Essays on Predictability of Stock Returns

by

Oleg Rytchkov

M.A., Economics, New Economic School, 2002
Ph.D., Physics, Steklov Mathematical Institute, 2001

Submitted to the Alfred P. Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2007

Author .....
Alfred P. Sloan School of Management
May 1, 2007

Certified by .....
Jiang Wang
Professor of Finance
Thesis Supervisor

Accepted by .....
Birger Wernerfelt
Professor of Management Science
Chair of PhD Program

Essays on Predictability of Stock Returns

by

Oleg Rytchkov

Submitted to the Alfred P. Sloan School of Management
on May 1, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Management

Abstract

This thesis consists of three chapters exploring predictability of stock returns.

In the first chapter, I suggest a new approach to analysis of stock return predictability. Instead of relying on predictive regressions, I employ a state space framework. Acknowledging that expected returns and expected dividends are unobservable, I use the Kalman filter technique to extract them from the observed history of realized dividends and returns. The suggested approach explicitly accounts for the possibility that dividend growth can be predictable. Moreover, it appears to be more robust to structural breaks in the long-run relation between prices and dividends than the conventional OLS regression. I show that for aggregate stock returns the constructed forecasting variable provides statistically and economically significant predictions both in and out of sample. The likelihood ratio test based on a simulated finite sample distribution of the test statistic rejects the hypothesis of constant expected returns at the 1% level.

In the second chapter, I analyze predictability of returns on value and growth portfolios and examine time variation of the value premium. As a major tool, I use the filtering technique developed in the first chapter. I construct novel predictors for returns and dividend growth on the value and growth portfolios and find that returns on growth stocks are much more predictable than returns on value stocks. Applying the appropriately modified state space approach to the HML portfolio, I build a novel forecaster for the value premium. Consistent with rational theories of the value premium, the expected value premium is time-varying and countercyclical.

In the third chapter, based on the joint work with Igor Makarov, I develop a dynamic asset pricing model with heterogeneously informed agents. I focus on the general case in which differential information leads to the problem of "forecasting the forecasts of others" and to non-trivial dynamics of higher order expectations. I prove that the model does not admit a finite number of state variables. Using numerical analysis, I compare equilibria characterized by identical fundamentals but different information structures and show that the distribution of information has substantial impact on equilibrium prices and returns. In particular, asymmetric information might generate predictability in returns and high trading volume.

Thesis Supervisor: Jiang Wang
Title: Professor of Finance

Acknowledgments

I would like to thank my thesis advisors, Professors Victor Chernozhukov, Leonid Kogan, Steve Ross, and Jiang Wang, for their guidance, support, and valuable advice over the years of my study at MIT. I am indebted to Professor Jiang Wang who initially sparked my interest in informational economics. As his Research Assistant, I learned immensely from his knowledge and experience. I am grateful to Professor Leonid Kogan who was always open to questions and discussions. Stimulating, inspiring, and thought-provoking conversations with him significantly shaped this dissertation. It is impossible to overestimate the importance of deep and insightful comments provided by Professor Steve Ross. I am especially thankful to Professor Victor Chernozhukov for his advice, encouragement, and friendship.

I would also like to express my warmest thanks to all faculty members who devoted their efforts and time to providing me with critical comments and suggestions. Their great feedback was invaluable for improving this dissertation.

I am thankful to my fellow PhD students Jiro Kondo, Anya Obizhaeva, Dimitris Papanikolaou, Jialan Wang, and especially to Igor Makarov for very instructive conversations and great suggestions. Their comments significantly improved the quality of this dissertation.

My years of study at MIT would have been much more stressful and confusing without administrative advice and support of Sharon Cayley, Cathy Ly, and Svetlana Sussman whose work was essential for making the whole process well-organized and very smooth. I wish to thank Sharon, Cathy, and Svetlana for doing great job.

I am especially indebted to my wife Liza for her unconditional support and encouragement while completing this dissertation. I wish to express my deep gratitude for her relentless proofreading, numerous critical remarks, and invaluable suggestions.

Contents

List of Figures

List of Tables

Chapter 1

Filtering Out Expected Dividends and Expected Returns

1.1 Introduction

After two decades of active academic research, there is still no consensus on time variability of expected aggregate stock returns. On the one hand, many studies have documented that returns are predictable.[1] On the other hand, there is an extensive literature that casts doubt on the possibility of predicting returns, arguing that there is no reliable statistical evidence for it.[2]

The standard approach to analysis of time variation in expected returns is to run OLS regressions of realized returns on forecasting variables. Although there exist many variables that have been argued to predict returns[3], the dividend-price ratio is the most popular of them. Indeed, all variation of the dividend-price ratio must come from the variation of expected returns, if dividend growth is unpredictable.[4] However, the regression approach to testing predictability has several drawbacks.

First of all, there is a set of econometric problems, which has received much attention in the literature. These problems are primarily caused by two facts. First, the dividend-price

[1]See Fama and French (1988,1989), Campbell and Shiller (1988), Campbell (1991), Hodrick (1992), Nelson and Kim (1992), Cochrane (1992), Lewellen (2004), Cochrane (2006) among others.

[2]See Goetzmann and Jorion (1993,1995), Lanne (2002), Valkanov (2003), Ferson, Sarkissian, and Simin (2003), Torous, Valkanov, and Yan (2004), Goyal and Welch (2005), Boudoukh, Richardson, and Whitelaw (2005) among others.

[3]Goyal and Welch (2005) provide one of the most comprehensive lists.

[4]See Cochrane (2005) for a textbook exposition of this argument.

ratio, like many other suggested predictors, is highly autocorrelated. Second, there is a contemporaneous negative correlation between innovations in returns and the dividend-price ratio, which makes the forecaster predetermined, but not exogenous. As demonstrated by Mankiw and Shapiro (1986), Stambaugh (1986, 1999) and others, in this case the OLS estimates of the regression slope are significantly biased upward in finite samples. As a result, the regression rejects the null hypothesis of no predictability too often. Moreover, in finite samples the t-statistic of the slope coefficient has a non-standard distribution, invalidating all tests based on the conventional quantiles. There is vast econometric literature on how to make inferences and test hypotheses in this case.[5]

Besides econometric problems, there are other important issues that could diminish the predictive power of the dividend-price ratio. Although it is widely recognized that the dividend-price ratio does not predict dividend growth, there is some evidence that the expected dividend growth is time varying. Lettau and Ludvigson (2005) demonstrate that the cointegration residual cdy_t of consumption, dividends and labor income predicts U.S. stock dividend growth at horizons from one to six years. Ang and Bekaert (2005) report robust cash flow predictability by earnings yields at horizons of up to one year. Ribeiro (2004) uses labor income to identify predictable variation in dividends. Predictability of dividend growth is important because it can have a significant impact on the ability of the dividend-price ratio to predict stock returns.[6] Fama and French (1988), Kothari and Shanken (1992), Goetzmann and Jorion (1995) pointed out that, when expected dividend growth is time varying, the dividend-price ratio is a noisy proxy for expected returns. Hence, the errors-in-variables problem arises and creates a downward bias in the slope coefficient. Moreover, as emphasized by Menzly, Santos, and Veronesi (2004) and Lettau and Ludvigson (2005), if expected returns are positively correlated with expected dividend growth, then the contribution of expected returns to variation of dividend-price ratio can be partially offset by variation in expected dividend growth. This effect also reduces the ability of the dividend-price ratio to forecast returns. Thus, the predictability of dividends could partially explain weak statistical evidence of the predictability of returns and calls for a testing procedure that accounts for the possibility of time variation in expected dividend

[5]See, for example, Cavanagh, Elliott and Stock (1995), Lanne (2002), Torous, Valkanov and Yan (2004), Campbell and Yogo (2006), Jansson and Moreira (2006) and many others.

[6]Time varying expected dividend growth rate is a key feature of asset pricing models with the long-run risk. See Bansal and Yaron (2004), Hanson, Heaton, and Li (2005), and others.

growth.[7]

Lastly, there is a growing concern that the standard linear predictive relation between returns and the dividend-price ratio is not stable over time. Lettau and Van Nieuwerburgh (2005) argue that there are statistically detectable shifts in the mean of the dividend-price ratio and these shifts are responsible for the poor forecasting power of the ratio. This means that the ex-ante robustness to this type of structural break is a highly desirable property of any inference procedure designed to uncover time variation of expected returns.

In this chapter, I suggest a new approach to the analysis of stock return predictability, which has several advantages relative to the predictive regression. Instead of looking at *ad hoc* linear regressions of returns on the dividend-price ratio, I employ a structural approach. I start with the assumption that both expected returns μ_t^r and expected dividend growth μ_t^d are time varying but *unobservable* to the econometrician. To keep the model parsimonious, I assume that in the benchmark case μ_t^r and μ_t^d follow $AR(1)$ processes with normal shocks which are allowed to be contemporaneously correlated. By definition, μ_t^r and μ_t^d are the best predictors of future returns and future dividend growth, respectively. However, since they are unobservable, the econometrician should use the observed data to uncover them. Realized returns and dividend growth are related to unobservable expectations as $r_{t+1} = \mu_t^r + \varepsilon_{t+1}^r$ and $\Delta d_{t+1} = \mu_t^d + \varepsilon_{t+1}^d$, where ε_{t+1}^r and ε_{t+1}^d are unexpected shocks to returns and dividends. Under a mild assumption on the joint behavior of prices and dividends, the present value relation imposes a restriction on shocks in this system making them mutually dependent. To maintain tractability, I use the log-linearized form of the present value relation suggested by Campbell and Shiller (1988).

This specification of returns and dividends has exactly the form of a state space model with state space variables μ_t^r and μ_t^d and observables r_{t+1} and Δd_{t+1}. Hence, the dynamics of the best estimates $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ of unobservable state variables are described by the Kalman filter. Note that by construction $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ optimally use the whole history of observed dividends and returns. In contrast, the predictive regression utilizes the dividend-price ratio, which is a specific combination of past dividends and returns. The Kalman filter

[7]Fama (1990) takes future growth rates of real activity as a proxy for the contribution of expected cash flows. Kothari and Shanken (1992) augmented this approach by using dividend yield and the growth rate of investments as additional proxies for expected dividend growth and by using future stock returns as a control variable. Although this approach helps to disentangle the expected returns and expected dividend growth, it uses future variables and thus is not appropriate for forecasting. Menzly, Santos, and Veronesi (2004) suggest to use the consumption-price ratio to disentangle the contributions of expected returns and expected dividend growth.

ignores the dividend-price ratio and allows data to form the best linear predictor.[8]

As in the case of predictive regression, the model parameters need to be estimated before making forecasts. Assuming that all shocks are normally distributed, I employ a maximum likelihood estimator (MLE) to obtain parameters of the Kalman filter. To test the time variability of expected dividends and expected returns, I use the log-likelihood ratio test based on the Kalman filter likelihood function. Since the predictability of returns is studied simultaneously with the predictability of dividend growth, the suggested testing procedure can be viewed as a generalization of Cochrane (2006) "the dog that did not bark" approach.

The empirical results of this chapter can be summarized as follows. I apply the described filtering technique to aggregate stock returns and demonstrate that the hypothesis of constant expected returns is rejected at the 1% level. I argue that for the empirical sample of annual aggregate stock returns and dividends the dividend-price ratio is not the best combination of prices and dividends that can be used for predicting returns and it is possible to construct a more powerful forecasting variable using dividends and returns separately. In particular, I show that the new forecast of future returns $\widehat{\mu}_t^r$ outperforms the dividend-price ratio both in and out of sample, providing a higher value of the R^2 statistic and smaller out-of-sample forecasting errors. Moreover, besides statistical significance, the constructed forecast is economically significant, allowing an investor who times the market to get higher return without taking an additional risk.

One might argue that if only information on returns and dividends is used, and if both expected returns and expected dividend growth are time varying and correlated with each other, it is impossible to distinguish their contributions to the variation of the dividend-price ratio. I show that, in general, this is not the case. For instance, there is no need for other variables correlated with expected returns and independent from expected dividend growth to separate their contributions. However, such additional information can increase the statistical quality of the inference. Furthermore, I demonstrate that, although in the benchmark model the information on dividends and returns is theoretically insufficient to identify variances and covariances of all innovations, the freedom is limited and can be described by a one-dimensional subset in the parameter space. More importantly, there is enough information for making inference about the correlation between expected returns and

[8]Conrad and Kaul (1988) also use the Kalman filter to extract expected returns, but only from the history of realized returns. Brandt and Kang (2004) model conditional mean and volatility as unobservable variables following a latent VAR process and filter them out from the observed returns.

expected dividend growth. Consistent with Lettau and Ludvigson (2005), I find evidence that this correlation is high and positive.

To ensure that my findings are not driven by particular details of the model specification, I perform a series of robustness checks. First, I examine several extensions of the benchmark model allowing expected returns and expected dividend growth to follow $AR(2)$ and general VAR processes and confirm most of the conclusions provided by the benchmark model. In particular, the predictors obtained from different models are highly correlated and all of them have comparable in-sample and out-of-sample performance. Second, I study an alternative measure of aggregate cash flows, which include dividends and repurchases, and show that my findings regarding time variation of expected returns do not change. Third, I examine the sensitivity of my test results to the distributional assumption by employing non-parametric bootstrap, and again demonstrate that the hypothesis of constant expected returns can be reliably rejected.

Although in my main analysis I use only the data on dividends and returns, I also study the possibility of adding other observables to the state space model. I adopt a simple framework in which a new observable serves as an additional proxy for the level of unobservable expected returns or expected dividend growth. As particular examples, I consider the book-to-market ratio BM_t and the equity share in total new equity and debt issues S_t. I demonstrate that although the book-to-market ratio does not bring new information about future returns, it helps to predict future dividend growth. On the contrary, the equity share variable does help to improve the predictive ability of the system both for dividends and returns. Surprisingly, this improvement comes from the ability of S_t to predict future dividend growth, but not returns.

The filtering approach has several advantages over the conventional predictive regression. First of all, it explicitly acknowledges that both expected returns and expected dividend growth can be time varying. As a result, the filtering approach is more flexible and allows us to disentangle the contributions of expected returns and expected dividend growth if dividends are predictable. This makes the prediction of returns more accurate and simultaneously gives us predictions of future dividend growth.

Next, the filtering approach employs a weaker assumption on the joint behavior of prices and dividends relative to the predictive regression, which implicitly assumes stationarity of the dividend-price ratio. As a result, the filtering approach is more robust to structural

breaks in the long run relation between prices and dividends, and this is the source of its superior forecasting performance. In particular, even a small change in the mean of dividend growth can produce a substantial shift in the dividend-price ratio, which destroys most of its forecasting power. The filtering approach is insensitive to such shifts. Since there is evidence supporting the presence of structural breaks in the empirical data, this robustness is very important and delivers more powerful tests of return predictability. Moreover, understanding why the filtering approach provides superior results is necessary to address the concern that my findings might be attributed to luck or data mining. Also, robustness to structural breaks makes the filtering approach more valuable from an *ex-ante* point of view when it is unclear whether structural breaks will occur.

By construction, the filtering approach does not grant a special role to the dividend-price ratio in predicting dividends and returns and allows data to form essentially new time series of the estimates $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$. Nevertheless, the correlation between $\widehat{\mu}_t^r$ and the dividend-price ratio is 0.69. This suggests that $\widehat{\mu}_t^r$ and the dividend-price ratio are likely to share the same predictive component. Ignoring the dividend-price ratio is also beneficial because, as widely recognized, the major problems of forecasting regressions come from a very high persistence of the forecasting ratios. The filtering approach emphasizes that high persistence of almost any predictive variable reflects high persistence of expected returns or indicates the presence of structural breaks.

The rest of the chapter is organized as follows. In Section 1.2, I formulate the state space model for time varying expected returns and expected dividend growth, and examine the identifiability of model parameters. Section 1.3 is devoted to empirical analysis of aggregate stock returns and dividend growth with the use of the suggested state space model. Specifically, I estimate the model parameters, examine the new forecasts provided by the model, and demonstrate the robustness of the filtering approach to structural breaks on this particular empirical sample. Section 1.4 contains the results of hypotheses testing, with the main focus on the hypothesis of constant expected returns. Section 1.5 studies the forecasting power of the constructed predictor out of sample. Section 1.6 provides several extensions. There I examine the economic significance of the discovered predictability by looking at optimal portfolios under different predictive strategies and study the implications of including repurchases into the definition of cash flows. Also, I analyze several extensions of the model. In particular, I show how to incorporate additional information and explore

alternative specifications for expected returns and expected dividend growth. Section 1.7 concludes, discussing several directions for future research.

1.2 Theory

1.2.1 State space model

Consider an economy in which both expected aggregate log returns μ_t^r and expected log dividend growth μ_t^d are time varying. I assume that their joint evolution can be described by a first-order VAR

$$\mu_{t+1} = \bar{\mu} + \Phi(\mu_t - \bar{\mu}) + \varepsilon_{t+1}, \tag{1.1}$$

where in general μ_t is a p-dimensional vector with μ_t^r and μ_t^d as the first and the second elements, respectively. Φ is a $(p \times p)$ matrix whose eigenvalues lie inside the unit circle. The VAR specification (1.1) is quite general and admits higher order autoregressive processes for expected returns and expected dividend growth as particular cases. To avoid unnecessary complications, I also assume that ε_{t+1} contains only two normally distributed shocks which I denote as $\varepsilon_{t+1}^{\mu r}$ and $\varepsilon_{t+1}^{\mu d}$: $\varepsilon_{t+1} = (\varepsilon_{t+1}^{\mu r}, \varepsilon_{t+1}^{\mu d}, 0, ..., 0)'$. These shocks can be interpreted as shocks to expected returns and expected dividend growth and, in general, are allowed to be correlated with each other: $cov(\varepsilon_{t+1}^{\mu r}, \varepsilon_{t+1}^{\mu d}) = \sigma_{\mu r \mu d}$. In contrast to the most of the literature studying time variation of expected returns, I assume that μ_t is an *unobservable* vector of state variables. In other words, only market participants know it, but not econometricians. Since by definition μ_t^r and μ_t^d are the best predictors of future returns and future dividend growth, an econometrician faces a problem of filtering them out of the empirical data. In the simplest case, he only observes realized log returns $r_{t+1} = \log(1 + R_{t+1})$ and realized log dividend growth $\Delta d_{t+1} = \log(D_{t+1}/D_t)$, which are related to unobservable expectations μ_t^r and μ_t^d as

$$r_{t+1} = \mu_t^r + \varepsilon_{t+1}^r, \qquad \Delta d_{t+1} = \mu_t^d + \varepsilon_{t+1}^d. \tag{1.2}$$

By definition, ε_{t+1}^r and ε_{t+1}^d are unexpected shocks to returns and dividend growth uncorrelated with the previous period expectations: $cov(\mu_t^r, \varepsilon_{t+1}^r) = cov(\mu_t^d, \varepsilon_{t+1}^d) = 0$.

To make the model economically meaningful, an additional restriction on the introduced shocks is needed. To motivate this restriction, I use the generalized Campbell - Shiller linearization of the present value relation. As demonstrated in Appendix A, it implies that

unexpected return ε^r_{t+1} can be decomposed as

$$\varepsilon^r_{t+1} = Q\varepsilon_{t+1} + \varepsilon^d_{t+1} + (E_{t+1} - E_t)\sum_{i=2}^{\infty} \rho^{i-1}\left(\rho \cdot dpr_{t+i} + \log(1 + \exp(-dpr_{t+i})\right), \qquad (1.3)$$

where dpr_t is the log dividend-price ratio: $dpr_t = \log(D_t/P_t)$, and ρ is a specified number close to 1 from below. The matrix Q is given by

$$Q = \rho e_{12}(1 - \rho\Phi)^{-1}, \qquad e_{12} = (-1, 1, 0, ..., 0). \qquad (1.4)$$

Eq. (1.3) is similar to the unexpected stock return decomposition of Campbell (1991). For instance, the first term $Q_1\varepsilon^{\mu r}_{t+1}$ corresponds to "news about future expected returns", the second and the third terms $Q_2\varepsilon^{\mu d}_{t+1} + \varepsilon^d_{t+1}$ are "news about future dividends". However, the decomposition (1.3) is more general, because the standard no-bubble condition is not imposed. The following assumption provides an analog of the no-bubble condition which I use to pin down specific empirical implications.

