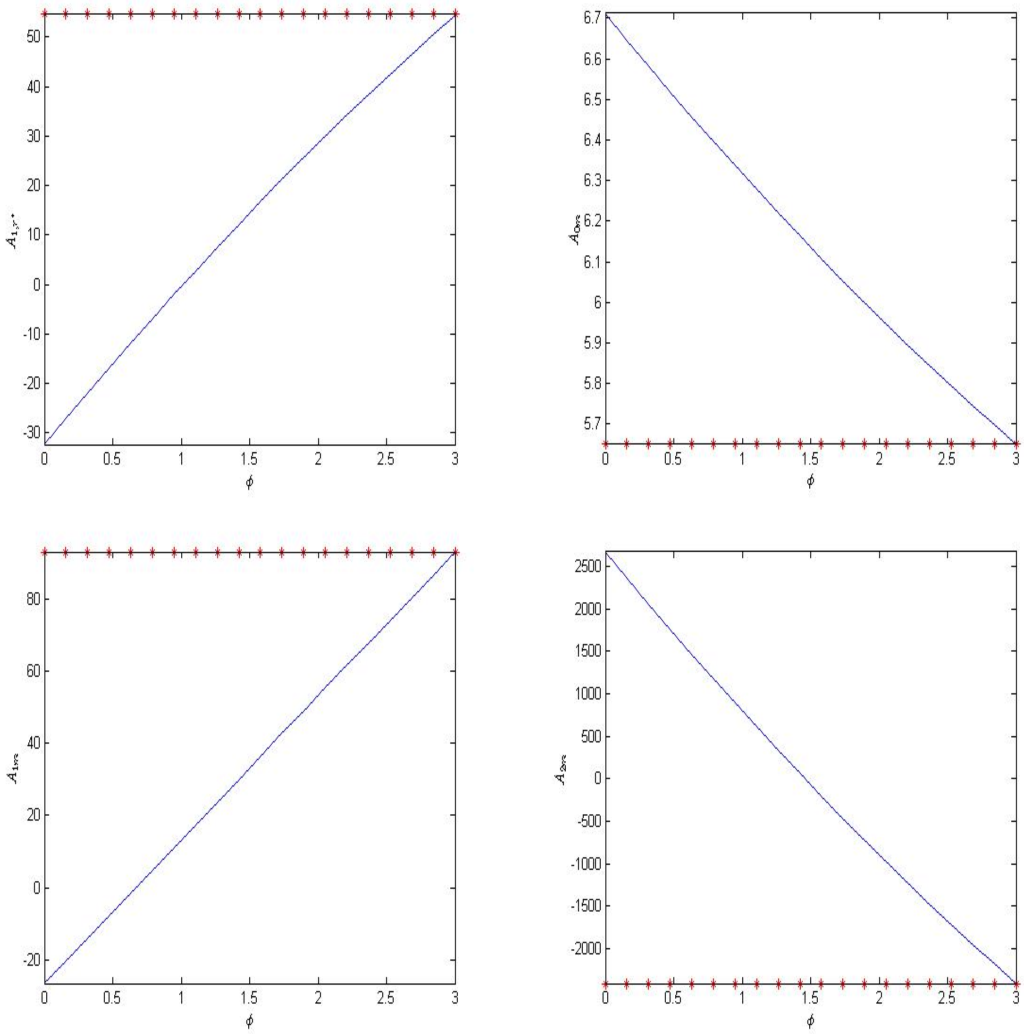

Fig. 2.2: Elasticity of loadings of expected excess returns and dividend-price ratio to variance of expected dividend growth

In this figure I plot the values of loadings of the expected excess returns and of the dividend-price ratio on the long run risks against the variance of the innovation to expected dividend growth. The values of the loadings which correspond to changing the value of ϕ from 0.000 to the original 3.000 are represented by a solid line while the values under the original calibration are indicated by a line in stars.

half, leads to a decrease in A_{1r^e} by more than two-thirds of its initial value. Similarly to the expected excess returns, a decrease in the variance of expected dividend growth leads to a lower predictability of returns, since a decrease in ϕ leads to a rapid increase in A_{2m}. Reducing ϕ by a half is met with a virtually zero A_{2m}.

Finally, in Figure 2.3 I plot the loadings while changing values of ρ from 0.000 to the original value of 0.979. Predictability of returns and excess returns is highly sensitive to a change in the persistence of expected consumption/dividend growth. Decreasing the value of the persistence from 0.979 to 0.900 leads to a negative and low in absolute value

A_{1r^e}. Additionally, for $\rho = 0.900$ A_{2m} is positive implying a counter intuitively positive effect of the time-varying volatility on prices.

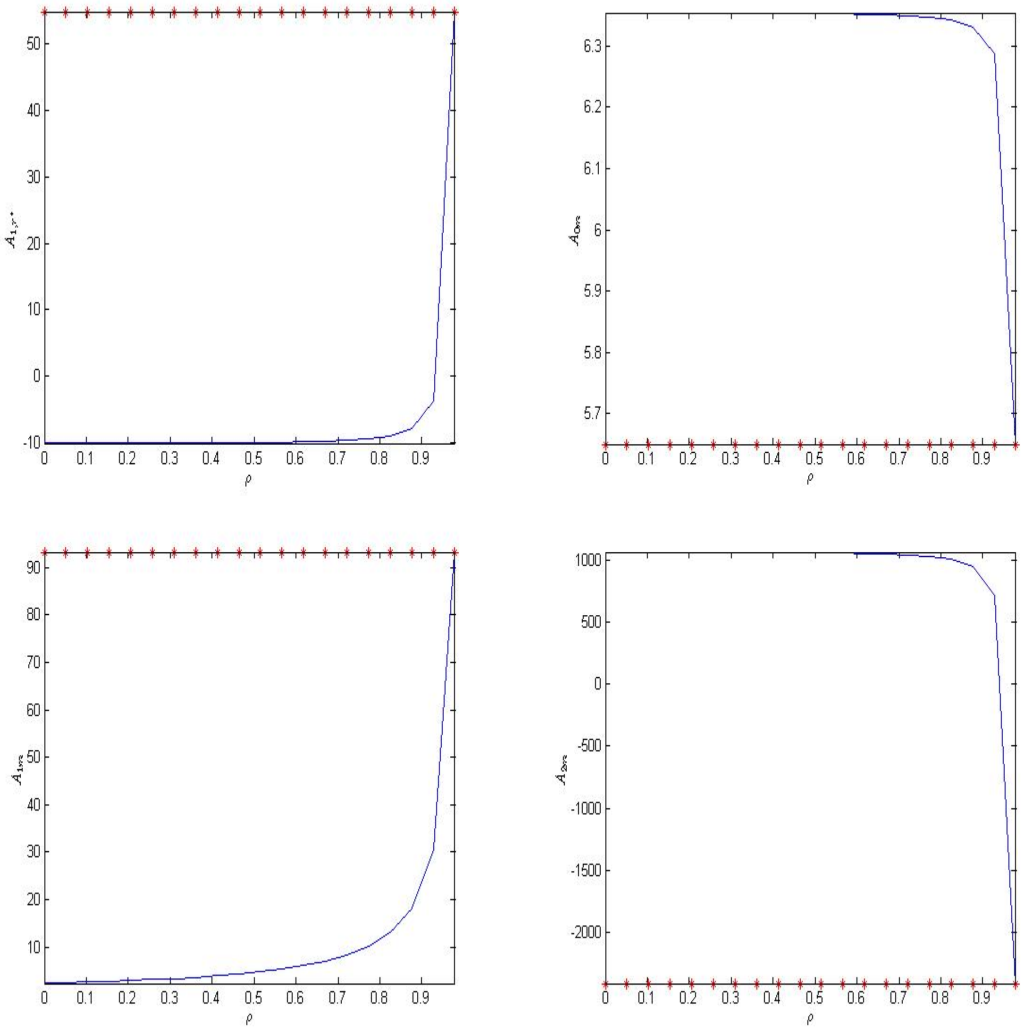

Fig. 2.3: Elasticity of loadings of expected excess returns and of dividend-price ratio to persistence of expected dividend growth

In this figure I plot the values of loadings of the expected excess returns and of the dividend-price ratio on the long run risks against persistence of expected dividend growth. The values of the loadings which correspond to changing the value of ρ from 0.000 to the original 0.979 are represented by a solid line while the values under the original calibration are indicated by a line in stars.

The analysis of Figure 2.1–Figure 2.3 reveals a high sensitivity of the parameters measuring the share of expected returns in the variance of the dividend-price ratio to a change in the share of expected dividend growth. In the BY model, a decrease in the share of expected dividend growth leads to a considerable decrease in the share of expected returns. This suggests that a large increase in ν_1 and σ_ω might be needed to compensate for the decrease in the share of expected dividend growth in the variance of the dividend-price ratio.

I switch to increasing the share of expected returns in the variance of the dividend-price ratio analyzing the effect of an increase σ_ω in Figure 2.4 and of an increase in ν_1 in Figure 2.5. Similarly to the previous figures, the solid line indicates the values of the loadings which correspond to changing the values of σ_ω or ν_1 while the line in stars, to the values under the original calibration.

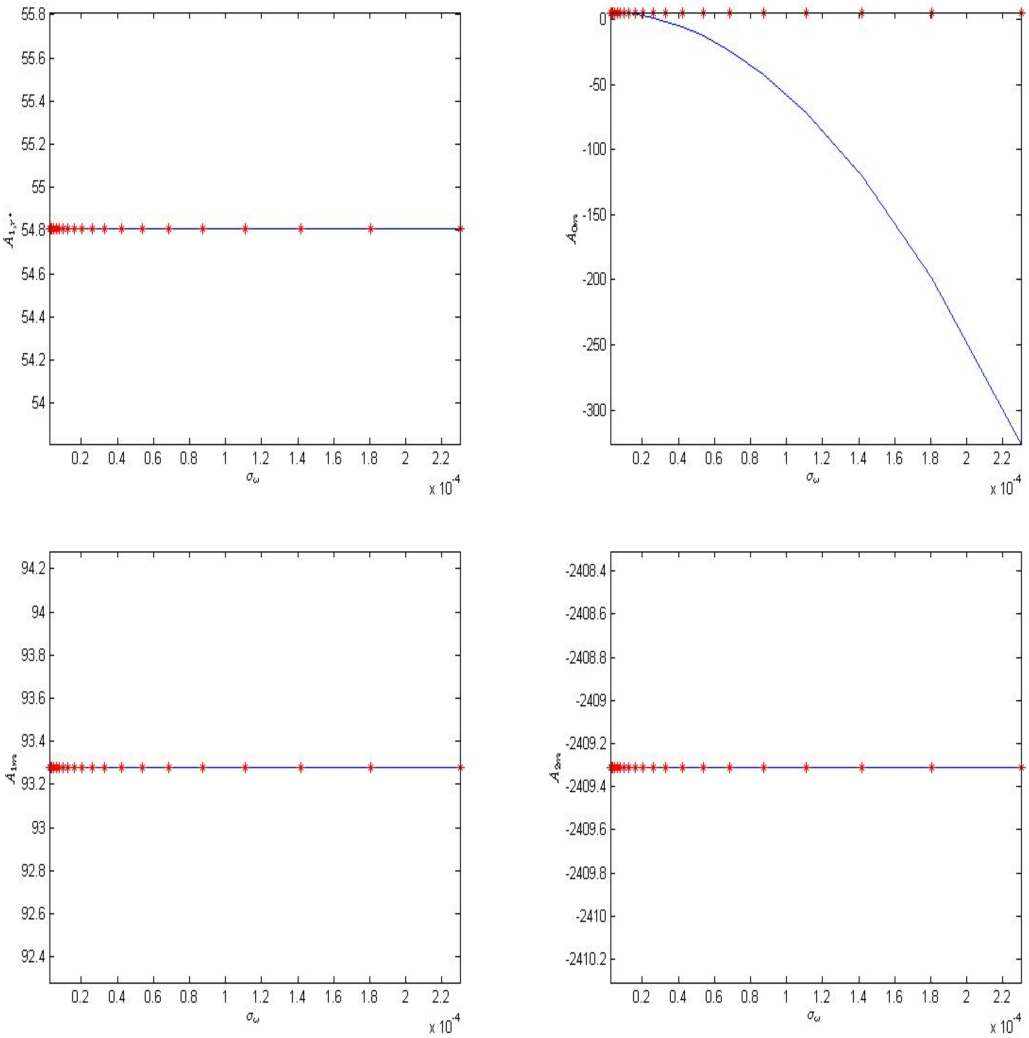

Fig. 2.4: Elasticity of loadings of expected excess returns and of dividend-price ratio to variance of innovations to time-varying volatility

In this figure I plot the values of loadings of the expected excess returns and of the dividend-price ratio on the long run risks against the variance of the innovations to the time-varying volatility. The values of the loadings which correspond to changing the value of σ_ω from the original 2.300×10^{-5} to 2.300×10^{-4} are represented by a solid line while the values under the original calibration are indicated by a line in stars.

Figure 2.4 shows that the mean of the dividend-price ratio is sensitive to the variance of the innovation to the time-varying volatility. A large increase in the variance leads to a negative and high in absolute value A_{0m}, implying that an increase in σ_ω may lead to a counterfactually high mean of the dividend price ratio. Additionally, since the mean of the dividend-price ratio determines the endogenous linearization constant k_{1m}, an excessively low or an excessively high value of A_{0m} might lead to an undefined k_{1m}. In this case the model does not have a solution.

In Figure 2.5 I plot the loadings of the expected excess returns and the dividend-price ratio on the long-run risks against the persistence of the time-varying volatility. The figure exposes a high elasticity of the share of expected returns in the variance of the dividend-price ratio to the persistence of the time-varying volatility. An increase in the persistence from the original value of 0.987 to 0.990 leads to a decrease in the value of A_{2m} from -2409.308 to -2921.314. A further increase in ν_1 leads to A_{2m} exceeding the value of -10000.000. Note that ν_1 has only a small effect on A_{0m} and does not affect A_{1m} or A_{1r^e}. This makes the persistence of the time-varying volatility a good candidate for compensating for a decrease in ρ, ϕ and φ_e.

To summarize, I found that specification of the long run risks processes leads to problems faced by the model in accommodating a decrease in the share of expected dividend growth in the variance of the dividend-price ratio. Heteroskedasticity in expected consumption growth leads to the dependence of the share of expected returns on the share of expected dividend growth. My analysis reveals that potentially large increase in σ_ω and in ν_1 is needed match the predictability patterns. An increase in these parameters, however, can lead to an excessively volatile consumption growth or to undefined linearization constants.

Finally, to get an idea of how large an increase in σ_ω and ν_1 should be and what are the consequences of that increase for the model, In Table 2.8 I recalibrate the model setting $\nu_1 = 0.9999$ and $\sigma_\omega = 3.500 \times 10^{-6}$. I intentionally keep the values of ρ, ϕ and φ_e on the original level. Since a decrease in the value of these parameters will lead to a decrease in the share of expected returns in the variance of the dividend-price ratio, keeping their original values and increasing σ_ω and ν_1 can give an idea about the ability of the model to generate required level of predictability. The assumed value of ν_1 less than one preserves stationarity of σ_t^2. I, however, tried to set ν_1 equal one. This leads to only a slight

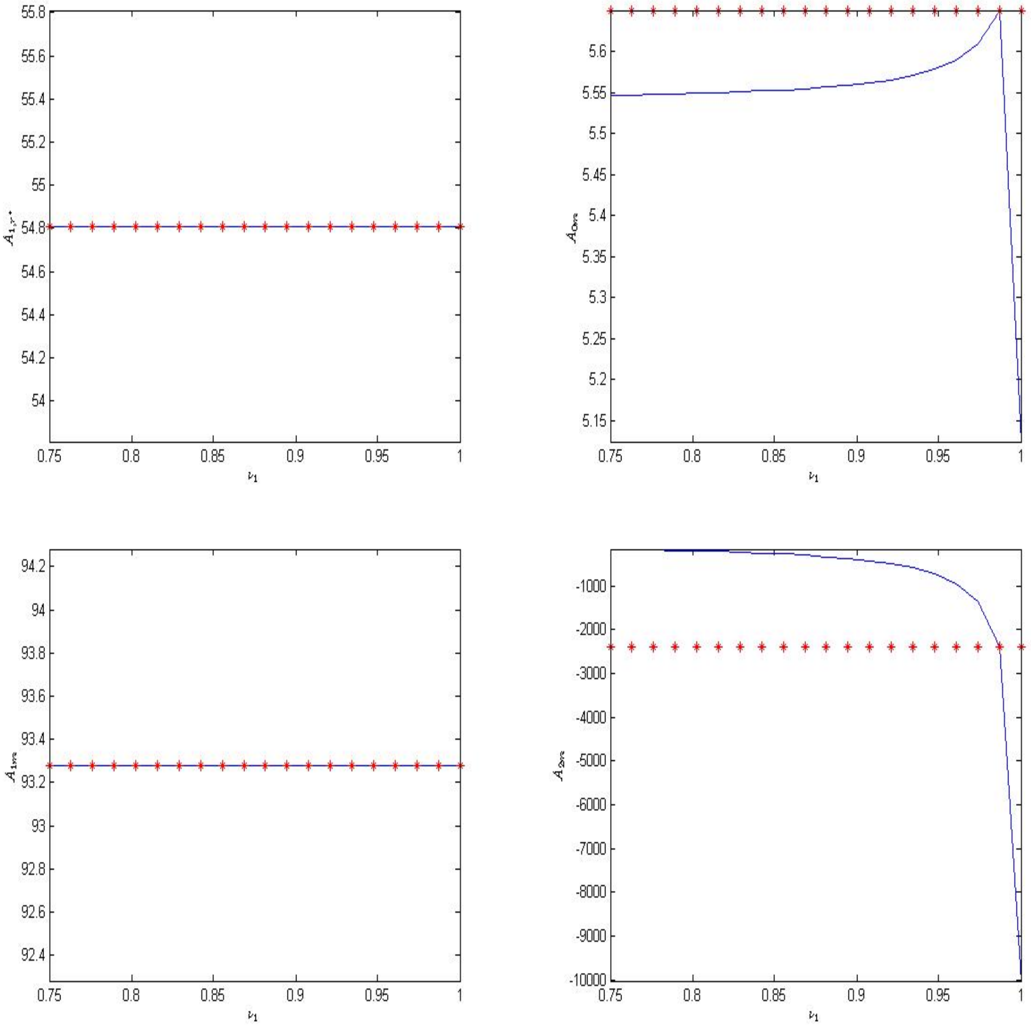

Fig. 2.5: Elasticity of loadings of expected excess returns and of dividend-price ratio to persistence of time-varying volatility

In this figure I plot the values of loadings of the expected excess returns and of the dividend-price ratio on the long run risks against persistence of the time-varying volatility. The values of the loadings which correspond to changing the value of ν_1 from 0.750 to 1.000 are represented by a solid line while the values under the original calibration are indicated by a line in stars.

improvement over the results reported in Table 2.8. Moreover, I also tried values of σ_ω above 3.500×10^{-6}. Without adjusting the mean of the time-varying volatility, in most cases this led to a problem with convergence while solving for linearization constant k_{1m}. The recalibrated in this way model has a high probability of getting a large number of negative realizations of σ_t^2. To address this issue, I also adjust $\bar{\sigma}^2$ to preserve the original mean/standard deviation ratio.

In Table 2.8 I report the results for the case when the recalibrated model produces a small number of negative σ_t^2 in *Model new (good)*, when the model produces almost all of σ_t^2 negative in *Model new (bad)* and adjusting the mean of the time-varying volatility by preserving the mean/standard deviation ratio of the original calibration in *Model new (adjusted).* For a comparison I also report the results obtained using the original calibration of the model in *Model old*, these results are identical to those reported in Table 2.2-Table 2.4.

Tab. 2.8: Increase in volatility of expected returns

This table reports the implications of the long run risks model when the persistence of the time-varying volatility (ν_1) and the variance of its innovation (σ_ω) is increased from the original value to $\nu_1 = 0.9999$ and $\sigma_\omega = 3.5000 \times 10^{-6}$. *Model new (good)* reports the results when only a small number of the realizations of the time-varying volatility process were negative. *Model new (bad)* reports the results when a large number of the realizations of the time-varying volatility process were negative. *Model new (adjusted)* reports the results when the mean of the time-varying volatility process is adjusted to preserve the mean/standard deviation of the original calibration. *Model old* reports the results obtained under the original calibration. Panel A reports the descriptive statistics. Panel B reports the implications for predictability. The numbers reported correspond to the medians of statistics taken across simulation paths. Standard deviations and R-square statistics are in %.

Panel A: Moments of the data							
Statistics		dp	r_m	r_f	$r_m - r_f$	Δd	Δc
Mean:	Model new (good)	−2.4721	0.0986	0.0135	0.0854	0.0172	0.0178
	Model new (bad)	−2.8641	0.0719	0.0251	0.0462	0.0178	0.0178
	Model new (adjusted)	−1.4004	0.2520	−0.1452	0.3939	0.0182	0.0176
	Model old	−3.0116	0.0674	0.0258	0.0413	0.0178	0.0178
Standard Deviation:	Model new (good)	26.18	18.76	1.45	18.63	12.47	3.09
	Model new (bad)	0.23	0.21	0.02	0.56	0.25	0.20
	Model new (adjusted)	63.41	65.21	5.46	64.96	51.05	12.76
	Model old	18.38	16.52	1.22	16.65	11.05	2.81
Panel B: Predictability							
Variable		dp	r_m	r_f	$r_m - r_f$	Δd	–
β	Model new (good)	0.7868	0.0513	−0.0465	0.0984	−0.2235	–
	Model new (bad)	0.6517	−0.0184	−0.0595	0.0581	−0.4187	–
	Model new (adjusted)	0.5978	−0.0404	−0.0721	0.0461	−0.5447	–
	Model old	0.6723	−0.0168	−0.0557	0.0419	−0.3739	–
t-stat	Model new (good)	10.8743	0.5972	−14.0588	1.1714	−4.8241	–
	Model new (bad)	7.5817	−0.1787	−15.5274	0.4323	−7.7244	–
	Model new (adjusted)	6.9038	−0.3672	−13.9529	0.4092	−8.4163	–
	Model old	8.2917	−0.1830	−14.3086	0.4253	−7.1629	–
R^2	Model new (good)	62.39	1.08	72.33	2.04	24.40	–
	Model new (bad)	42.45	0.88	75.01	5.66	42.76	–
	Model new (adjusted)	35.60	0.64	69.56	0.61	46.49	–
	Model old	45.08	00.65	71.06	0.63	38.88	–

In Panel A I report the descriptive statistics of the data obtained from the models. A good realization of the recalibrated model matches the moments of the data well. Comparing with the original calibration, the recalibrated model performs slightly worse in matching the means of the dividend-price ratio and market returns and performs somewhat better in matching the standard deviations of market returns, the interest rate, the excess returns and dividend growth. More importantly, Panel B shows that the performance of the model significantly improves in matching the predictability patterns, providing additional evidence in favor of the larger role of the time-varying volatility channel in the BY model. The predictive relation between returns and the dividend-price ratio now is positive and the slope coefficient and R^2 are significantly higher than under the original calibration. There is also an improvement in the predictability of the interest rate and the dividend growth. The new model, however, still produces excessive predictability of these series.

Calibrating $\nu_1 = 0.9999$ and $\sigma_\omega = 3.500 \times 10^{-6}$ it is also possible to get the results reported under *Model new (bad)*. This are the results when almost all the realizations of σ_t^2 were negative. Negative realizations of σ_t^2 are substituted by a small positive number. Not surprisingly, therefore, the model severely understates the volatilities of the data. Additionally, Panel B shows that in this case the model produces similar to the original calibration predictability patterns.

Adjusting the mean of the time-varying volatility to decrease the probability of getting negative realizations of σ_t^2 leads to excessively high volatility of the generated data. The variance of the consumption growth under the adjusted new model is far over 12 %, indicating that this new volatility process does not describe the variance of the consumption growth well. Moreover, as Panel B shows, the recalibrated model does not improve the original calibration in matching the predictability patterns observed in the data.

2.5 Conclusion

Bansal and Yaron (2004) offer the long run risks model which is capable of solving the equity premium puzzle with a reasonable level of the risk aversion and elasticity of intertemporal substitution more than one. The model is based on the Epstein and Zin (1989) preferences and on two highly persistent processes which govern the conditional mean and the conditional variance of consumption growth–the long run risks.

In this article, I have analyzed predictability patterns generated by the long run risks model. First, I conducted a Monte Carlo simulation exercise and compared predictability of dividend growth, market returns, the riskless interest rate and the excess returns by the dividend-price ratio generated by the model with the one observed in the data. Since there is evidence that dividend-price ratio is subject to structural breaks and the model does not include this channel, I also included into the analysis the dividend-price ratio adjusted for the structural breaks. I found that the model produces very different from the data predictability patterns. Contrary to the data, in the model dividend growth is more predictable than returns with the slope coefficient being negative in the both regressions. These results indicate that in the model most part of the variation in the dividend-price ratio is from expected dividend growth rather than expected returns and that the model does not produce the discount rate effect. Additionally, it produces a highly predictable by the dividend-price ratio interest rate contradicting the evidence from the data. The model is rejected even stronger if the dividend-price ratio adjusted for the structural breaks is used in the predictive regressions.

Using the state-space representation for returns of Cochrane (2008a), I connected the predictability patterns to the processes of expected returns and expected dividend growth. I found that the long run risks model implies that both processes are highly persistent, positively correlated and that the innovation to expected dividend growth is much more volatile than the innovation to expected returns. This implies that expected dividend growth is much more volatile than expected returns and leads to an excessive fraction of the variance of the dividend-price ratio attributed to expected dividend growth.

Furthermore, using the state-space representation I have connected the predictability patterns generated by the model to the specific parameters of the calibration. I found that the model ability to reproduce the predictability patterns is limited by the specification of the long run risks processes.

3. EXPECTED STOCK RETURNS AND AGGREGATED LONG RUN RISKS

Introduction

One of the most successful models proposed to solve the equity premium puzzle is the long-run risks model by Bansal and Yaron (2004) (BY). This model generates the level of the mean and the variance of the equity premium and the risk-less interest rate relying on the plausible values of the risk aversion. Other dimensions of the performance of the model, however, are a subject of a debate. Some research clam that the BY model is not consistent with the cross-sectional and time-series behavior of stock returns, that it implies too much predictability in consumption growth and that it relies on a value of the intertemporal substitution which is rejected by the data (See Beeler and Campbell (2009), Constantinides and Ghosh (2010), Bonomo, Garcia, Meddahi and Tedongap (2010)). Most of the research which tests the long-run risks model, however, is based on the annual data, leaving the question of the effect of the aggregation on the conclusions opened.

The aggregation issue in the BY model arises since the decision interval of the agent is monthly, but the data observed is annual. The problem is that aggregating consumption growth requires a nonlinear transformation of the monthly series. This creates a difficulty with interpretation of the estimation results based on the annual data in terms of the equations describing the evolution of the economy in the monthly frequency. The aggregation problem in the context of asset pricing was addressed by Hansen and Sargent (1983) and by Heaton (1995). Even more relevant for this study, Bansal, Kiku and Yaron (2007a) show that the value of an estimate of the elasticity of intertemporal substitution less than one obtained from the Generalized Method of Moments applied to the annual data is totally consistent with its value of more than one in the monthly BY economy.

Moreover, using an approximation to the aggregated consumption growth, they show that the aggregation introduces a moving average component to the consumption growth process and changes a correlation structure of the basis stochastic equations describing the model. In this paper I extend the analysis of Bansal, Kiku and Yaron (2007a) to the effect the aggregation has on predictability of returns in the long-run risks model.

Aggregating dividend and consumption growth requires taking a natural log of the averages of dividend/consumption growth over the last two years in levels. Bansal, Kiku and Yaron (2007a) approximate the log of a sum by a sum of logs. Such an approximation leads to an excessively volatile series. Instead, I approximate the log of a sum by regressing it on the sum of logs. Simulation results show that this approximation works very well–the resulting approximated series is almost indistinguishable from the actual one. Using this approximation I connect the aggregated expected dividend/consumption growth with the monthly economy. As the result I find that a part of expected dividend/consumption growth is observable. The observable part is completely characterized by the realized monthly dividend/consumption growth. Differently from dividend growth, aggregation of returns is done through a simple summation of the monthly series. Thus, the aggregated expected returns are fully characterized by the latent long run risks processes. This implies that the aggregation breaks the tight link between expected dividend growth, expected returns and expected consumption growth which is a characteristic of the monthly economy. Moreover, confirming Bansal, Kiku and Yaron (2007a), my analysis exposes problems with the usual approaches to estimation of the elasticity of intertemporal substitution.

To characterize the effect of the aggregation on predictability of returns, I proceed by deriving the state-space representation for returns implied by the aggregated economy. Comparing with the monthly state-space representation I find that the aggregation leads to an additional state variable and a moving average component for expected dividend growth process. Moreover, a correlation structure of the errors also changes. Differently to the monthly frequency, the innovations to both, expected returns and expected dividend growth are contemporaneously correlated with the innovations to the realized dividend growth. This implies additional to the dividend-price ratio variables might help to predict returns and exposes some possible statistical difficulties faced by the predictive regression. Finally, using the calibration of BY, I find that in both, the monthly and the annual frequencies, most part of the variation in the dividend-price ratio in the model stems from expected dividend growth and that the dividend-price ratio tracks the variation in the poor economic growth risk component in expected returns and expected dividend growth.

The paper proceeds as follows. Section 1 summarizes the main features of the long-run risks model. Section 2 describes the aggregation procedure. Section 3 presents the analysis of the effect of the aggregation on expected returns and expected dividend growth. Section 4 analyses the effect of the aggregation on predictability of returns. Section 5 concludes.

3.1 The long run risks model specification

Denoting Δc_{t+1} and Δd_{t+1} as log-consumption and log-dividend growth, the BY model is described by the following dynamics of the monthly series:

$$\Delta c_{t+1} = m + x_t + \sigma_t \eta_{t+1}, \tag{3.1.1}$$

$$\Delta d_{t+1} = m_d + \phi x_t + \varphi_d \sigma_t u_{d,t+1}, \tag{3.1.2}$$

$$x_{t+1} = \rho x_t + \varphi_e \sigma_t e_{t+1}, \tag{3.1.3}$$

$$\sigma^2_{t+1} = \bar{\sigma}^2(1-\nu_1) + \nu_1 \sigma^2_t + \sigma_\omega \omega_{t+1}. \tag{3.1.4}$$

From Eq.(3.1.1), $m+x_t$ equals to the conditional mean and σ^2_t equals to the conditional variance of consumption growth. These processes in the BY economy are highly persistent and, therefore, are the sources of the long run risks: the risk of poor economic growth and the risk of high economic uncertainty. Dividend growth has a leveraged exposure to the long run risks. Its exposure to the poor economic growth risk is governed by the parameter ϕ and to the high economic uncertainty risk by the parameter φ_d.

Each month a representative agent chooses her consumption stream by maximizing the Epstein and Zin (1989) utility function given the feasibility and non-negativity of the consumption choice. Recall that Epstein and Zin (1989) utility function leads to the following Euler Equation:

$$E_t\left[exp\left(\theta log\beta - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1)r_{c,t+1} + r_{i,t+1}\right)\right] = 1, \tag{3.1.5}$$

with $r_{c,t+1}$ being log-returns on the claim to the aggregate consumption at time $t+1$ and $r_{i,t+1}$ being log return on any other asset at time $t+1$. Moreover, β is the subjective time discount factor and θ is related to the relative risk aversion γ and the elasticity of intertemporal substitution ψ through: $\theta = \dfrac{1-\gamma}{1-\frac{1}{\psi}}$.

To solve the model, BY use the Campbell and Shiller (1988) linearization. Denoting the consumption-price ratio at time t as cp_t, the linearization implies the following expression for returns on the claim to aggregate consumption:

$$r_{c,t+1} = k_0 - k_1 cp_{t+1} + cp_t + \Delta c_{t+1}. \tag{3.1.6}$$

$k_1 = \dfrac{e^{-\bar{cp}}}{1+e^{-\bar{cp}}}$ and $k_0 = -(1-k_1)log(1-k_1) - k_1 log(k_1)$ are linearization constants which both, depend and determine the mean of the consumption-price ratio $\bar{cp}_t$.

BY show that the solution for the consumption-price ratio in the monthly long run risks economy takes the following form:

$$cp_t = -A_0 - A_1 x_t - A_2\sigma_t^2. \tag{3.1.7}$$

The exact expressions for the loadings A_0, A_1 and A_2 are provided by many authors including Beeler and Campbell (2009), Bansal, Kiku and Yaron (2007a) and Constantinides and Ghosh (2009). A_0 together with $A_2\bar{\sigma}^2(1-\nu_1)$ determine the mean of the consumption-price ratio and is dependent on the parameters which govern the unconditional means of the processes described by Eq.(3.1.1)-Eq.(3.1.4). A_1 summarizes a dependence of returns on the claim to aggregate consumption and consumption growth on x_t while A_2 reflects heteroskedasticity of expected and realized consumption growth.

The long run risks economy is characterized by the following innovation to the pricing kernel:

$$m_{t+1} - E_t\left[m_{t+1}\right] = -\gamma\sigma_t\eta_{t+1} - (1-\theta)k_1 A_1\varphi_e\sigma_t e_{t+1} - (1-\theta)k_1 A_2\omega_{t+1}. \tag{3.1.8}$$

From Eq.(3.1.8), the innovation to the pricing kernel has three independent components stemming from the innovation to the realized consumption growth–the short run risk, from the innovation to expected consumption growth– the economic growth risk and from the innovation to the time-varying volatility– the economic uncertainty risk. Since the long run risks affect the innovation to the pricing kernel, they are the sources of the risk premium.

3.2 Aggregation

Most of the papers analyze the BY model using the monthly equations. An exception is Bansal, Kiku and Yaron (2007a) who claim that aggregation is responsible for a downward bias in estimates of the intertemporal elasticity of substitution resulting from a GMM. They show that the aggregation procedure employed in the BY model introduces an MA component into consumption growth and, consequently to the pricing kernel. Moreover, the aggregation changes the correlation structure of the errors in Eq.(3.1.1)-Eq.(3.1.4).

To be more specific, aggregation dividends and consumption in the BY model involves taking averages of the levels of the monthly series. Using a superscript a for the annualized series, a subscript *12t* to indicate that the annualized variable is measured over 12 month and large letters for the levels, consumption is aggregated through the following expression:

$$\Delta c^a_{12(t+1)} \equiv log\frac{C^a_{12(t+1)}}{C^a_{12t}} = log\frac{\sum_{j=0}^{11} C_{12(t+1)-j}}{\sum_{j=0}^{11} C_{12t-j}} \tag{3.2.1}$$

By approximating $\Delta c^a_{12(t+1)}$ as:

$$\Delta c^a_{12(t+1)} \approx \sum_{j=0}^{22} \chi_j \Delta c_{12(t+1)-j}, \tag{3.2.2}$$

with:

$$\chi_j = \begin{cases} \dfrac{j+1}{12}, j < 12 \\ \dfrac{23-j}{12}, j \geq 12, \end{cases}$$

Bansal, Kiku and Yaron (2007a) show that the aggregated consumption is related to the monthly expected consumption growth through:

$$\Delta c^a_{12(t+1)} \approx \phi^a_c m + \phi^a_c x^a_{12(t+1)-23} + \eta^a_{12(t+1)}, \tag{3.2.3}$$

with:

$$\phi^a_c = \sum_{j=0}^{22} \chi_j \rho^j,$$

$$\eta^a_{12(t+1)} = \sum_{j=0}^{22} \chi_j \left[\sigma_{12(t+1)-j-1} \eta_{12(t+1)-j} + \varphi_e \sum_{i=1}^{22-j} \rho^{i-1} \sigma_{12(t+1)-j-1-i} e_{12(t+1)-j-i} \right]$$

According to Eq.(3.2.3) the aggregation introduces an MA structure to the consumption growth process. Note additionally, that while the errors to expected consumption growth were independent from the errors to the realized consumption growth, it is not the case after the aggregation. Consequently, Bansal, Kiku and Yaron (2007a) model the aggregated consumption growth as:

$$\Delta c_{t+1} = m + x_t + \alpha \sigma_{t-1} \eta_t + \sigma_t \eta_{t+1}.$$

Moreover, from Eq.(3.2.3), the new long run risks components coincide with the December counterpart, so that the aggregated risk of poor economic growth is described by:

$$x_{12t} = \rho^{12} x_{12(t-1)} + \varphi_e \sum_{j=0}^{11} \rho^j \sigma_{12(t+1)-j-1} e_{12(t+1)-j}, \tag{3.2.4}$$

while the aggregated risk of high economic uncertainty follows:

$$\sigma^2_{12t} = \bar{\sigma}^2 (1 - \nu_1^{12}) + \nu_1^{12} \sigma^2_{12(t-1)} + \sigma_\omega \sum_{j=0}^{11} \omega_{12(t+1)-j} \tag{3.2.5}$$

The innovation to the pricing kernel after the aggregation is:

$$\begin{gathered} m_{t+1} - E_t\left[m_{t+1}\right] = -\frac{\alpha}{\psi}\sigma_{t-1}\eta_t - \left((\gamma - \frac{1}{\psi})k_1\alpha + \gamma\right)\sigma_t\eta_{t+1} - \\ -(1-\theta)k_1 A_1 \varphi_e e_{t+1} - (1-\theta)k_1 A_2 \sigma_\omega \omega_{t+1}. \end{gathered} \tag{3.2.6}$$

Comparing with Eq.(3.1.8), the aggregation procedure employed in the BY model introduces a lagged error of the realized consumption growth into the innovation to the pricing kernel. This additional term and the correlation between the errors to expected and to the realized consumption growth are new sources of the volatility of the pricing kernel and the risk premium.

In this paper I extend the analysis of Bansal, Kiku and Yaron (2007a) to predictability generated by the long-run risks model. To do that I modify the approximation given by Eq.(3.2.2). The new approximation works well and allows a derivation of the closed form solutions for dividend growth in terms of monthly processes for the long run risks.

Similarly to consumption growth, the annualized dividend growth is obtained from:

$$\Delta d^a_{12(t+1)} = log \frac{\sum_{j=0}^{11} D_{12(t+1)-j}}{\sum_{j=0}^{11} D_{12t-j}}. \tag{3.2.7}$$

In the Appendix, it is shown that Eq.(3.2.7) can be approximated as:

$$\Delta d^a_{12(t+1)} = b_1 \sum_{j=0}^{11} (j+1)\Delta d_{12(t+1)-j} + b_2 \sum_{j=0}^{10} (11-j)\Delta d_{12t-j}. \tag{3.2.8}$$

with b_1 and b_2 being picked up by a regression.

In Table 3.1 I estimate the coefficients of Eq.(3.2.8) and a similar equation for consumption growth. I report the means and the medians of the finite sample distribution of the statistics obtained from simulating 80 annual observations 1000 times and the population estimates obtained from a single simulation of 5000 annual observations. Note that the standard errors are GMM-adjusted using the codes downloaded from Prof. John H. Cochrane web page.

Tab. 3.1: Approximation

Panel A of this table provides the results of the regression:

$$\Delta d^a_{12(t+1)} = c_0 + c_1 \sum_{j=0}^{11} (j+1)\Delta c_{12(t+1)-j} + c_2 \sum_{j=0}^{10} (11-j)\Delta c_{12t-j} + \epsilon^c_{12(t+1)}. \ ,$$

while Panel B, that of:

$$\Delta d^a_{12(t+1)} = b_0 + b_1 \sum_{j=0}^{11} (j+1)\Delta d_{12(t+1)-j} + b_2 \sum_{j=0}^{10} (11-j)\Delta d_{12t-j} + \epsilon^d_{12(t+1)}.$$

Population values are calculated from a single simulation of 5000 annual observations. The minimum, the maximum, the mean and the median of the finite sample distribution of the statistics are calculated based on 80 annual observations generated 1000 times. Standard errors are GMM-adjusted to account for heteroskedasticity. t-statistics are reported in the round brackets. R-squared statistics are reported in %.

	Pop. Value	Min	Max	Mean	Median
Panel A: Consumption growth					
$\hat{c_0}$	0.0000	−0.0000	0.0003	0.0000	0.0000
	(0.5110)	(−1.4631)	(2.0082)	(0.0335)	(0.0094)
$\hat{c_1}$	0.0835	0.0824	0.0845	0.0835	0.0835
	(83.96×10^2)	(12.23×10)	(73.37×10)	(33.61×10)	(32.99×10)
$\hat{c_2}$	0.0831	0.0820	0.0847	0.0831	0.0831
	(67.57×10^2)	(99.98)	(62.05×10)	(27.79×10)	(27.06×10)
R^2	100.00	99.90	99.99	99.97	99.99
Panel B: Dividend growth					
$\hat{b_0}$	−0.0000	−0.0003	0.0004	0.0000	0.0000
	(−0.0081)	(−1.6049)	(2.8780)	(0.0182)	(0.0078)
$\hat{b_1}$	0.0835	0.0821	0.0847	0.0835	0.0835
	(20.75×10^2)	(12.30×10)	(66.48×10)	(30.63×10)	(30.01×10)
$\hat{b_2}$	0.0831	0.0815	0.0847	0.0831	0.0831
	(17.02×10^2)	(92.99)	(62.07×10)	(25.38×10)	(24.83×10)
R^2	99.97	99.90	99.99	99.97	99.97

The constant term in the regression equals to zero and statistically insignificant for the population, the mean and the median of the finite sample distribution. The slope coefficients for both of the elements of the approximation are highly statistically significant with the values having low dispersion across the simulation paths. Moreover, the means and the medians of the slope coefficients are equal to the population values. Importantly, the R^2 close to one indicates that the two elements of the decomposition capture virtually all the time-variation of the aggregated dividend growth and of the aggregated consumption growth. Note that the approximation used in Bansal, Kiku and Yaron (2007a) is obtained by restricting b_1 and b_2 equal to one. My simulation results show that restricting the coefficients to one leads to an excessively volatile approximated dividend growth process, with the standard deviation equal to 137.64 % v.s. the standard deviation of 11.48 % of the actual dividend growth.

3.3 Expected returns and expected dividend growth in the aggregated economy

In the monthly frequency expected returns and expected dividend growth are related to expected consumption growth and the time-varying volatility through:

$$\mu_t^r = a^{\mu r} + \frac{1}{\psi}x_t - (1 - k_{1m}\nu_1)A_{2m}\sigma_t^2; \tag{3.3.1}$$

$$\mu_t^d = m_d + \phi x_t; \tag{3.3.2}$$

$$\Delta d_t = \mu_{t-1}^d + \epsilon_{t+1}^d, \tag{3.3.3}$$

with $a^{\mu,r} = k_{0m} + k_{1m}(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)) - A_{0m} + m_d$ and A_{2m} being a loading of the dividend-price ratio on the time-varying volatility. Eq.(3.3.1) lies behind a commonly used procedure for estimating the elasticity of intertemporal substitution. If the time-varying volatility does not vary much, the elasticity of intertemporal substitution can be recovered from a regression of returns or the interest rate on consumption growth[1]. Note that while expected returns are related to the both long run risks, expected dividend growth is determined only by expected consumption growth. Additionally, since the innovations

[1]Such a regression may suffer from the endogeneity problem. Therefore, an IV estimation is usually used

to x_t are independent from the innovations to Δd_t, the innovations to expected dividend growth are uncorrelated with the innovations to the realized dividend growth.

Expected annualized dividend growth is related to the monthly series through:

$$\mu_{12t}^{d,a} = 78b_1 m_d + b_1 \sum_{j=0}^{11} (j+1)\rho^{11-j}\phi x_{12t} + b_2\phi \sum_{j=0}^{10} (11-j)\Delta d_{12t-j}. \tag{3.3.4}$$

Similarly, expected annualized consumption growth is:

$$\mu_{12t}^{c,a} = 78c_1 m + c_1 \sum_{j=0}^{11} (j+1)\rho^{11-j} x_{12t} + c_2 \sum_{j=0}^{10} (11-j)\Delta c_{12t-j}. \tag{3.3.5}$$

Part of expected dividend growth is observable. From Eq.(3.3.4) expected aggregated dividend growth consists of a latent x_{12t} and a weighted sum of the observable realized monthly dividend growth series. Eq.(3.3.5) shows that expected consumption growth has a similar to expected dividend growth form, consisting of a part due to x_{12t} and a part stemming from the realized monthly consumption growth. This implies that the aggregation breaks a tight link between expected dividend growth and expected consumption growth. Each of the processes is affected by an independent component given by the innovations to the realized dividend growth and to the realized consumption growth. To elaborate, in Panel A of Table 3.2 I decompose the variation in expected dividend growth and in expected consumption growth into the components related to the aggregated state variable x_{12t} and to the monthly realized dividend/consumption growth. In Panel B, in addition, I calculate the correlations of each of the components with expected dividend growth and expected consumption growth.

Panel A of Table 3.2 shows that x_{12t} accounts for only 13 % of the variation in expected dividend growth and only 18 % of the variation in expected consumption growth. The other component accounts for over 67 % of the variance of expected dividend growth and over 51 % of the variance of expected consumption growth. Moreover, Panel B of Table 3.2 shows that the correlations of expected dividend growth and expected consumption growth with the other component is over 90 %. This evidence implies that x_{12t} has a

Tab. 3.2: Variation in expected consumption/dividend growth

This table reports the fractions of the variance of expected dividend growth and expected consumption growth attributed to the aggregated long-run risk x_{12t} and to the weighted sum of the monthly realized dividend growth over the previous year in the case of expected dividend growth or to the weighted sum of the monthly realized consumption growth over the previous year in the case of expected consumption growth, in Panel A. In Panel B the table reports correlations of expected dividend growth and expected consumption growth with the components. All the values are in %.

Series	x_{12t}	other component
Panel A: Fractions of the variance		
μ_t^d	12.81	67.31
μ_t^c	18.22	51.31
Panel B: Correlations		
μ_t^d	63.56	94.16
μ_t^c	78.37	92.90

somewhat different from x_t interpretation. Though a small predictable component x_{12t} does capture some part of the variation in expected consumption growth it does not completely describe dynamics of expected consumption growth.