Assumption. Shocks $\varepsilon^{\mu r}_{t+1}$, $\varepsilon^{\mu d}_{t+1}$, ε^r_{t+1} and ε^d_{t+1} are subject to the following linear constraint:

$$\varepsilon^r_{t+1} = Q\varepsilon_{t+1} + \varepsilon^d_{t+1}. \qquad (1.5)$$

I put the assumption in terms of the model disturbances, although as obvious from (1.3) it can be equivalently stated in terms of the dividend-price ratio dpr_t.

To make a better sense of Eq. (1.5), consider a case in which the dividend-price ratio is stationary. This stationarity is crucial for predictability of returns by the dividend-price ratio and is a conventional assumption in the literature. If dpr_t is stationary, then Eq. (1.3) immediately gives (1.5) since in the linear approximation the Taylor expansion around the mean level of the dividend-price ratio $\overline{dpr}$ yields $\rho \cdot dpr_{t+i} + \log(1+\exp(-dpr_{t+i}) \approx -k$. Thus, the above assumption is consistent with the previous literature on predictability. However, Eq. (1.5) is also valid under more general conditions and allows a mild non-stationarity of the dividend-price ratio.[9] In particular, it might also be valid if the dividend price ratio has

[9]The generality of the Assumption highlights a simple but interesting observation. Namely, it is not necessary for the dividend-price ratio to be stationary to be consistent with stationary expected returns.

a deterministic growth component with the growth rate less then $1/\rho$. Hence, the suggested assumption can be viewed as a relaxed version of the no-bubble condition.

The exogenous parameter ρ plays two roles. On one hand, it is related to the origin of the Taylor expansion $\overline{dpr}$ and must be chosen such that the difference between dpr_t and $\overline{dpr}$ is small justifying the linear approximation. On the other hand, ρ controls the contribution of future values of the dividend-price ratio into the innovation of returns. If ρ is sufficiently small, it effectively suppresses far future terms, so even if the dividend-price ratio is mildly non-stationary it will not break the validity of Eq. (1.5). However, if ρ is close to 1 then almost all terms in Eq. (1.3) are important and the Assumption is valid only if dpr_t is very close to a stationary process. Thus, the parameter ρ can be viewed as a measure of allowed non-stationarity in the dividend-price ratio.

It must be emphasized that allowing dpr_t to be non-stationary does not mean that there is an economic rationale behind it. The purpose of this assumption is to make the model more flexible since in a finite sample a non-stationary process is a good approximation to a stationary process with structural breaks. Correspondingly, the parameter ρ incorporates our beliefs of how quickly the dividend-price ratio is allowed to explode in the finite sample. Importantly, the obtained flexibility makes the model more robust since the new assumption can be warranted in a wider range of circumstances. In particular, structural breaks in the dividend-price ratio invalidate all standard arguments regarding the predictive power of the dividend-price ratio, but the suggested assumption is still valid and the model is expected to produce reasonable forecasts.

The flexibility brought by the relaxation of the no-bubble condition comes at a cost. If the dividend-price ratio is stationary and does not experience structural breaks, a model based on assumption (1.5) delivers less precise forecasts with lower R^2 in comparison with any similar model incorporating the stationarity assumption. In particular, if additionally expected dividend growth is constant then the forecast based on the conventional predictive regression of returns on the dividend-price ratio will be more precise than the forecast provided by the suggested model. This is a well known econometric tradeoff between robustness of the estimation procedure and its efficiency.

The imposed linear relation (1.5) leaves three independent shocks $\varepsilon^{\mu r}_{t+1}$, $\varepsilon^{\mu d}_{t+1}$, and ε^{d}_{t+1}. From the beginning, I assume that they have a general correlation structure with the fol-

lowing covariance matrix:

$$\Sigma = Var \begin{pmatrix} \varepsilon_{t+1}^{\mu r} \\ \varepsilon_{t+1}^{\mu d} \\ \varepsilon_{t+1}^{d} \end{pmatrix} = \begin{pmatrix} \sigma_{\mu r}^2 & \sigma_{\mu r \mu d} & \sigma_{\mu r d} \\ \sigma_{\mu r \mu d} & \sigma_{\mu d}^2 & \sigma_{\mu d d} \\ \sigma_{\mu r d} & \sigma_{\mu d d} & \sigma_{d}^2 \end{pmatrix}.$$

Although this generality is appealing, it may provide too much freedom, leaving some of the parameters unidentified. Indeed, there would be no problems if Σ were known exactly. However, in practice all model parameters must be estimated from the empirical data. If some sets of parameters are non-identifiable, they provide exactly the same observables and even an infinite history of data does not allow us to say which set of parameters we deal with. For the moment, I assume that all parameters are known and postpone a detailed analysis of identifiability to Section 1.2.2.

Abusing notation, it is convenient to denote *demeaned* expected returns and *demeaned* expected dividend growth by μ_t^r and μ_t^d. Then, the system (1.1) reduces to:

$$\mu_{t+1} = \Phi \mu_t + \varepsilon_{t+1}. \tag{1.6}$$

Correspondingly, the demeaned observables take the following form:

$$\Delta d_t = \mu_{t-1}^d + \varepsilon_t^d, \tag{1.7}$$

$$r_t = \mu_{t-1}^r + Q\varepsilon_t + \varepsilon_t^d. \tag{1.8}$$

From the representation (1.6), (1.7), and (1.8) it is clear that the model for dividends and returns has a time homogenous state space form[10], where Eqs. (1.7) and (1.8) are measurement equations and Eq. (1.6) is the transition equation for μ_t. Since all parameters of the state space system (1.6) - (1.8) are assumed to be known, the solution to the forecasting problem is provided by Proposition 1.

Proposition 1. Let x_t be a vector combining past state variables μ_{t-1} with current shocks $\varepsilon_t^{\mu r}$, $\varepsilon_t^{\mu d}$, and ε_t^d: $x_t = (\mu_{t-1}, \varepsilon_t^{\mu r}, \varepsilon_t^{\mu d}, \varepsilon_t^d)'$. Denote the current observables as $y_t = (r_t, \Delta d_t)'$. Given the state space system (1.1) with observables (1.7) and (1.8) the best

[10]Durbin and Koopman (2001) give a review of state space methods applied to time series analysis.

linear estimates of expected returns $\widehat{\mu}_t^r$ and expected dividend growth $\widehat{\mu}_t^d$ are given by the first two components of the vector $\hat{\mu}_t$ such that

$$\hat{\mu}_t = \left(\begin{array}{c|ccc} & 1 & 0 & 0 \\ & 0 & 1 & 0 \\ \Phi & 0 & 0 & 0 \\ & & \cdots & \\ & 0 & 0 & 0 \end{array}\right) \widehat{x}_t,$$

where the best linear estimate $\widehat{x}_t$ is provided by the Kalman filter

$$\hat{x}_t = (I - KM)F\hat{x}_{t-1} + Ky_t.$$

The Kalman gain matrix K is determined from the set of matrix equations

$$U = (I - KM)(FUF' + \Gamma\Sigma\Gamma'),$$

$$K = (FUF' + \Gamma\Sigma\Gamma')M'[M(FUF' + \Gamma\Sigma\Gamma')M']^{-1}.$$

The matrices M, F, and Γ are constant and defined in Appendix B. I is the $(p+3)\times(p+3)$ identity matrix.

Proof. See Appendix B.

Proposition 1 states that each period the best forecast of future dividends and returns should be updated with new information consisting of realized dividends and returns. The recursive structure of the Kalman filter means that solving the predictability problem is equivalent to extracting market expected returns from the whole history of observable data. Clearly, this problem cannot be reduced to a simple OLS regression of returns on some forecasting variables such as the dividend-price ratio.

The process for expected returns and expected dividend growth is quite general so far, but for the empirical work it must be specified more precisely. As a benchmark model, I consider the simplest form of the VAR system (1.1) in which μ_t^r and μ_t^d follow $AR(1)$ processes with the persistence parameters ϕ_r and ϕ_d, respectively. In this case the matrix

Q takes the following form:

$$Q = \begin{pmatrix} -\frac{\rho}{1-\rho\phi_r} & \frac{\rho}{1-\rho\phi_d} \end{pmatrix}. \tag{1.9}$$

The $AR(1)$ specification has two major benefits. First, it is quite parsimonious and has the minimum number of parameters sufficient to describe time varying expected returns and expected dividend growth. Second, this specification captures almost all interesting facts. I demonstrate it in Section 1.6 where I examine several extensions of the benchmark model including a general first order VAR with state variables μ_t^r and μ_t^d and $AR(2)$ processes for expected returns and expected dividend growth.

1.2.2 The identification problem

Proposition 1 gives a solution to the forecasting problem under the assumption that the parameters of the state space model are known. However, in practice all parameters should be estimated. Clearly, the question about parameter identifiability is very important at this stage. More formally, let $F(y, \theta)$ be a distribution function of all observables y_t, $t = 0, 1, ..., T$ generated by a state space model with parameters $\theta \in \Theta$. I will say that the vector of parameters θ_0 is identifiable if for any $\theta \in \Theta$, $\theta \neq \theta_0$ there exists a vector of observables y such that $F(y, \theta) \neq F(y, \theta_0)$. If all $\theta \in \Theta$ are identifiable, the state space system will be referred to as identifiable. In general, the identifiability of the model (1.6)-(1.8) depends on the particular specification of the state variables μ_t and the VAR matrix Φ. In this Section I examine the identification problem in the benchmark case where μ_t^r and μ_t^d follow $AR(1)$ processes.

In this model, the observables y are represented by the history of dividends and returns $\{(r_t, \Delta d_t), t = 0, 1, ..., T\}$ and the set of unknown parameters is $\theta = (\phi_r, \phi_d, \sigma^2_{\mu r}, \sigma^2_{\mu d}, \sigma^2_d, \rho_{\mu r \mu d}, \rho_{\mu r d}, \rho_{\mu d d})$, $\theta \in I_o^2 \times R_+^3 \times I_c^3$ where $I_o = (-1, 1)$, $I_c = [-1, 1]$. Note that the correlations $\rho_{\mu r \mu d}$, $\rho_{\mu r d}$, and $\rho_{\mu d d}$ are used as parameters instead of covariances which are subject to sophisticated constraints since the matrix Σ must be restricted to be positive definite.[11] Since the model (1.6)-(1.8) has three shocks with a general correlation structure and only two observables we may suspect that there are more parameters than can be identified with the available data even putting aside a limited sample size. While the

[11] More rigorously, the parameter space is $I_o^2 \times R_+^3 \times I_c^3$ with three sets of identified points:
1) if $\sigma^2_{\mu r} = 0$ then $(\phi_r, \phi_d, 0, \sigma^2_{\mu d}, \sigma^2_d, \rho_{\mu r \mu d}, \rho_{\mu r d}, \rho_{\mu d d}) \sim (0, \phi_d, 0, \sigma^2_{\mu d}, \sigma^2_d, 0, 0, \rho_{\mu d d})$;
2) if $\sigma^2_{\mu d} = 0$ then $(\phi_r, \phi_d, \sigma^2_{\mu r}, 0, \sigma^2_d, \rho_{\mu r \mu d}, \rho_{\mu r d}, \rho_{\mu d d}) \sim (\phi_r, 0, \sigma^2_{\mu r}, 0, \sigma^2_d, 0, \rho_{\mu r d}, 0)$;
3) if $\sigma^2_d = 0$ then $(\phi_r, \phi_d, \sigma^2_{\mu r}, \sigma^2_{\mu d}, 0, \rho_{\mu r \mu d}, \rho_{\mu r d}, \rho_{\mu d d}) \sim (\phi_r, \phi_d, \sigma^2_{\mu r}, \sigma^2_{\mu d}, 0, \rho_{\mu r \mu d}, 0, 0)$.

identification problem indeed exists, the freedom is quite limited and data still place tight restrictions on the model. The sets of indistinguishable parameters are characterized by Proposition 2.

Proposition 2. The persistence parameters ϕ_r and ϕ_d are fully identifiable. Two positive definite covariance matrices Σ and $\widetilde{\Sigma}$ are observationally indistinguishable if there exists $\lambda \in R$ such that $\Sigma - \widetilde{\Sigma} = \lambda\Omega$, where

$$\Omega = \begin{pmatrix} -\frac{(1-\phi_r^2)(1-\rho\phi_r)^2}{(1-\rho\phi_d)^2} & -\frac{(1-\rho\phi_r)(1-\phi_d\phi_r)}{1-\rho\phi_d} & \frac{\phi_r(1-\rho\phi_r)}{1-\rho\phi_d} \\ -\frac{(1-\rho\phi_r)(1-\phi_d\phi_r)}{1-\rho\phi_d} & -(1-\phi_d^2) & \phi_d \\ \frac{\phi_r(1-\rho\phi_r)}{1-\rho\phi_d} & \phi_d & 1 \end{pmatrix}.$$

Proof. See Appendix C.

Proposition 2 is very important for empirical analysis. Basically, it says that we are unable to recover the whole covariance structure of shocks even if we are given an infinitely long history of returns and dividends. However, only one element of Σ must be fixed for recovering the whole matrix. It means that the space of parameters can naturally be decomposed into one-dimensional subsets whose points are observationally indistinguishable. Moreover, as demonstrated below, the natural restriction that Σ is positive definite also imposes strict limitations on admissible parameters. Thus, although being unable to identify all parameters exactly, we can say much about them.

In several cases below, to resolve the described ambiguity in parameters for reporting purposes I use the following rule: among all empirically indistinguishable points of the parameter space I choose the one with the lowest absolute value of $\rho_{\mu dd}$. Given the structure of the matrix Ω, this procedure unambiguously fixes one point in each set of equivalence. In particular, when this set hits the hyperplane $\rho_{\mu dd} = 0$ the point on this hyperplane is chosen. Note that the suggested rule is a matter of convenience only and is a concise way to specify a particular locus. Clearly, there are a number of many other ways which are absolutely equivalent to the selected one. Although this rule of fixing uncertainty is used in the subsequent empirical work, I mostly focus on those results which hold for all indistinguishable points of the parameter space and consider those results the most interesting and the most important.

1.3 Empirical analysis

1.3.1 Data

The data used in this chapter come from the Center for Research in Securities Prices (CRSP) and consist of the annual value-weighted CRSP index of stocks traded on the NYSE, AMEX and NASDAQ. Since returns on the index are provided both with and without dividends, it is easy to build the time series of dividend growth. The annual CRSP data set used in my research covers the period from 1926 to 2004. To calculate the real values for all variables I use the CPI index also available from CRSP.

In the subsequent empirical analysis, I use logs of returns and dividend growth. First, it is consistent with the theory, which also operates with logs. Second, the distributions of logs are closer to normal. This is particularly important because the ML approach used for estimation and hypotheses testing essentially relies on the distributional assumption. Third, logs are more homoscedastic.

In the empirical work, I focus on the one-year horizon. Indeed, consideration of shorter horizons is complicated by seasonality in the dividend growth. Consideration of horizons longer than one year unavoidably leads to overlapping returns. However, as demonstrated by Boudoukh, Richardson and Whitelaw (2005), overlapping returns along with high persistency of predictive variable produce high correlation across multiple horizon estimators. Furthermore, Valkanov (2003) argues that in case with overlapping returns even asymptotic distribution of test statistics has a non-standard form and this can partially explain the reported evidence in favor of predictability. To avoid such critique, I do not use overlapping returns and focus on the one year horizon only.

1.3.2 Parameter estimation

The first step of my empirical analysis is the estimation of the model parameters. Since it is assumed that all shocks of the state space system are normally distributed, the most efficient estimator of parameters is provided by the maximum likelihood estimator (MLE). The log-likelihood function for the model is

$$\log L(\theta) = -T\log(2\pi) - \frac{T}{2}\log(\det\Lambda) - \frac{1}{2}\sum_{t=1}^{T}(y_t - \widehat{y}_t)'\Lambda^{-1}(y_t - \widehat{y}_t), \qquad (1.10)$$

where $y_t, t = 1, 2, ..., T$ is a set of observables and $\widehat{y}_t = MF\hat{x}_{t-1}$. This structure of the log-likelihood function is termed a prediction error decomposition.[12] It originates from the fact that conditional forecasting errors $y_t - \widehat{y}_t$ are serially independent by construction and have the covariance matrix $\Lambda = M(FUF' + \Gamma\Sigma\Gamma')M'$, where all constituent matrices are defined in Appendix B. The set of unknown parameters is $\theta = (\phi_r, \phi_d, \sigma^2_{\mu r}, \sigma^2_{\mu d}, \sigma^2_d, \rho_{\mu r \mu d}, \rho_{\mu r d}, \rho_{\mu d d})$. By construction, $\phi_r \in I_o$, $\phi_d \in I_o$, $\sigma^2_{\mu r} \in R_+$, $\sigma^2_{\mu d} \in R_+$, $\sigma^2_d \in R_+$, $\rho_{\mu r \mu d} \in I_c$, $\rho_{\mu r d} \in I_c$, $\rho_{\mu d d} \in I_c$, where $I_o = (-1, 1)$, $I_c = [-1, 1]$. Note that stationarity of expected returns and expected dividend growth is imposed explicitly.

The number of parameters is worth a comment. There might be concern that having 8 parameters in the model relative to 1 slope parameter in the predictive regression automatically puts the filtering approach into a favorable position with its better predictive ability following almost mechanically. However, this is not true. By construction, the slope in the predictive regression is a number minimizing the variance of the residual, so we choose it to maximize the predictive ability of the regressor. This is not the case in the filtering approach. Maximizing the log-likelihood function of the model we are looking for the set of parameters which provides the best fit to all data. In particular, the model with the estimated parameters must match the empirical values of major moments such as variances and correlations of observables. Obviously, this target is not identical to boosting predictability, hence it is not clear from the outset that more parameters help to increase the predictive ability of the model.

As demonstrated in Section 1.2, not all parameters of the state space system (1.6)-(1.8) with $AR(1)$ processes are identifiable. It means that there are many points in the parameter space where the log-likelihood function takes its maximum value and these points can be found one from another with the use of Proposition 2. I scale the identified set to fit the unit interval $[0, 1]$ and plot the estimated parameters in Figure 1-1.

Figure 1-1 provides several observations. First of all, the ML estimates of the parameters make a good sense. Indeed, the expected returns are very persistent with the mean reversion coefficient of 0.8005 and this is consistent with the intuition based on the predictive regressions. This coefficient is identifiable and thus does not change along the estimated set. Next, the correlation $\rho_{\mu r \mu d}$ between shocks to expected returns and expected dividend

[12] Originally it was suggested in Schweppe (1965). For textbook discussion of the Kalman filter estimation see Hamilton (1994).