The expression relating expected aggregated returns to the monthly economy is:

$$\begin{aligned}\mu_{12t}^{r,a} &= 12a^{\mu r} - (1 - k_{1m}\nu_1)A_{2m}\bar{\sigma}^2 \frac{12\nu_1 - \nu_1^{12} - 11}{\nu_1 - 1} + \\ &+ \frac{\rho^{12} - 1}{\psi(\rho - 1)} x_{12t} - (1 - k_{1m}\nu_1)A_{2m}\frac{\nu_1^{12} - 1}{(\nu_1 - 1)}\sigma_{12t}^2\end{aligned} \tag{3.3.6}$$

Differently to dividend growth, returns are aggregated by a simple summation of the monthly series. Consequently, expected returns in the monthly and in the annual frequency are completely characterized by the monthly long run risks processes. Note that Eq.(3.3.6) exposes problems of the common approach to estimation of the elasticity of intertemporal substitution. Specifically, the aggregation implies that expected returns are related to expected consumption growth only indirectly, through x_{12t}. Moreover, even if x_{12t} were perfectly correlated with expected consumption growth, expected returns have a loading of $\frac{\rho^{12} - 1}{\psi(\rho - 1)}$ on x_{12t}. Therefore, Eq.(3.3.6) confirms the evidence for the bias in the estimate of the intertemporal substitution which arises from the aggregation provided in Bansal, Kiku and Yaron (2007a).

Cochrane (2008a) shows that the state-space representation–a system which includes the equations describing the dynamics of expected returns, expected dividend growth and the innovation to the realized dividend growth is sufficient to fully recover predictive relations between returns and the dividend-price ratio. Using dynamics of the state variables given by Eq.(3.1.3) and Eq.(3.1.4) for the monthly series and by Eq.(3.2.4) and Eq.(3.2.5) for the annual series in the appendix I derive the state-space representation for returns for the monthly and for the aggregated economy.

In the monthly frequency, the state-space representation is:

$$\mu^r_{t+1} = \alpha^{\mu r} + \delta^{\mu r}\mu^r_t + \tau\mu^d_t + \epsilon^{\mu r}_{t+1}; \tag{3.3.7}$$

$$\mu^d_{t+1} = \alpha^{\mu d} + \delta^{\mu d}\mu^d_t + \epsilon^{\mu d}_{t+1}; \tag{3.3.8}$$

$$\Delta d_{t+1} = \mu^d_t + \epsilon^d_{t+1}. \tag{3.3.9}$$

While in the aggregated economy the state space-representation is described by the following system:

$$\mu^{r,a}_{12(t+1)} = \alpha^{\mu r,a} + \delta^{\mu r,a}\mu^{r,a}_{12t} + \tau^a x^a_{12t} + \epsilon^{\mu r,a}_{12(t+1)}; \tag{3.3.10}$$

$$\mu^{d,a}_{12(t+1)} = \alpha^{\mu d,a} + \delta^{\mu d,a}\mu^{d,a}_{12t} + \vartheta^a \epsilon^X_{12t} + \epsilon^{\mu d,a}_{12(t+1)}; \tag{3.3.11}$$

$$x^a_{12(t+1)} = \delta^{x,a} x^a_{12t} + \epsilon^{x,a}_{12(t+1)}; \tag{3.3.12}$$

$$\Delta d^a_{12(t+1)} = \mu^{d,a}_{12t} + \epsilon^{d,a}_{12(t+1)}. \tag{3.3.13}$$

A perfectly correlated expected consumption growth and expected dividend growth implies that lagged expected dividend growth affects expected returns in the monthly frequency. Since the aggregation weakens the relation between expected dividend growth with the long-run risks process, an additional state variable–x_{12t}, imperfectly correlated with expected dividend growth affects expected returns. Moreover, while in the monthly frequency expected dividend growth follows an AR(1) specification, expected aggregated dividend growth has an MA component.

Bansal, Kiku and Yaron (2007a) show that aggregation affects a correlation structure of errors in the economy. I present the correlation structure of the errors of the monthly and the annual state-space representation in Table 3.3.

Tab. 3.3: State-space representation for returns: correlation structure

This table reports the variance-covariance matrix of the state-space representation for returns for the monthly economy in Panel A and for the annualized economy in Panel B. The values on the diagonal are the standard deviations while those on the off-diagonal are the correlations. The values are computed from a single simulation of 60000 monthly or 5000 annual observations. All the values are in %.

Panel A: Monthly economy					
Errors	$\epsilon^{\mu r}$	$\epsilon^{\mu d}$	ϵ^{d}		
$\epsilon^{\mu r}$	00.00	95.22	00.00		
$\epsilon^{\mu d}$	95.22	00.10	00.00		
ϵ^{d}	00.00	00.00	03.51		
Panel B: Aggregated economy					
Errors	$\epsilon^{\mu r,a}$	$\epsilon^{\mu d,a}$	$\epsilon^{d,a}$	$\epsilon^{x,a}$	ϵ^{X}
$\epsilon^{\mu r,a}$	00.84	37.98	08.15	91.65	11.85
$\epsilon^{\mu d,a}$	37.98	07.27	49.08	41.68	−84.88
$\epsilon^{d,a}$	08.15	49.08	07.49	08.78	−48.43
$\epsilon^{x,a}$	91.65	41.68	08.78	00.11	12.66
ϵ^{X}	11.85	−84.88	−48.43	12.66	05.17

I start from the monthly system. From Panel A of Table 3.3 innovations to expected returns in the BY model are characterized by a very low volatility. Their standard deviation is lower than four digits after the decimal point reported in the table. The volatility of the innovations to expected dividend growth, on the other hand, is very high. Their standard deviation exceeds the standard deviation of the innovations to expected returns by many times. Moreover, the innovations to expected returns and expected dividend growth are highly positively correlated and both are uncorrelated with innovations to realized dividend growth.

After aggregation, two additional shocks appear in the variance-covariance matrix of the state-space representation. The first is due to the additional state variable x_{12t}, the second is from the MA structure of expected aggregated dividend growth. Similarly to the monthly frequency case, the volatility of the innovations to expected returns is low and the volatility of the innovations to expected dividend growth is high. The corresponding numbers from Panel B of Table 3.3 show that the standard deviation of the innovations to expected dividend growth exceeds the standard deviation of the innovations to expected returns by a factor of more than eight. The correlation between the innovations to expected returns and expected dividend growth, however, is much lower than in the monthly frequency, being less than 40 % against over 95 % for the monthly frequency. Moreover, innovations to the realized dividend growth are now correlated with the innovations to expected dividend growth. The correlation between the innovations, according to Panel B of the table is 49 %. In the monthly frequency, the innovations to expected dividend growth were perfectly correlated with the risk of poor economic growth given by ϵ_t^x. After aggregation, this correlation is lower than 42 %. Note that the innovations to expected aggregated returns are highly correlated with the innovations to the long run risks process x in both the monthly and the annual frequency.

I devote the next section to the predictive relations implied by the aggregated BY economy.

3.4 Aggregation and predictability of returns in the long run risks model

Predicting returns using a regression frequently involves examining dynamics of the explanatory variable. This approach is justified since, as shown in Stambaugh (1999), a highly persistent regressor may lead to a bias and incorrect inference in a predictive regression for returns. Additionally, following Cochrane (2008b) it is useful to analyze predictability of returns jointly with predictability of dividend growth. Therefore, in this section I analyze the following VAR:

$$dp_{t+1} = \alpha^{dp} + \beta^{dp} dp_t + \epsilon_{t+1}^{dp}; \tag{3.4.1}$$

$$r_{m,t+1} = \alpha^{r} + \beta^{r} dp_t + \epsilon_{t+1}^{r}; \tag{3.4.2}$$

$$\Delta d_{t+1} = \alpha^{d} + \beta^{d} dp_t + \epsilon_{t+1}^{d}. \tag{3.4.3}$$

Running a predictive regression for returns, we estimate expected returns using information contained in the regressor. To examine how good the dividend-price ratio is a proxy for expected returns, in the appendix I derive the following expressions relating the dividend-price ratio to expected returns in the monthly and in the aggregated economy:

$$dp_t = B_0 + \frac{\mu_t^r}{1 - k_{1m}\delta^{\mu r}} - \frac{(1 - (\tau + \delta^{\mu r})k_{1m})\mu_t^d}{(1 - k_{1m}\delta^{\mu r})(1 - k_{1m}\delta^{\mu d})} \tag{3.4.4}$$

and

$$dp^a_{12t} = B^a_0 + \frac{\mu^{r,a}_t}{1-k^a_{1m}\delta^{\mu r,a}} - \frac{\mu^{d,a}_t}{1-k^a_{1m}\delta^{\mu d,a}} + \frac{\tau^a k^a_{1m} x^a_{12t}}{(1-k^a_{1m}\delta^{\mu r,a})(1-k^a_{1m}\delta^{\mu d,a})} - \\ - \frac{\vartheta^a \epsilon^X_{12t}}{\delta^{\mu d,a}(1-k^a_{1m}\delta^{\mu d,a})} \tag{3.4.5}$$

In the monthly frequency, the variation in the dividend-price ratio reflects the variation in both expected returns and expected dividend growth. Eq.(3.4.5) shows that, after aggregation, additionally to expected returns and expected dividend growth, variation in the dividend-price ratio contains information about the state variable affecting expected returns–x_{12t} and about the lagged innovation to expected dividend growth–ϵ^X_{12t}. In both, the monthly and the annual frequency, the dividend-price ratio is a noisy proxy for expected returns since it contains information about the variation in the other variables. Because two more latent processes affect dividend-price ratio, controlling for the noise is more difficult in the annual frequency.

The state-space representation allows to obtain predictive relations implied by the model. Specifically, the following dynamics of the dividend-price ratio characterizes the monthly and the annual frequencies:

$$dp_{t+1} = a^{dp} + \delta^{\mu r} dp_t + \frac{\tau + \delta^{\mu r} - \delta^{\mu d}}{1-k_{1m}\delta^{\mu d}} + \epsilon^{dp}_{t+1}; \tag{3.4.6}$$

$$dp^a_{12(t+1)} = a^{dp,a} + \delta^{\mu r,a} dp^a_{12t} + \frac{\delta^{\mu r,a} - \delta^{\mu d,a}}{1-k^a_{1m}\delta^{\mu d,a}} \mu^{d,a}_{12t} + \left[\frac{\tau^a}{1-k^a_{1m}\delta^{x,a}}\right] x^a_{12t} + \\ + \frac{\vartheta^a}{1-k^a_{1m}\delta^{\mu d}} \left(\frac{\delta^{\mu r,a}}{\delta^{\mu d,a}} - 1\right) \epsilon^X_{12t} + \epsilon^{dp,a}_{12(t+1)}. \tag{3.4.7}$$

Eq.(3.4.6) and Eq.(3.4.7) show that the dividend-price ratio follows a more complicated than AR(1) dynamics. Additionally to its lagged value, expected dividend growth enters the implied monthly relation while expected dividend growth, the long run risk process–x_{12t} and the MA component of expected dividend growth enter the implied annualized relation. If the dividend-price ratio is contemporaneously correlated with any of the additional variables, fitting Eq.(3.4.1) will result in an inconsistent estimate of the slope coefficient. This inconsistency can be important for applying the Stambaugh (1999) correction to the predictive regression for returns since it accounts only for a finite sample bias and ignores a possible bias from the additional variables.

The implied predictive relation between returns and the dividend-price ratio in the monthly and the aggregated economies are:

$$r_{m,t+1} = a^r + (1 - k_{1m}\delta^{\mu r})dp_t + \left[1 - k_{1m}\left(\frac{\tau}{1 - k_{1m}\delta^{\mu d}} - \frac{(\delta^{\mu d} - \delta^{\mu r})}{1 - k_{1m}\delta^{\mu d}}\right)\right]\mu_t^d + \epsilon_{t+1}^r. \tag{3.4.8}$$

$$\begin{aligned} r^a_{m,12(t+1)} = a^{r,a} + (1 - k^a_{1m}\delta^{\mu r,a})dp^a_{12t} + \left[1 - \frac{k^a_{1m}(\delta^{\mu r,a} - \delta^{\mu d,a})}{1 - k^a_{1m}\delta^{\mu d,a}}\right]\mu^{d,a}_{12t} - \frac{k^a_{1m}\tau^a}{1 - k^a_{1m}\delta^{x,a}}x^a_{12t} - \\ - \frac{k^a_{1m}\vartheta^a}{1 - k^a_{1m}\delta^{\mu d,a}}\left(\frac{\delta^{\mu r,a}}{\delta^{\mu d,a}} - 1\right)\epsilon^X_{12t} + \epsilon^{r,a}_{12(t+1)}. \end{aligned} \tag{3.4.9}$$

Following a version of Eq.(3.4.8) which is based on unequal persistence of expected returns and expected dividend growth, some authors have shown that it is important to control for variation in expected dividend growth while predicting returns[2]. According to Eq.(3.4.9), two additional variables–the aggregated long run risk process and the MA component of expected dividend growth may be important for predicting returns. Moreover, an exclusion of the relevant variables from the predictive regression may lead to a difficulty with tests for predictability of returns by the dividend-price ratio. The omitted variable problem can also be viewed from a slightly different angle. Additional variables affect an interpretation of the slope coefficient resulting from the predictive regression. Using the monthly state-space representation, the slope coefficient in Eq.(3.4.2) incorporates information about the correlation between the dividend-price ratio with expected dividend growth. Using the annual state-space representation, however, the slope coefficient is also affected by the correlation of the dividend-price ratio with the additional state variable and with the MA component of expected dividend growth.

Finally, the implied predictive relation between dividend growth and the dividend-

[2] See Binsbergen and Koijen (2010) and Ferreira and Santa-Clara (2011).

price ratio is:

$$\begin{aligned}\Delta d_{t+1} = B_0\frac{(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})}{(1-(\tau+\delta^{\mu r})k_{1m})} + \frac{(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})}{(1-(\tau+\delta^{\mu r})k_{1m})}dp_t - \\ -\frac{1-k_{1m}\delta^{\mu d}}{(1-(\tau+\delta^{\mu r})k_{1m})}\mu_t^r + \epsilon_{t+1}^d;\end{aligned} \tag{3.4.10}$$

and

$$\begin{aligned}\Delta d_{12(t+1)} = B_0^a(1-k_{1m}^a\delta^{\mu d,a}) + (1-k_{1m}^a\delta^{\mu d,a})dp_{12t}^a - \frac{1-k_{1m}^a\delta^{\mu d,a}}{(1-k_{1m}^a\delta^{\mu r,a})}\mu_{12t}^{r,a} - \\ -\frac{\tau^a k_{1m}^a}{(1-k_{1m}^a\delta^{\mu r,a})}x_{12t}^a + \frac{\vartheta^a}{\delta^{\mu d,a}}\epsilon_{12t}^X + \epsilon_{t+1}^d\end{aligned} \tag{3.4.11}$$

Similarly to the case of fitting an AR(1) dynamics to the dividend-price ratio and the predictive regression for returns, predictive regression for dividend growth described by Eq.(3.4.3) may suffer from the omitted variable problem. In addition to the lagged dividend-price ratio, expected returns enter the implied monthly predictive relation and expected returns, the additional state variable and the MA component of expected dividend growth enter the implied annual predictive relation.

To summarize, aggregation implies that additional variables enter the predictive relations and, consequently, may help to predict returns. Moreover, a presence of the additional variables may affect statistical inference and the interpretation of estimated coefficients resulting from the predictive regression. I devote the rest of this section to the predictability implied by the BY calibration of the long run risks model.

In Table 3.4 I estimate the predictive VAR given by Eq.(3.4.1)–Eq.(3.4.3) using a single simulation path of 6000 annual observations generated by the BY model.

Tab. 3.4: Predictive regressions

This table reports the results of predicting the dividend-price ratio, returns and dividend growth by the dividend-price ratio using a single simulation path of 5000 annual observation generated from the long run risk model. Standard errors are GMM-adjusted to account for heteroskedasticity. R-squared statistics is in %.

Regression	β	*t-stat*	R^2
$dp_{t+1} = \alpha^{dp} + \beta^{dp}dp_t + \epsilon_{t+1}^{dp}$	0.7822	84.5614	61.19
$r_{m,t+1} = \alpha^r + \beta^r dp_t + \epsilon_{t+1}^r$	−0.0376	−2.8821	00.17
$\Delta d_{t+1} = \alpha^d + \beta^d dp_t + \epsilon_{t+1}^d$	−0.2933	−39.2850	24.29

Table 3.4 shows that the distinguishing feature of the BY model is that the slope coefficient for the predictive regression for returns and for dividend growth is negative. Moreover, predictability of dividend growth by the dividend-price ratio is excessively strong and predictability of returns by the dividend-price ratio is excessively weak. The R^2 in the dividend growth regression is more than 24 % indicating that the dividend-price ratio captures 24 % of the variation in expected dividend growth. While a virtually zero R^2 for the returns' regression indicates a low variability of expected returns in the model.

How different is interpretation of the results presented in Table 3.4 using the monthly or the annual state-space representation? I address this question in Table 3.5. In Panel A of Table 3.5 I decompose the variation in the dividend-price ratio into the components following Eq.(3.4.5) and in Panel B of Table 3.5 I calculate the cross-correlations between the variables of the state-space representation and the dividend-price ratio.

Tab. 3.5: Variation in dividend-price ratio

Panel A of this table reports the fractions of the standard deviation of the dividend-price ratio attributed to the elements of the state-space representation. Panel B, the standard deviations on the diagonal and the cross-correlations on the off-diagonal of the elements of the state-space representation. All the values are in %.

Frequency		dp	μ^r	μ^d	x	ϵ^X
Panel A: Fractions of the variation of the dividend-price ratio						
monthly		100.00	46.25	138.01	–	–
annual		100.00	42.06	173.01	12.92	135.12
Panel B: Correlations and standard deviations						
monthly	dp	16.13	−76.94	−97.68	–	–
	μ^r	−76.94	00.13	88.84	–	–
	μ^d	−97.68	88.84	00.51	–	–
annual	dp	19.50	−72.17	−84.46	−92.49	23.55
	μ^r	−72.17	01.37	56.03	87.27	08.20
	μ^d	−84.46	56.03	08.51	63.38	−70.30
	x	−92.49	87.27	63.38	00.17	10.44
	ϵ^X	23.55	08.20	−70.30	10.44	05.13

Consistently with the evidence on higher predictability of dividend growth generated by the BY model, from Panel A of Table 3.5, variation in expected dividend growth is the major contributor to the volatility of the dividend-price ratio in the BY model both in the monthly and in the annual frequency. The ratio of the standard deviation of expected dividend growth to the standard deviation of the dividend-price ratio in the monthly frequency is 138 % and in the annual frequency 173 %. Moreover, in the annual frequency, the MA component of expected dividend growth also accounts for a significant part of the variation in the dividend price ratio. Variation in expected returns in the both monthly and annual frequency, account only for a small fraction of the variation in the dividend-price ratio. The ratio of the standard deviation of expected returns to the standard deviation of the dividend-price ratio does not exceed 50 %. The monthly and the annual state-space representation imply the following expressions for the correlation of the dividend-price ratio with expected returns:

$$
\begin{aligned}
corr(dp_t, \mu_t^r) &= \frac{1}{1-k_{1m}\delta^{\mu r}}\frac{\sigma^{\mu r}}{\sigma^{dp}} - \frac{(1-(\tau+\delta^{\mu r})k_{1m})}{(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})}\frac{\sigma^{\mu d}}{\sigma^{dp}}corr(\mu^r,\mu^d) \quad c \\
&= 0.4625 - 1.2261;
\end{aligned}
$$

$$
\begin{aligned}
corr(dp^a, \mu^{r,a}) &= \frac{1}{1-k_1\delta^{\mu r,a}}\frac{\sigma^{\mu r,a}}{\sigma^{dp,a}} - \frac{1}{1-k_1\delta^{\mu d,a}}\frac{\sigma^{\mu d,a}}{\sigma^{dp,a}}corr(\mu^{r,a},\mu^{d,a}) + \\
&+ \frac{\tau^a k_1}{(1-k_1\delta^{\mu r,a})(1-k_1\delta^{\mu d,a})}\frac{\sigma^{x,a}}{\sigma^{dp,a}}corr(x^a,\mu^{r,a}) - \frac{\vartheta^a}{\delta^{\mu d,a}}\frac{1}{1-k_1\delta^{\mu d,a}}\frac{\sigma^{\epsilon^X}}{\sigma^{dp,a}}corr(\epsilon^X,\mu^{r,a}) \\
&= 0.4206 - 0.9694 - 0.1128 - 0.1108
\end{aligned}
$$

The Panel B shows that expected returns are positively correlated with expected dividend growth in both the monthly and the annual frequency. This positive correlation together with a high fraction of expected dividend growth in the variation of the dividend-price ratio is responsible for the negative correlation between expected returns and the dividend-price ratio in the both frequencies. Additional variables which appear in the annual predictive regression reinforce a negative correlation between the dividend-price ratio and expected returns.

In line with the evidence on a larger predictable component of dividend growth in Table 3.4, Panel B of Table 3.5 shows that expected dividend growth is highly volatile in the BY economy in both monthly and annual frequencies. The standard deviation of expected dividend growth exceeds the standard deviation of expected returns by more than a factor of four in the monthly frequency and by more than a factor of six in the annual frequency implying a much higher predictability of dividend growth than that of returns.

Moreover, the dividend-price ratio is highly negatively correlated with the long run risk variable x_t in the monthly frequency and with x_{12t} in the annual one. In the monthly frequency 100 % of the variation of expected dividend growth comes from x_t. Therefore, a correlation of 97.68 % between the dividend-price ratio and expected dividend growth reported in the table translates into an equal correlation between the dividend-price ratio and x_t. The correlation between the annualized dividend-price ratio and the long run risk process x_{12t} is only slightly less than in the monthly frequency. This evidence implies that the dividend-price ratio is a good proxy for x and that it tracks variation in this state variable in the predictive regressions, Eq.(3.4.1), Eq.(3.4.2) and Eq.(3.4.3).

3.5 Conclusion

Bansal, Kiku and Yaron (2007a) stress the importance of accounting for the aggregation problem while testing the long-run risks model. In this paper I extended their analysis. Specifically, I developed an improved version of their approximation in order to connect the aggregated dividend growth with the monthly economy and analyzed the effect of the aggregation on the predictability of returns in the model.

I found that the aggregation implies that a part of expected dividend growth is observable, since it is determined by a weighted sum of the monthly realized dividend growth. Additionally, the aggregation breaks the tight link between expected returns, expected dividend growth and expected consumption growth which characterizes the monthly economy. Moreover, my analysis exposes problems with the common procedures used to estimate the elasticity of intertemporal substitution.

After aggregation, an additional variable enters the state-space representation and the process for expected dividend growth is augmented by a moving average structure. Furthermore, even though in the monthly frequency the innovations to expected dividend growth are uncorrelated with the innovations to realized dividend growth, they become correlated after aggregation. These changes in the state-space representation imply additional to the dividend-price ratio variables should help to predict returns and needed to be controlled for in the predictive regression whenever they are correlated with the dividend-price ratio.

Using the calibration of Bansal and Yaron (2004), I found that the counterfactually negative relation between returns and the dividend-price ratio is due to excessive variability of expected dividend growth and a high correlation between expected returns and expected dividend growth. The presence of the additional variables in the implied predictive relation reinforces the negative correlation between expected returns and the dividend-price ratio.

4. SIMULATION STUDY OF PREDICTIVE REGRESSIONS

[1]

Introduction

Goyal and Welch (2008) note that the dividend-price ratio, together with the other popular variables, predicts returns poorly out-of-sample. The out-of-sample R-squared of the conventional predictive regression is small and frequently negative, indicating that the historical mean produces a forecast with an equal or even lower mean square error than the conventional predictive regression.

There is still a debate on an interpretation of this result. Goyal and Welch (2008) argues that a poor out-of-sample performance might indicate non-stability of the models used to predict stock returns. Cochrane (2008b) and Campbell and Thompson (2008) attribute it to a high persistence of the dividend-price ratio and finite samples, while Inoue and Kilian (2004) documents a low statistical power of out-of-sample tests.

In this article we examine statistical properties of the conventional predictive regression which involves regressing stock returns on a lagged dividend-price ratio. We confirm and extend the results of Cochrane (2008b) and Ferreira and Santa-Clara (2011) that the conventional predictive regression performs poorly out-of-sample even in the environment when the dividend-price ratio is designed to predict returns. Moreover, we provide more evidence on low power of the conventional predictive regression, confirming Inoue and Kilian (2004). Additionally, we show that a large unpredictable component in realized returns leads to a small out-of-sample R-squared ever possible to obtain. Using the true expected returns in place of an estimate, we calculate that the maximum out-of-sample R-squared attainable in our economies only slightly exceeds 10 %. Moreover, this unpredictable component conceals high correlations of the estimates by the prediction

[1]This essay represents joint work with Pedro Sana Clara

methods with the true expected returns.

Furthermore, we examine statistical properties of alternatives to the conventional predictive regression. Several possible alternatives are offered by the literature. Stambaugh (1999) suggests a correction based on an approximation of the finite sample bias in the persistence of the dividend-price ratio. Lewellen (2004) uses evidence against explosive bubbles in stock prices to set up an upper bound on the bias in the coefficients resulting from predictive regression. Ashley (2006) develops a shrinkage based on the coefficient of variation due to the sampling error. Differently to these methods, Connor (1997) and Ferreira and Santa-Clara (2011) approach the problem through economic rather than pure statistical perspective. Connor (1997) advocates a Bayesian adjustment to the coefficients. Ferreira and Santa-Clara (2011) defines expected returns by a combination of the dividend-price ratio and dividend growth forecasts. None of the existing works, however, compares the performance of these methods.

For our analysis we set up a simulation exercise. We follow Cochrane (2008b), Ferreira and Santa–Clara (2011) and Moon and Velasco (2010) and base our simulation design on the Campbell and Shiller (1988) identity. Our approach recognizes a full dependence of the observed return predictability patterns on the processes of expected returns, expected dividend growth and an innovation to the realized dividend growth and, therefore, allows to obtain a fully controlled environment for studying a predictive regression.

From the existing literature, we find four different systems–economies determining the processes for expected returns, expected dividend growth and an innovation to the realized dividend growth. One economy, while being consistent with many empirical facts, implies that the dividend-price ratio does not predict returns but, instead, predicts dividend growth. Therefore, we view this economy as an economically justified Null of no predictability in returns.

While in all the other three alternative economies dividend-price ratio predicts returns, they differ in the specification of the dynamics of expected returns and expected dividend growth and thus all the three economies have different implications for the predictive regressions. In one economy, expected returns and expected dividend growth have different persistence which implies an additional variable in the predictive relation between returns and the dividend-price ratio. In another, expected dividend growth has an MA structure leading to an ARMA process for the dividend-price ratio.

Our results can be summarized as follows. We find that the conventional predictive regression is likely to produce a higher mean square error than the historical mean in all the alternative economies. Furthermore, it has a low power to reject our null economy. We attribute these disadvantages to a high variance of its expected returns estimate. This evidence, however, does not imply a superiority of the forecast based on a historical mean. Our analysis shows that its estimate has a low correlation with the true expected returns and is, possibly, biased.

From the alternatives to the predictive regression, applying Stambaugh (1999) correction, leads to a deterioration of the out-of–sample performance, the methods of Lewellen (2004), Ashley (2006) and Connor (1997) show some improvement and the method of Ferreira and Santa-Clara (2011) is a considerable improvement over the predictive regression. The differences in the performance are largely attributed to a change in the noisiness of the estimate and to the covariance of the estimate with the realized returns.

Furthermore, we show that, even though the conventional predictive regression and the method of Stambaugh (1999) have a bad performance judging from the out-of-sample R-squared, these methods have a high correlation with the true expected returns explaining a significant part of their time variation.

Overall, the sum-of-parts method of Ferreira and Santa-Clara (2011) has a good performance producing a high out-of-sample R-squared and having a good power. Our analysis shows that this method produces a less noisy expected returns estimate which is highly correlated with the true expected returns.

Section 1 describes the economies we use for the simulation. Section 2 summarizes the conventional predictive regression and its alternatives. Section 3 presents the results. Section 4 concludes.

4.1 Economies

In this section we discuss distinguishing features of economic models we use for the simulation design. These economic models consist of the processes for expected returns, expected dividend growth and the innovation to the realized dividend growth. Cochrane (2008a) shows that this environment is sufficient to completely characterize a predictive relation between the market return and the dividend-price ratio.

Denote r_{t+1} and Δd_{t+1} as log of stock returns and log of dividend growth at time $t+1$. Let dp_{t+1} be the dividend-price ratio at time $t+1$. Furthermore, let:

$$\mu_t^r \equiv E_t[r_{t+1}],$$

$$\mu_t^d \equiv E_t[\Delta d_{t+1}],$$

We consider four different specifications for the processes for the expected returns, expected dividend growth–four different economies. In our Null economy, the dividend-price ratio does not predict returns, while in the three alternative economies which we refer to as Economy I, Economy II and Economy III, the dividend-price ratio predicts returns.

A model for expected returns and expected dividend growth which incorporates all the economies as special cases is:

$$\mu_{t+1}^r = a^{\mu,r} + b^{\mu,r}\mu_t^r + c^{\mu,r}x_t + \epsilon_{t+1}^{\mu,r}, \tag{4.1.1}$$

$$\mu_{t+1}^d = a^{\mu,d} + b^{\mu,d}\mu_t^d + c^{\mu,d}\epsilon_t^M + \epsilon_{t+1}^{\mu,d}, \tag{4.1.2}$$

$$x_{t+1} = a^x + b^x x_t + \epsilon_{t+1}^x, \tag{4.1.3}$$

$$\Delta d_{t+1} = \mu_t^d + \epsilon_{t+1}^d. \tag{4.1.4}$$

Table 4.1 reports parameter values we use to calibrate the economies.

Null Economy. Our Null economy corresponds to the Bansal and Yaron (2004) model. In the simulations, we use monthly calibration of the model and aggregate the data to the annual frequency. We use approximate annual processes, however, to demonstrate the differences between this economy and the alternative economies. In the annual frequency, the Null economy is characterized by the following system:

Tab. 4.1: Calibration

This table presents the values of the calibration of:

$$
\begin{aligned}
\mu^r_{t+1} &= a^{\mu,r} + b^{\mu,r}\mu^r_t + c^{\mu,r}x_t + \epsilon^{\mu,r}_{t+1},\\
\mu^d_{t+1} &= a^{\mu,d} + b^{\mu,d}\mu^d_t + c^{\mu,d}\epsilon^M_t + \epsilon^{\mu,d}_{t+1},\\
x_{t+1} &= a^x + b^x x_t + \epsilon^x_{t+1},\\
\Delta d_{t+1} &= \mu^d_t + \epsilon^d_{t+1}.
\end{aligned}
$$

corresponding to the Null Economy (second column), Economy I (third column), Economy II (fourth column) and Economy III (fifth column). Note that parameter values presented for the Null economy are approximate. In the simulation we use monthly calibration and follow the aggregation procedure described in Beeler and Campbell (2009).

	Null Economy	Economy I	Economy II	Economy III
$a^{\mu,r}$	0.0031	0.0029	0.0061	0.0037
$a^{\mu,d}$	0.0018	7.5220×10^{-4}	0.0401	0.0217
a^x	0.0000	–	–	–
$b^{\mu,r}$	0.8547	0.9410	0.9320	0.9570
$b^{\mu,d}$	0.7752	0.9410	0.3540	0.6380
b^x	0.7752	–	–	–
$c^{\mu,r}$	−0.5676	–	–	–
$c^{\mu,d}$	1.0000	–	–	0.6380
σ^{ϵ^d}	0.0761	0.1400	0.0020	0.0542
$\sigma^{\epsilon^{\mu,r}}$	0.0084	0.0155	0.0160	0.0160
$\sigma^{\epsilon^{\mu,d}}$	0.0738	0.0012	0.0580	0.0721
σ^{ϵ^x}	0.0011	–	–	–
σ^{ϵ^M}	0.0520	–	–	0.0540
$\rho^{\epsilon^{\mu,r}}_{\epsilon^{\mu,d}}$	0.3951	1.0000	0.4170	0.4170
$\rho^{\epsilon^{\mu,r}}_{\epsilon^x}$	0.9167	–	–	–
$\rho^{\epsilon^{\mu,r}}_{\epsilon^M}$	0.0957	–	–	−0.4369
$\rho^{\epsilon^{\mu,r}}_{\epsilon^d}$	0.0754	0.0750	−0.1470	−0.1470
$\rho^{\epsilon^{\mu,d}}_{\epsilon^x}$	0.4276	–	–	–
$\rho^{\epsilon^{\mu,d}}_{\epsilon^M}$	−0.8524	–	–	0.1728
$\rho^{\epsilon^{\mu,d}}_{\epsilon^d}$	0.4851	0.0750	0.0000	0.0000
$\rho^{\epsilon^x}_{\epsilon^M}$	0.9167	–	–	–
$\rho^{\epsilon^x}_{\epsilon^d}$	0.0773	–	–	–
$\rho^{\epsilon^M}_{\epsilon^d}$	−0.4888	–	–	0.0963

$$
\begin{aligned}
\mu^r_{t+1} &= 0.0031 + 0.8547\mu^r_t - 0.5676x_t + \epsilon^{\mu,r}_{t+1},\\
\mu^d_{t+1} &= 0.0018 + 0.7752\mu^d_t + \epsilon^M_t + \epsilon^{\mu,d}_{t+1},\\
x_{t+1} &= 0.7752x_t + \epsilon^x_{t+1},
\end{aligned}
$$

The Null economy is the most general one. Expected returns at time $t+1$, in addition to their lagged value depend on a state variable x_t. This state variable corresponds to the December expected consumption growth in the Bansal and Yaron (2004) model. Moreover, expected dividend growth has an MA structure. The additional state variable and the MA component are absent in a monthly formulation of the Bansal and Yaron

(2004) model and arise due to the aggregation to the annual frequency.

From Table 4.1, comparing with the alternative economies, the Null economy is characterized by the lowest persistence of expected returns and a high persistence of expected dividend growth. An additional distinguishing feature of the Null economy is a very low standard deviation of the error to expected returns. Its value of 0.0084 is almost twice lower than the value of 0.0160 in the alternative economies. The standard deviation of the innovation to expected dividend growth is also slightly higher than in the other economies.

It will be clear from the results presented below that those differences will translate into very different predictability patterns observed in the Null economy if compared with the alternatives.

Economy I. Economy I is obtained from a predictive VAR estimated in Cochrane (2008b). Expected returns and expected dividend growth in this economy follow:

$$\mu^r_{t+1} = 0.0029 + 0.9410\mu^r_t + \epsilon^{\mu,r}_{t+1},$$
$$\mu^d_{t+1} = 7.5220 \times 10^{-4} + 0.9410\mu^d_t + \epsilon^{\mu,d}_{t+1},$$

Economy I is the most restrictive. The additional state variable and the MA component in this economy are absent. Both processes follow are AR(1) and are equally persistent. Moreover, Table 1 shows that standard deviation of the innovation to expected dividend growth is only 0.0012 which is many times lower than the corresponding value in the other economies.

Economy II. Our Economy II is obtained from the Binsbergen and Koijen (2010) cash-invested dividends system. It is characterized by the following processes:

$$\mu^r_{t+1} = 0.0061 + 0.9320\mu^r_t + \epsilon^{\mu,r}_{t+1},$$
$$\mu^d_{t+1} = 0.0401 + 0.3540\mu^d_t + \epsilon^{\mu,d}_{t+1},$$

Economy II is less restrictive than Economy I. While in the both economies expected returns and expected dividend growth are an AR(1) process, in Economy II expected returns are more persistent than expected dividend growth. Another difference is that, from Table 1, in Economy II the innovation to expected returns is less volatile than the innovation to expected dividend growth.

Economy III. Economy III is the Binsbergen and Koijen (2010) market-invested dividends case. In this economy, expected returns and expected dividend growth follow:

$$\mu_{t+1}^{r} = 0.0037 + 0.9570\mu_t^r + \epsilon_{t+1}^{\mu,r},$$
$$\mu_{t+1}^{d} = 0.0217 + 0.6380\mu_t^d + 0.6380\epsilon_t^M + \epsilon_{t+1}^{\mu,d},$$

Reinvesting dividends in the stock market leads to an additional MA structure for expected dividends. Note that, while persistence of both, expected returns and expected dividend growth, is higher in Economy III than in Economy II, persistence of expected dividend growth increases by more. Additionally, reinvesting dividends in the stock market increases volatility of the innovation to expected dividend growth and of the innovation to the realized dividend growth.

In simulations, data on returns and the dividend-price ratio is obtained from the Campbell and Shiller (1988) identity which states:

$$r_{t+1} = k_0 - k_1 dp_{t+1} + dp_t + \Delta d_{t+1}. \tag{4.1.5}$$

2

It follows from Eq.(4.1.5) that the dividend-price ratio is connected with the state variables through the present value identity:

$$dp_t = B_0 + B_1\mu_t^r + B_2\mu_t^d + B_3 x_t + B_4\epsilon_t^M. \tag{4.1.6}$$

Coefficients B_0, B_1, B_2, B_3 and B_4 are determined by the structural parameters of Eq.(4.1.1)–Eq.(4.1.4). The exact expressions, for the interested reader, are provided in the Appendix.

For each of the economies we generate 80 observations which correspond to the 80 annual observations from the stock market data. In Table 4.2 we report the means taken

[2] In Eq.(5) $k_1 = \dfrac{e^{-\bar{dp}}}{1+e^{-\bar{dp}}}$ and $k_0 = -(1-k_1)log(1-k_1) - k_1 log(k_1)$.

across the 10000 simulation paths of the descriptive statistics of the generated data.

Tab. 4.2: Descriptive statistics

This table presents descriptive statistics of realized returns, dividend growth and the dividend-price ratio implied by each of the economies. The calculations are based on 80 observations simulated 10000 times. The Null economy (Bansal and Yaron (2004) model) is simulated using a monthly system and then aggregated to the annual frequency using the procedure of Beeler and Campbell (2009). All the entries of the table correspond to the means of corresponding statistics taken across the simulation paths.

Descriptive Statistics	Null Economy	Economy I	Economy II	Economy III
dp:				
Mean	-3.0075	-3.2626	-3.4428	-3.8296
Std	0.1892	0.3415	0.3774	0.5358
Min	-3.4398	-3.9806	-4.2280	-4.9083
Max	-2.5504	-2.5490	-2.6582	-2.7496
r:				
Mean	0.0682	0.0509	0.1011	0.0692
Std	0.1693	0.1987	0.1534	0.2076
Min	-0.3493	-0.4340	-0.2719	-0.4351
Max	0.4888	0.5370	0.4736	0.5739
Δd:				
Mean	0.0178	0.0120	0.0622	0.0584
Std	0.1130	0.1388	0.0609	0.1006
Min	-0.2604	-0.3279	-0.0858	-0.1853
Max	0.2967	0.3521	0.2099	0.3033

The dividend-price ratio in the Null economy is characterized by the highest mean and the lowest volatility if compared with the other economies. The difference is especially pronounced for the standard deviation which is more than twice lower than the standard deviation of the dividend-price ratio in Economy III. Economy III, which corresponds to the case of dividends reinvested into the stock market, implies the most volatile dividend-price ratio and returns. The market reinvested dividend–price ratio is characterized by the volatility which is more than 1,5 times higher than the volatility of the dividend-price ratio in Economy I and Economy II. Finally, cash-reinvestment strategy, Economy II, produces a low volatility of dividend growth.

A conventional approach to predict stock returns involves estimating a VAR:

$$dp_{t+1} = \alpha^{dp} + \beta^{dp} dp_t + \epsilon^{dp}_{t+1}, \tag{4.1.7}$$

$$r_{t+1} = \alpha^{r} + \beta^{r} dp_t + \epsilon^{r}_{t+1}. \tag{4.1.8}$$

Eq.(4.1.1)-Eq.(4.1.4) together with Eq.(4.1.5) allow to derive the implied predictive relations in each economy. Using Eq.(4.1.1)–Eq.(4.1.4) in Eq.(4.1.6) leads to the following dynamic description of the dividend-price ratio:

$$dp_{t+1} = K_0 + b^{\mu,r} dp_t + K_1 \mu_t^d + K_2 x_t + K_3 \epsilon_t^M + B_1 \epsilon_{t+1}^{\mu,r} + B_2 \epsilon_{t+1}^{\mu,d} + B_3 \epsilon_{t+1}^{x} + B_4 \epsilon_{t+1}^{M}. \tag{4.1.9}$$

In Eq.(4.1.9) $K_0 = (1 - b^{\mu,r})B_0 + a^{\mu,r}B_1 + a^{\mu,d}B_2 + a^x B_3$, $K_1 = (b^{\mu,d} - b^{\mu,r})B_2$, $K_2 = (b^x - b^{\mu,r})B_3$ and $K_3 = (c^{\mu,d}B_2 - b^{\mu,r}B_4)$.

Furthermore, substituting Eq.(4.1.9) into Eq.(4.1.5) leads to the following implied predictive relation between the returns and the dividend-price ratio:

$$r_{t+1} = (1 - k_1)K_0 + (1 - k_1 b^{\mu,r})dp_t + (1 - k_1 K_1)\mu_t^d - k_1 K_2 x_t - k_1 K_3 \epsilon_t^M + \epsilon_{t+1}^d - k_1 \epsilon_{t+1}^{dp}, \tag{4.1.10}$$

with:

$$\epsilon_{t+1}^{dp} = B_1 \epsilon_{t+1}^{\mu,r} + B_2 \epsilon_{t+1}^{\mu,d} + B_3 \epsilon_{t+1}^{x} + B_4 \epsilon_{t+1}^{M}.$$

Eq.(4.1.9) and Eq.(4.1.9) expose several problems with the predictive VAR, Eq.(4.1.7)–Eq.(4.1.8). The error to the dividend-price ratio from Eq.(4.1.9) enters negatively into Eq.(4.1.10). If ϵ_{t+1}^d and ϵ_{t+1}^{dp} are uncorrelated this implies that the prediction error to returns is negatively correlated with the error to the dividend-price ratio leading to a violation of the strict exogeneity assumption. As the result, Stambaugh (1999) shows that predicting returns with the dividend-price ratio leads to a finite sample bias in the estimated coefficients and standard errors.

There are several other potential problems with the predictive VAR. Eq.(4.1.9) indicates that estimating Eq.(4.1.7) we omit the other variables. Similarly, Eq.(4.1.10) shows that the additional variables may help to predict returns. Both equations imply that if the dividend-price ratio tracks other variables in addition to the expected returns, estimates of the persistence of the dividend-price ratio by fitting an AR(1) process or predicting returns with only a lagged dividend-price ratio may produce inconsistent coefficient estimates.

Table 4.3 provides the values of the coefficients of Eq.(4.1.9) and Eq.(4.1.10). It shows that additional variables enter the implied predictive relations in Economy II, Economy III and the Null economy.

Tab. 4.3: Implied predictive relations

This table summarizes the implications of each of the economies on the predictive relations between the returns and the dividend-price ratio. All the values are computed using the calibration parameters. The fractions and the correlations are in %.

Parameter	Null Economy	Economy I	Economy II	Economy III
Panel A: Shares of the variance of the dividend-price ratio				
$\frac{std(B_1\mu_t^r)}{std(dp_t)}$	47.81	108.54	102.40	94.63
$\frac{std(B_2\mu_t^d)}{std(dp_t)}$	192.71	8.57	21.36	34.34
Panel B: $dp_{t+1} = K_0^{dp} + K_1^{dp} dp_t + K_2^{dp} \mu_t^d + K_3^{dp} x_t + K_4^{dp} \epsilon_t^M + \epsilon_{t+1}^{dp}$				
K_1^{dp}	0.08547	0.9410	0.9320	0.9570
K_2^{dp}	0.3042	–	0.8798	0.8342
K_3^{dp}	−2.1710	–	–	–
K_4^{dp}	0.3924	–	–	0.8342
Panel C: $r_{t+1} = K_0^r + K_1^r dp_t + K_2^r \mu_t^d + K_{3\,t}^{r\,x} + K_4^r \epsilon_t^M + \epsilon_{t+1}^r$				
K_1^r	0.1857	0.0931	0.0969	0.0736
K_2^r	0.7102	–	−0.8525	0.1925
K_3^r	2.0684	–		
K_4^r	−0.5264	–		−0.8075
$corr(\epsilon_{t+1}^r, \epsilon_{t+1}^{\mu,r})$	58.53	−72.80	−85.05	−74.41

Evidence in favor of predictability of returns by the dividend-price ratio implies that a fraction of the variation of the dividend-price ratio which is attributed to expected returns is larger than the fraction attributed to expected dividend growth. Panel A of Table 4.3 shows that in all but one economy the fraction of the standard deviation due to expected returns exceeds the fraction due to expected dividend growth. The fraction of the standard deviation of the dividend-price ratio due to expected returns in Economy I, Economy II and Economy III is around 100 % which exceeds the fraction due to expected dividend growth by many times. In the Null economy, the pattern reverses. In this economy, the contribution of expected dividend growth is three times higher than that of expected returns. In other words, while in Economy I, Economy II and Economy

III the dividend-price ratio predicts returns, in the Null economy it predicts dividend growth.

Many return predictability patterns, like for example low autocorrelation and an increasing R^2 in the long horizon predictive regressions are explained by a presence of the discount rate effect. The discount rate effect states that a contemporaneous correlation between the innovations to expected returns and to the realized returns is negative. We check the presence of the discount rate effect in our economies in the last row of the Panel C. While the discount rate effect is present in Economy I, Economy II and Economy III, it is absent in the Null economy.

We estimate the predictive VAR in each of the economies in Table 4.4. While all the coefficients for the dividend growth regression follow from Eq.(4.1.7)-Eq.(4.1.8) and Eq.(4.1.5), to facilitate the comparison of the predictability patterns across the economies, we also include the results of the predictive regression for dividend growth.

Tab. 4.4: In-sample performance of conventional predictive regression

This table presents the results of the predictive VAR, Eq.(4.1.7)–Eq.(4.1.8) in the text, augmented by a predictive regression for dividend growth. The results are based on 80 observations simulated 10000 times. All the entries of the table correspond to the means of corresponding statistics taken across the simulation paths. The R^2-statistics is in %.