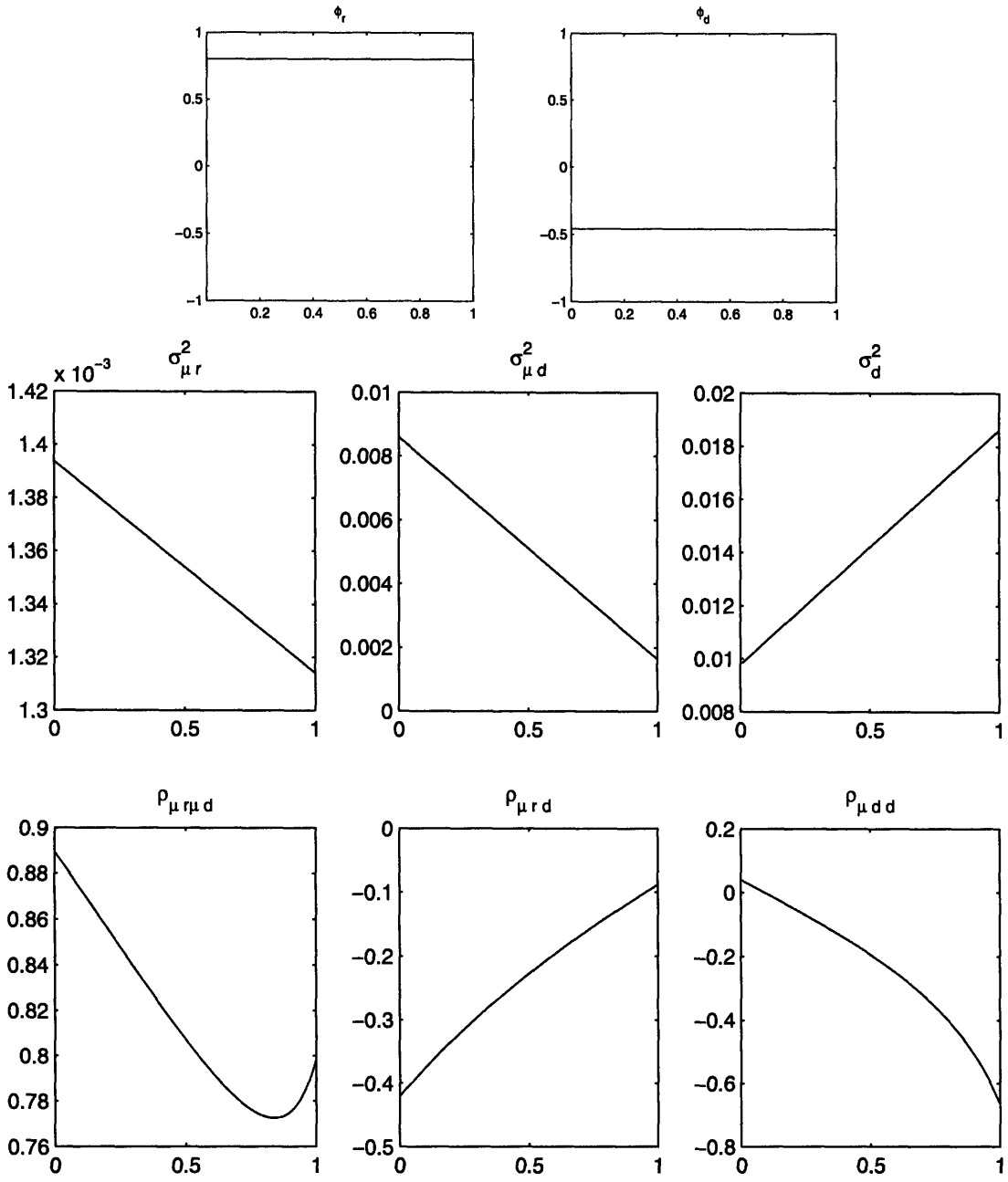


Figure 1-1: Maximum likelihood estimates of the benchmark model parameters along the identified set in the parameter space. The identified set is scaled to fit the interval $[0, 1]$.

growth is very high and positive for all points in the identified set and doesn't fall below 0.77. It means that although we cannot identify $\rho_{\mu r \mu d}$ exactly, we can say much about it. Note that high and positive value of $\rho_{\mu r \mu d}$ is consistent with Menzly, Santos and Veronezi (2004) and Lettau and Ludvigson (2005) who argue that high correlation between expected returns and expected dividend growth might be responsible for the mediocre predictive ability of the dividend-price ratio.

The other correlations $\rho_{\mu r d}$ and $\rho_{\mu d d}$ vary more significantly along the identified set, but they still have reasonable signs. In particular, negative values of $\rho_{\mu r d}$ are consistent with the intuition that in good times when dividends do up expected returns go down. The negative sign of $\rho_{\mu d d}$ indicates the mean reverting nature of the dividend growth: when dividends increase expected dividend growth decreases.

To get better interpretation of the obtained parameter estimates, I compute several statistics implied by the model with the estimated parameters and draw them in Figure 1-2. Since none of them are observable directly, their values are not identifiable and vary along the optimal locus.

First, Figure 1-2 shows standard deviations of expected returns and expected dividend growth. The obtained estimates indicate that both expected returns and expected dividend growth are time varying and have comparable volatility although innovations to expected dividend growth appear to be much more volatile than innovations to expected returns (cf. Figure 1-1). The reconciliation comes from high persistence of expected returns which explains why even small variance of shocks to expected returns might have important implications such that a dominating contribution into the variation of the dividend-price ratio.[13] Also, because of high positive autocorrelation of expected returns and negative autocorrelation of expected dividend growth the unconditional correlation between μ_t^r and μ_t^d is lower than the conditional correlation between their innovations, but it is still sufficiently high and positive.

Second, I examine the model implied innovations to returns and the dividend-price ratio. Their standard deviations reported in the second row of Figure 1-2 almost exactly coincide with the corresponding values obtained from the predictive regression errors (see, for example, Cochrane (2005)) Importantly, the implied correlation between innovations to returns and the dividend-price ratio is high and negative, and this is consistent with the

[13]This was initially emphasized by Campbell (1990, 1991).

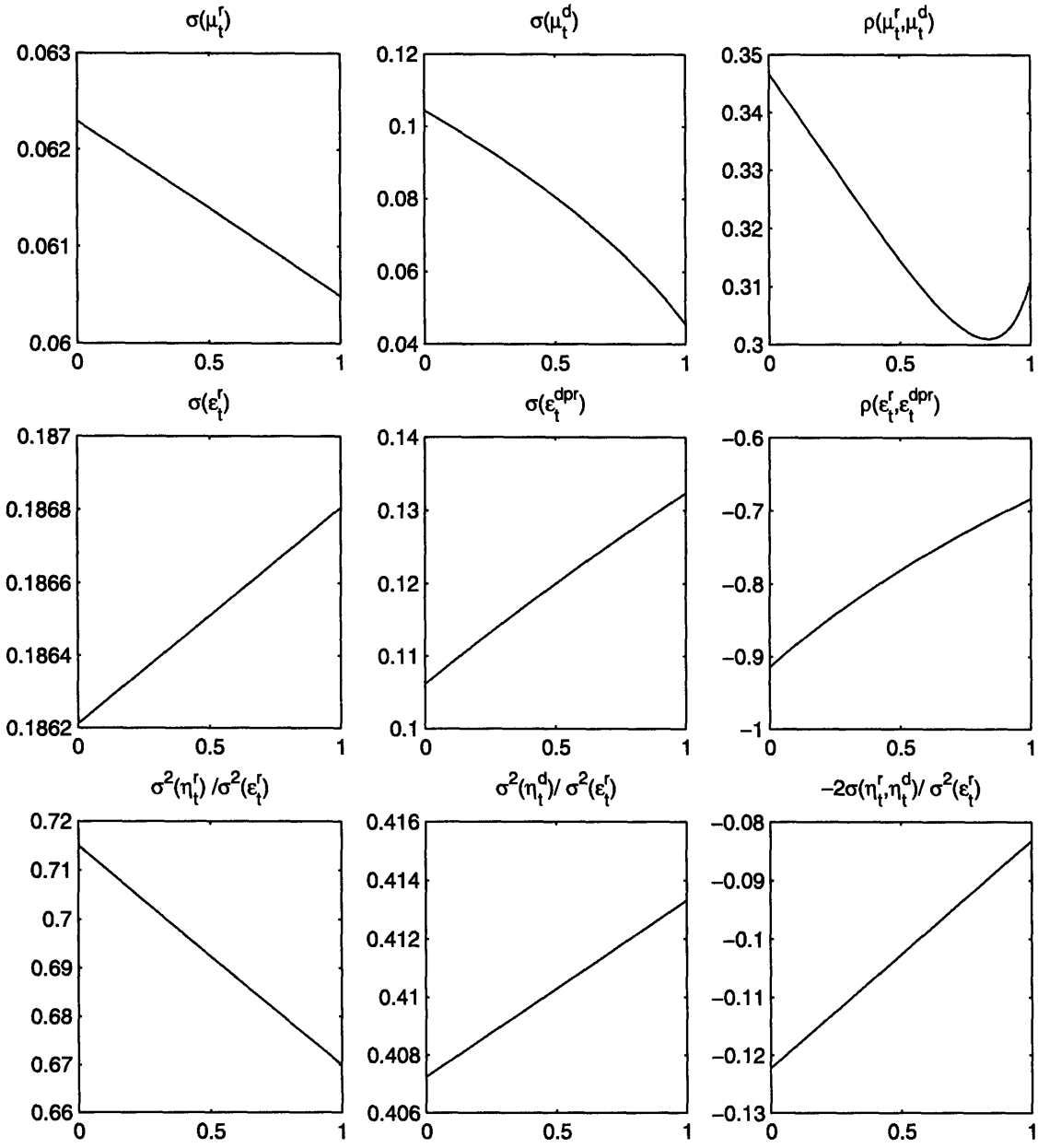


Figure 1-2: Various statistics and variance decomposition of unexpected returns along the optimal identified set in the parameter space.

The locus is scaled to fit the interval $[0,1]$. $\sigma(\mu_t^r)$ and $\sigma(\mu_t^d)$ are standard deviations of expected returns and expected dividend growth, $\rho(\mu_t^r, \mu_t^d)$ is the correlation between them. $\sigma(\varepsilon_t^r)$ and $\sigma(\varepsilon_t^{dpr})$ are standard deviations of unexpected return and the innovation to the dividend-price ratio, $\rho(\varepsilon_t^r, \varepsilon_t^{dpr})$ is the correlation between them. $\sigma^2(\eta_t^r)\sigma^2(\epsilon_t^r)$, $\sigma^2(\eta_t^d)\sigma^2(\epsilon_t^r)$, and $-2\sigma(\eta_t^r, \eta_t^d)\sigma^2(\epsilon_t^r)$ represent the Campbell (1991) decomposition of unexpected stock returns.

literature on predictability of stock returns by the dividend-price literature.

Three last graphs at the bottom of Figure 1-2 represent Campbell (1991) variance decomposition of unexpected returns ε_t^r along the estimated identified set. $Var(\eta_t^r)/Var(\varepsilon_t^r)$ and $Var(\eta_t^d)/Var(\varepsilon_t^r)$ are the contributions of "news about future expected returns" and "news about future dividends" into ε_t^r, $-2Cov(\eta_t^r, \eta_t^d)/Var(\varepsilon_t^r)$ is the covariance term. Although these terms change along the locus, we can conclude that the major part of unexpected return volatility is generated by time varying expected stock return: its contribution falls in the range from 67% to 72%. The impact of "news about future dividends" also does not vary significantly along the locus and about 41% of the variance can be attributed to it. The role of the covariance term is smaller and it gives negative contribution of around 10%.

Although the obtained estimates are quite reasonable, they should be evaluated for their statistical precision. However, there is no easy way to do it and there are several reasons for that. First, we deal not with an identified point estimate but with an identified set, and in this case conventional methods do not work.[14] Moreover, the empirical sample is not large enough to make asymptotic values sufficiently reliable. Second, a number of estimated parameters are defined only on a bounded set in R^n. In particular, $\sigma_{\mu r}^2 \in R_+$, $\sigma_{\mu d}^2 \in R_+$, $\sigma_d^2 \in R_+$, $\rho_{\mu r \mu d} \in [-1, 1]$, $\rho_{\mu r d} \in [-1, 1]$, $\rho_{\mu d d} \in [-1, 1]$. It is well known in the econometric literature that the ML estimate has a non-standard asymptotic distribution and the conventional inference procedure is not applicable if the true parameter lies on a boundary.[15] Finally, the statistical significance of many parameters can be evaluated only jointly. For example, the hypothesis of constant expected returns should be stated as $H_0 : \sigma_{\mu r} = 0$, $\rho_{\mu r \mu d} = 0$, $\rho_{\mu r d} = 0$, $\phi_r = 0$. Rigorous tests of hypotheses are performed in Section 1.4.

To partially circumvent these difficulties and give a sense of the precision of estimates, I run Monte-Carlo simulations. In particular, I take one of the point estimates from the identified set and simulate 1000 samples with 79 observations. Note that any point from the identified set can be used, and all of them will produce identical simulated samples. Next, for each sample I find the ML estimates of the model parameters. For expositional purposes, instead of looking at the whole identified sets I choose only one point from each of them using the identification rule formulated in Section 1.2.2. Namely, I search for a

[14] Chernozhukov, Hong, and Tamer (2004) develop estimators and confidence regions for identified sets.
[15] See, for example, Andrews (1999).

ρ	ϕ_r	ϕ_d	$\sigma^2_{\mu r}$	$\sigma^2_{\mu d}$	σ^2_d	$\rho_{\mu r \mu d}$	$\rho_{\mu r d}$	R^2_r	R^2_d
0.964	0.8005	−0.4584	0.0014	0.0079	0.0106	0.8746	−0.3797	0.050	0.034
0.98	0.8085	−0.4648	0.0011	0.0079	0.0107	0.8774	−0.3892	0.058	0.035
0.95	0.7931	−0.4544	0.0016	0.0080	0.0106	0.8720	−0.3690	0.043	0.034
0.90	0.7572	−0.4468	0.0028	0.0079	0.0107	0.8621	−0.3180	0.029	0.032

Table 1.1: Maximum likelihood estimates of the benchmark state space model.

This table collects maximum likelihood estimates of the state space model (1.6) - (1.8) with $AR(1)$ processes. Different rows correspond to different values of the Campbell - Shiller linearization parameter ρ. In the first row the current empirical value of this parameter is chosen and other values are considered for robustness check. The identification strategy of Section 1.2 yields $\rho_{\mu dd} = 0$ for all examined ρ. R^2_r and R^2_d measure the ability of $\widehat{\mu}^r_t$ and $\widehat{\mu}^d_t$ to predict future returns and future dividend growth in sample.

combination of parameters that gives the best fit for the observed data and has the lowest correlation $\rho_{\mu dd}$ between the shock to expected dividend growth and unexpected dividend growth.

The simulated distributions of the obtained point estimates are presented in Figure 1-3, which allows us to make several observations. Thus, we can conclude that the standard deviations of the estimates are rather high and the distributions of many estimates are far from normal. However, all qualitative inferences based on the point estimates are still valid. Moreover, the simulations allow us to evaluate the finite sample bias of estimates, which is one of the major econometric problems of conventional predictive regressions. It is not clear from the outset whether the obtained ML estimates suffer from the similar drawbacks. I calculate biases of estimates as differences between the average of the simulated estimates and the population parameters. The results are reported in Figure 1-3. To visualize the conclusions, the values of population parameters are indicated by black bars. It is remarkable, that the obtained biases are tiny, so we can conclude that the ML estimates are almost unbiased.[16]

Recall that besides the parameters listed above the state space model (1.6)-(1.8) also contains the Campbell-Shiller linearization parameter ρ. This parameter is taken as exogenous in MLE, and the ML estimates might be sensitive to it. To illustrate that this is not the case, Table 1.1 gives the point estimates of parameters for different values of ρ. The first row corresponds to the sample value of ρ, whereas other three rows are computed for

[16]In general, bias depends on the particular values of population parameters. Strictly speaking, I demonstrated unbiasedness for only one point in the parameter space. However, there is no reason to think that for other points with ϕ_r and ϕ_d reasonably less than one the conclusion would be different.

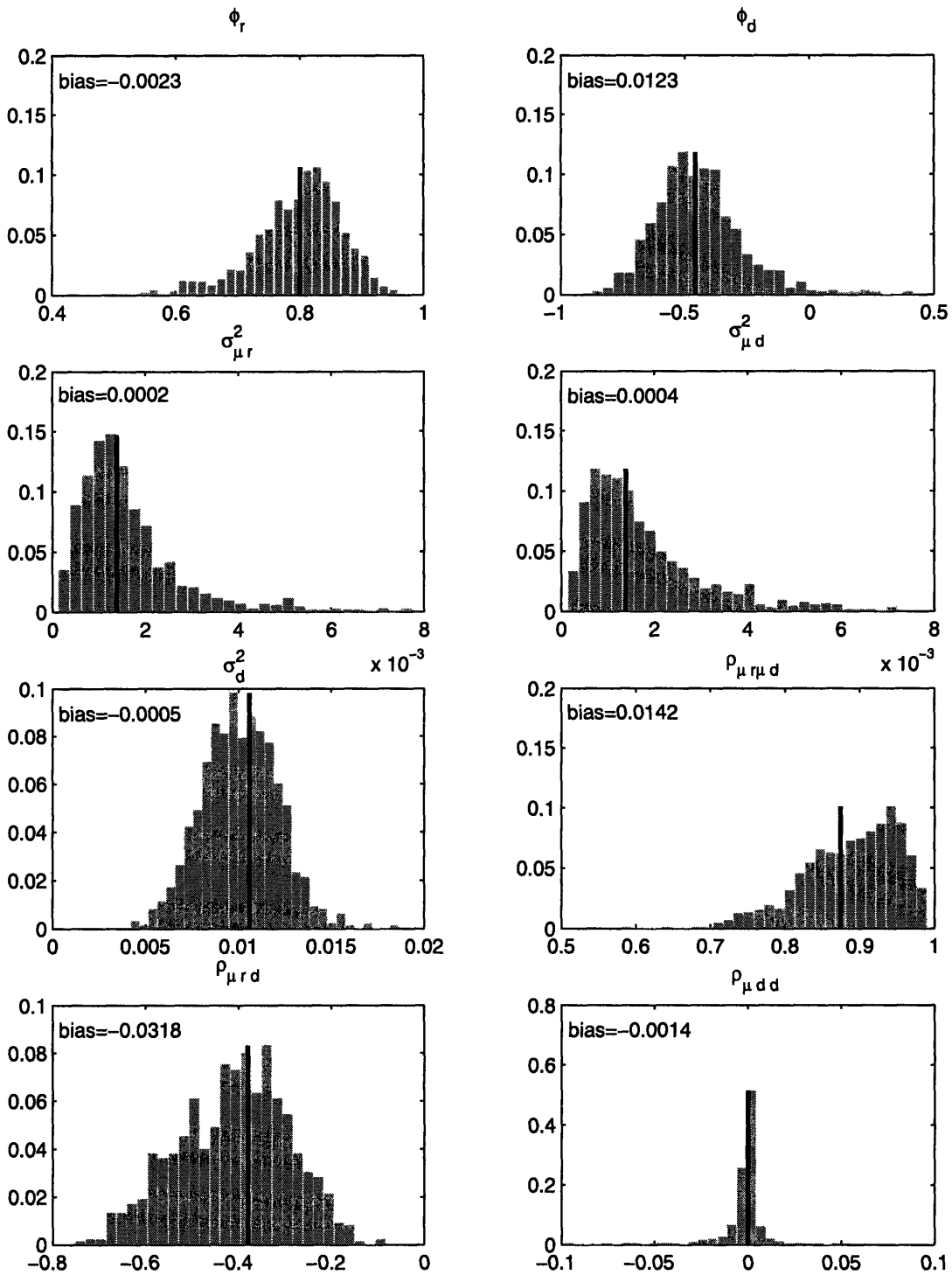


Figure 1-3: Simulated distributions of parameter estimates.

For the benchmark set of parameters from Table 1.1 I simulate 1000 samples of length equal to the length of the empirical sample and find the ML estimates of parameters for each sample. Black bars show the population values of parameters.

arbitrary chosen values 0.98, 0.95 and 0.9.[17] Again, instead of the whole identified set I report only one point with the lowest absolute value of $\rho_{\mu dd}$. Notably, the values in different rows are almost the same. Taking into account how noisy the obtained estimates are, we can conclude that there is almost no sensitivity to the choice of ρ.

1.3.3 Forecasts of dividends and returns

Given the estimated parameters the econometrician can use the Proposition 1 to construct the forecasts $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ and evaluate their in-sample performance by R^2. Note that the identification problem is irrelevant at this stage since all points from the identified set produce exactly the same values of $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$. Figure 1-4(a) plots the realized stock returns along with the forecast $\widehat{\mu}_t^r$. Also for comparison I plot the forecast based on the conventional predictive regression. Although the dividend-price ratio has some forecasting power in sample and can explain 3.8% of variation in stock returns, it is outperformed by the constructed predictor $\widehat{\mu}_t^r$ which provides R^2 of 5%. Although this improvement might seem very small, it is quite important because even tiny increase in ability to forecast future returns leads to significant effect on the optimal portfolios of long term investors. In Section 1.6 I provide a detailed analysis of economic significance of the obtained improvement.

Similarly, Figure 1-4(b) plots the realized dividend growth and the constructed forecasting variable $\widehat{\mu}_t^d$. The dash-dot line represents the forecast by the dividend-price ratio. It is well-known that the dividend-price ratio has no predictive power for the dividend growth and Figure 1-4(b) clearly supports this result. However, the constructed predictor for the dividend growth $\widehat{\mu}_t^d$ works much better and can explain about 3% of dividend growth.

To get additional insights about time variation of $\widehat{\mu}_t^r$, I juxtapose it with other variables which were found to be proxies for expected returns. As such variables, I choose the book-to-market ratio BM_t advocated by Kothari and Shanken (1997), Pontiff and Schall (1997) and others, the equity share in total new equity and debt issues S_t proposed by Baker and Wurgler (2000), and a cointegration residual between log consumption, log asset wealth and log labor income cay_t constructed in Lettau and Ludvigson (2001). I choose these particular variables because according to Goyal and Welch (2005) they demonstrate the strongest

[17] In the case of stationary dpr_t, the parameter ρ is usually chosen to be related to unconditional means of returns and dividend growth $\overline{\mu}_r$ and $\overline{\mu}_d$ as $\rho = \exp(\overline{\mu}_d - \overline{\mu}_r)$. The sample value of ρ is computed from the sample analogs of $\overline{\mu}_r$ and $\overline{\mu}_d$.

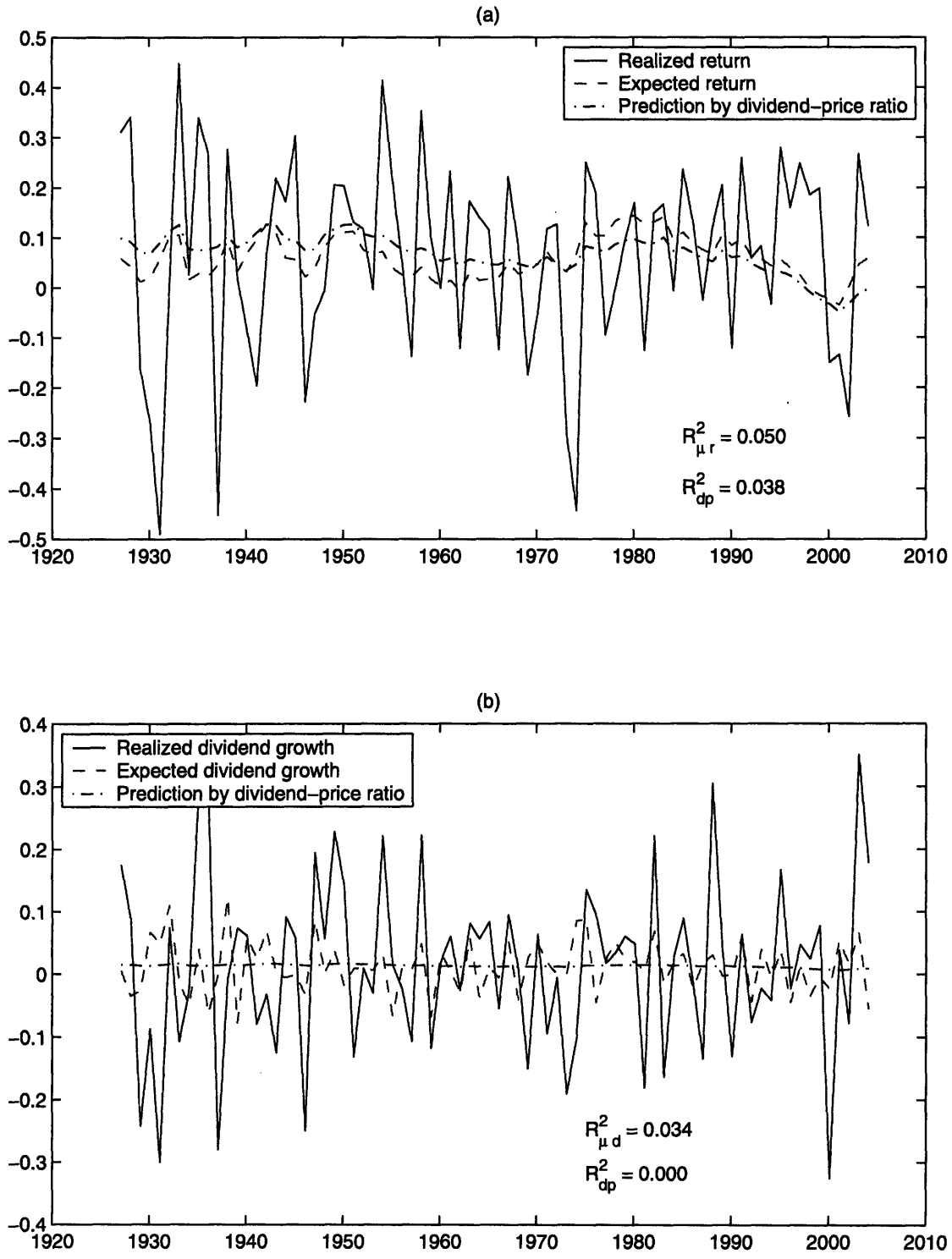


Figure 1-4: (a) Realized aggregate annual stock returns along with the expected returns and the returns predicted by the divided-price ratio. (b) Realized aggregate annual dividend growth along with the expected dividend growth and the dividend growth predicted by the dividend-price ratio.

	$\hat{\mu}_t^r$						$\hat{\mu}_t^d$					
dpr	0.07						0.01					
	(6.00)						(1.24)					
DEF		1.87						1.21				
		(2.03)						(2.52)				
NBER			0.02						0.05			
			(2.12)						(6.65)			
BM				0.12						0.04		
				(5.21)						(3.36)		
S					−0.01						0.06	
					(−0.09)						(2.16)	
cay						1.57						
						(5.26)						
cdy												0.43
												(2.45)
Adj-R^2	0.40	0.10	0.02	0.50	−0.01	0.25	0.00	0.04	0.23	0.05	0.02	0.05
N	79	79	79	79	79	54	79	79	79	79	79	54

Table 1.2: Regressions of expected dividends and returns on business cycle variables.

This table collects regression results of filtered expected return $\hat{\mu}_t^r$ and filtered expected dividend growth $\hat{\mu}_t^d$ on several business cycle variables. dpr_t is the log dividend-price ratio; DEF_t is the default premium, defined as the yield spread between Moody's Baa and Aaa corporate bonds; $NBER_t$ is the NBER recession dummy; BM_t is the aggregate book-to-market ratio; S_t is the equity share in total new equity and debt issues; cay_t is a cointegration residual between log consumption, log asset wealth and log labor income constructed by Lettau and Ludvigson (2001); cdy_t is a cointegration residual between log consumption, log dividends and log labor income constructed by Lettau and Ludvigson (2005). $\hat{\mu}_t^r$, $\hat{\mu}_t^d$, dpr_t, BM_t, and S_t are based on the annual sample which covers the period 1926 - 2004; cay_t and cdy_t are constructed from the annual data for the period 1948 - 2001. t-statistics in parentheses are computed using the Newey-West standard errors.

ability to predict returns in sample.[18] Also, I include several proxies for business cycles to examine quantitatively whether the filtered expected returns vary counter-cyclically. In particular, I add the NBER recession dummy which equals to 1 for the particular year if the December belongs to the NBER recession period and the default premium DEF_t defined as the yield spread between Moody's Baa and Aaa corporate bonds.[19]

Table 1.2 reports estimates from OLS regressions of $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ on all variables of interest. First, as one can presume from looking at Figure 1-4(a) the forecast $\widehat{\mu}_t^r$ and dpr_t are highly and positively correlated. The adjusted R^2 statistic is 0.4 and it corresponds to the sample correlation coefficient of 0.68. On one hand, it is not surprising, since both variables are proxies for expected stock return and it is natural that the correlation between them is high. On the other hand, $\widehat{\mu}_t^r$ was constructed absolutely independently from dpr_t and uses less restrictive assumption on the joint behavior of prices and dividends.

Next, as reported in Table 1.2, $\widehat{\mu}_t^r$ appears to be substantially correlated with other proxies for expected returns. Indeed, the slope coefficients in regressions on DEF_t, BM_t, and cay_t are statistically significant and the adjusted R^2 statistics correspond to the correlation coefficients of 0.34, 0.71, and 0.51, respectively. The latter number is especially high given that cay_t is constructed from a very different data set including aggregate consumption and labor income. These high correlations demonstrate that DEF_t, BM_t, cay_t, and $\widehat{\mu}_t^r$ are likely to share the same predictable component of stock returns. Although the regression of $\widehat{\mu}_t^r$ on $NBER_t$ also has a statistically significant coefficient, R^2 is less impressive. Nevertheless, the high correlations between $\widehat{\mu}_t^r$ and the countercyclical variables BM_t and DEF_t indicate that $\widehat{\mu}_t^r$ is also countercyclical. This is consistent both with the previous empirical findings in the literature and theoretical models explaining time variation of expected stock returns.

Table 1.2 also contains regression results for $\widehat{\mu}_t^d$. In particular, it reports the OLS regression on cdy_t, which is is a cointegration residual between log consumption, log dividends and log labor income. Lettau and Ludvigson (2005) argue that cdy_t captures a predictable component of aggregate dividend growth and it is interesting to compare it with $\widehat{\mu}_t^d$. Notably, the slope coefficient is positive and significant, although the R^2 statistic is only 0.05. Predictive regression of realized dividend growth on both cdy_t and $\widehat{\mu}_t^d$ (not reported) shows that these

[18] I am grateful to Martin Lettau for providing the data on cay_t and cdy_t on his website http://pages.stern.nyu.edu/~mlettau/data/ and to Jeffrey Wurgler for making available the data on S_t on his website http://pages.stern.nyu.edu/~jwurgler/.

[19] I use the data on corporate bond yields provided by Global Financial Data.

variables contain different pieces of information about future dividends and adding each of them to the regression improves the result. Namely, R^2's of univariate regression of Δd_{t+1} on cdy_t and $\widehat{\mu}_t^d$ for the period 1948 - 2001 are 18% and 13% correspondingly, whereas R^2 of the multivariate regression is 24%.

Table 1.2 also reports the regressions of $\widehat{\mu}_t^d$ on countercyclical variables BM_t and DEF_t. Interestingly, expected dividend growth also appears to be countercyclical, and the high correlation with the NBER dummy supports this conclusion. This is consistent with the co-movement of expected returns and expected dividend growth discussed above.

Figure 1-5 visualizes the relation between filtered expected returns, filtered expected dividend growth and business cycles. Gray bars indicate the periods of the NBER recessions. Again, as the business cycle variables I choose the default premium DEF_t, the dividend-price ratio dpr_t, and the book-to-market ratio BM_t which are known to be countercyclical. Figure 1-5(a) clearly demonstrates that the filtered expected returns are also countercyclical and go up when other variables indicating trough go up. Also, in consistency with the correlation table discussed above, a similar pattern arises for expected dividend growth. As follows from Figure 1-5(b), expected dividend growth is also countercyclical jumping upward almost every recession and quickly bouncing back afterwards.

To get better understanding of the structure of $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ it is instructive to consider their decomposition over observables $r_{t-\tau}$, $\Delta d_{t-\tau}$, $\tau = 0, 1, ...$ and over disturbances $\varepsilon_{t-\tau}^{\mu r}$, $\varepsilon_{t-\tau}^{\mu d}$ and $\varepsilon_{t-\tau}^{d}$, $\tau = 0, 1, ...$ Given the estimated parameters, the corresponding coefficients immediately follow from Proposition 1. Figure 1-6(a, b) presents the decomposition of $\widehat{\mu}_t^r$ over observables for up to 50 lags. It is remarkable that the coefficients of this decomposition decline slowly, and even very far observations have a significant effect on the current forecast of future returns. For instance, current dividend growth of 1% above average increases the projected returns by 17 basis points whereas dividend growth of 1% 30 years ago has an effect of 5 basis points. Similarly, the realized returns above average have a long lasting negative impact on the expected returns.

The effects of realized returns and dividends on $\widehat{\mu}_t^r$ admit an intuitive interpretation. Indeed, high realized return predicts negative future return due to the well-known discount rate effect of Fama and French (1988). A positive shock to returns can indicate a negative shock to expected returns or positive shocks to expected or unexpected dividend growth. Because of high persistence of expected returns and the resulting dominance of "news about

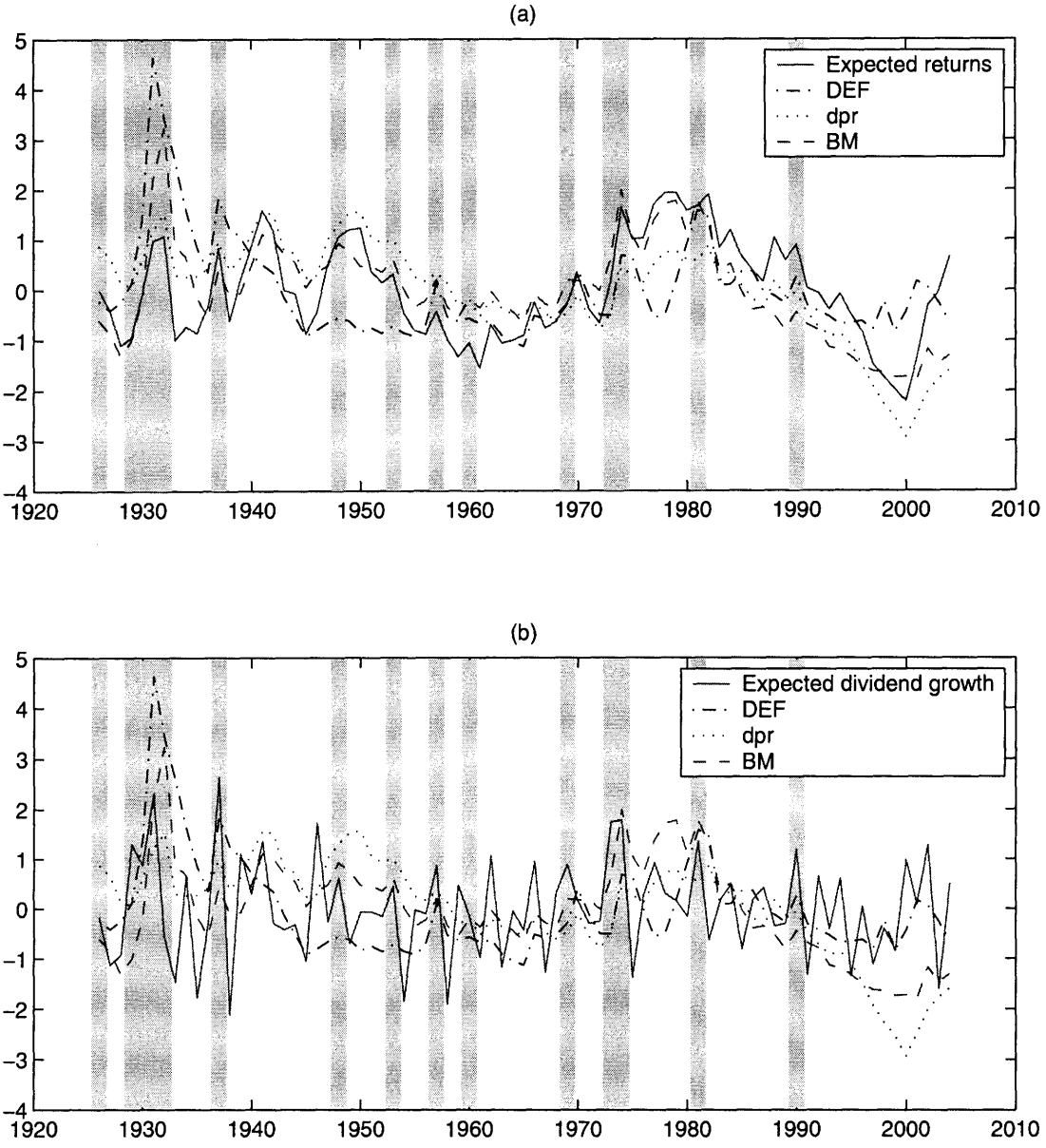


Figure 1-5: Filtered expected stock returns (a), filtered expected dividend growth (b), and business cycles.

The shaded areas indicate NBER recession dates. DEF_t is the default premium, defined as the yield spread between Moody's Baa and Aaa corporate bonds; dpr_t is the log dividend-price ratio; BM_t is the aggregate book-to-market ratio. All variables are standardized to unit variance.

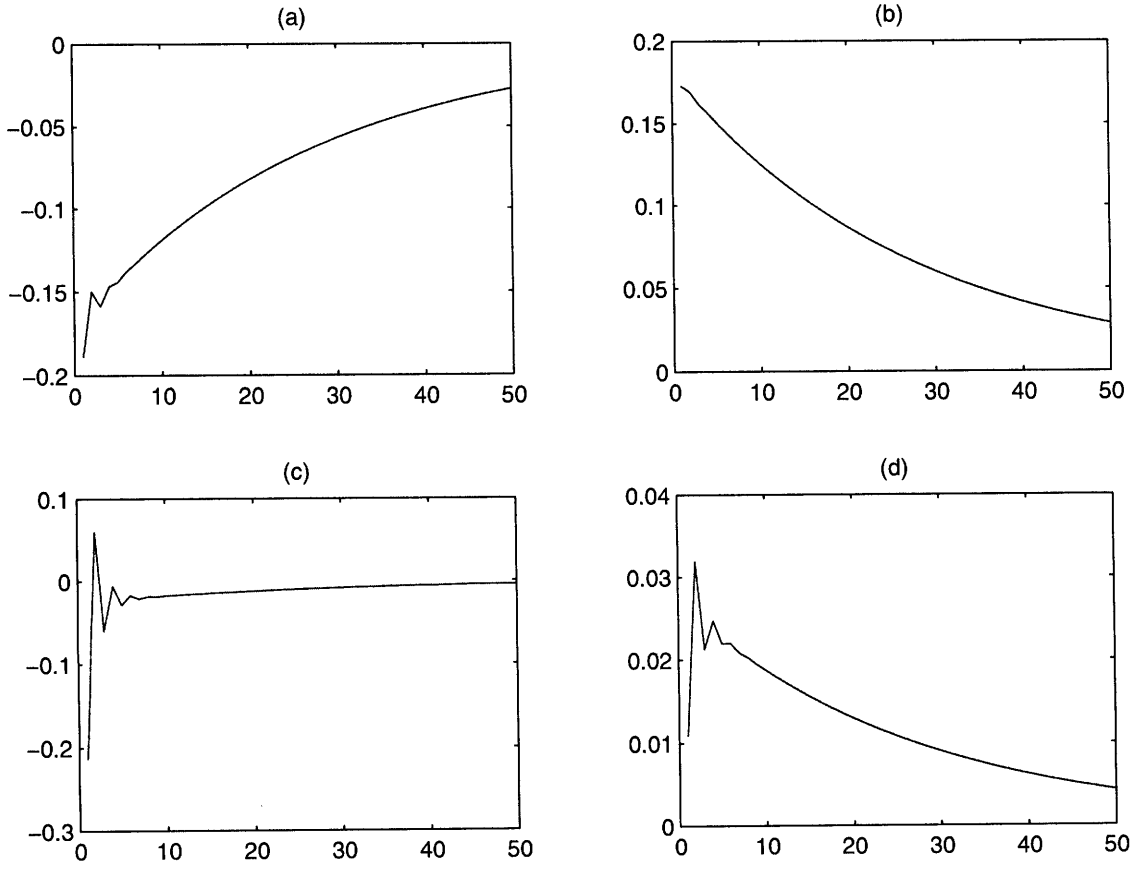


Figure 1-6: Decomposition of filtered expected return and filtered expected dividend growth over the history of observables.

(a) $\widehat{\mu}_t^r$ over $r_{t-\tau}$; (b) $\widehat{\mu}_t^r$ over $\Delta d_{t-\tau}$; (c) $\widehat{\mu}_t^d$ over $r_{t-\tau}$; (d) $\widehat{\mu}_t^d$ over $\Delta d_{t-\tau}$.