Parameter	Null Economy[a]	Economy I	Economy II	Economy III
Panel A: $dp_{t+1} = \alpha^{dp} + \beta^{dp} dp_t + \epsilon^{dp}_{t+1}$				
b	0.6566	0.8812	0.8841	0.9151
t-stat	8.00	17.70	18.77	22.53
R^2	44.17	77.12	79.29	83.95
Panel B: $r_{t+1} = \alpha^r + \beta^r dp_t + \epsilon^r_{t+1}$				
b	−0.0058	0.1553	0.1353	0.1151
t-stat	−0.15	2.21	2.89	2.47
R^2	1.31	6.55	10.08	7.86
Panel C: $\Delta d_{t+1} = \alpha^d + \beta^d dp_t + \epsilon^d_{t+1}$				
b	−0.3810	0.0044	−0.0081	0.0017
t-stat	−7.28	0.11	−0.34	0.12
R^2	40.07	1.31	2.47	2.35

The Null economy is characterized by very different from the alternatives predictability patterns. The estimated AR coefficient for the dividend-price ratio is 0.65 and the R^2 in this regression is 44 %, much lower than in all the alternative economies. More importantly, dividend growth is highly predictable by the dividend-price ratio whereas returns are not. The R^2 in the dividend growth regression indicates that current dividend-price ratio captures 40 % of the variation in future dividend growth.

While additional variables enter the implied predictive relations in Economy II and Economy III and are absent in Economy I, all alternative economies produce very similar results. In all of these economies, fitting an AR(1) process to the dividend-price ratio leads to estimates of the slope coefficient close to 0.9 and there is more evidence that returns are predictable by the dividend-price ratio and the dividend growth is not.

4.2 Predicting returns out-of-sample

In this section we briefly summarize the methods to predict stock returns. We start from the conventional predictive regression.

The conventional predictive regression involves estimation of:

$$r_s = \alpha^r + \beta^r dp_{s-1} + \epsilon_s^r. \tag{4.2.1}$$

An out-of-sample prediction of the returns for the period $s+1$ is formed by using the estimated parameters from Eq.(4.2.1) and a current observation of the dividend-price ratio. In other words, the out-of-sample estimate of the current expected returns is:

$$\hat{\mu}_s^{PR} = \hat{\alpha^r} + \hat{\beta^r} dp_s. \tag{4.2.2}$$

Note that each period, the information used to estimate Eq.(4.2.1) expands leading to different $\hat{\alpha^r}$ and $\hat{\beta^r}$.

As was already emphasized in the previous section, the conventional predictive regression suffers from several drawbacks. Using the dividend-price ratio as a predictor of returns leads to a violation of the strict exogeneity assumption and, consequently, to the Stambaugh (1999) problem. Moreover, the uncertainty about the processes for expected returns and expected dividend growth leaves a possibility of the additional variables we omit while estimating Eq.(4.2.1) .

Almost none of the methods we consider address the problem of the omitted variables. Only the sum–of–parts method by Ferreira and Santa–Clara (2011) takes into account the possibility of expected dividend growth entering the implied predictive relation for returns.

Stambaugh (1999) shows that the bias in the estimated coefficient in Eq.(4.2.1) obeys:

$$E\left[\hat{\beta}^r - \beta^r\right] = \frac{cov(\epsilon^r_{t+1}, \epsilon^{dp}_{t+1})}{var(\epsilon^{dp}_{t+1})} E\left[\hat{\beta}^{dp} - \beta^{dp}\right]. \tag{4.2.3}$$

The methods–alternatives to the conventional predictive regression try to solve the Stambaugh (1999) problem through either applying corrections to the bias to the coefficients from the predictive regression directly, or following Eq.(4.2.3) through correcting the estimate of the persistence of the dividend-price ratio. [3]

We consider the following alternatives to the conventional predictive regression.

Stambaugh (1999). Stambaugh (1999) explores a possibility of correcting the bias in coefficients of the predictive regression by using an approximation to the bias in the persistence of the dividend-price ratio. Following Marriott and Pope (1954) and Kendall (1954) the approximation to the bias in the persistence of the dividend-price ratio is:

$$E\left[\hat{\beta}^{dp} - \beta^{dp}\right] = -\frac{1 + 3\beta^{dp}}{T} + O(1/T).$$

Since the true persistence is unobserved, Stambaugh (1999) replaces it with an estimate leading to the following correction to the bias:

$$\hat{\beta}^{r,Stambaugh} = \hat{\beta}^r + \frac{cov(\epsilon^r_{t+1}, \epsilon^{dp}_{t+1})}{var(\epsilon^{dp}_{t+1})} \left(\frac{1 + 3\hat{\beta}^{dp}}{T}\right). \tag{4.2.4}$$

defining the constant term as:

[3] Our simulations show that $\frac{cov(\epsilon^r_{t+1}, \epsilon^{dp}_{t+1})}{var(\epsilon^{dp}_{t+1})}$ can be precisely estimated.

$$\hat{\alpha}^{r,Stambaugh} = \bar{r} - \hat{\beta}^{r,Stambaugh}\bar{dp}, \tag{4.2.5}$$

we form the estimates of the expected returns using the Stambaugh (1999) correction method as:

$$\hat{\mu}_s^{Stambaugh} = \hat{\alpha}^{r,Stambaugh} + \hat{\beta}^{r,Stambaugh} dp_s. \tag{4.2.6}$$

Lewellen (2004). Lewellen (2004) avoids the problem of the unobserved persistence of the dividend-price ratio in a different way. Eq.(4.2.3) implies a negative relation between the bias in the coefficients resulting from the predictive regression and the bias in the estimate of the persistence of the dividend-price ratio. A lower bound on the bias in the persistence, then, provides an upper bound on the bias in the predictive coefficient. In order to obtain the lower bound, Lewellen (2004) refers to the evidence against explosive bubbles in the stock prices.

To elaborate, note that the conditional bias in the slope coefficient of the conventional predictive regression is:

$$E\left[\hat{\beta}^r - \beta^r | \hat{\beta}^{dp}\right] = \frac{cov(\epsilon_{t+1}^r, \epsilon_{t+1}^{dp})}{var(\epsilon_{t+1}^{dp})}\left[\hat{\beta}^{dp} - \beta^{dp}\right]$$

By fixing $\beta^{dp} \approx 1$, the adjusted $\hat{\beta}^r$ is:

$$\hat{\beta}^{r,Lewellen} = \hat{\beta}^r - \frac{cov(\epsilon_{t+1}^r, \epsilon_{t+1}^{dp})}{var(\epsilon_{t+1}^{dp})}\left[\hat{\beta}^{dp} - 0.999\right] \tag{4.2.7}$$

Similarly to the Stambaugh (1999) method case, we let:

$$\hat{\alpha}^{r,Lewellen} = \bar{r} - \hat{\beta}^{r,Lewellen}\bar{dp}, \tag{4.2.8}$$

Consequently, we form the estimate of expected returns by:

$$\hat{\mu}_s^{Lewellen} = \hat{\alpha}^{r,Lewellen} + \hat{\beta}^{r,Lewellen} dp_s. \tag{4.2.9}$$

Connor (1997). Connor (1997) uses a Bayesian adjustment to coefficients of the conventional predictive regression.

The main idea is that, under the efficient market hypothesis, a predictable variation in returns from their long-run mean should be distributed around zero. Setting up this prior, a Bayesian adjustment for the slope has the following form:

$$\hat{\beta}^{r,Connor} = \left[\frac{Tvar(dp_t)/var(\epsilon_{t+1}^r)}{Tvar(dp_t)/var(\epsilon_{t+1}^r) + 1/var(\beta^r)}\right]\hat{\beta}^r$$

$var(\beta^r)$ is the prior variance of the unknown β^r. Though $var(\beta^r)$ is unobserved, Connor (1997) shows that the Bayesian adjustment for the predictive coefficient can be written as a function of the sample size and of the expected R^2 of the predictive regression. Therefore, letting:

$$\rho = E\left[\frac{R^2}{1-R^2}\right],$$

An adjusted slope coefficient is:

$$\hat{\beta}^{r,Connor} = \left[\frac{T}{T+1/\rho}\right]\hat{\beta}^r. \tag{4.2.10}$$

As usual, we set the constant to be equal to:

$$\hat{\alpha}^{r,Connor} = \bar{r} - \hat{\beta}^{r,Connor}\bar{dp} \tag{4.2.11}$$

Consequently, we estimate expected returns as:

$$\hat{\mu}_s^{Connor} = \hat{\alpha}^{r,Connor} + \hat{\beta}^{r,Connor} dp_s. \tag{4.2.12}$$

In our analysis we follow Ferreira and Santa-Clara (2011) and set $1/\rho = 100$ corresponding to the expected R^2 of approximately 1%.

Ashley (1990), (2006). Ashley (1990) and (2006) shows that in the case of a highly noisy forecast which results from finite samples the mean square error-minimizing forecast is always a forecast that is shrunk towards zero. The method is appealing for predicting returns since it addresses the issue of a noisy forecast and, as it is shown in the next section, it is a noise in the forecast which contributes considerably to the mean square error of the conventional predictive regression. To demonstrate, let:

$$\hat{\mu}_t - \mu_t = \nu_t,$$

Note that ν_t is an error that the forecast by the predictive regression makes while estimating expected returns. Recall that:

$$r_{t+1} = \mu_t + \epsilon^r_{t+1}.$$

Denote λ as the shrinkage factor. Following Ashley (2006), the optimal shrinkage factor is:

$$\lambda = \frac{1}{1 + \frac{E[\nu_t^2]}{E[\mu_t^2]}}.$$

$\dfrac{E[\nu_t^2]}{E[\mu_t^2]}$ is the coefficient of variation of $\hat{\mu}_t$ due to the sampling error ν_t.
The corrected slope coefficient is:

$$\hat{\beta}^{r,Ashley} = \lambda\hat{\beta}^r. \tag{4.2.13}$$

Following Ashley (2006), we estimate the coefficient of variation as:

$$\hat{cv}_t = \frac{s_t^2[1\ dp_t]'\left([\iota\ dp]'[\iota\ dp]\right)^{-1}[1\ dp_t]}{\hat{\mu_t}^2},$$

where ι is a vector of ones, $dp = [dp_1\ dp_2 \ldots\ dp_{t-1}]'$, $\hat{\mu}_t$ is the forecast resulting from the conventional predictive regression and s^2 is an estimate of the variance of ϵ^r.

Defining the constant term as:

$$\hat{\alpha}^{r,Ashley} = \bar{r} - \hat{\beta}^{r,Ashley}\bar{dp}. \tag{4.2.14}$$

We estimate expected returns using the method of Ashley (2006) as:

$$\hat{\mu}_s^{Ashley} = \hat{\alpha}^{r,Ashley} + \hat{\beta}^{r,Ashley}dp_s. \tag{4.2.15}$$

Ferreira and Santa-Clara (2011). Another alternative to the conventional predictive regression is the sum–of–parts method suggested by Ferreira and Santa-Clara (2011). To demonstrate their method, denote R_{t+1} as level of returns. Then:

$$1 + R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t} = \left(1 + \frac{D_{t+1}}{P_{t+1}}\right)\frac{P_{t+1}/D_{t+1}}{P_t/D_t}\frac{D_{t+1}}{D_t},$$

implying:

$$r_{t+1} = log\left(1 + R_{t+1}\right) = log\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) - \Delta dp_{t+1} + \Delta d_{t+1}.$$

Moreover, taking the expectation conditional on information available at time t:

$$\mu_t = E_t\left[log\left(1 + \frac{D_{t+1}}{P_{t+1}}\right)\right] - E_t\left[\Delta dp_{t+1}\right] + E_t\left[\Delta d_{t+1}\right]. \tag{4.2.16}$$

Eq.(4.2.16) relates expected returns to the expectation of a function of the future dividend-price ratio and the expectation of the future dividend growth. Consequently, a proxy for expected returns can be obtained by forecasting separately the dividend-price ratio and dividend growth. We follow Ferreira and Santa Clara (2011) and use a current dividend-price ratio as an estimate of the future dividend-price ratio and a twenty years moving average of the past dividend growth to estimate expected dividend growth.

We denote an estimate of expected returns formed using the method of Ferreira and

Santa-Clara (2011) as:

$$\hat{\mu}_s^{F\&SC} = log\left(1 + \frac{D_s}{P_s}\right) + \sum_{t=s-19}^{s} \Delta d_t. \tag{4.2.17}$$

Note that this approach avoids regression of returns on a valuation ratio and, therefore, does not face the Stambaugh (1999) problem. Additionally, it controls for a possibility that expected dividend growth enters the implied predictive relation for returns.

4.3 Results

4.3.1 Relative performance of the methods

Following Goyal and Welsh (2008) and Campbell and Thomson (2005) we use the out-of-sample R-squared to evaluate the performance of each method. The out-of-sample R-squared compares the mean square error (MSE) of a given method with the MSE of the historical mean and is calculated as:

$$R^2_{OOS} = 1 - \frac{\sum_{s=s_0}^{T-1} \left[\hat{\mu}_s - r_{s+1}\right]^2}{\sum_{s=s_0}^{T-1} \left[\bar{r}_s - r_{s+1}\right]^2}, \tag{4.3.1}$$

with $\bar{r}_s$ being the expected returns estimate by the historical mean and $\hat{\mu}_s$ the estimate produced by each of the methods. R^2_{OOS} is negative whenever the prediction by the historical mean produces a lower root-mean-squared error than the prediction by an alternative method.

Performance of each of the methods is evaluated in Table 4.5. Panel A of the table presents the results for the Null economy. For this economy, we report the mean, the median as well as the 95^{th} percentile of the distribution of the R^2_{OOS}. In Panel B, Panel C and Panel D we evaluate performance of the methods in the alternative economies. For these economies we report the percentage of the negative realizations of the R^2_{OOS} in the *%-negative* and the power of each method to reject the Null economy in the *Power*. We define the power as the percentage of the R^2_{OOS} produced by each of the methods with a value greater or equal to the 95^{th} percentile under the Null economy.

Tab. 4.5: Out-of-sample performance of methods

This table presents the results of predicting returns out–of–sample by each of the methods. In Panel A we report the mean, the median and the 95^{th} percentile of the distribution of R^2_{OOS} in % in the Null economy. In Panel B, Panel C and Panel D, the *%-negative* corresponds to the percentage of negative realizations of the R^2_{OOS} and the *Power*– the percentage of realizations of R^2_{OOS} greater or equal to the 95^{th} percentile of the distribution of R^2_{OOS} in the Null economy.

Statistics	Predictive regression	Stambaugh (1999)	Lewellen (2004)	Ashley (2006)	Connor (1997)	F & SC (2011)	True expected returns
Panel A: Null economy							
Mean	-3.08	-4.18	-15.18	-1.39	-0.21	-1.33	2.82
Median	-2.70	-3.18	-14.57	-1.23	-0.49	-1.03	2.54
95^{th}*-percentile*	2.62	2.14	00.01	2.77	2.15	2.50	9.55
Panel B: Economy I							
Mean	0.46	-2.60	1.08	1.01	2.63	4.32	7.02
Median	00.59	-1.79	0.85	0.96	2.28	4.21	7.33
%-negative	46.32	58.49	42.93	39.78	14.65	3.09	5.44
Power	36.88	31.99	57.02	32.93	52.38	79.21	–
Panel C: Economy II							
Mean	2.17	-3.72	2.80	3.02	3.84	6.49	10.31
Median	2.84	-2.16	2.53	3.54	3.60	5.97	10.65
%-negative	37.34	56.26	29.92	29.15	6.68	1.83	4.11
Power	51.13	37.61	70.05	54.81	71.08	91.82	–
Panel D: Economy III							
Mean	1.21	-5.34	3.32	1.93	3.30	5.43	8.67
Median	1.41	-4.37	2.95	1.92	2.97	4.91	8.67
%-negative	42.00	65.18	25.20	35.18	9.94	5.18	3.27
Power	42.96	27.44	74.74	43.43	62.76	79.41	–

Using Table 4.5 we confirm the result of Cochrane (2008b) who documents a poor out-of-sample performance of the conventional predictive regression even under the predictable returns null. In Economy I, Economy II and Economy III the number of negative realizations of the R^2_{OOS} for the conventional predictive regression ranges from 37.34 % to 46.32 %.

Additionally, Table 4.5 reports a high fraction of negative realizations of the R^2_{OOS} for the methods by Stambaugh (1999), Lewellen (2004) and Ashley (2006). Note that Stambaugh (1999) correction produces even a larger value of the *%-negative* than the conventional predictive regression. For this method, the fraction of the negative realizations is so high, that it leads to a negative mean and a negative median of the distribution of the R^2_{OOS} .

While the value of the *%-negative* for the methods by Lewellen (2004) and Ashley (2006) is somewhat smaller than for the conventional predictive regression, the percentage of the negative relaizations of R^2_{OOS} for these methods is still high, ranging from 25.20 % to 42.93 % for Lewellen (2004) and from 29.15 % to 39.78 % for the method by Ashley (2006). The *%-negative* is smaller than for the above mentioned methods for the shrinkage of Connor (1997). The number of the negative realizations of the R^2_{OOS} for this method ranges from 6.68 % to 14.65 %.

The method of Ferreira and Santa–Clara (2011) performs significantly better than the other methods by producing only from 1.83 % to 5.18% of negative realizations of the R^2_{OOS}. Note that the mean and the median of the R^2_{OOS} for this method is higher than for any other method across all the alternative economies.

Furthermore, we document a low power of the conventional predictive regression to reject our Null economy. From Table 4.5 the *Power* for the conventional predictive regression ranges from 36.88 to 51.13, indicating that only from 37% to 51% of realizations of the R^2_{OOS} in the alternative economies were higher than the value of the 95^{th} percentile in the Null economy. Moreover, the Stambaugh (1999) method is less powerful and the method by Ashley (2006) has approximately the same power as the conventional predictive regression.

The methods which are more powerful than the conventional predictive regression are

Lewellen (2004), Connor (1997) and Ferreira and Santa-Clara (2011)[4]. The method of Lewellen (2004) is more powerful than the method of Connor (1997) in Economy I and Economy III while the method of Connor (1997) is only slightly more powerful than the method of Lewellen (2004) in Economy II.

The most powerful method from those we consider is the sum-of-parts method of Ferreira and Santa-Clara (2011). This method produces around 80 % in Economy I and Economy III and over 90 % in Economy II of the R^2_{OOS} larger than the value of the 95^{th} percentile.

In short, the results presented in Table 4.5 indicate that the conventional predictive regression and the methods of Stambaugh (1999), Lewellen (2004) and Ashley (2006) tend to produce a negative R^2_{OOS}. For these methods, observing a negative R^2_{OOS} does not indicate an absence of return predictability or instability of a predictive relation between returns and the dividend-price ratio. Moreover, the conventional predictive regression and the methods of Stambaugh (1999) and Ashley (2006) have a very low power to detect a presence of predictability in returns.

It is the sum-of-parts method of Ferreira and Santa-Clara (2011) which produces the best results. This method is characterized by a low number of negative realizations of the R^2_{OOS} and has a comparatively high power to distinguish the null from the alternative economies.

Finally, our simulation design allows calculation of the maximum R^2_{OOS} which is attainable by replacing an estimate with the true expected returns in Eq.(4.3.1). We present the value of the maximum R^2_{OOS} for each of the economies in the last column of Table 4.5. Our results indicate that the maximum possible R^2_{OOS} only slightly exceeds 10 %.

In Table 4.6 we apply the methods to the data from the original papers used to construct the economies. Below the values of R^2_{OOS}, in the round brackets, we report the fraction of the values of R^2_{OOS} larger or equal to the value in the simulations. Note that in this exercise we simulate the data sets with the sample sizes which equal to the sample sizes of the papers. The Null economy is based on the data sample which starts in 1929 and ends in 1998. Thus, it covers the pre-WWII years and the second

[4]High power of the Lewellen (2004) method has been already documented in Goyal and Welch (2008)

half of 90's-the periods where the predictive relation between the dividend-price ratio and returns was weak[5]. The data sample used for Economy I starts in 1926 end ends in 2004. Therefore, it includes a period of a good predictive performance of the dividend-price ratio occurring during early 2000's. The main characteristics of the data set used in Economy II and Economy III, is that it excludes the pre-WWII period-the period of the weak predictability of returns. The data for Economy II uses dividends reinvested into the one month T-bill while the data for Economy III employs dividends reinvested into S&P 500. As Section 1 of this paper shows, reinvestment of dividends into S&P 500 leads to an MA structure for the dividend-price ratio and, consequently, to an additional variable entering the predictive relation between returns and the dividend-price ratio which is omitted by the methods we consider.

Tab. 4.6: Test of economies

This table presents the results of applying different predictability methods to the stock market data. The data samples correspond to the ones used in Cochrane (2008b), Binsbergen and Koijen (2010) and Bansal and Yaron (2004)– the theoretical frameworks we use to construct our economies. The first row reports the sample size, the rest report R^2_{OOS} in %. Numbers in round brackets are the percentage of the simulation paths with the value of R^2_{OOS} of more or equal to its value obtained in the data.

Method	Null economy	Economy I	Economy II	Economy III
N	70	79	62	62
Predictive regression	−14.21	−9.96	−0.26	0.44
	(97)	(94)	(64)	(51)
Stambaugh (1999)	−26.07	−21.41	−8.57	−6.04
	(99)	(96)	(67)	(51)
Lewellen (2004)	−1.36	2.49	6.94	6.24
	(7)	(36)	(20)	(27)
Ashley (2006)	0.60	1.12	10.96	10.48
	(16)	(50)	(12)	(10)
Connor (1997)	−0.43	2.12	3.30	3.24
	(45)	(53)	(55)	(43)
Ferreira and Santa-Clara (2011)	3.56	3.91	6.98	6.75
	(3)	(60)	(36)	(33)

Table 4.6 confirms evidence for the sensitivity of predictability of returns to the data samples. Predictability of returns is sensitive to the inclusion of the pre-WWII period, since all the methods produce lower R^2_{OOS} in the Null economy and Economy I. Comparing with the data set used in the Null economy, an inclusion of the early 2000s into the sample leads to some improvement, while an exclusion of the pre-WWII data leads to a considerable improvement in the performance of all the methods. Moreover, existence

[5]See, for example, Goyal and Welch (2008).

of the additional variable in the predictive relation leads to a slight deterioration of the predictive performance of Lewellen (2004), Ashley (2006), Connor (1997) and Ferreira and Santa-Clara (2011) since for these methods the values of the $R^2_{OOS}s$ observed in Economy II are larger than the ones in Economy III.

Furthermore, Table 4.6 reveals a large dispersion of the distribution of the R^2_{OOS} produced by the method of Stambaugh (1999) and by the conventional predictive regression. The dispersion for the method of Stambaugh (1999) is very large. The difference between the highest and the lowest value of the R^2_{OOS} produced by this method is over 20 %. Note that in all the data sets this method is over performed by the historical mean as indicated by negative and high in absolute value R^2_{OSS}s. For the conventional predictive regression, the dispersion of the distribution of the R^2_{OOS} is considerably smaller, since the difference between the lowest and the highest value of the R^2_{OOS} is over 14 %. This method produces an only one slightly positive value of the R^2_{OOS} confirming a systematic failure of the conventional predictive regression to predict returns out-of-sample reported by Goyal and Welch (2008). The method of Stambaugh (1999) and the conventional predictive regression are followed by the shrinkage of Ashley (2006) with the difference between the highest and the lowest value of the R^2_{OOS} of around 10 %. This method does not produce any negative R^2_{OOS}. Moreover, it performs especially well in the post-WWII data producing the value of the R^2_{OOS} of over 10 % in Economy II and Economy III–the maximum across all the methods.

The method of Lewellen (2004) seems to be characterized by a lower dispersion of the R^2_{OOS}. This method produces a negative R^2_{OOS} in the data used in the Null economy and an R^2_{OOS} of almost 7 % in the post-WWII data of Economy II. The lowest dispersion, however, is observed for the methods of Connor (1997) and of Ferreira and Santa-Clara (2011). The shrinkage of Connor (1997) produces a negative R^2_{OOS} in the data set used to calibrate the Null economy and R^2_{OOS} of over 3 % in the post-WWII data of Economy II and Economy III. The sum-of-parts method of Ferreira and Santa-Clara (2011) does not produce a negative R^2_{OOS} in any data sample used. Note that the method of Ferreira and Santa-Clara (2011) is the only one significantly outperforming the historical mean in the data set of the Null economy. In the post-WWII the value of the R^2_{OOS} it produces is also quite high–almost 7 %.

Below each of the values of the R^2_{OOS}, in the round brackets we report the fraction of the simulation paths with realizations of the R^2_{OOS} more or equal to the value obtained in the data. Recall from Section 1 that the Null economy is characterized by the unusual predictability patterns. Comparing with the data, in this economy the dividend-price ratio is less persistent, less volatile and predicts dividend growth rather than returns with the slope coefficient in both regressions being negative.

Judging from the results for the conventional predictive regression, the Null economy is not rejected by the data. The numbers in Table 4.6 indicate that the probability of observing the value of R^2_{OOS} as high as -14.21 in the Null economy is 97 %. The method of Stambaugh (1999) produces even stronger evidence in favour of the model by having 99 % of observing the value of the R^2_{OOS} larger or equal to the one obtained in the data.

Interestingly, the method of Connor (1997) also does not have an ability to reject the Null economy. In fact, it gives roughly equal evidence in favor of each of the economies. A relatively high power of this method to reject the Null economy in favor of the alternative economies reported in Table 4.5, then might signal that all of the economies miss some features of the data. Indeed, Lettau and Van Nieuwerburgh (2007) and Koijen and Van Nieuweburgh (2010) provide evidence in favor of structural breaks in the stock market data that are not modeled by any of the economies. The evidence for a weak power of the method of Connor (1997) from Table 4.6 might then indicate vulnerability of the performance of this method to structural breaks.

The method of Lewellen (2004) and of Ferreira and Santa-Clara (2011) appear to be the most powerful in detecting the unusual predictability patterns of the Null economy. Results presented in Table 4.6 show that applying methods of Lewellen (2004) and Ferreira and Santa-Clara (2011) in the Null economy produces only 7 % and 3 % of R^2_{OOS} larger than the value observed in the stock market data.

We devote the next subsection to explaining the differences in the performance of the methods in the alternative economies.

4.3.2 Decomposition of the MSE

Ferreira and Santa-Clara (2011) using a simulation design which corresponds to our Economy II find that it is an excessive variance of the estimate of expected returns that is responsible for the poor performance of the conventional predictive regression and that it is a low variance of the estimate which leads to a lower mean square error produced by the historical mean. Additionally, the estimate of expected returns produced by the historical mean has a low correlation with the true expected returns. Their sum-of-parts method achieves a better combination of the variance of the estimate and the covariance of the estimate with true expected returns resulting in a lower than that of the historical mean mean square error and a high correlation of the estimate with the true expected returns. This section confirms and extends their results using the other economies for the simulation design and the other methods to predict stock returns.

Following Ferreira and Santa-Clara (2011) we start the analysis with scatter plots of the estimates of expected returns produced by each of the methods against the true expected returns. Figure 4.1 presents the scatter plot for Economy III. We find that the relative performance of the methods is similar in all the alternative economies. To save the space, therefore, we present the scatter plot only for Economy III.

From Figure 4.1, the historical mean does not capture well the true expected returns. The graph is parallel to the x-coordinate implying that the estimate by the historical mean severely understates high and overstates low realizations of the true expected returns.

More points are concentrated on the 45-degree line for the conventional predictive regression. The points, however, are also more dispersed with some extreme outliers. Campbell and Thompson (2008) document an improvement of the out-of-sample performance of the conventional predictive regression if negative estimates of expected returns produced by this method are replaced by zero. Figure 4.1 explains why. The conventional predictive regression produces a large number of extreme negative estimates not matched by the true expected returns realizations. Additionally, this method performs badly in capturing high realizations of the true expected returns.

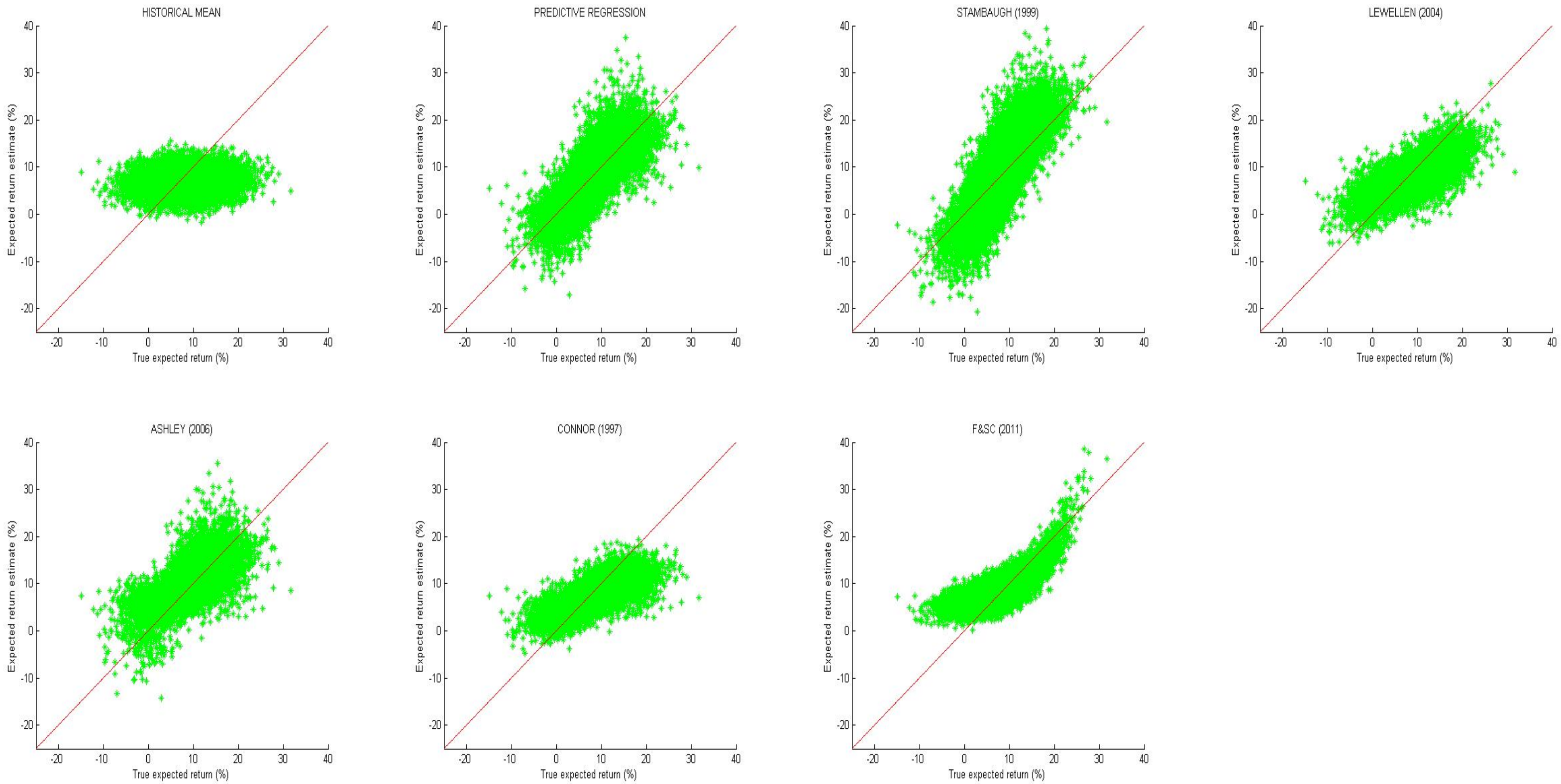

Fig. 4.1: Relative performance of methods, Economy III

In this figure we plot the estimates of expected returns produced by each method against the true expected returns in Economy III. The solid line is a 45 degree line.

The bad performance of the conventional predictive regression in the upper tail of the realizations of the true expected returns worsens after applying Stambaugh (1999) correction. Comparing with the scatter plot for the predictive regression, more points are concentrated along the 45-degree line. However, there is even a higher number of extreme positive estimates.

The method of Lewellen (2004) is a visible improvement over the previous methods. The points on the scatter plot are less dispersed and this method produces significantly lower number of extreme negative and extreme positive estimates not matched by the true realizations. The method of Ashley (2006) produces a higher number of the extreme estimates than the method by Lewellen (2004), while the method by Connor (1997) understates large realizations of expected returns.

More points are located close to the 45-degree line for the method of Ferreira and Santa-Clara (2011). The points on the scatter plot are less dispersed. Moreover, this method captures better than the alternatives high realizations of the true expected returns and does not produce extreme negative estimates.

We support the above analysis in Table 4.7 by decompose the mean square error (MSE) as follows:

$$MSE = E\left[\frac{1}{T-s_0}\sum_{s=s_0}^{T-1}(\hat{\mu}_s - r_{s+1})^2\right] = \frac{1}{T-s_0}\sum_{s=s_0}^{T-1}[E(\hat{\mu}_s - r_{s+1})]^2 + \frac{1}{T-s_0}\sum_{s=s_0}^{T-1} var(\hat{\mu}_s) + \\ + \frac{1}{T-s_0}\sum_{s=s_0}^{T-1} var(r_{s+1}) - 2\frac{1}{T-s_0}\sum_{s=s_0}^{T-1} cov(\hat{\mu}_s, r_{s+1}). \tag{4.3.2}$$

We refer to the first term in the decomposition as to the *square bias*, the second component is the *variance of estimate*, the third is the *variance of target* and the forth, the *covariance term*–corresponds to covariance of the estimate of expected returns with the realized returns. The value of the *variance of target* is the same across all the methods. The sum of the other elements can add to or subtract from the value of the *variance of target* resulting in a higher or lower MSE. In Table 4.7, we combine the *variance of estimate* and the *square bias* with the *covariance term* into a single element which we call the *difference* in order to measure the net effect of a change in the values of these components on MSE. Additionally, we report the values of the elements of the decomposition for the true expected returns. Since using the true expected return instead of an estimate leads to the highest mean of the R^2_{OOS}, it can serve as the benchmark.

Table 4.7 shows that it is the variance of the realized returns which contributes the most to the MSE. Applying each method results in a value of the MSE which only slightly deviates from the value of the *variance of target*. The values for the true mean set up a lower bound on the MSE. Therefore, the last column of Table 4.7 implies that the minimum value of the *difference* attainable in the alternative economies ranges from 5 % to 8 % of the value of the variance of the realized returns.

As it has been already mentioned in Ferreira and Santa–Clara (2011) the main advantage of the historical mean is a low variance of the estimate of expected returns it produces. The *variance of estimate* is the lowest comparing with all the methods in all the economies. However, the magnitude of the covariance of the estimate with the realized returns is also the smallest and negative.

Note that the sign of the covariance of the estimate by the historical mean depends on the presence and the strength of the discount rate effect. Using Economy III, a simple algebra yields:

$$cov\left[\frac{1}{s}\sum_{j=1}^{s} r_j, r_{s+1}\right] = cov\left[\frac{1}{s}\sum_{j=1}^{s} r_j, \mu_s\right] = cov\left[\frac{1}{s}\sum_{j=1}^{s}\left(\mu_{j-1}+\epsilon^r_j\right), a^\mu + b^\mu \mu_{s-1} + \epsilon^\mu_s\right] =$$
$$= \frac{1}{s}\left[b^\mu var(\mu_{j-1}) + cov(\epsilon^\mu_s, \epsilon^r_s)\right] = \frac{1}{s}\left[\frac{b^\mu var(\epsilon^\mu_s)}{1-(b^\mu)^2} + cov(\epsilon^\mu_s, \epsilon^r_s)\right].$$

Therefore, an estimate of expected returns produced by the historical mean will have a positive correlation with the true expected returns if:

$$\frac{b^\mu}{1-(b^\mu)^2} > -\beta^{\epsilon^\mu,\epsilon^r}. \tag{4.3.3}$$

Tab. 4.7: Decomposition of MSE

This table reports values of the elements of the decomposition of the mean square error (MSE) following Eq.(4.3.2) in the text.
The *square bias* is given by $\frac{1}{T-s_0}\sum_{s=s_0}^{T-1}[E(\hat{\mu}_s - r_{s+1})]^2$, the *variance of estimate* is give by $\frac{1}{T-s_0}\sum_{s=s_0}^{T-1} var(\hat{\mu}_s)$, the *variance of target* is given by $\frac{1}{T-s_0}\sum_{s=s_0}^{T-1} var(r_{s+1})$ and the *covariance term* is given by $2\frac{1}{T-s_0}\sum_{s=s_0}^{T-1} cov(\hat{\mu}_s, r_{s+1})$.
The *difference* is defined as:

$$difference = \sum_{s=s_0}^{T-1}[E(\hat{\mu}_s - r_{s+1})]^2 + \sum_{s=s_0}^{T-1} var(\hat{\mu}_s) - 2\sum_{s=s_0}^{T-1} cov(\hat{\mu}_s, r_{s+1}).$$

In columns headed *Value* we report a value of each element multiplied by $100 \times (T - s_0)$. In columns headed % we report a value of each element dividend by the value of $var(r_{s+1})$ in %.

	Historical mean		Predictive regression		Stambaugh (1999)		Lewellen (2004)		Ashley (2006)		Connor (1997)		Ferreira and Santa-Clara (2011)		**True mean**	
	Value	%	Value	%	Value	%	Value	%	Value	%	Value	%	Value	%	**Value**	**%**
Panel A: Economy I																
Square bias	0.0006	0	0.0001	0	0.0009	0	0.0002	0	0.0161	0	0.0002	0	0.0005	0	**0.0000**	**0**
Variance of estimate	0.0609	2	0.4282	11	0.7554	19	0.1533	4	0.2305	6	0.0770	2	0.0915	2	**0.1991**	**5**
Variance of target	3.9607	100	3.9607	100	3.9607	100	3.9607	100	3.9607	100	3.9607	100	3.9607	100	**3.9607**	**100**
Covariance term	-0.0240	1	0.3640	9	0.5687	14	0.1139	3	0.2032	5	0.0993	3	0.1831	5	**0.4049**	**10**
MSE	4.0456	102	4.0196	101	4.1397	105	3.9987	101	4.0010	101	3.9373	99	3.8685	98	**3.7549**	**95**
Difference	**0.0849**	**2**	**0.0589**	**1**	**0.1790**	**5**	**0.0380**	**1**	**0.0403**	**1**	**-0.0234**	**1**	**-0.0922**	**2**	**-0.2058**	**5**
Panel B: Economy II																
Square bias	0.0244	1	0.0016	0	0.0317	1	0.0087	0	0.0133	1	0.0102	0	0.0007	0	**0.0000**	0
Variance of estimate	0.0264	1	0.2697	12	0.5475	24	0.0525	2	0.1295	6	0.0366	2	0.0511	2	**0.1918**	**8**
Variance of target	2.3177	100	2.3177	100	2.3177	100	2.3177	100	2.3177	100	2.3177	100	2.3177	100	**2.3177**	**100**
Covariance term	-0.0201	1	0.2970	13	0.4870	21	0.0889	4	0.1862	8	0.0824	4	0.1531	7	**0.3811**	**16**
MSE	2.3880	103	2.2881	99	2.4031	104	2.2892	99	2.2720	98	2.2813	98	2.2157	96	**2.1283**	**92**
Difference	**0.0704**	**3**	**-0.0296**	**1**	**0.0855**	**4**	**-0.0284**	**1**	**-0.0457**	**2**	**-0.0364**	**2**	**-0.1019**	**4**	**-0.1894**	**8**
Panel C: Economy III																
Square bias	0.0290	1	0.0043	0	0.0634	1	0.0045	0	0.0126	0	0.0103	0	0.0107	0	**0.0000**	**0**
Variance of estimate	0.0688	2	0.4994	12	1.0771	25	0.1614	4	0.2934	7	0.0901	2	0.1402	3	**0.2946**	**7**
Variance of target	4.2851	100	4.2851	100	4.2851	100	4.2851	100	4.2851	100	4.2851	100	4.2851	100	**4.2851**	**100**
Covariance term	-0.0036	0	0.4552	11	0.8056	19	0.2115	5	0.2899	7	0.1431	3	0.2899	7	**0.5741**	**13**
MSE	4.3855	102	4.3270	101	4.6074	108	4.2372	99	4.2972	100	4.2407	99	4.1439	97	**4.0003**	**93**
Difference	**0.1005**	**2**	**0.0420**	**1**	**0.3224**	**8**	**-0.0478**	**1**	**0.0122**	**0**	**-0.0444**	**1**	**-0.1412**	**3**	**-0.2847**	**7**

The *square bias* is of some importance for the historical mean. Comparing with the other methods, the *difference* is high in Economy II and Economy III and the *squared bias* contributes over a third of its value.

Confirming the result of Ferreira and Santa-Clara (2011), the estimate by the conventional predictive regression is characterized by an excessive variance. The value of the *covariance term* for the predictive regression is high, close to the corresponding value for the true expected returns. A high value of the *variance of estimate*, however, eliminates the benefit of a high covariance.

We find that the method of Stambaugh (1999) is even noisier than the conventional predictive regression and is possibly biased. Comparing with the conventional predictive regression, both the *covariance term* and the *variance of estimate* increase. The *variance of estimate*, however, increases by more, resulting in a high value of the *difference* and significantly increasing MSE. The value of the bias is the largest across all the methods. The *square bias* accounts for more than a third of the value of the *difference* in Economy II and over one fifth of the value of the *difference* in Economy III.

Applying the methods of Lewellen (2004), Ashley (2006) and Connor (1997) results in an improvement over the conventional predictive regression. The methods of Lewellen (2004) and Connor (1997) results in a negative *difference*, which decreases the value of the *variance of target* to deliver a 2 % lower MSE than the MSE for the conventional predictive regression. The covariance, however, is somewhat low for these methods if compared with the others. For the method by Ashley (2006), the value of the *covariance term* is roughly equal to the *variance of estimate* plus the *square bias* resulting in a zero net effect on the *variance of target*. For this method, the values of the *variance of estimate* and the *square bias* are somewhat high. The MSE for this method is 1 % lower than the MSE produced by the conventional predictive regression.

The sum-of-parts method by Ferreira and Santa-Clara (2011) leads to the lowest *difference* if compared with all the other methods. This method achieves a significant decrease in the *variance of estimate* while keeping a moderate value of the *covariance term*. The resulting MSE of the estimate is 4 % lower than the MSE of the conventional predictive regression.

The elements of the decomposition can be combined to produce the squared correlation coefficient:

$$corr(\hat{\mu}_s, r_{s+1})^2 = \sum_{s=s_0}^{T-1} \left[\frac{cov(\hat{\mu}_s, r_{s+1})}{std(\hat{\mu}_s) std(r_{s+1})} \right]^2. \tag{4.3.4}$$

The values of the squared correlation coefficient are reported in Table 4.8. Our simulation set up allows a calculation of the squared correlation of an estimate with the true expected returns. In Panel B of Table 4.8 we report values of the squared correlation replacing the realized returns with the true expected returns in Eq.(4.3.4). Note that in this case the squared correlation can be interpreted as the R^2 in the regression of the true expected returns on their estimate. It measures the fraction of the variation in the true expected returns captured by the estimate.

Tab. 4.8: Variation in expected returns

In Panel A of this table we present the values in % of $corr(\hat{\mu}_s, r_{s+1})^2$ calculated as:

$$corr(\hat{\mu}_s, r_{s+1})^2 = \sum_{s=s_0}^{T-1} \left[\frac{cov(\hat{\mu}_s, r_{s+1})}{std(\hat{\mu}_s) std(r_{s+1})} \right]^2.$$

In Panel B, we present the values in % of $corr(\hat{\mu}_s, \mu_s)^2$ calculated as:

$$corr(\hat{\mu}_s, \mu_s)^2 = \sum_{s=s_0}^{T-1} \left[\frac{cov(\hat{\mu}_s, \mu_s)}{std(\hat{\mu}_s) std(\mu_s)} \right]^2.$$

Method	Economy I	Economy II	Economy III
Panel A: Correlation the estimate with the realized returns			
Historical mean	0.09	0.22	0.10
Predictive regression	2.10	3.78	2.63
Stambaugh (1999)	2.82	4.86	3.66
Lewellen (2004)	0.76	2.05	1.89
Ashley (2006)	1.28	3.05	1.92
Connor (1997)	1.10	2.51	1.70
Ferreira and Santa-Clara (2011)	2.41	5.05	3.56
Panel B: Correlation the estimate with the true expected returns			
Historical mean	1.64	4.19	1.09
Predictive regression	39.90	41.79	40.76
Stambaugh (1999)	54.08	56.38	56.78
Lewellen (2004)	14.39	17.81	29.75
Ashley (2006)	24.27	30.19	29.77
Connor (1997)	20.63	21.41	26.40
Ferreira and Santa-Clara (2011)	45.94	58.24	56.02

Only a small fraction of realized returns is predictable as implied by the large contribution of the variance of the realized returns into the MSE. In Table 4.7, the covariance term constitutes from 1% to 21 % of the value of the variance of the realized returns across the methods. This leads to small values of the squared correlation coefficient reported in Panel A of Table 4.8, indicating that the methods we consider are able to capture 5.05 % of the variation in the realized returns at maximum. Panel B, however, reveals that the methods capture a significant part of the variation in expected returns. All the predictive methods except the historical mean produce values of more than 20 % , while for some methods the values exceed 50 % in all the alternative economies.

Comparing the values across the methods, the historical mean captures from 1.09 % to 4.19 % of the variation of the true expected returns. These values are very low not only in absolute but also in relative terms. Thus, the results of Table 4.8 indicate that if expected returns vary, the historical mean explains little of their variation.

Though frequently producing MSE higher than the MSE of the historical mean, the conventional predictive regression captures a significant part of the variation in expected return. The value of the squared correlation for the conventional regression is also high in comparison with the other methods.

The methods of Lewellen (2004), Ashley (2006) and Connor (1997) have mediocre performance while the methods of Stambaugh (1999) and the sum-of-parts method of Ferreira and Santa–Clara (2011) show the best performance in capturing variation in the true expected returns. The methods of Stambaugh (1999) and of Ferreira and Santa–Clara (2011) allow to explain from 46 % to 57 % of their variation.

To summarize this section, a low MSE produced by the historical mean is entirely due to a low variance of its estimate of expected returns. The estimate of expected returns produced by the historical mean is biased and has a low correlation with the true expected returns. Extremely large and extremely low estimates produced by the conventional predictive regression lead to an excessively high variance of its expected returns estimate. This results in a poor out-of-sample performance. The method of Stambaugh (1999) produces even a higher number of extreme positive estimates than the predictive regression. This leads to even higher noise and a possible bias in the estimate of expected returns, substantially increasing MSE. While those methods fail to beat the historical mean if judged by R^2_{OOS}, they produce an estimates which explain a substantial part of variation in the true expected returns.