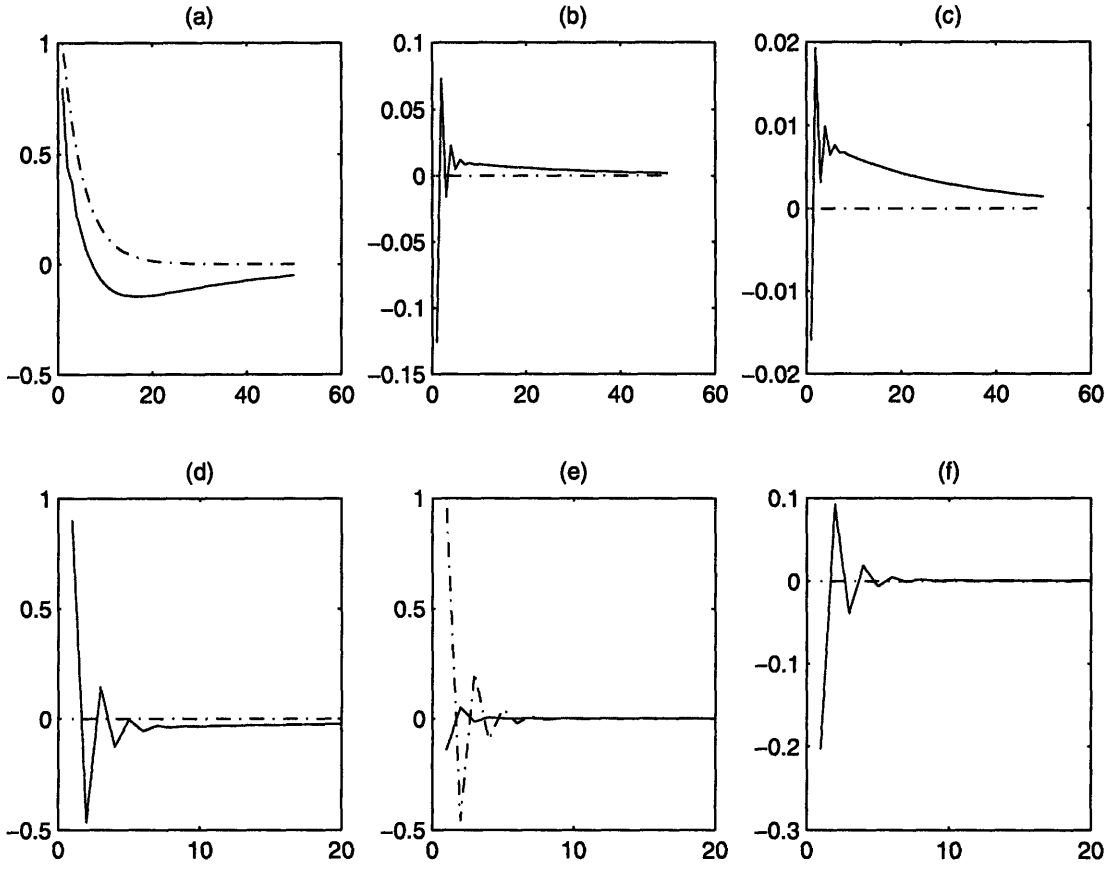


Figure 1-7: Decomposition of the filtered expected return and the filtered expected dividend growth over shocks.

(a) $\widehat{\mu}_t^r$ over $\varepsilon_{t-\tau}^{\mu r}$; (b) $\widehat{\mu}_t^r$ over $\varepsilon_{t-\tau}^{\mu d}$; (c) $\widehat{\mu}_t^r$ over $\varepsilon_{t-\tau}^{d}$; (d) $\widehat{\mu}_t^d$ over $\varepsilon_{t-\tau}^{\mu r}$; (e) $\widehat{\mu}_t^d$ over $\varepsilon_{t-\tau}^{\mu d}$; (f) $\widehat{\mu}_t^d$ over $\varepsilon_{t-\tau}^{d}$. Dashed-dot lines represent the corresponding decompositions of μ_t^r and μ_t^d.

future returns" in the variance of unexpected returns, the discount rate effect prevails and explains negative impact of realized returns on expectations. On the contrary, a positive shock to realized dividend growth shows that either expected dividend growth is high or unexpected shock to dividends is high. Conditioning on observable returns expected return also goes up to offset the effect of $\varepsilon_t^{\mu d}$ or ε_t^d. So it is natural that the expectations of future returns are revised upward.

The intuition behind the decomposition of $\widehat{\mu}_t^d$ is more complex. Indeed, observing positive innovation to returns while trying to make inference about the future dividend growth the econometrician admits that either $\varepsilon_t^{\mu d}$ or ε_t^d is positive or $\varepsilon_t^{\mu r}$ is negative. Since the disturbances $\varepsilon_t^{\mu r}$ and $\varepsilon_t^{\mu d}$ are highly positively correlated, the latter possibility implies that $\varepsilon_t^{\mu d}$ is negative. For the given set of parameters the second effect is stronger and it causes the downward revision of $\widehat{\mu}_t^d$. Next, the increase in Δd_t is mostly attributed to ε_t^d. Conditioning on realized returns it means that either $\varepsilon_t^{\mu d}$ is negative or $\varepsilon_t^{\mu r}$ is positive. Once again, because of high correlation between $\varepsilon_t^{\mu d}$ and $\varepsilon_t^{\mu r}$ the second option means higher $\varepsilon_t^{\mu d}$. This explains positive revision in $\widehat{\mu}_t^d$.

Figure 1-7 provides decompositions of constructed $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ and unobservable μ_t^r and μ_t^d over shocks $\varepsilon_{t-\tau}^{\mu r}$, $\varepsilon_{t-\tau}^{\mu d}$ and $\varepsilon_{t-\tau}^d$. As it can be expected, $\varepsilon_{t-\tau}^{\mu r}$ affects less $\widehat{\mu}_t^r$ than μ_t^r since the econometrician cannot perfectly distinguish a positive disturbance to expected returns and negative disturbance to expected dividend growth. So he updates $\widehat{\mu}_t^r$ less than he would do under full information. Also, impossibility to separate the impact of $\varepsilon_{t-\tau}^{\mu r}$ and $\varepsilon_{t-\tau}^{\mu d}$ ($\varepsilon_{t-\tau}^d$) leads to mistakes and non-zero weights on $\varepsilon_{t-\tau}^{\mu d}$ ($\varepsilon_{t-\tau}^d$).

Trying to predict future dividend growth the econometrician also assigns non-zero weights to shocks which do not affect unobservable μ_t^d. Thus, observing shifts in dividends and returns resulting from positive shock $\varepsilon_t^{\mu r}$ the econometrician is uncertain whether they are generated by positive shock $\varepsilon_t^{\mu r}$ or negative shock $\varepsilon_t^{\mu d}$. Since the higher weight is put on $\varepsilon_t^{\mu r}$ and shocks to expected returns and expected dividend growth are highly correlated, $\widehat{\mu}_t^d$ is revised upward. Similarly, positive shocks to $\varepsilon_t^{\mu d}$ or ε_t^d result in considering $\varepsilon_t^{\mu r}$ as negative and, again, due to high correlation between $\varepsilon_t^{\mu r}$ and $\varepsilon_t^{\mu d}$ lead to negative revision of $\widehat{\mu}_t^d$.

1.3.4 Filtering approach vs. predictive regression: comparison of robustness

One of the most important advantages of the filtering approach is its robustness to structural breaks in the long run relation between prices and dividends. In this Section I demonstrate this robustness on the sample of aggregate dividends and returns and claim that, whereas the forecasting power of the dividend-price ratio might be destroyed by structural breaks, the filtering approach can still provide accurate predictions of future returns.[20]

The intuition behind the relative robustness of the filtering approach is quite simple and appealing. It hinges on the fact that small shifts in the means of returns or dividend growth translate into noticeable breaks in the dividend-price ratio dpr_t. Indeed, if dpr_t is stationary then the mean level of the dividend-price ratio $\overline{dpr}$ in the linear approximation is related to the mean levels of returns and dividend growth $\overline{\mu}^r$ and $\overline{\mu}^d$ as

$$\overline{dpr} = \log\left(\exp(\overline{\mu}^r - \overline{\mu}^d) - 1\right).$$

For example, if $\overline{\mu}^r = 0.07$, $\overline{\mu}^d = 0.01$, then $\overline{dpr} = -2.78$. Now assume that $\overline{\mu}^d$ changes to 0.04. This shift is quite small relative to the standard deviation of Δd_t, which is 0.14. Consequently, if we examine the series r_t and Δd_t only, detecting this break will take some time and ignoring it will not produce a large error. Moreover, the estimated model coefficients will not be strongly affected by the break, and the estimate of $\overline{\mu}^d$ will gradually change from 0.01 to 0.04. This is what we can expect to observe in the filtering approach. However, the new mean level of the dividend-price ratio is -3.49, which corresponds to the break of 0.71. Since even the sample standard deviation of dpr_t is 0.41, a break of this size is easily noticeable and cannot be ignored. Effectively, this break will increase the sample variance and the sample autocorrelation of the dividend-price ratio. Moreover, the shift in the mean level of the dividend-price ratio translates into the shift in the intercept of the predictive regression, which in turn makes the slope coefficient biased downward.

I start the demonstration of the filtering approach robustness with comparison of the model parameters estimated in subsamples. Along with the whole sample, which spans the period 1926-2004, I also consider three subsamples. The first one is based on the post

[20]The idea that the forecasting power of the dividend-price ratio is ruined by structural breaks in its level was recently emphasized by Lettau and Van Nieuwerburgh (2006).

	ϕ_r	ϕ_d	$\sigma^2_{\mu r}$	$\sigma^2_{\mu d}$	σ^2_d	$\rho_{\mu r \mu d}$	$\rho_{\mu r d}$	R^2_r	R^2_d
$1926-2004$	0.8005	-0.4584	0.0014	0.0079	0.0106	0.8746	-0.3797	0.050	0.034
$1946-2004$	0.8489	-0.3968	0.0007	0.0088	0.0062	0.9108	-0.4109	0.061	0.138
$1926-1990$	0.7757	-0.5855	0.0017	0.0060	0.0117	0.8811	-0.3582	0.046	0.004
$1946-1990$	0.7957	-0.4819	0.0011	0.0078	0.0053	0.9061	-0.4210	0.082	0.140

Table 1.3: ML estimates of the benchmark state space model for various subsamples.

This table gives maximum likelihood estimates of the state space model (1.6)-(1.8) with $AR(1)$ processes for various subsamples. The identification strategy of Section 2 yields $\rho_{\mu dd} = 0$ for all subsamples. R^2_r and R^2_d measure the ability of $\widehat{\mu}^r_t$ and $\widehat{\mu}^d_t$ to predict future returns and future dividend growth in-sample.

World War II data and this choice acknowledges that the Great Depression and the WWII period might be "special". The second subsample is 1926-1990, and it is based on the concern that the Internet bubble period is "special". Also, as shown by Lettau and Van Nieuwerburgh (2005), it is likely that in the early 90's the dividend-price ratio experienced a structural break and this is another motivation to consider subsamples without the last 14 years. In the third analyzed subsample both suspicious periods are eliminated. The parameter estimates for different time periods are reported in Table 1.3.

Although the estimated parameters are not identical in all subsamples, the variation across different periods is strikingly small for most of them. In particular, all qualitative conclusions about parameters drawn for the whole sample are also valid in subsamples.This robustness is consistent with the provided intuition that the estimates of the Kalman filter parameters are not sensitive to structural breaks. Furthermore, the in-sample R^2 statistics for $\hat{\mu}^r_t$ in different subsamples are also very close to each other, suggesting that the filtering approach reasonably works in all periods, and its ability to predict returns was not ruined by structural breaks.

R^2 statistics for $\hat{\mu}^d_t$ exhibit a different pattern. If the Great Depression period is included, then they are relatively small, but if the period starts in 1946, $\hat{\mu}^d_t$ has much higher predictive ability with R^2 of 14%. This observation indicates that probably the period from 1926 to 1946 was special regarding the way how dividends were announced and paid, and the model does not capture this specificity correctly. However, this does not prevent the filtering procedure to predict returns adequately, although in periods with high R^2 for $\hat{\mu}^d_t$ the R^2 statistic for $\hat{\mu}^r_t$ is higher as well.

Next manifestation of the robustness of the filtering approach comes from the comparison of several empirical statistics with their model implied counterparts. In particular, it is

	$\sigma(r_t)$	$\sigma(\Delta d_t)$	$\sigma(dpr_t)$	$\rho(r_t)$	$\rho(\Delta d_t)$	$\rho(dpr_t)$	β_r	β_d	R_r^2	R_d^2
Panel A: 1926-2004										
I	0.1984	0.1400	0.4156	0.0461	−0.1399	0.9299	0.0947	0.0056	0.038	0.000
II	0.1964	0.1438	0.2567	−0.0589	−0.2236	0.8311	0.2341	0.0351	0.094	0.004
Panel B: 1946-2004										
I	0.1709	0.1322	0.4228	0.0180	−0.2326	0.9475	0.1005	0.0314	0.063	0.011
II	0.1605	0.1287	0.2660	−0.0366	−0.2484	0.8739	0.1826	0.0277	0.092	0.003
Panel C: 1926-1990										
I	0.2041	0.1380	0.2555	0.0383	−0.1520	0.8070	0.2699	0.0406	0.113	0.006
II	0.2017	0.1444	0.2411	−0.0706	−0.2557	0.8004	0.2614	0.0284	0.098	0.002
Panel D: 1946-1990										
I	0.1712	0.1266	0.2519	−0.0075	−0.2990	0.8531	0.2877	0.1062	0.191	0.049
II	0.1571	0.1244	0.2226	−0.0581	−0.3173	0.8194	0.2399	0.0272	0.116	0.002

Table 1.4: Empirical and implied statistics for the benchmark state space model.

This table summarizes the empirical values of statistics (row I) and the statistics implied by the estimated state space model (1.6)-(1.8) with $AR(1)$ processes (row II). $\sigma(r_t)$, $\sigma(\Delta d_t)$ and $\sigma(dpr_t)$ are standard deviations of aggregate stock returns, dividend growth and the log price-dividend ratio, respectively. $\rho(r_t)$, $\rho(\Delta d_t)$ and $\rho(dpr_t)$ are their autocorrelations. β_r, β_d, R_r^2 and R_d^2 are slopes and the R^2 statistics from predictive regressions of returns and dividend growth on the dividend-price ratio.

interesting to examine the characteristics of the dividend-price ratio and its real and implied abilities to predict future returns. Recall, that neither the estimation of the model, nor the construction of forecasters uses dpr_t directly. Hence, the analysis of the dividend-price ratio can provide independent insights about the model and the estimated parameters. As for the underlying model parameters, I study several subsamples.

The results are reported in Table 1.4. First of all, the empirical variances of returns and dividend growth almost coincide with the corresponding variances implied by the ML estimates of the parameters in all subsamples. Also, there is a reasonable fit for autocorrelations of returns and dividends. This means that although the model parameters are not fully identifiable, and although their estimates are not very precise, the model with obtained parameter values fits the major data characteristics quite well.

Next, I compute the variance and the autocorrelation of dpr_t implied by the state space model and compare them with their empirical counterparts. The estimated parameters imply that the dividend-price ratio should be highly autocorrelated and this is obviously supported by empirical data. However, the quantitative mismatches between the observed and implied statistics are different in various subsamples. Thus, in those subsamples that do not include the last 14 years, the implied variance and autocorrelation of the dividend-price ratio are almost identical to the corresponding sample values. On the contrary, when

the Internet bubble period is included, the model-implied and sample values are strikingly different. In particular, the observed dividend-price ratio is more persistent than the model dictates. Moreover, the sample variance of the dividend-price ratio is significantly higher than its implied value, which clearly fails to match its empirical counterpart. Note that exactly this discrepancy we can expect if a stationary time series experiences a structural break in its level: structural breaks in general increase the variance of the process and make it seem more persistent. Hence, the obtained mismatch is the first indication that probably there exist shifts in the model parameters which are too small to impact the filter parameter estimates, but which are amplified in the dividend-price ratio.

To provide further intuition, I also compute the empirical and implied statistics for conventional predictive regressions of returns and dividend growth on the dividend-price ratio. From Table 1.4 we get that in subsamples without the 90's, the implied and empirical slopes in predictive regressions almost coincide and dpr_t has a predictive power with quite high R^2, as suggested by the model. However, in the samples with the last 14 years, the sample regression coefficient goes down and a large gap between its empirical and implied values appears. For example, the empirical slope coefficient in the regression of returns on the dividend-price ratio in the whole sample is 0.0947, which is significantly less than the model implied value 0.2341. Simultaneously the ability to predict future returns goes down as well, and is smaller than it should be for the estimated set of parameters. Recalling a well-known econometric result that a structural break in the regression intercept leads to a downward bias in the slope estimator, we get one more indication that there was a structural break in the early 90's, which destroyed the predictive power of the dividend-price ratio. This is exactly what was reported by Lettau and Van Nieuwerburgh (2006).

Table 1.4 allows us to draw several other conclusions. Quite high value of the theoretical R^2 statistic says that the variation of the dividend growth is probably not a valid reason for poor predictive power of the dividend-price ratio. In other words, it means that for the estimated parameters predictability of dividend growth does not affect significantly the quality of the standard predictive regression. Thus, although variation of expected dividend growth and its positive correlation with expected returns works against the ability of dpr_t to predict returns, the predictive power of the ratio is still warranted by high persistence of expected returns relative to the persistence of expected dividend growth.

Along with testing predictability of returns by dividend-price ratio, there always was

interest to potential predictability of dividend growth by dpr_t. However, opposite to the case of returns, there is a consensus among researchers that dividend growth is unpredictable by the dividend-price ratio. Consistent with that, the observed slope of regression of dividend growth on the log dividend-price ratio is almost zero, signalling of no dividend growth predictability despite time varying expected dividend growth. Furthermore, the model-based regression slope is also very close to 0, so the absence of predictability here is not a consequence of structural breaks. Noteworthy, the dividend-price ratio fails to uncover variability of expected dividend growth even if the predictability of dividend growth is there. Comparison of theoretical and empirical R^2 statistics only supports this result.

To demonstrate the robustness of the filtering approach relative to the predictive regression I also run a Monte-Carlo experiment. I take the parameters of the model estimated for the whole sample and simulate 10 groups of artificial samples with 79 observations each. In all simulations the average expected dividend growth has a structural break in the middle of the sample, and the size of the break varies from 0.01 in the first group to 0.055 in the last group. Each group contains 600 simulated samples. For each sample, I estimate the model parameters and find in-sample R^2 statistics for $\hat{\mu}_t^r$ and for the forecast, based on the conventional predictive regression of returns on the dividend-price ratio. The average R^2 statistics for each group are plotted in Figure 1-8. Clearly, if the break is small enough the dividend-price ratio provides better forecast than $\hat{\mu}_t^r$. Indeed, the use of simulated dpr_t for predictions implies that the dividend-price ratio is stationary and this makes the forecaster more powerful. However, the higher power in cases with small breaks comes at cost. If the break is big, the dividend-price ratio looks like non-stationary in the finite sample, and it looses its ability to predict returns. On the contrary, the filter based forecast $\hat{\mu}_t^r$ does not rely on the stationarity assumption. Without this additional constraint on prices and dividends, $\hat{\mu}_t^r$ is less efficient providing lower R^2 statistics, but it is more robust to structural breaks. Indeed, its R^2 decreases more slowly than the R^2 of the dividend-price ratio as the break size goes up, and ultimately for sufficiently large breaks $\hat{\mu}_t^r$ outperforms dpr_t.

In summary, the suggested filtering approach is more robust to structural breaks in the long run relation between prices and dividends than the conventional predictive regression. This explains why the filtering approach works better in the whole sample of aggregate dividends and returns, which is likely to contain such breaks. Note that this property being interesting ex-post is particularly valuable ex-ante when it is not clear whether the

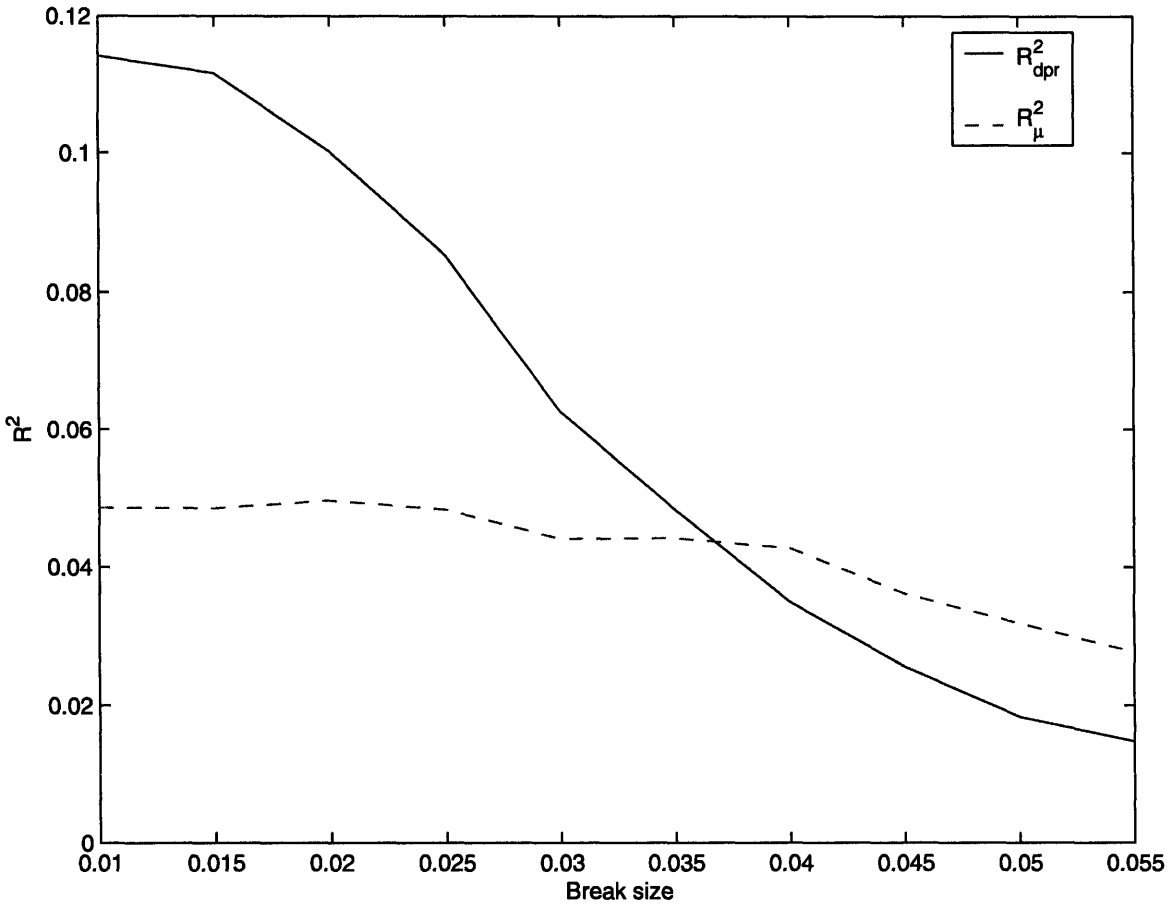


Figure 1-8: The effect of structural breaks on the predictive power of conventional regression and the filtering approach.