Applying the methods of Lewellen (2004), Ashley (2006) and Connor (1997) leads to a somewhat lower MSE than applying the conventional predictive regression due to a lower noise in the estimate of expected returns. These methods, however, are characterized by a lower than the conventional predictive regression correlation with the true expected returns. The method of Ferreira and Santa-Clara (2011) performs better than the conventional predictive regression producing a significantly lower MSE and a higher correlation with the true expected returns. Only a small part of the realized returns is predictable. The variance of the realized returns dominates the other element of the decomposition and hides the high correlations with the true expected returns produced by the forecasting methods.

4.4 Conclusion

In this paper, we study the out-of-sample predictability of stock returns. We use the models for expected returns, expected dividend growth together with the innovation to the realized dividend growth to set up a simulation exercise. Imposing the Campbell and Shiller (1988) present value identity then leads to a completely controlled environment to analyze the conventional predictive regression and its alternatives proposed in the literature.

Our results show that all the methods but Connor (1997) and Ferreira and Santa - Clara (2011) tend to perform poorly out of sample and all the methods except of Lewellen (2004), and Ferreira and Santa-Clara (2011) have a low statistical power. Further analysis indicates that most of the methods either overestimate or underestimate true expected returns producing too much or too little variation in the estimates. In other words, most of the methods produce estimates of expected returns which are characterized by an excessively high variance or by an excessively low covariance with the true expected returns.

Furthermore, the poor out-of-sample performance of a given method does not imply a superiority of the expected returns estimate by the historical mean. Whenever expected returns vary, the historical mean is not able to capture their variation. Moreover, it possibly produces a biased estimate of the true expected returns.

Additionally, we show that a large unpredictable component of the realized returns conceals high correlations of the estimates produced by the forecasting methods with the true expected returns.

The sum-of-parts method of Ferreira and Santa-Clara (2011) is superior to the other methods we consider. This method avoids predicting returns by running a regression on a valuation ratio and, therefore, does not face the Stambaugh (1999) problem and incorporates an additional proxy for the expected dividend growth entering the implied predictive relations whenever expected returns and expected dividend growth are characterized by a different persistence. The good performance of the method comes from a low volatility of the estimate of expected returns which is highly correlated with the true expected returns.

BIBLIOGRAPHY

Ashley, R., 2006. Beyond optimal forecasting. Unpublished working paper. Virginia Polytechnic Institute and State University; Blackburs VA.

Bansal, R., Yaron, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Bansal, R., Gallant, R., Tauchen, G., 2007. Rational pessimism, rational exuberance, and asset pricing models. Review of Economic Studies 74, 1005–1033.

Bansal, R., Kiku, D., Yaron, A., 2007a. Risks for the long run: estimation and inference. Unpublished working paper.

Bansal, R., Kiku, D., Yaron, A., 2007b. A note on the economics and statistics of predictability: a long run risks perspective. Unpublished working paper.

Bansal, R., Kiku, D., Yaron, A., 2009. An empirical evaluation of the long—run risks model for asset prices. Unpublished working paper.

Bansal, R., Kiku, D., Yaron, A., 2010. Long-run risks, the macroeconomy, and asset prices. American Economic Review 100(2), 542–546.

Bonomo, M., Garcia, R., Meddahi, N., Tedongap, R., 2010. Generalized disappointment aversion, long-run volatility risk, and asset prices. Forthcoming in The Review of Financial Studies.

Beeler, J., Campbell, J., 2009. The long–run risks model and aggregate asset prices: an empirical assessment. Unpublished working paper.

Binsbergen, J., Koijen, R., 2010. Predictive regressions: a present-value approach. Journal of Finance 65, 1439–1471.

Campbell, J., 1991. A Variance decomposition of stock returns. Economic Journal 405, 157–179.

Campbell, J., 1991. Intertemporal asset pricing without consumption data. American Economic Review 83(3), 487–512.

Campbell, J., 1996. Understanding risk and return. Journal of Political Economy 104(2), 298–345.

Campbell, J., 1988, Asset prices, consumption and the business cycle.

Campbell, J., Shiller, R., 1988. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Campbell, J., Thompson, S., 2008. Predicting the equity premium out of sample: can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Campbell, J., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.

Cavanagh, C., Elliott, G., Stock, J., 1995. Inference in models with nearly integrated regressors. Econometric Theory 11, 1131–1147.

Cochrane, J., 2008a. State–space v.s. VAR models for stock returns. Unpublished working paper.

Cochrane, J., 2008b. The dog that did not bark: a defense of return predictability. Review of Financial Studies 21, 1533–1575.

Connor, G., 1997. Sensible return forecasting for portfolio management. Financial Analysts Journal 53, 44–51.

Constantinides G., Ghosh, A., 2009. Asset pricing tests with long run risks in consumption growth. Unpublished working paper.

Fama, E., French, K., 1988a. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.

Fama, F., French, K., 1988b. Permanent and temporary components of stock prices. Journal of Political Economy 96(2), 246–273.

Fama, E., French, K., 1989. Business conditions and expected returns on stock and bonds. Journal of Financial Economics 25, 23–49.

Fama, E., French, K., 2002. The equity premium. Journal of Finance 57, 637–659.

Ferson, W., Sarkissian, S., Simin, T., 2003. Spurious regressions in financial economics. Journal of Finance 58, 1393–1413.

Ferreira, M., Santa-Clara, P., 2011. Forecasting stock market returns: the sum of the parts is more than the whole. Journal of Financial Economics 100, 514–537.

Foster, F. D., Smith, T., Whaley, R., 1997. Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R^2. Journal of Finance 52, 591–607.

Fulvio, O., Tamoni, A., Tebaldi, C., 2010. Long run risk and the persistence of consumption shocks. Unpublished working paper.

Hansen, L., Sargent, T., 1983. Aggregation over time and the inverse optimal predictor problem for adaptive expectations in continuous time. International Economic Review 1, 1–20.

Heaton, J., 1995. An empirical investigation of asset pricing with temporally dependent preference specifications. Econometrica 3, 681–717.

Kandel, S., Stambaugh, R., 1991. Asset returns and intertemporal preferences. Journal of Monetary Economics 27, 39–71.

Koijen., R., Van Nieuwerburgh, S., 2010. Predictability of returns and cash flows. Unpublished working paper.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815–849.

Lettau, M., Van Nieuwerburgh, S., 2008, Reconciling the return predictability evidence. Review of Financial Studies 21(4), 1607–1652.

Lettau, M., Ludvigson, S., Wachter, J., 2008, The declining equity premium: what role does the macroeconomic risk play? Review of Financial Studies 21(4), 1653–1687.

Lewellen, J., 2004. Predicting returns with financial ratios. Journal of Financial Economics 74, 209–235.

Pastor, L., Stambaugh, R., 2009. Predictive systems: living with imperfect predictors. Journal of Finance 64, 1583–1628.

Pastor, L., Stambaugh, R., 2009. The equity premium and structural breaks. Journal of Finance 56(4), 1207–1239.

Rapach, D., Strauss, J., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. The Review of Financial Studies 23(2), 821–862.

Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68, 201–232.

APPENDIX

A. CHAPTER 2

A.1 Derivation of the state-space representation for returns in the long run risks model

This Appendix provides derivations of the state-space representation for returns in the BY model, Eq.(2.3.1)-Eq.(2.3.3) in the text. The derivations employ Eq.(2.1.2)-Eq.(2.1.4), Eq.(2.1.7) and Eq.(2.1.9). More specifically the following system of equations is used:

$$x_{t+1} = \rho x_t + \varphi_e \sigma_t e_{t+1}, \tag{A.1.1}$$

$$\sigma_{t+1}^2 = \bar{\sigma}^2(1-\nu_1) + \nu_1 \sigma_t^2 + \sigma_\omega \omega_{t+1}, \tag{A.1.2}$$

$$\Delta d_{t+1} = m_d + \phi x_t + \varphi_d \sigma_t u_{d,t+1}, \tag{A.1.3}$$

$$r_{m,t+1} = k_{0m} - k_{1m} dp_{t+1} + dp_t + \Delta d_{t+1}, \tag{A.1.4}$$

$$dp_{t+1} = -A_{0m} - A_{1m} x_{t+1} - A_{2m} \sigma_{t+1}^2. \tag{A.1.5}$$

Substituting Eq.(A.1.1) and Eq.(A.1.2) into Eq.(A.1.5):

$$dp_{t+1} = -A_{0m} - A_{2m}\bar{\sigma}^2(1-\nu_1) - A_{1m}\rho x_t - A_{2m}\nu_1\sigma_t^2 - A_{1m}\varphi_e\sigma_t e_{t+1} - A_{2m}\sigma_\omega\omega_{t+1}. \tag{A.1.6}$$

Eq.(A.1.3) implies:

$$\mu_t^d = E_t\left[\Delta d_{t+1}\right] = m_d + \phi x_t. \tag{A.1.7}$$

Using Eq.(A.1.1), then leads to the following dynamics for expected dividend growth (Eq.(2.3.2) in the text):

$$\mu_t^d = (1-\rho)m_d + \rho\mu_{t-1}^d + \phi\varphi_e\sigma_{t-1}e_t, \tag{A.1.8}$$

Using Eq.(A.1.3), Eq.(A.1.5) and Eq.(A.1.6) in Eq.(A.1.4) leads:

$$\begin{gathered}
r_{m,t+1} = k_{0m} - k_{1m}\left[-A_{0m} - A_{2m}\bar{\sigma}^2(1-\nu_1) - A_{1m}\rho x_t - A_{2m}\nu_1\sigma_t^2\right] + \\
+k_{1m}\left[A_{1m}\varphi_e\sigma_t e_{t+1} + A_{2m}\sigma_\omega\omega_{t+1}\right] - A_{0m} - A_{1m}x_t - A_{2m}\sigma_t^2 + m_d + \phi x_t + \varphi_d\sigma_t u_{d,t+1} = \\
= \left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d\right] + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]x_t - \\
-A_{2m}(1-k_{1m}\nu_1)\sigma_t^2 + \left[\varphi_d\sigma_t u_{d,t+1} + k_{1m}A_{1m}\varphi_e\sigma_t e_{t+1} + k_{1m}A_{2m}\sigma_\omega\omega_{t+1}\right]
\end{gathered} \tag{A.1.9}$$

This implies:

$$\begin{gathered}
\mu_t^r = E_t\left[r_{m,t+1}\right] = k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d - \\
-A_{2m}(1-k_{1m}\nu_1)(1-\nu_1)\bar{\sigma}^2 + \rho\left[\phi - A_{1m}(1-k_{1m}\rho)\right]x_{t-1} - \nu_1 A_{2m}(1-k_{1m}\nu_1)\sigma_{t-1}^2 + \\
+\left[\phi - A_{1m}(1-k_{1m}\rho)\right]\varphi_e\sigma_{t-1}e_t - A_{2m}(1-k_{1m}\nu_1)\sigma_\omega\omega_t = \\
= (1-\nu_1)\left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d - A_{2m}(1-k_{1m}\nu_1)\bar{\sigma}^2\right] + \\
+\nu_1\mu_{t-1}^r + (\rho-\nu_1)\left[\phi - A_{1m}(1-k_{1m}\rho)\right]x_{t-1} + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]\varphi_e\sigma_{t-1}e_t - \\
-A_{2m}(1-k_{1m}\nu_1)\sigma_\omega\omega_t,
\end{gathered}$$

Noting that:

$$\phi - A_{1m}(1-k_{1m}\rho) = \frac{1}{\psi};$$

and using Eq.(A.1.7):

$$\begin{gathered}
\mu_t^r = (1-\nu_1)\left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d - A_{2m}(1-k_{1m}\nu_1)\bar{\sigma}^2\right] - \\
-\frac{(\rho-\nu_1)m_d}{\phi\psi} + \nu_1\mu_{t-1}^r + \frac{(\rho-\nu_1)}{\phi\psi}\mu_{t-1}^d + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]\varphi_e\sigma_{t-1}e_t - A_{2m}(1-k_{1m}\nu_1)\sigma_\omega\omega_t,
\end{gathered}$$

establishing Eq.(2.3.1) in the text.

A.2 Derivation of the present value identity

To obtain the relation between the dividend-price ratio and expected returns and expected dividend growth explored in Table 2.6, start with the SSR:

$$\mu^r_{t+1} = \alpha^{\mu r} + \delta^{\mu r}\mu^r_t + \tau\mu^d_t + \epsilon^{\mu r}_{t+1},$$
$$\mu^d_{t+1} = \alpha^{\mu d} + \delta^{\mu d}\mu^d_t + \epsilon^{\mu d}_{t+1},$$
$$\Delta d_{t+1} = \mu^d_t + \epsilon^d_{t+1}.$$

Iterating Eq.(A.1.4) for dp_t:

$$dp_t = -\frac{k_{0m}}{1-k_{1m}} + \sum_{j=0}^{\infty}(k_{1m})^j E_t\left[r_{m,t+1+j} - \Delta d_{t+1+j}\right] - (k_{1m})^{\infty} dp_{\infty}. \tag{A.2.1}$$

The no asset bubbles condition states:

$$(k_{1m})^{\infty} dp_{\infty} = 0. \tag{A.2.2}$$

$$E_t\left[r_{m,t+1+j}\right] = E_t\left[E_{t+j}\left(r_{m,t+1+j}\right)\right] = E_t\left[\mu^r_{t+j}\right] = E_t\left[\alpha^{\mu r} + \delta^{\mu r}\mu^r_{t+j-1} + \tau\mu^d_{t+j-1} + \epsilon^{\mu,r}_{t+j}\right] =$$
$$= \alpha^{\mu r}\sum_{i=0}^{j-1}(\delta^{\mu r})^i + (\delta^{\mu r})^j\mu^r_t + \tau\sum_{i=0}^{j-1}(\delta^{\mu r})^i E_t\left[\mu^d_{t+j-1}\right].$$

Using the dynamics of $\mu^d_{t+j-1-i}$:

$$E_t\left[\mu^d_{t+j-1-i}\right] = \alpha^{\mu d}\sum_{y=0}^{j-2-i}(\delta^{\mu d})^y + (\delta^{\mu d})^{j-1-i}\mu^d_t.$$

This implies:

$$\tau\sum_{i=0}^{j-1}(\delta^{\mu r})^i E_t\left[\mu^d_{t+j-1}\right] = \frac{\tau\alpha^{\mu d}}{1-\delta^{\mu d}}\left[\frac{1-(\delta^{\mu r})^j}{1-\delta^{\mu r}} - \frac{(\delta^{\mu d})^j - (\delta^{\mu r})^j}{\delta^{\mu d}-\delta^{\mu r}}\right] + \tau\frac{(\delta^{\mu d})^j - (\delta^{\mu r})^j}{\delta^{\mu d}-\delta^{\mu r}}\mu^d_t.$$

Therefore:

$$E_t\left[\mu^r_{t+1}\right] = \alpha^{\mu r}\frac{1-(\delta^{\mu r})^j}{1-\delta^{\mu r}} + (\delta^{\mu r})^j\mu^r_t + \frac{\tau\alpha^{\mu d}}{1-\delta^{\mu d}}\left[\frac{1-(\delta^{\mu r})^j}{1-\delta^{\mu r}} - \frac{(\delta^{\mu d})^j-(\delta^{\mu r})^j}{\delta^{\mu d}-\delta^{\mu r}}\right] + \\ + \tau\mu^d_t\frac{(\delta^{\mu d})^j-(\delta^{\mu r})^j}{\delta^{\mu d}-\delta^{\mu r}}. \tag{A.2.3}$$

In a similar fashion:

$$E_t\left[\mu^d_{t+j}\right] = \alpha^{\mu d}\frac{1-(\delta^{\mu d})^j}{1-\delta^{\mu d}} + (\delta^{\mu d})^j\mu^d_t. \tag{A.2.4}$$

To obtain the identity linking the dividend-price ratio with expected returns and expected dividend growth used in Table 2.6 I plug Eq.(A.2.2), Eq. (A.2.3) and Eq.(A.2.4) into Eq.(A.2.1):

$$\begin{aligned} dp_t &= a^{dp} + \sum_{j=0}^{\infty}(k_{1m})^j\left[(\delta^{\mu r})^j\mu^r_t + \tau\frac{(\delta^{\mu d})^j-(\delta^{\mu r})^j}{\delta^{\mu d}-\delta^{\mu r}}\mu^d_t - (\delta^{\mu d})^j\mu^d_t\right] = \\ &= a^{dp} + \frac{\mu^r_t}{1-k_{1m}\delta^{\mu r}} - \frac{1-(\tau+\delta^{\mu r})k_{1m}}{(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})}\mu^d_t, \end{aligned} \tag{A.2.5}$$

with

$$a^{dp} = -\frac{k_{0m}}{1-k_{1m}} + \frac{\alpha^{\mu r}k_{1m}}{(1-k_{1m})(1-k_{1m}\delta^{\mu r})} + \frac{\tau\alpha^{\mu d}k_{1m}}{(1-\delta^{\mu d})(1-k_{1m})(1-k_{1m}\delta^{\mu r})} - \\ - \frac{\tau\alpha^{\mu d}k_{1m}}{(1-\delta^{\mu d})(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})} - \frac{\alpha^{\mu d}k_{1m}}{(1-k_{1m})(1-k_{1m}\delta^{\mu d})},$$

A.3 Derivation of predictive relations implied by the model

To obtain Eq.(2.3.7) start with Eq.(A.1.5) and use the dynamics of expected consumption growth, Eq.(A.1.1) and the dynamics of the time-varying volatility, Eq.(A.1.2):

$$
\begin{aligned}
dp_{t+1} &= -A_{0m} - A_{1m}x_{t+1} - A_{2m}\sigma_{t+1}^2 = -A_{0m} - A_{1m}(\rho x_t + \varphi_e \sigma_t e_{t+1}) - \\
&\quad -A_{2m}(\bar{\sigma}^2(1-\nu_1) + \nu_1\sigma_t^2 + \sigma_\omega\omega_{t+1}) = -A_{0m} - A_{2m}\bar{\sigma}^2(1-\nu_1) - \\
&\quad -A_{1m}\rho x_t - A_{2m}\nu_1\sigma_t^2 - A_{1m}\varphi_e\sigma_t e_{t+1} - A_{2m}\sigma_\omega\omega_{t+1}.
\end{aligned}
\tag{A.3.1}
$$

The second equality in Eq.(2.3.7) follows from regrouping the elements of Eq.(A.3.1). Using Eq.(A.1.7) then leads to the first equality in Eq.(2.3.7):

$$
\begin{aligned}
dp_{t+1} &= (1-\nu_1)\left[-A_{0m} - A_{2m}\bar{\sigma}^2\right] + \nu_1 dp_t - (\rho-\nu_1)\frac{\phi - \frac{1}{\psi}}{1-k_{1m}\rho}x_t + \\
&\quad + \left[-A_{1m}\varphi_e\sigma_t e_{t+1} - A_{2m}\sigma_\omega\omega_{t+1}\right] = \\
&= (1-\nu_1)\left[-A_{0m} - A_{2m}\bar{\sigma}^2\right] + (\rho-\nu_1)\frac{A_{1m}}{\phi}m_d + \nu_1 dp_t - (\rho-\nu_1)\frac{A_{1m}}{\phi}\mu_t^d + \\
&\quad + \left[-A_{1m}\varphi_e\sigma_t e_{t+1} - A_{2m}\sigma_\omega\omega_{t+1}\right].
\end{aligned}
\tag{A.3.2}
$$

In order to receive the first equality in Eq.(2.3.8) use the second equality of Eq.(A.3.2) in Eq.(A.1.4):

$$
\begin{aligned}
r_{m,t+1} &= k_{0m} - k_{1m}\left[(1-\nu_1)(-A_{0m} - A_{2m}\bar{\sigma}^2) + (\rho-\nu_1)\frac{A_{1m}}{\phi}m_d\right] - \\
&\quad -k_{1m}\nu_1 dp_t + k_{1m}(\rho-\nu_1)\frac{A_{1m}}{\phi}\mu_t^d - k_{1m}\epsilon_{t+1}^{dp} + dp_t + \mu_t^d + \epsilon_{t+1}^d = \\
&= \left[k_{0m} - k_{1m}\left((1-\nu_1)(-A_{0m} - A_{2m}\bar{\sigma}^2) + (\rho-\nu_1)\frac{A_{1m}}{\phi}m_d\right)\right] + \\
&\quad +(1-k_{1m}\nu_1)dp_t + \left(1 + k_{1m}(\rho-\nu_1)\frac{A_{1m}}{\phi}\right)\mu_t^d - k_{1m}\epsilon_{t+1}^{dp} + \epsilon_{t+1}^d.
\end{aligned}
$$

The second equality of Eq.(2.3.8) is obtained from the first by using Eq.(A.1.3) and:

$$
x_t = -\frac{A_{0m}}{A_{1m}} - \frac{1}{A_{1m}}dp_t - \frac{A_{2m}}{A_{1m}}\sigma_t^2. \tag{A.3.3}
$$

To obtain the first equality in Eq.(2.3.9), note that Eq.(A.2.5) implies:

$$
\mu_t^d = \alpha^{dp}\frac{(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})}{1+(\tau+\delta^{\mu r})k_{1m}} + \frac{(1-k_{1m}\delta^{\mu d})}{1-(\tau+\delta^{\mu r})k_{1m}}\mu_t^r - \frac{(1-k_{1m}\delta^{\mu r})(1-k_{1m}\delta^{\mu d})}{1+(\tau+\delta^{\mu r})k_{1m}}dp_t
$$

and substitute this expression into Eq.(2.3.3). Moreover, the second equality in Eq.(2.3.9) is obtained by using Eq.(A.3.3) in Eq.(A.1.3).

B. CHAPTER 3

B.1 Calibration of the long run risks model

This appendix provides the parameters used to calibrate the long run risks model. The calibration is the one used in Bansal and Yaron (2004). The following table provides the values of the parameters used to calibrate the model.

Tab. B.1: Calibration

This table presents the values of the parameters of the calibration of the processes describing the long run risks model. Panel A provides the calibration of the conditional means of the processes. Panel B, the calibration of the conditional variances. Panel C, the values of the preference parameters.

Panel A: Conditional means						
Parameter:	m	m_d	$\bar{\sigma}$	ρ	ν_1	ϕ
Value:	0.0015	0.0015	0.0078	0.9790	0.9870	3.0000
Panel B: Conditional variances						
Parameter:	φ_e	φ_d	σ_ω			
Value:	0.0440	4.5000	2.3000×10^{-6}			
Panel C: Preference parameters						
Parameter:	β	γ	ψ			
Value:	0.9980	10.0000	1.5000			

B.2 Derivation of the approximation to the aggregated dividend growth

This appendix provides derivation of the approximation of the aggregated dividend growth in terms of the monthly series, Eq.(3.2.8) in the text. The approximated consumption growth is obtained in a similar way and, to save the space, its derivation is omitted.

Eq.(3.2.7) in the text implies:

$$\Delta d^a_{12(t+1)} = log\frac{\sum_{j=0}^{11} D_{12(t+1)-j}}{\sum_{j=0}^{11} D_{12t-j}} = log\frac{\frac{\sum_{j=0}^{11} D_{12(t+1)-j}}{D_{12t}}}{\frac{\sum_{j=0}^{11} D_{12t-j}}{D_{12t}}}.$$

Note:

$$log\sum_{j=0}^{11}\frac{D_{12(t+1)-j}}{D_{12t}} = log\sum_{j=0}^{11}\prod_{k=0}^{11-j}\frac{D_{12(t+1)-j-k}}{D_{12(t+1)-j-k-1}}\approx b_1\sum_{j=0}^{11}\sum_{k=0}^{11-j}\Delta d_{12(t+1)-j-k}.$$

Similarly:

$$log\sum_{j=0}^{11}\frac{D_{12t-j}}{D_{12t}} = log\sum_{j=0}^{11}\prod_{k=0}^{j-1}\frac{D_{12t-j+k}}{D_{12t-j+k+1}}\approx -b_2\sum_{j=0}^{11}\sum_{k=0}^{j-1}\Delta d_{12t-j-k+1}.$$

Combining both:

$$\Delta d^a_{12(t+1)} \approx b_1\sum_{j=0}^{11}(j+1)\Delta d_{12(t+1)-j} + b_2\sum_{j=0}^{10}(11-j)\Delta d_{12t-j}, \tag{B.2.1}$$

B.3 Derivation of the state-space representation for the long run risks model

This appendix provides derivations of the monthly and the annual state-space representation for the BY model. The model is characterised by the following system of equations:

$$x_{t+1} = \rho x_t + \varphi_e\sigma_t e_{t+1}, \tag{B.3.1}$$

$$\sigma^2_{t+1} = \bar{\sigma}^2(1-\nu_1) + \nu_1\sigma^2_t + \sigma_\omega\omega_{t+1}, \tag{B.3.2}$$

$$\Delta d_{t+1} = m_d + \phi x_t + \varphi_d\sigma_t u_{d,t+1}, \tag{B.3.3}$$

$$r_{m,t+1} = k_{0m} - k_{1m}dp_{t+1} + dp_t + \Delta d_{t+1}, \tag{B.3.4}$$

$$dp_{t+1} = -A_{0m} - A_{1m}x_{t+1} - A_{2m}\sigma^2_{t+1}. \tag{B.3.5}$$

I first derive the monthly state-space representation. Substituting Eq.(B.3.1) and Eq.(B.3.2) into Eq.(B.3.5):

$$dp_{t+1} = -A_{0m} - A_{2m}\bar{\sigma}^2(1-\nu_1) - A_{1m}\rho x_t - A_{2m}\nu_1\sigma^2_t - A_{1m}\varphi_e\sigma_t e_{t+1} - A_{2m}\sigma_\omega\omega_{t+1}. \tag{B.3.6}$$

Eq.(B.3.3) implies:

$$\mu_t^d = E_t\left[\Delta d_{t+1}\right] = m_d + \phi x_t. \tag{B.3.7}$$

Using Eq.(B.3.1), then leads to the following dynamics for expected dividend growth:

$$\mu_t^d = (1-\rho)m_d + \rho\mu_{t-1}^d + \phi\varphi_e\sigma_{t-1}e_t, \tag{B.3.8}$$

Using Eq.(B.3.3), Eq.(B.3.5) and Eq.(B.3.6) in Eq.(B.3.4) leads:

$$\begin{gathered} r_{m,t+1} = k_{0m} - k_{1m}\left[-A_{0m} - A_{2m}\bar{\sigma}^2(1-\nu_1) - A_{1m}\rho x_t - A_{2m}\nu_1\sigma_t^2\right] + \\ + k_{1m}\left[A_{1m}\varphi_e\sigma_t e_{t+1} + A_{2m}\sigma_\omega\omega_{t+1}\right] - A_{0m} - A_{1m}x_t - A_{2m}\sigma_t^2 + m_d + \phi x_t + \varphi_d\sigma_t u_{d,t+1} = \\ = \left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d\right] + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]x_t - \\ - A_{2m}(1-k_{1m}\nu_1)\sigma_t^2 + \left[\varphi_d\sigma_t u_{d,t+1} + k_{1m}A_{1m}\varphi_e\sigma_t e_{t+1} + k_{1m}A_{2m}\sigma_\omega\omega_{t+1}\right] \end{gathered} \tag{B.3.9}$$

This implies:

$$\begin{gathered} \mu_t^r = E_t\left[r_{m,t+1}\right] = k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d - \\ - A_{2m}(1-k_{1m}\nu_1)(1-\nu_1)\bar{\sigma}^2 + \rho\left[\phi - A_{1m}(1-k_{1m}\rho)\right]x_{t-1} - \nu_1 A_{2m}(1-k_{1m}\nu_1)\sigma_{t-1}^2 + \\ + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]\varphi_e\sigma_{t-1}e_t - A_{2m}(1-k_{1m}\nu_1)\sigma_\omega\omega_t = \\ = (1-\nu_1)\left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d - A_{2m}(1-k_{1m}\nu_1)\bar{\sigma}^2\right] + \\ + \nu_1\mu_{t-1}^r + (\rho-\nu_1)\left[\phi - A_{1m}(1-k_{1m}\rho)\right]x_{t-1} + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]\varphi_e\sigma_{t-1}e_t - \\ - A_{2m}(1-k_{1m}\nu_1)\sigma_\omega\omega_t, \end{gathered}$$

Noting that:

$$\phi - A_{1m}(1-k_{1m}\rho) = \frac{1}{\psi};$$

and using Eq.(B.3.7):

$$\mu_t^r = (1-\nu_1)\left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d - A_{2m}(1-k_{1m}\nu_1)\bar{\sigma}^2\right] - \\ - \frac{(\rho-\nu_1)m_d}{\phi\psi} + \nu_1\mu_{t-1}^r + \frac{(\rho-\nu_1)}{\phi\psi}\mu_{t-1}^d + \left[\phi - A_{1m}(1-k_{1m}\rho)\right]\varphi_e\sigma_{t-1}e_t - A_{2m}(1-k_{1m}\nu_1)\sigma_\omega\omega_t,$$

Finally, the last equation of Eq.(3.3.1)-Eq.(3.3.3) follows from Eq.(B.3.3) and the definition of expected dividend growth given by Eq.(B.3.7).

To derive the annual state-space representation for returns, write:

$$\Delta d_{12(t+1)-j} = \mu^d_{12(t+1)-j-1} + \epsilon^d_{12(t+1)-j} = m_d + \rho^{11-j}\left(\mu^d_{12t} - m_d\right) + \sum_{k=0}^{10-j}\rho^k\epsilon^{\mu d}_{12(t+1)-j-1-k} + \epsilon^d_{12(t+1)-j}.$$

Then, since:

$$E_{12t}\left[\Delta d_{12t-j}\right] = \Delta d_{12t-j},$$

from Eq.(B.2.1) the expression for the aggregated expected dividend growth is:

$$\mu^{d,a}_{12t} = b_1 m_d \sum_{j=0}^{11}(j+1) + b_1\sum_{j=0}^{11}(j+1)\rho^{11-j}\left(\mu^d_{12t} - m_d\right) + b_2\sum_{j=0}^{10}(11-j)\Delta d_{12t-j}. \tag{B.3.10}$$

Note that, from Eq.(B.3.7):

$$\mu^d_{12t} - m_d = \phi x_{12t}. \tag{B.3.11}$$

Using Eq.(B.3.11) in Eq.(B.3.10) leads to:

$$\mu^{d,a}_{12t} = 78 b_1 m_d + b_1\sum_{j=0}^{11}(j+1)\rho^{11-j}\phi x_{12t} + b_2\sum_{j=0}^{10}(11-j)\Delta d_{12t-j},$$

which is Eq.(3.3.4) in the text.

From Eq.(B.3.1):

$$x_{12t} = \rho^{12} x_{12(t-1)} + \sum_{i=0}^{11} \rho^i \varphi_e \sigma_{12t-1-i} e_{12t-i},$$

which implies:

$$\mu_{12t}^{d,a} = 78 b_1 m_d + b_1 \sum_{j=0}^{11} (j+1)\rho^{11-j} \phi \left[\rho^{12} x_{12(t-1)} + \sum_{i=0}^{11} \rho^i \varphi_e \sigma_{12t-1-i} e_{12t-i} \right] +$$
$$+ b_2 \sum_{j=0}^{10} (11-j) \Delta d_{12t-j} = (1-\rho^{12}) 78 b_1 m_d + \rho^{12} \mu_{12(t-1)}^{d,a} - \rho^{12} b_2 \sum_{j=0}^{10} (11-j) \Delta d_{12(t-1)-j} +$$
$$+ b_2 \sum_{j=0}^{10} (11-j) \Delta d_{12t-j} + b_1 \phi \sum_{j=0}^{11} (j+1) \rho^{11-j} \sum_{i=0}^{11} \rho^i \varphi_e \sigma_{12t-1-i} e_{12t-i}. \tag{B.3.12}$$

Noting that:

$$\Delta d_{12t-j} - \rho^{12-j} \Delta d_{12(t-1)} = \sum_{i=0}^{11-j} \rho^i \phi \varphi_e \sigma_{12t-j-2-i} e_{12t-j-1-i} - \tag{B.3.13}$$
$$- \sum_{i=0}^{11-j} \rho^i \left[\rho \varphi_d \sigma_{12t-j-2-i} u_{d,12t-j-1-i} - \varphi_d \sigma_{12t-j-1-i} u_{d,12t-j-i} \right], \tag{B.3.14}$$

the process describing the dynamics of expected dividend growth is:

$$\mu_{12t}^{d,a} = (1-\rho^{12}) 78 b_1 m_d + \rho^{12} \mu_{12(t-1)}^{d,a} + \rho^{12} b_2 \sum_{j=0}^{10} (11-j) \left[\rho^{-j} \Delta d_{12(t-1)} - \Delta d_{12(t-1)-j} \right] +$$
$$+ b_2 \sum_{j=0}^{10} (11-j) \left[\Delta d_{12t-j} - \rho^{12-j} \Delta d_{12(t-1)} \right] + b_1 \phi \sum_{j=0}^{11} (j+1) \rho^{11-j} \sum_{i=0}^{11} \rho^i \varphi_e \sigma_{12t-1-i} e_{12t-i}$$

Aggregation of returns is done through a simple summation. The aggregated series is connected to the monthly one through:

$$r_{m,12(t+1)}^a = \sum_{j=0}^{11} r_{m,12(t+1)-j} = \sum_{j=0}^{11} \mu_{12(t+1)-j-1}^r + \sum_{j=0}^{11} \epsilon_{12(t+1)-j}^r.$$

Note that Eq.(B.3.9) implies:

$$\mu^r_{12(t+1)-j-1} = \left[k_{0m} + k_{1m}\left(A_{0m} + A_{2m}\bar{\sigma}^2(1-\nu_1)\right) - A_{0m} + m_d\right] + \frac{1}{\psi}x_{12(t+1)-j-1} -$$
$$-(1-k_{1m}\nu_1)A_{2m}\sigma^2_{12(t+1)-j-1} = a^{\mu r} + \frac{1}{\psi}x_{12(t+1)-j-1} - (1-k_{1m}\nu_1)A_{2m}\sigma^2_{12(t+1)-j-1},$$

$$E_{12t}\left[\mu^r_{12(t+1)-j-1}\right] = a^{\mu r} + \frac{1}{\psi}E_{12t}\left[x_{12(t+1)-j-1}\right] - (1-k_{1m}\nu_1)A_{2m}E_{12t}\left[\sigma^2_{12(t+1)-j-1}\right].$$

Moreover:

$$E_{12t}\left[x_{12(t+1)-j-1}\right] = \rho^{11-j}x_{12t};$$

and

$$E_{12t}\left[\sigma^2_{12(t+1)-j-1}\right] = \bar{\sigma}^2 + \nu_1^{11-j}(\sigma_{12t} - \bar{\sigma}^2).$$

Therefore:

$$\mu^{r,a}_{12t} = E_{12t}\left[r^a_{m,12(t+1)}\right] = \sum_{j=0}^{11}\left[a^{\mu r} - (1-k_{1m}\nu_1)A_{2m}\bar{\sigma}^2\right] + \frac{1}{\psi}\sum_{j=0}^{11}\rho^{11-j}x_{12t} - \qquad \text{(B.3.15)}$$
$$-(1-k_{1m}\nu_1)A_{2m}\sum_{j=0}^{11}\nu_1^{11-j}(\sigma_{12t} - \bar{\sigma}^2).$$

An expression relating expected returns to its lagged value is obtained by using:

$$x_{12t} = \rho^{12}x_{12(t-1)} + \sum_{i=0}^{11}\rho^i\varphi_e\sigma_{12t-1-i}e_{12t-i}$$

and

$$\sigma^2_{12t} = \bar{\sigma}^2 + \nu_1^{12}(\sigma^2_{12(t-1)} - \bar{\sigma}^2) + \sum_{i=0}^{11}\nu_1^i\sigma_\omega\omega_{12t-i}$$

in Eq.(B.3.15) and regrouping the terms.

B.4 Derivation of the present value identity implied by the state-space representation for the aggregated economy

In this appendix I derive the present value relation implied by the aggregated state-space representation, Eq.(3.4.5) in the text. The present value relation in the monthly frequency, Eq.(3.4.4) is obtained by restricting corresponding coefficients of the aggregated system. To save the space, its derivation is not included.

The annualized state-space representation for returns is described by the following system of equations:

$$\begin{aligned}
\mu^{r,a}_{12(t+1)} &= \alpha^{\mu r,a} + \delta^{\mu r,a}\mu^{r,a}_{12t} + \tau^a x^a_{12t} + \epsilon^{\mu r,a}_{12(t+1)}, \\
\mu^{d,a}_{12(t+1)} &= \alpha^{\mu d,a} + \delta^{\mu d,a}\mu^{d,a}_{12t} + \vartheta^a \epsilon^X_{12t} + \epsilon^{\mu d}_{12(t+1)}, \\
x^a_{12(t+1)} &= \delta^{x,a} x^a_{12t} + \epsilon^{x,a}_{12(t+1)}, \\
\Delta d^a_{12(t+1)} &= \mu^{d,a}_{12t} + \epsilon^{d,a}_{12(t+1)},
\end{aligned}$$

together with the Campbell and Shiller (1988) linearization:

$$r^a_{m,12(t+1)} = k^a_{0m} - k^a_{1m} dp^a_{12(t+1)} + dp^a_{12t} + \Delta d^a_{12(t+1)}.$$

The Campbell and Shiller (1988) linearization implies:

$$dp^a_{12t} = -\frac{k^a_{0m}}{1-k^a_{1m}} + \sum_{j=0}^{\infty} (k^a_{1m})^j E_{12t}\left[r^a_{12(t+1+j)} - \Delta d^a_{12(t+1+j)}\right] - (k^a_{1m})^{\infty} dp^a_{\infty}. \tag{B.4.1}$$

The no asset bubbles condition states:

$$(k^a_{1m})^{\infty} dp^a_{\infty} = 0. \tag{B.4.2}$$

Note that:

$$
\begin{aligned}
E_{12t}\left[r_{12(t+1+j)}\right] &= E_{12t}\left[E_{12(t+j)}\left(r_{12(t+1+j)}\right)\right] = E_{12t}\left[\mu^{r,a}_{12(t+j)}\right] = \\
&= E_{12t}\left[\alpha^{\mu r,a} + \delta^{\mu r,a}\mu^{r,a}_{12(t+j-1)} + \tau^a x^a_{12(t+j-1)} + \epsilon^{\mu r,a}_{12(t+j)}\right] = \\
&= \alpha^{\mu r,a}\sum_{i=0}^{j-1}(\delta^{\mu r,a})^i + (\delta^{\mu r,a})^j\mu^{r,a}_{12t} + \vartheta^a\sum_{i=0}^{j-1}(\delta^{\mu r,a})^i E_{12t}\left[x^a_{12(t+j-1-i)}\right].
\end{aligned}
$$

Moreover, using the dynamics of $x_{12(t+j-1-i)}$:

$$
E_{12t}\left[x_{12(t+j-1-i)}\right] = (\delta^{x,a})^{j-1-i}x_{12t},
$$

implying that:

$$
\tau^a\sum_{i=0}^{j-1}(\delta^{\mu r,a})^i E_{12t}\left[x^a_{12(t+j-1-i)}\right] = \tau^a\frac{(\delta^{x,a})^j - (\delta^{\mu r,a})^j}{\delta^{x,a} - \delta^{\mu r,a}}x^a_{12t}
$$

Therefore:

$$
E_{12t}\left[\mu^{r,a}_{12(t+j)}\right] = \alpha^{\mu r,a}\frac{1-(\delta^{\mu r,a})^j}{1-\delta^{\mu r,a}} + (\delta^{\mu r,a})^j\mu^{r,a}_{12t} + \tau^a\frac{(\delta^{x,a})^j - (\delta^{\mu r,a})^j}{\delta^{x,a} - \delta^{\mu r,a}}x^a_{12t} \tag{B.4.3}
$$

In a similar fashion:

$$
E_{12t}\left[\mu^{d,a}_{12(t+j)}\right] = \alpha^{\mu d,a}\frac{1-(\delta^{\mu d,a})^j}{1-\delta^{\mu d,a}} + (\delta^{\mu d,a})^j\mu^{d,a}_{12t} + \tau^a(\delta^{\mu d,a})^{j-1}\epsilon^X_{12t}. \tag{B.4.4}
$$

Using Eq.(B.4.2), Eq.(B.4.3) and Eq.(B.4.4) in Eq.(B.4.1) leads to:

$$
\begin{aligned}
dp^a_{12t} = B^a_0 + \frac{\mu^{r,a}_t}{1-k^a_{1m}\delta^{\mu r,a}} - \frac{\mu^{d,a}_t}{1-k^a_{1m}\delta^{\mu d,a}} + \frac{\tau^a k^a_{1m}x^a_{12t}}{(1-k^a_{1m}\delta^{\mu r,a})(1-k^a_{1m}\delta^{\mu d,a})} - \\
- \frac{\vartheta^a\epsilon^X_{12t}}{\delta^{\mu d,a}(1-k^a_{1m}\delta^{\mu d,a})},
\end{aligned} \tag{B.4.5}
$$

with:

$$
B^a_0 = -\frac{k^a_{0m}}{1-k^a_{1m}} + \frac{\alpha^{\mu r,a}k^a_{1m}}{(1-k^a_{1m})(1-k^a_{1m}\delta^{\mu r,a})} - \frac{\alpha^{\mu d,a}k^a_{1m}}{(1-k^a_{1m})(1-k^a_{1m}\delta^{\mu d,a})},
$$

which is Eq.(3.4.5) in the text.

C. CHAPTER 4

C.1 Derivation of the constants in the present-value identity

This Appendix provides derivations of constants in Eq.(4.1.6). The derivations employ the generalized economy described by Eq.(4.1.1)-Eq.(4.1.4) :

$$\begin{aligned}
\mu^r_{t+1} &= a^{\mu,r} + b^{\mu,r}\mu^r_t + c^{\mu,r}x_t + \epsilon^{\mu,r}_{t+1}, \\
\mu^d_{t+1} &= a^{\mu,d} + b^{\mu,d}\mu^d_t + c^{\mu,d}\epsilon^M_t + \epsilon^{\mu,d}_{t+1}, \\
x_{t+1} &= a^x + b^x x_t + \epsilon^x_{t+1}, \\
\Delta d_{t+1} &= \mu^d_t + \epsilon^d_{t+1},
\end{aligned}$$

together with the Campbell and Shiller (1988) linearization, Eq.(4.1.5):

$$r_{t+1} = k_0 - k_1 dp_{t+1} + dp_t + \Delta d_{t+1}.$$

The Campbell and Shiller (1988) linearization implies:

$$dp_t = -\frac{k_0}{1-k_1} + \sum_{j=0}^{\infty}(k_1)^j E_t\left[r_{t+1+j} - \Delta d_{t+1+j}\right] - (k_1)^\infty dp_\infty. \tag{C.1.1}$$

The no asset bubbles condition states:

$$(k_1)^\infty dp_\infty = 0. \tag{C.1.2}$$

Note that:

$$E_t\left[r_{t+1+j}\right] = E_t\left[E_{t+j}\left(r_{t+1+j}\right)\right] = E_t\left[\mu^r_{t+j}\right] = E_t\left[a^{\mu,r} + b^{\mu,r}\mu^r_{t+j-1} + c^{\mu,r}x_{t+j-1} + \epsilon^{\mu,r}_{t+j}\right] =$$
$$= a^{\mu,r}\sum_{i=0}^{j-1}(b^{\mu,r})^i + (b^{\mu,r})^j\mu^r_t + c^{\mu,r}\sum_{i=0}^{j-1}(b^{\mu,r})^i E_t\left[x_{t+j-1-i}\right].$$

Moreover, using the dynamics of $x_{t+j-1-i}$:

$$E_t\left[x_{t+j-1-i}\right] = a^x\sum_{y=0}^{j-2-i}(b^x)^y + (b^x)^{j-1-i}x_t,$$

implying that:

$$c^{\mu,r}\sum_{i=0}^{j-1}(b^{\mu,r})^i E_t\left[x_{t+j-1-i}\right] = \frac{a^x c^{\mu,r}}{1-b^x}\left[\frac{1-(b^{\mu,r})^j}{1-b^{\mu,r}} - \frac{(b^x)^j-(b^{\mu,r})^j}{b^x-b^{\mu,r}}\right] + c^{\mu,r}\frac{(b^x)^j-(b^{\mu,r})^j}{b^x-b^{\mu,r}}x_t$$

Therefore:

$$E_t\left[\mu^r_{t+j}\right] = a^{\mu,r}\frac{1-(b^{\mu,r})^j}{1-b^{\mu,r}} + (\mu,r)^j\mu^r_t + \frac{a^x c^{\mu,r}}{1-b^x}\left[\frac{1-(b^{\mu,r})^j}{1-b^{\mu,r}} - \frac{(b^x)^j-(b^{\mu,r})^j}{b^x-b^{\mu,r}}\right] +$$
$$+c^{\mu,r}\frac{(b^x)^j-(b^{\mu,r})^j}{b^x-b^{\mu,r}}x_t \tag{C.1.3}$$

In a similar fashion:

$$E_t\left[\mu^d_{t+j}\right] = a^{\mu,d}\frac{1-(b^{\mu,d})^j}{1-b^{\mu,d}} + (b^{\mu,d})^j\mu^d_t + c^{\mu,d}(b^{\mu,d})^{j-1}\epsilon^M_t. \tag{C.1.4}$$

Using Eq.(C.1.2), Eq.(C.1.3) and Eq.(C.1.4) in Eq.(C.1.1) leads to the Eq.(4.1.6) in the text with:

$$B_0 = -\frac{k_0}{1-k_1} + \frac{a^{\mu,r}k_1}{(1-k_1)(1-k_1b^{\mu,r})} + \frac{a^x c^{\mu,r}}{1-b^x}\left[\frac{k_1}{(1-k_1)(1-k_1b^{\mu,r})} - \frac{k_1}{(1-k_1b^x)(1-k_1b^{\mu,r})}\right] -$$
$$-\frac{a^{\mu,d}k_1}{(1-k_1)(1-k_1b^{\mu,d}))};$$
$$B_1 = \frac{1}{1-k_1b^{\mu,r}};$$
$$B_2 = \frac{1}{1-k_1b^{\mu,d}};$$
$$B_3 = \frac{c^{\mu,r}}{b^x - b^{\mu,r}}\left[\frac{1}{1-k_1b^x} - \frac{1}{1-k_1b^{\mu,r}}\right]$$
$$B_4 = \left(\frac{c^{\mu,d}}{b^{\mu,d}}\right)\frac{1}{1-k_1b^{\mu,d}}.$$