The horizontal axis shows the values of shifts in $\bar{\mu}^d$. I simulate 10 groups of artificial data, each group contains 600 samples with 79 observations each. R^2_{dpr} and R^2_{μ} are in-sample R^2 statistics for the dividend-price ratio and $\widehat{\mu}^r_t$, respectively, computed as an average of R^2s in each group.

structural breaks will occur.

1.4 Testing Hypotheses

Although the point estimates of parameters obtained in the previous section indicate that expected returns and expected dividend growth are time varying, only rigorous statistical tests can confirm it reliably. This section is devoted to such tests.

In general, there are three hypotheses of major interest. The first one is that expected returns are constant. In terms of the state space model parameters it can be stated as $\phi_r = \sigma_{\mu r} = \rho_{\mu r \mu d} = \rho_{\mu r d} = 0$. Clearly, testing this hypothesis is equivalent to examining predictability of stock returns. The second hypothesis is that expected dividend growth is constant, i. e. dividends are not predictable: $\phi_d = \sigma_{\mu d} = \rho_{\mu r \mu d} = \rho_{\mu d d} = 0$. Again, it is an advantage of the filtering approach that the test of this hypothesis draws on both returns and dividends and this helps to increase the power of the test. The third hypothesis is that the correlation between expected returns and expected dividend growth is negative: $\rho_{\mu r \mu d} < 0$. This hypothesis is inspired by recent discussions in the literature.

Given the log-likelihood function (1.10), it is natural to employ the maximum likelihood ratio test as the major tool for testing the above hypotheses. The test statistic has the following form:

$$LR(y) = \max_{\theta \in \Theta} \log L(\theta, y) - \max_{\theta \in \Theta_0} \log L(\theta, y), \tag{1.11}$$

where Θ_0 is the restricted set of parameters specified by the null hypothesis. The rejection region for the test is $\{y : LR(y) > C_\alpha\}$, i.e the null hypothesis is rejected if the value of the LR statistic computed for the empirical sample exceeds an appropriate threshold. The value of C_α is determined by the desired test level α and the distribution of the test statistic.

There are several complications related to practical realization of the maximum likelihood ratio test in our case. First, the inference based on the standard asymptotic distribution of the log likelihood ratio may be incorrect. Indeed, at least for the first two hypotheses we have a so called parameter-on-the-boundary problem since the values of parameters under null are on the boundary of the parameter set. As a consequence, the distribution of the test statistic might be non-standard even asymptotically. Moreover, the annual sample is too small to make the inference based on asymptotic distribution sufficiently reliable and, as

a result, I have to consider the finite sample distribution of the test statistic. Unfortunately, this distribution is not analytically feasible and Monte Carlo simulations are needed.

Second, all null hypotheses stated above are composite, and this immediately leads to the nuisance parameters problem. It means that the distribution of the test statistic depends on several unknown parameters. For example, for the null hypothesis of constant expected returns these parameters are ϕ_d, $\sigma_{\mu d}$, $\rho_{\mu r \mu d}$, and $\rho_{\mu d d}$. To warrant the desired test level α, the threshold of the rejection region C_α should satisfy the following inequality:

$$\sup_{\theta \in \Theta_0} P_\theta(LR(y) \geq C_\alpha) \leq \alpha,$$

where P_θ is a probability measure under particular set of parameters $\theta \in \Theta_0$. It means that the rejection region for a composite hypothesis is an intersection of rejection regions for all possible simple hypotheses from Θ_0. In general, construction of this intersection is an extremely formidable task which in most cases cannot be solved even numerically.

To avoid the last problem and make the inference feasible, I do not consider the whole space Θ_0 but focus only on a neighborhood of one particular point θ_0 such that $\theta_0 = \arg\max_{\theta \in \Theta_0} \log L(\theta, y_0)$, where y_0 is the given empirical sample.[21] The intuition behind this simplification is straightforward. Indeed, if the null hypothesis is true, then by construction it is most likely that the sample y_0 was drawn from P_{θ_0}. As the point representing a simple null hypothesis moves away from θ_0, the value of $LR(y_0)$ increases since the first term in (1.11) is the same but the second term decreases. However, the distribution of $LR(y)$ should not change significantly. Indeed, for each particular realization of the sample y the value of $LR(y)$ depends on benefits from tuning 8 parameters instead of 4. It is not likely that the reduction in the optimized log likelihood function depends significantly on the point in the four dimensional parameter space Θ_0. Consequently, the critical quantiles of $LR(y)$ also should not significantly change from point to point. Hence, if the simple hypothesis θ_0 is rejected, then all other simple hypotheses will also be rejected and the rejection region C_α is determined by P_{θ_0}: $P_{\theta_0}(LR(y) \geq C_\alpha) \leq \alpha$.

Equipped with this methodology, we are ready to test the particular hypotheses. The fist one is that the expected returns are constant. This is the most interesting hypothesis which attracted much attention. Alternatively, in our case it can be stated as impossibility

[21] In econometrics this procedure is termed as parametric bootstrap.

to predict future returns given historic records of past returns and dividends. Clearly, this is a composite hypothesis with Θ_0 parameterized by $(\phi_d, \sigma^2_{\mu d}, \sigma^2_d, \rho_{\mu dd})$ where Θ_0 is defined as

$$\Theta_0 = \{\theta \in \Theta : \phi_r = \sigma_{\mu r} = \rho_{\mu r \mu d} = \rho_{\mu r d} = 0 \}.$$

Note that opposite to the case with an unconstrained parameter space, there is no identification problem under null hypothesis and all parameters of the model are unambiguously determined by available observations. This fact simply comes from the structure of the matrix Ω defined in Proposition 2.

According to the described methodology, I first estimate the set of nuisance parameters θ_0 for the actual data sample y_0. The obtained estimates show that expected dividend growth must be very persistent with the autocorrelation coefficient of 0.961 if expected returns are constant. The value of the log likelihood function drops from 90.03 to 80.02, so the LR statistic for y_0 is 10.01. Next, I test the simple null hypothesis represented by the single point θ_0. Since the distribution of the LR statistic under the null is not analytically feasible, I take θ_0 as population parameters, simulate 2000 samples with 79 observations each, and compute the test statistic for each sample. Since the null hypothesis is simple, there is no maximization in the second term of (1.11). The simulated distribution of the LR statistic is presented in Figure 1-9.

To visualize the conclusion, the empirical value of the test statistic $LR(y_0)$ is indicated by the black bar. Clearly, its value is too high to be justified by a statistical error. In other words, it is very unlikely that under the null θ_0 the minimum of the restricted log likelihood function exceeds the unrestricted minimum value by 10.01. Thus, the simple null hypothesis θ_0 is rejected at least at the 1% level.

As discussed earlier, the rejection of θ_0 in general does not imply the rejection of the composite Θ_0. So, it is necessary to confirm the intuition that as θ moves away from θ_0 the value of the LR statistic for y_0 increases whereas the quantiles of its distribution remain approximately the same. For this purpose, I randomly take 50 points $\theta_i \in \Theta_0$, $i = 1...50$ which are in the neighborhood of θ_0 and compute the empirical values of the test statistic $LR_i = \max_{\theta \in \Theta} \log L(\theta, y_0) - \log L(\theta_i, y_0)$. As it can be expected, the far from θ_0 the point is, the larger the value of LR_i becomes. Next, I simulate the distribution of the LR statistic at each point θ_i and find the 5% quantiles q_i.[22] The pairs (q_i, LR_i), $i = 1...50$ are plotted

[22]Simulation of the distribution of the LR statistic is computationally intensive. To reduce the compu-

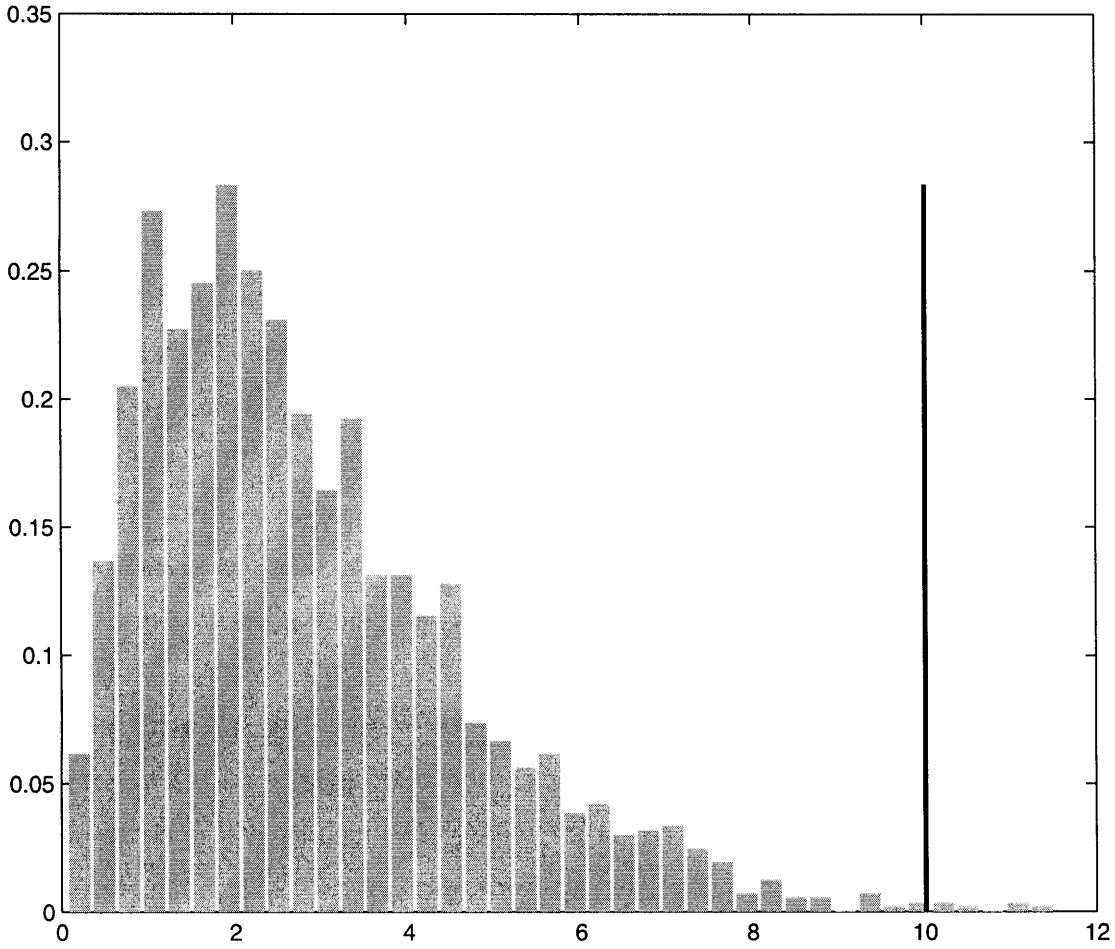


Figure 1-9: Simulated finite sample distribution of the LR statistic under the null hypothesis of constant expected returns.

The likelihood ratio test statistic is computed for 2000 simulated samples with 79 observations each. The black bar stands at the empirical value of the LR statistic.

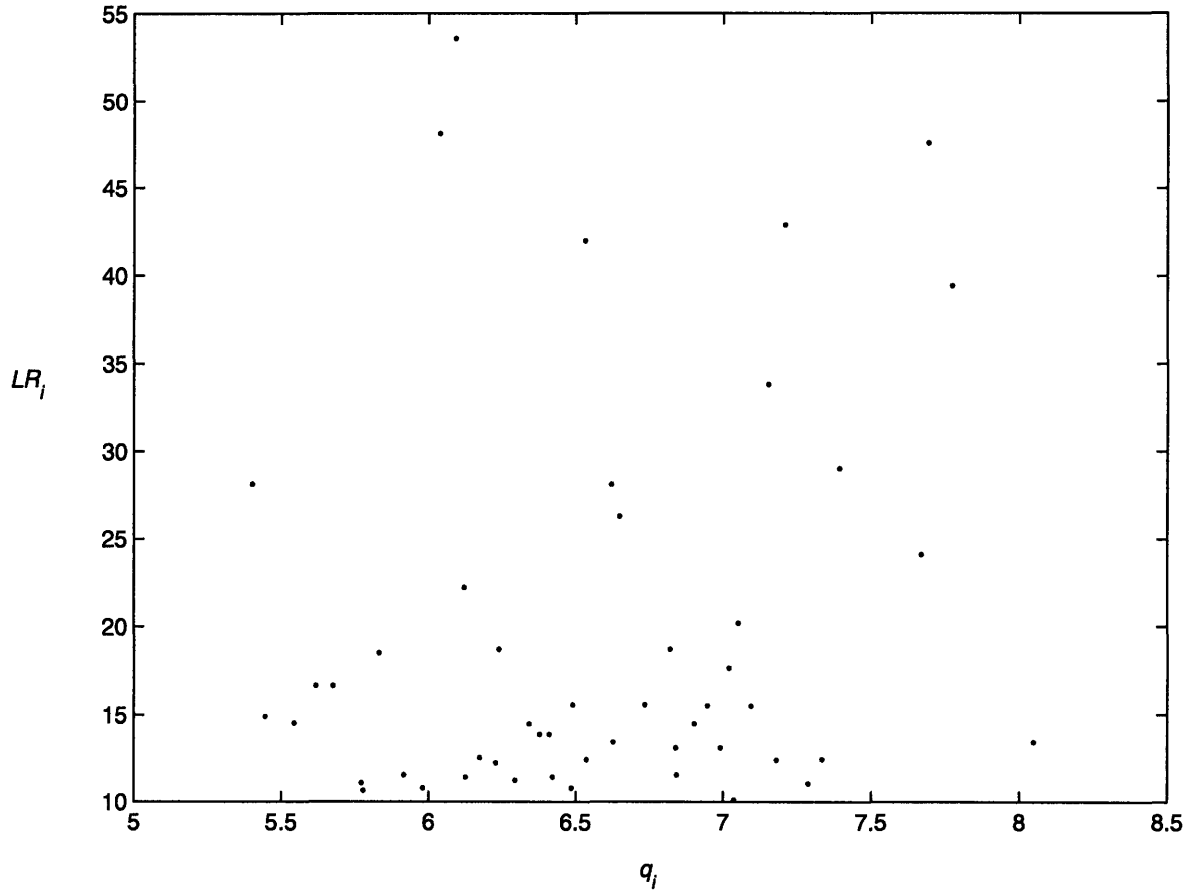


Figure 1-10: Scatter plot of the pairs (q_i, LR_i), $i = 1, 2, ..., 50$.
Here 50 points $\theta_i \in \Theta_0$, $i = 1, 2, ..., 50$ which are in the neighborhood of θ_0 are taken randomly. q_i, $i = 1, 2, ..., 50$ are 5% quantiles of the simulated distribution of the LR statistic at each point θ_i. LR_i, $i = 1, 2, ..., 50$ are empirical values of the test statistic at the points θ_i: $LR_i = \max_{\theta \in \Theta} \log L(\theta, y_0) - \log L(\theta_i, y_0)$.

in Figure 1-10.

It is easy to see that although values of the LR vary a lot and for sufficiently distant points they are around 50, the variation of quantiles q_i is limited to a compact range from 5 to 8.5. Thus, Figure 1-10 thoroughly supports the intuition provided above. Moreover, at every point the value of LR_i significantly exceeds the value of q_i. It means that the simple null hypotheses represented by θ_i are rejected at the conventional statistical level. Although

tational time, I simulate only 100 draws for each point θ_i, $i = 1, 2, ..., 50$ and take the fifth largest draw as q_i. According to Dufour (2005), the error in q_i computed in this way is relatively small and the simulated critical region has the level of 6/101. I do not consider 1% quantiles since the estimation error is quite high for them given the simulated sample of 100 points.

Figure 1-10 does not provide rigorous proof that all other simple hypotheses $\theta \in \Theta_0$ are rejected, it demonstrates that it is unlikely to find $\theta \in \Theta_0$ that will not be rejected by the suggested test. Overall, we have a solid statistical evidence that the empirical data is not consistent with constant expected returns.

The results of testing other hypotheses are less impressive. The parametric bootstrap allows me to reject the hypothesis of constant dividend growth only at the 12% level which is lower than conventional levels. Similarly, the hypothesis $\rho_{\mu r \mu d} < 0$ can be rejected only at the 15% level. This implies that either parametric bootstrap is too conservative and does not have enough power, or we really deal with constant expected dividend growth.

1.5 Out-of-Sample Analysis

Recently Goyal and Welch (2005) pointed out that, although conventional predictors of returns have some predictive power in sample, most of them underperform the naive historical average prediction out of sample. In particular, the dividend-price ratio gives poor out-of-sample forecasts. In this Section, it is demonstrated that the constructed predictive variables $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$ not only outperform the dividend-price ratio out of sample, but also provide forecasts superior to historical average.

A natural way to quantify the out-of-sample behavior of any predictive variable is to compare the error of the forecast it provides with the error of the simplest forecast based on the historical average. In particular, for all forecasting variables including the historical mean I calculate the mean absolute error (MAE) and the root mean squared error ($RMSE$) defined as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \widehat{y}_t|, \qquad RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y}_t)^2},$$

where y_t is the realized value of the forecasted variable and $\widehat{y}_t$ is its prediction. Here I compare three types of forecasts. The first one is based on the historical mean. The second one is generated by the standard predictive linear regression. In particular, to get the best estimate of the return $r_{\tau+1}$ at time τ given the history of returns r_t and the dividend-price ratios dpr_t up to time τ, I run the regression $r_{t+1} = a + bdpr_t + \varepsilon_{t+1}$ and obtain the estimates $\hat{a}$ and $\hat{b}$. Then, the best forecast is $\hat{r}_{\tau+1} = \hat{a} + \hat{b}dpr_\tau$. The third type of forecast is

	σ	MAE	ΔMAE	$RMSE$	$\Delta RMSE$
	Panel A: Prediction of returns				
Historical average of returns	0.0089	0.1409	0.0000	0.1715	0.0000
Log dividend-price ratio	0.0702	0.1485	-0.0076	0.1775	-0.0061
Filtered expected return	0.0236	0.1363	0.0047	0.1689	0.0025
	Panel B: Prediction of dividend growth				
Historical average of dividend growth	0.0052	0.1023	0.0000	0.1331	0.0000
Log dividend-price ratio	0.0145	0.1048	-0.0025	0.1371	-0.0040
Filtered expected dividend growth	0.0348	0.1004	0.0019	0.1313	0.0018

Table 1.5: Out-of-sample analysis.

This table reports out-of-sample forecasting power of the historical mean, the log dividend-price ratio, and the constructed forecasts $\hat{\mu}_t^r$ and $\hat{\mu}_t^d$. σ is a standard deviation of the predictor, MAE is a mean absolute error of prediction. $\Delta MAE = MAE_{hist} - MAE$, where MAE_{hist} is a mean absolute error of prediction based on historical average. $RMSE$ is a root mean squared error of prediction, $\Delta RMSE = RMSE_{hist} - RMSE$, where $RMSE_{hist}$ is a root mean squared error of prediction based on historical average.

provided by the constructed variables $\widehat{\mu}_t^r$ and $\widehat{\mu}_t^d$. To avoid a look-ahead bias, I estimate the parameters of the state space model (1.6) - (1.8) only on the data available to a fictitious observer at the moment τ and use the estimated parameters for predicting dividends and returns one year ahead. Thus, the parameters of the model are reestimated each year. To form the first forecast, I use 25 years of data, so the evaluation period starts in 1950.

The measures of predictive power for all discussed forecasting variables are reported in Table 1.5. First, both MAE and $RMSE$ indicate that the log dividend-price ratio lacks any ability to forecast returns out of sample and only adds noise to the naive predictor based on the historical mean. This is a replication of the Goyal and Welch (2005) result. However, the filtered expected return $\widehat{\mu}_t^r$ performs much better. Not only it gives smaller prediction error than the dividend-price ratio, but also beats the historical average. In particular, the $RMSE$ of the filtered expected return is 0.1689 reflecting a noticeable improvement relative to the $RMSE$ of the historical mean which is 0.1715. This error reduction approximately corresponds to the out-of-sample R^2 statistic of 3%.

Panel B of Table 1.5 provides the same out-of-sample statistics for different variables forecasting dividend growth. Since the dividend-price ratio cannot forecast dividends even in sample, one would not expect to see any out-of-sample predictability. Unsurprisingly, Table 1.5 confirms that there is absolutely no indication of ability of the dividend-price ratio to forecast dividend growth. Nevertheless, the filtered expected dividend growth $\widehat{\mu}_t^d$ possesses the predictive power even out-of-sample. Specifically, it decreases $RMSE$ from 0.1331 to 0.1313 and provides the forecast with the out-of-sample R^2 statistic of 2.7%.

Although the obtained out-of-sample results are noteworthy, their importance should not be overestimated. Campbell and Thompson (2005) and Cochrane (2006) show that given the limited sample size, the out-of-sample statistics do not say much about the real predictive power of forecasting variables. In particular, they demonstrate that even if all parameters are known, only expected returns are time varying, and the dividend-price ratio is a perfect forecaster of future returns, it is not surprising to get values of the Goyal-Welch statistic indicating poor out-of-sample performance of the forecasting variable. Indeed, in a finite sample it could happen that the historical mean prediction has lower forecasting error than the established forecaster because of simple bad luck. However, an opposite error can also occur: a variable which does not have any predictive power out of sample might produce lower $RMSE$ relative to the historical mean due to good luck. Although it is not likely that the above results are driven by this finite sample error, we must be aware of this possibility.

1.6 Extensions

1.6.1 Stock repurchases

In the previous analysis, I mostly focus on stock dividends as cash flows from the corporate sector to equity holders. However, dividends is not the only way to disgorge cash to investors and it is instructive to consider alternative measures of aggregate payout. In this section I study the implications of including stock repurchases into the definition of cash flows.[23] This modification might have a strong impact on predictability of returns. Thus, Boudoukh, Michaely, Richardson, and Roberts (2004) demonstrate that opposite to the dividend-price ratio the total payout ratio, defined as dividends plus repurchases over price, has statistically significant forecasting power for future stock returns.

In the filtering approach the redefinition of aggregate cash flows is equivalent to consideration of different trading strategy implemented by investors. Indeed, identifying dividends with cash flows we get a simple "buy and hold one share forever" strategy, which value is by definition the price per share. However, if we measure the transfer of resources from the corporate sector to investors as a sum of dividends and repurchases we get a "repurchasing"

[23] As reported by Jagannathan, Stephens, and Weisbach (2000), Fama and French (2001), Grullon and Michaely (2002) and others, stock repurchases are becoming a popular channel of returning cash to investors.

strategy: buy a number of shares, get dividends, and sell a part of them to the firm if the firm buys them back. This new strategy provides exactly the same return as the "buy and hold one share forever" strategy, but has different cash flows and, consequently, a different valuation ratio.[24]

Clearly, different strategies can be more or less appropriate for analysis of expected returns. Indeed, valuation ratios of different strategies might have different sources of time variation. In particular, some of them can be driven by changes in expected returns whereas others are mostly affected by changes in expected cash flows. There is abundant evidence, also corroborated in the previous section, that changes in the valuation ratio of the "buy and hold one share forever" strategy are mostly due to changes in expected returns, but not in expected cash flows. However, for alternative definitions of aggregate payout the conclusion can be the opposite. Bansal, Khatchatrian, and Yaron (2005) consider earnings as an alternative measure of cash flows and demonstrate that the major part of fluctuations in the price-earnings ratio is explained by fluctuations in the earnings growth rate. Bansal and Yaron (2006) examine an aggregate transfer of resources from all firms to all equity holders, which consists of dividends and repurchases net of stock issuance. The authors show that in this case at least 50% of asset price variability is explained by predictability of aggregate payout growth. Larrain and Yogo (2006) define aggregate cash flow even more generally including not only stock dividends and repurchases net of issuance, but also interest and debt repurchases net of issuance. They show that changes in expected cash flows account for the major part of the asset valuation ratio.

Here I demonstrate that the filtering approach applied to the "buy and hold one share forever" strategy and the "repurchasing" strategy provides comparable results regarding time variation of expected returns. I take the data on repurchases collected in Grullon and Michaely (2002) and adjust the testing period accordingly. The data in Grullon and Michaely (2002) cover the period from 1972 to 2000. Prior to 1972 the contribution of repurchases into aggregate cash flows is negligible and can be ignored.

Table 1.6 collects the estimation results. Panel (a) shows that the parameter estimates and the related qualitative conclusions are very similar for both strategies. In particular, the in-sample R^2 statistics are almost identical and the strategies provide the return forecasts of

[24]For the "repurchasing" strategy the cash flows are $CF_t = N_{t-1}D_t + P_t(N_{t-1} - N_t)^+$ where N_t is the number of shares held at time t.

(a)

	ϕ_r	ϕ_d	$\sigma^2_{\mu r}$	$\sigma^2_{\mu d}$	σ^2_d	$\rho_{\mu r \mu d}$	$\rho_{\mu r d}$	R^2_r	R^2_d
I	0.8751	−0.6348	0.0006	0.0049	0.0119	0.9017	−0.3475	0.031	0.025
II	0.8433	−0.7307	0.0008	0.0031	0.0159	0.8657	−0.1246	0.034	0.025

(b)

	$\sigma(r_t)$	$\sigma(\Delta d_t)$	$\sigma(dpr_t)$	$\rho(r_t)$	$\rho(\Delta d_t)$	$\rho(dpr_t)$	β_r	β_d	R^2_r	R^2_d
I	0.1973	0.1364	0.2813	0.0299	−0.1387	0.8407	0.1869	0.0184	0.0662	0.0012
II	0.1949	0.1499	0.2790	−0.0391	−0.2155	0.8513	0.1896	0.0086	0.0736	0.0003

Table 1.6: Repurchasing strategy.

(a) Maximum likelihood estimates of the state space model and in-sample R^2 for different strategies: I - "buy and hold one share forever" strategy; II - "repurchasing" strategy. (b) The empirical values of statistics (row I) and the statistics implied by the estimated state space model (1.6)-(1.8) with $AR(1)$ processes (row II) for the "repurchasing" strategy.

equal quality. This is one more manifestation of the filtering approach robustness. Panel (b) of Table 1.6 compares the empirical statistics and those implied by the estimated parameters in the case of the "repurchasing" strategy and demonstrates that the discrepancy between the empirical statistics and the implied statistics is notably less than for the "buy and hold one share forever" strategy (cf. Table 1.4).[25] In particular, the variance and the autocorrelation of the payout ratio are less than of the dividend-price ratio and are consistent with their implied counterparts. In addition, the R^2 statistic of the predictive regression is 6.6% which is higher than for the "buy and hold one share forever" strategy and is very close to the theoretical value of 7.4%. This supports the result of Boudoukh, Michaely, Richardson, and Roberts (2004) who argue that the total payout ratio but not the dividend-price ratio can reliably predict future returns.

The ML ratio test on time variation of expected return also provides very similar conclusions for both strategies. The simulated distribution of the LR statistic under the "repurchasing" strategy allows us to reject the null hypothesis of constant expected returns at the 3.5% level. Although this result is not as strong as under the "buy and hold one share forever" strategy reported in Section 1.4, it indicates the predictability of returns at the conventional statistical level.

[25] In the first row of Table 6 Panel (b) dpr_t stands for the payout ratio, which is dividends plus repurchases over price. In the second row of Table 6 Panel (b) dpr_t denotes the implied valuation ratio of the "repurchasing" strategy. Although these two objects are formally different, it is unlikely that the difference is essential so it is meaningful to compare their statistics.

1.6.2 Implications for asset allocation

Statistical evidence that aggregate stock returns are predictable can have important economic implications. For example, an investor who splits her assets between the stock market and the risk-free treasury bills must try to time the market on the basis on her changing expectations. As demonstrated by Kandel and Stambaugh (1996), Campbell and Viceira (1999), Barberis (2000), Campbell and Thompson (2005) and others, even small predictability of stock returns leads to a substantial effect on optimal portfolio weights. It means that even a small increase in the forecasting power might be very important for investors. In particular, the slight improvement brought by the filtering approach can lead to large welfare gains relative to the naive strategy ignoring the predictability.

To evaluate the benefits of the filtering procedure, I need to consider an investor who reestimates the model parameters each period, forms new forecasts of future stock returns, and allocates wealth to maximize her expected utility. In general, the trading policy of such investor is rather complicated and does not admit a tractable solution for several reasons. Indeed, due to a mean reverting nature of expected returns and expected dividend growth, the values of μ_t^r and μ_t^d determine the current investment opportunities. Since they change over time, a trading strategy of a multiperiod investor contains a hedging component which is complicated by the unobservability of expected returns and dividend growth to the investor.[26] Moreover, each period the investor reestimates the parameters of her model taking into account new information revealed in this period. Thus, a rational investor would also hedge changes in parameter estimates and this further complicates the portfolio problem.[27]

To illustrate the effect of the filtering approach on asset allocation avoiding the above complications, I assume that the investor has mean-variance preferences and at each moment cares only about the portfolio return one period ahead. The risk aversion parameter is set to 5.

I start with comparison of real wealth accumulated by investors who follow different strategies to form their expectations regarding future stock returns. The simplest approach is to ignore the predictability and to take an average of past returns as the best forecast.

[26] See, for example, Wang (1993) for solution to the portfolio problem of a multiperiod investor who filters out information about unobservable state variables from the history of available data.

[27] Xia (2001) solves the portfolio problem of an investor who takes into account the uncertainty in the parameters of the predictive relation.

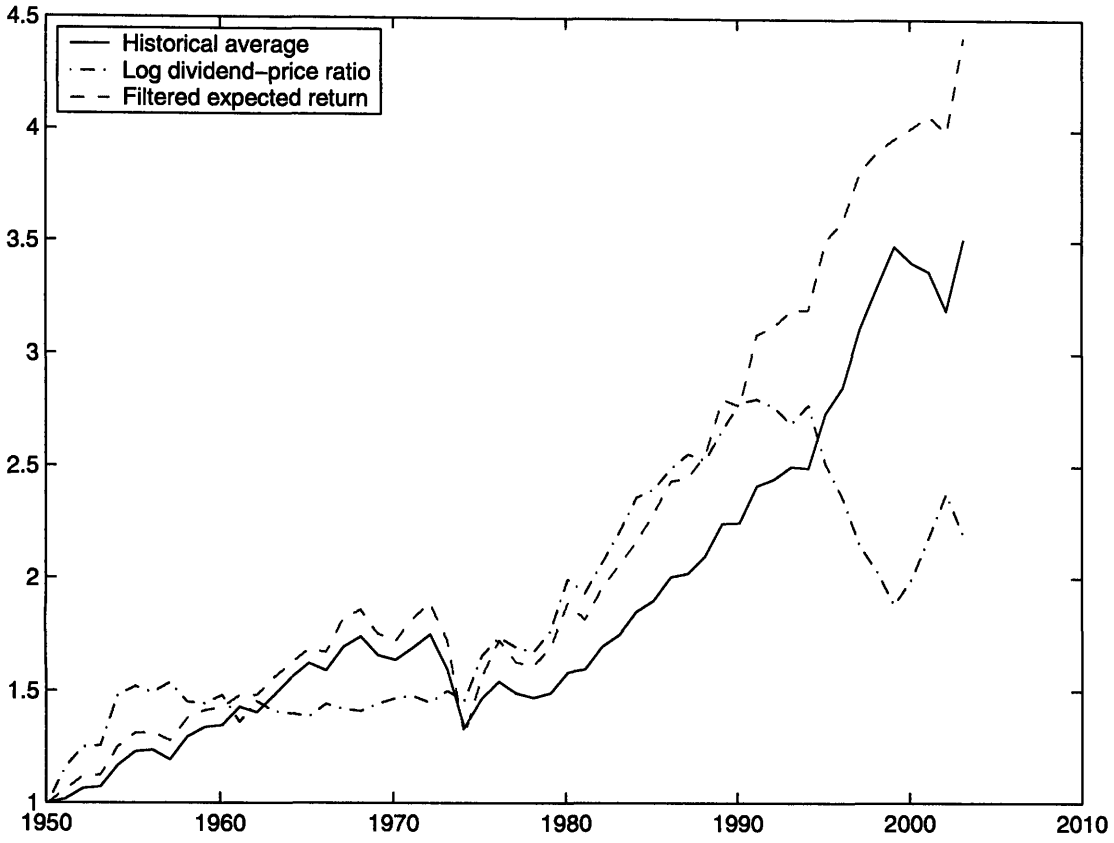


Figure 1-11: Accumulated real wealth of investors with one period horizon and mean-variance preferences.

Initial wealth is \$1. The solid line indicates wealth of the investor predicting future returns as an average of past returns. The dash-dot line shows the wealth accumulation by investors who use the dividend-price ratio to predict stock returns. The dashed line represents wealth of investors who follow the filtering procedure to form their expectations.

Another strategy is to use the dividend-price ratio as a predictive variable. Obviously, the most interesting question is about the relative performance of investors following the filtering approach.

Wealth accumulation of these three types of investors is depicted on Figure 1-11. It shows that the investors who try to filter out expected returns (dashed line) consistently outperform the investors who ignore the time variation in expected returns (solid line). Ultimately, those investors who do not time the market earn 2.5 dollars on each dollar invested in 1950 whereas the investors who follow the filtering strategy earn 3.5 dollars on the same initial investment. The strategy based on the dividend-price ratio (dash-dot line) also provides noteworthy results. Thus, in 2004 the investors who form their forecasts on

the conventional predictive regression earn 1.2 dollars on each dollar of initial investment and underperform not only those investors who use the Kalman filter to predict returns, but also naive investors who do not try to time the market at all. This disastrous result is mostly due to the last 15 years which seriously questioned the ability of the dividend-price ratio to predict stock returns. However, in the interim the strategy based on the dividend-price ratio outperforms not only naive strategy but sometimes even the filtering strategy. This is consistent with the conclusion that the conventional predictive regression provides more precise forecast and, consequently, higher returns in case of the stationary dividend-price ratio but is not robust to certain shifts in parameters.

Although accumulated wealth provides a clear metric for comparison of different predictive strategies, it is possible that some strategies are more risky than others and their higher return is not an indication of their superior predictive performance but a compensation for the risk. To study this possibility, I compute certainty equivalents of investors using different methods to predict returns. It appears that investors ignoring predictability get 1.46 whereas the certainty equivalent of investors who rely on the filtering approach is 1.79. Clearly, the latter approach is beneficial since it not only provides higher return but also increases the utility level.

Overall, the analysis of asset allocation under different approaches to predictability of stock returns demonstrates that the filtering approach not only provides statistically significant evidence in favor of return predictability, but also has economically important implications for portfolios of long term investors.

1.6.3 Additional robustness checks

Sensitivity to distributional assumptions

The parametric bootstrap used for hypothesis testing essentially hinges on the specified distribution. Throughout the analysis, I maintain the assumption that all variables are normal and draw all simulated shocks from the normal distribution. Here I study the impact of relaxing this assumption and perform the test on constant expected returns with the non-parametric bootstrap. In particular, I estimate the model parameters under the null hypothesis and use the realized returns and dividends to infer the realized values of $\varepsilon_t^{\mu d}$ and ε_t^d. Clearly, under the null all $\varepsilon_t^{\mu r}$ are zeros, and this allows us to reconstruct all other

shocks.[28] Then, instead of drawing simulated shocks from the normal distribution with the estimated covariance matrix, I draw the bootstrap sample with replacement from the set of realized $\varepsilon_t^{\mu d}$ and ε_t^d and construct the pseudo-sample of dividends and returns. This sample, in turn, is used for the computation of the ML ratio test statistic LR. I repeat these steps 2000 times and use the obtained distribution of LR for finding its quantiles and comparing them with the empirical value of the LR statistic. To save the space, I do not report the resulting distribution since qualitatively it is very similar to the simulated LR distribution under the normality assumption.

The results of the non-parametric bootstrap are very similar to the results obtained under normally distributed shocks. Again, the hypothesis of constant expected returns can be rejected at the significance level of 1.5%. This similarity of results indicates that the normality assumption is quite reasonable and does not drive the results of hypotheses testing.

Alternative process specifications

The benchmark model with expected returns and expected dividend growth specified as $AR(1)$ processes has a virtue of simplicity and a small number of parameters. In this section, I examine several extensions of the benchmark model and show that although more complicated models fit data slightly better, providing in general higher in-sample R^2 statistics, the simplest model captures most of the interesting effects. I do not run a horse race among different specifications, my purpose is to show that all of them provide comparable results, thus it is possible to rely on the simplest model with the minimum number of parameters.

I consider several modifications of the benchmark model. In particular, I examine specifications in which one or both state variables μ_t^r and μ_t^d follow $AR(2)$ processes. For brevity, I denote such models as $AR(1)/AR(2)$, $AR(2)/AR(1)$, and $AR(2)/AR(2)$, where the first and the second parts indicate the processes for μ_t^r and μ_t^d, correspondingly. Obviously, in this notation the benchmark model is $AR(1)/AR(1)$. The extended models have one or two additional parameters ϕ_{2r} and ϕ_{2d} corresponding to the second lags of $AR(2)$:

$$\mu_{t+1}^r = \phi_{1r}\mu_t^r + \phi_{2r}\mu_{t-1}^r + \varepsilon_{t+1}^{\mu r},$$

[28] I also take the unconditional mean $\bar{\mu}^d$ as the initial value of expected dividend growth.

$$\mu_{t+1}^d = \phi_{1d}\mu_t^d + \phi_{2d}\mu_{t-1}^d + \varepsilon_{t+1}^{\mu d}.$$

Opposite to the benchmark model, all models with at least one $AR(2)$ process are completely identifiable. In other words, not only the new parameters but also the covariance matrix Σ can be unambiguously recovered from the data.[29] Also, I consider the model with general VAR process for expected dividends and expected returns. Specifically, I assume that μ_t^r and μ_t^d evolve as

$$\begin{pmatrix} \mu_{t+1}^r \\ \mu_{t+1}^d \end{pmatrix} = \Phi \begin{pmatrix} \mu_t^r \\ \mu_t^d \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1}^{\mu r} \\ \varepsilon_{t+1}^{\mu d} \end{pmatrix},$$

where Φ is a (2×2) matrix, whose eigenvalues are all inside the unit circle. The identifiability of this model is very similar to the identifiability of the $AR(1)/AR(1)$ model. In particular, the matrix Φ in general is fully identifiable and its non-diagonal elements give two additional parameters to be estimated.[30] However, the covariance matrix Σ is not fully determined by data and an analog of Proposition 2 holds.

The estimated parameters for different models are quantitatively different, but they provide very similar conclusions, which are mostly the same as obtained from the benchmark $AR(1)/AR(1)$ model. Thus, expected returns are more persistent than expected dividends, but variance of shocks to expected dividends is higher. Again, except the VAR specification, the correlation between $\varepsilon_{t+1}^{\mu r}$ and $\varepsilon_{t+1}^{\mu d}$ is high and positive and in all models Campbell (1991) decomposition attributes the major part of variation in returns to the variation in expected returns.

The results of comparison the model implied parameters with their corresponding empirical values are almost identical to the benchmark case and are not reported for the sake of brevity. Again, the observed statistics of returns and dividends are closely matched, although the observed variance and autocorrelation of the dividend-price ratio are con-

[29] For example, the identifiability of new parameters in the $AR(2)/AR(2)$ model follows from

$$cov(r_t - \phi_{1r}r_{t-1} - \phi_{2r}r_{t-2}, r_{t-3}) = 0, \qquad cov(\Delta d_t - \phi_{1d}\Delta d_{t-1} - \phi_{2d}\Delta d_{t-2}, \Delta d_{t-3}) = 0,$$

$$cov(r_t - \phi_{1r}r_{t-1} - \phi_{2r}r_{t-2}, r_{t-4}) = 0, \qquad cov(\Delta d_t - \phi_{1d}\Delta d_{t-1} - \phi_{2d}\Delta d_{t-2}, \Delta d_{t-4}) = 0.$$

The identifiability of the covariance matrix Σ results from non-trivial second order autocorrelations of $\tilde{r}_t = r_t - \phi_{1r}r_{t-1} - \phi_{2r}r_{t-2}$ and $\widetilde{\Delta d}_t = \Delta d_t - \phi_{1d}\Delta d_{t-1} - \phi_{2d}\Delta d_{t-2}$.

[30] Identifiability follows from

$$cov\left[\begin{pmatrix} r_t \\ \Delta d_t \end{pmatrix} - \Phi \begin{pmatrix} r_{t-1} \\ \Delta d_{t-1} \end{pmatrix}, \begin{pmatrix} r_{t-2} \\ \Delta d_{t-2} \end{pmatrix}\right] = 0.$$

	R_r^2	R_d^2	ΔMAE	$\Delta RMSE$	ΔW	ΔW^{ce}
$AR(1)/AR(1)$	0.050	0.038	0.0047	0.0025	0.88	0.33
$AR(1)/AR(2)$	0.048	0.068	0.0044	0.0014	1.52	0.52
$AR(2)/AR(1)$	0.055	0.033	−0.0007	−0.0009	0.98	0.61
$AR(2)/AR(2)$	0.056	0.084	0.0038	−0.0010	1.03	0.54
VAR	0.072	0.042	0.0037	−0.0019	1.38	0.85

Table 1.7: Alternative specifications of expected dividends and expected returns.

This table gives in-sample and out-of-sample predictive power of $\hat{\mu}_t^r$ and $\hat{\mu}_t^d$ for alternative specifications of expected dividends and expected returns. R_r^2 and R_d^2 are in-sample R^2 statistics for $\hat{\mu}_t^r$ and $\hat{\mu}_t^d$. $\Delta MAE = MAE_{hist} - MAE$, where MAE_{hist} and MAE are mean absolute errors of predictions based on historical average and $\hat{\mu}_t^r$, respectively. $\Delta RMSE = RMSE_{hist} - RMSE$, where $RMSE_{hist}$ and $RMSE$ are root mean squared errors of prediction based on historical average and $\hat{\mu}_t^r$, respectively. ΔW is the difference of terminal wealths accumulated by investors, who predict returns with $\hat{\mu}_t^r$ and with the historical mean, ΔW^{ce} is the difference in their certainty equivalents.

spicuously higher than their model values. Also, as in the benchmark case, the empirical predictive power of the dividend-price ratio is weaker than dictated by model.

Table 1.7 allows to compare the in-sample fit of different models as well as to evaluate their out-of-sample behavior. In general, allowing two additional parameters I get better in-sample fit of the model since in all cases I use the same amount of data, but the increase in the number of model parameters provides additional flexibility. Interestingly, allowing $AR(2)$ processes mostly improves the ability of the model to predict future dividend growth whereas the captured variation in returns is almost the same as in the benchmark case. For instance, if both μ_t^r and μ_t^d are modelled as $AR(2)$ processes the R^2 statistic for $\hat{\mu}_t^r$ is 0.056, which is slightly higher than 0.05 obtained for the benchmark model, but the R^2 statistic for $\hat{\mu}_t^d$ goes up from 0.038 to 0.084. Importantly, the predictors obtained from different models are highly correlated with the correlation coefficients around 0.9. Although it does not prove that they all share the same predictive component, it is likely to be the case.

Similarly to the benchmark case, to evaluate the predictive ability of $\hat{\mu}_t^r$ out-of-sample I look at two metrics: the mean absolute error (MAE) and the root mean squared error ($RMSE$). Table 1.7 reports the differences ΔMAE and $\Delta RMSE$, which show how $\hat{\mu}_t^r$ from different models helps to decrease the error of prediction relative to the naive forecast based on the historical mean. The results for different models are mixed. On one hand, ΔMAE is positive except for $AR(2)/AR(1)$ specification saying that in general the constructed forecast is valuable. On the other hand, $\Delta RMSE$ is negative for three out of five specifications. Although the obtained out-of-sample performance does not unambiguously indicate that the models help to reduce the forecasting error, the results are not inconsistent with the

return predictability and a priory, given small sample size, the chance to get mixed results is quite high even if returns are really predictable.[31] Table 1.7 also provides the difference in wealth earned by a market timer, who used the filtering approach, and an agnostic investor, who used historical mean to form her best expectation of returns. Similar to the simplest $AR(1)/AR(1)$ model all extensions have positive ΔW, thus the market timing is rewarded. To make sure that this reward is not a compensation for extra risk I also compute the differences in certainty equivalents earned by the same investors ΔW^{ce}. For all models the investor who time the market does not reduce her certainty equivalent relative to the agnostic peer, thus she does not take an extra risk that can justify higher return.

Alternative data periods

In my analysis, I use annual data on dividends and returns where the period coincides with the calendar year, i.e. each forecast is constructed at the end of December. Due to seasonality of dividends, I cannot use higher frequency data directly in my approach, however the availability of monthly data can be used for an additional robustness check. Specifically, using monthly CRSP data on returns with and without dividends I construct new annual samples covering other calendar periods such as from February to January, from March to February, and etc. As in the benchmark case, I estimate the model parameters for each sample and construct one-year-ahead forecasts, which now are made at the end of each month. Obviously, the parameter estimates as well as the constructed forecasts in different samples are not independent one from another, since they effectively use overlapping data.[32] However, looking at 12 samples instead of 1 means that I use much more information, and this information helps to examine the robustness of the obtained results and even get new insights about predictability.

Overall, the obtained parameter estimates for other calendar periods are qualitatively similar to the benchmark January to December case and for the sake of brevity I do not report them. Again, expected returns are highly persistent with the parameter ϕ_r lying in the range from 0.75 to 0.82, the correlation $\rho_{\mu r \mu d}$ is above 0.8, the correlations $\rho_{\mu r d}$ and $\rho_{\mu d d}$ are negative. Hence, the model parameters obtained in the January to December sample are quite representative and their reasonable values are not the result of an accidental

[31] See Campbell and Thompson (2005) and Cochrane (2006).

[32] An interesting question for future research is how to test the hypothesis of constant expected returns simultaneously using all samples and accounting for their overlaps.

coincidence.

Additionally, I compare the empirical and model implied statistics for the dividend-price ratio in different calendar samples. In all of them I get exactly the same pattern as in the benchmark case: empirical variances and autocorrelations of the dividend-price ratio are higher than their model implied values and the predictive regression R^2 statistics are significantly lower than suggested by the model. It means that this mismatch is an intrinsic property of data supporting its interpretation as an indication of a structural break.

The most interesting behavior is demonstrated by the R^2 statistics for $\hat{\mu}_t^r$ and $\hat{\mu}_t^d$ in different calendar periods. To avoid confusion with the R^2 statistics in predictive regressions, I denote them as $R^2_{\mu^r}$ and $R^2_{\mu^d}$. As shown in Figure 1-12, they significantly change from period to period and their highest levels far exceed the values in the January to December sample. For example, for annual returns measured from July to June $R^2_{\mu^r}$ is around 7% and $R^2_{\mu^d}$ approaches amazing 25%. This indicates that the information about future dividends and returns containing in their history varies from sample to sample.

The statistics $R^2_{\mu^r}$ and $R^2_{\mu^d}$ exhibit an interesting pattern. Thus, they have their highest values in the calendar periods in which much information about dividends is released in the last month making the observable returns most informative. High value of $R^2_{\mu^d}$ is very intuitive in this case. Indeed, better information about future returns increases variation of μ_t^d and improves the predictive ability of $\hat{\mu}_t^d$ which manifests itself in higher $R^2_{\mu^d}$. More importantly, the predictive power of μ_t^r also goes up for these calendar periods. When there is much information about future dividends, the filtering approach is especially useful since it allows to disentangle expected dividends and expected returns providing a cleaner and thus more efficient forecaster for returns.

Notably, the R^2 statistic in the predictive regression for returns $R^2_{dpr}(r)$ exhibits a completely different pattern being around 4% for all samples. The predictive relation between the dividend-price ratio and returns is not that sensitive to additional information about future dividends and, consequently, the predictive power of the dividend-price ratio does not change from period to period. However, in the predictive regression of dividend growth on the dividend-price ratio the $R^2_{dpr}(\Delta d)$ statistic partially resembles the pattern of $R^2_{\mu^d}$. It means that despite its very low power to predict dividend growth, dpr_t still contains some information about future dividends, but it is possible to obtain much more information using dividends and returns separately.

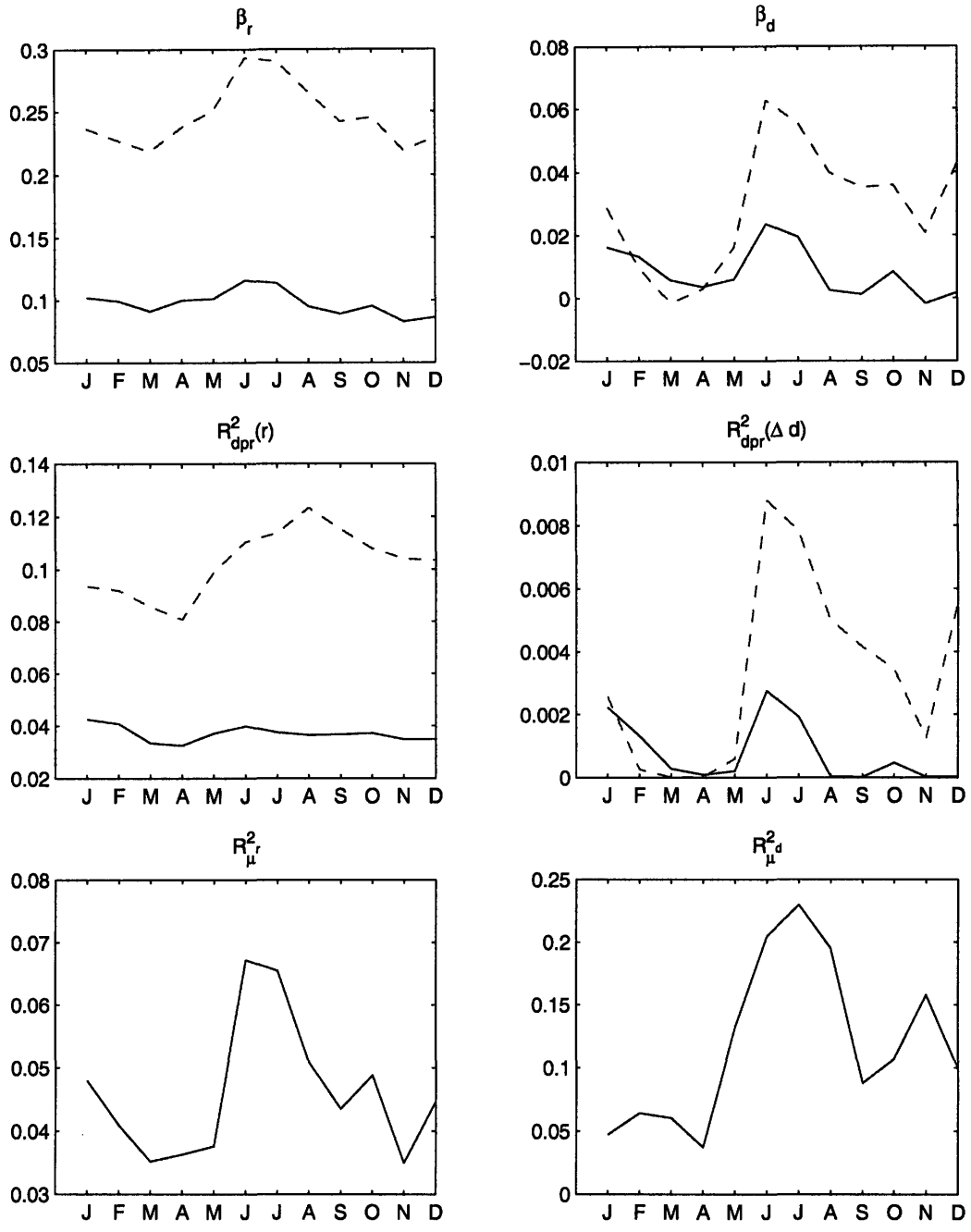


Figure 1-12: Predicting dividends and returns for alternative annual calendar periods. β_r and β_d are slope coefficients in the predictive regressions of returns and dividend growth on the dividend-price ratio; $R^2_{dpr}(r)$ and $R^2_{dpr}(\Delta d)$ are the corresponding R^2 statistics. Solid lines represent the empirical values, dashed lines indicate the model implied values. $R^2_{\mu^r}$ and $R^2_{\mu^d}$ are the R^2 statistics for filtered expected returns and expected dividend growth. Month abbreviations along the horizontal line denote the first month in each sample.

1.6.4 Adding other observables

The state space model of Section 1.2 uses only the history of dividends and returns. However, there might exist other useful information about future dividends and returns which would allow to improve the quality of the forecasts. In this section I examine an extension of the benchmark model with an additional observable q_t and show how to incorporate the information contained in the history of q_t in order to get the most powerful predictors for future dividends and returns.

In general, q_t can be any predictor of future dividends or returns suggested in the literature. Note that available predictors for Δd_{t+1} are also helpful for predicting returns in the filtering framework, and even if q_t contains information only about expected dividend growth, it still can be useful for predicting returns allowing to distinguish the shocks $\varepsilon_t^{\mu r}$ and $\varepsilon_t^{\mu d}$ and, as a result, making the forecast $\hat{\mu}_t^r$ more precise.

There is some ambiguity in how to model the relation of q_t to μ_t^r, μ_t^d, $\varepsilon_t^{\mu r}$ and $\varepsilon_t^{\mu d}$. For instance, we can think about q_t as an additional persistent state variable following $AR(1)$ process with the persistence parameter ϕ_q and the innovation ε_t^q, which is correlated with other innovations $\varepsilon_t^{\mu r}$, $\varepsilon_t^{\mu d}$, and ε_t^d. This approach was recently exploited by Pástor and Stambaugh (2006) in their predictive system framework. Effectively, the predictor q_t and unobservable expected returns μ_t^r share the same innovation, and the innovation ε_t^q allows to make inference about the unobservable shock $\varepsilon_t^{\mu r}$, which in turn helps to improve the forecast $\hat{\mu}_t^r$. In particular, if ε_t^q is perfectly correlated with $\varepsilon_t^{\mu r}$ and $\phi_q = \phi_r$, then q_t is a perfect proxy for μ_t^r.

Another way to treat q_t is to consider it as an additional observable linearly related to expected returns: $q_t = a\mu_t^r + \varepsilon_t^q$. This specification captures the idea that q_t is a proxy for the *level* of μ_t^r, but not for the innovations of μ_t^r. However, in this case we force μ_t^r to have exactly the same persistence as q_t, and this seems to be an unwarrantedly strong restriction.

In my analysis, I adopt an alternative framework. I assume that q_t is a linear combination of μ_t^r, μ_t^d and μ_t^q, where μ_t^q is an additional unobservable state variable, which follows $AR(1)$ process with the persistence parameter ϕ_q and the innovation $\varepsilon_t^{\mu q}$. On one hand, this approach allows to relate q_t to the level of μ_t^r, on the other hand, it is sufficiently flexible and does not impose any constrains on the autocorrelations of μ_t^r and q_t. The basic $AR(1)$

model of Section 1.2 augmented with the new observable has the following form:

$$\mu^r_{t+1} = \phi_r \mu^r_t + \varepsilon^{\mu r}_{t+1}, \qquad \mu^d_{t+1} = \phi_d \mu^d_t + \varepsilon^{\mu d}_{t+1}, \qquad \mu^q_{t+1} = \phi_q \mu^q_t + \varepsilon^{\mu q}_{t+1},$$

$$r_{t+1} = \mu^r_t - \frac{\rho}{1-\rho\phi_r}\varepsilon^{\mu r}_{t+1} + \frac{\rho}{1-\rho\phi_d}\varepsilon^{\mu d}_{t+1} + \varepsilon^d_{t+1}, \tag{1.12}$$

$$\Delta d_{t+1} = \mu^d_t + \varepsilon^d_{t+1},$$

$$q_{t+1} = a\mu^r_{t+1} + b\mu^d_{t+1} + \mu^q_{t+1}.$$

To reduce the number of parameters, I assume that $cov(\varepsilon^{\mu q}_{t+1}, \varepsilon^{\mu r}_{t+1}) = cov(\varepsilon^{\mu q}_{t+1}, \varepsilon^{\mu d}_{t+1}) = 0$. It means that the shock to expected returns $\varepsilon^{\mu r}_{t+1}$ affects q_{t+1} only directly through μ^r_{t+1}, but not through the correlation with $\varepsilon^{\mu q}_{t+1}$. If q_t represents only those variables that are supposed to be proxies for the level of expected returns, than the assumptions on the covariances is quite reasonable[33]. Thus, the model (1.12) has only one new correlation $\rho_{\mu qd} = cov(\varepsilon^{\mu q}_{t+1}, \varepsilon^d_{t+1})$ to be estimated.

As additional variables providing new information about future returns, I choose the book-to-market ratio BM_t and the equity share in total new equity and debt issues S_t proposed by Baker and Wurgler (2000). On the one hand, according to Goyal and Welch (2005), these variables are among the best in-sample predictors of future returns. On the other hand, S_t has the lowest correlation with the price-related predictors such as dpr_t and BM_t, and presumably it contains more additional information relative to the history of dividends and returns. Thus, we can expect to see the strongest impact from adding these particular variables as new observables.

To estimate the parameters of the model (1.12) I use MLE and the results are reported in Table 1.8. Adding new observables does not significantly change most of the benchmark model parameters. The only major difference is a positive sign of ϕ_d, and it can be explained either by misspecification of the process for expected dividend growth, or imprecision of parameter estimates.

Table 1.8 shows that the book-to-market ratio BM_t has high and positive loading of 3.501 on expected returns, but small and insignificant loading of -0.326 on expected divi-

[33]This assumption also helps to identify all other parameters of the system (1.12). Unlike the benchmark case, in the extended model the moments of observable variables non-linearly depend on unknown parameters. This makes the analysis of identifiability much more difficult since identifiability is now equivalent to uniqueness of the solution to a complicated system of non-linear equations. Although I don't have an analytical proof of identifiability in this case, my numerical analysis indicates that the system is identifiable.

	BM_t	S_t		BM_t	S_t		BM_t	S_t
ϕ_r	0.8324	0.8013	$\sigma^2_{\mu q}$	0.0097	0.0020	$\rho_{\mu qd}$	−0.475	−0.922
ϕ_d	0.2516	0.6932	σ^2_d	0.0182	0.0166	a	3.501	0.739
ϕ_q	0.9050	−0.0795	$\rho_{\mu r \mu d}$	0.6116	0.8465	b	−0.326	−3.314
$\sigma^2_{\mu r}$	0.0011	0.0026	$\rho_{\mu r d}$	−0.3226	−0.0070	R^2_r	0.047	0.132
$\sigma^2_{\mu d}$	0.0022	0.0011	$\rho_{\mu d d}$	−0.8451	−0.2137	R^2_d	0.087	0.111

Table 1.8: ML estimates of extended state space model.

Maximum likelihood estimates of the state space model (1.12) with the book-to-market ratio BM_t and the equity share in total new equity and debt issues S_t. R^2_r and R^2_d measure the ability of $\widehat{\mu}^r_t$ and $\widehat{\mu}^d_t$ to predict future returns and future dividend growth in-sample.

dend growth. It means that the book-to-market ratio is indeed a proxy for the level of μ^r_t and the contribution of μ^d_t is small enough not to ruin the predictive power of BM_t in the predictive regression. Table 1.8 also gives the R^2 statistics for $\hat{\mu}^r_t$ and $\hat{\mu}^d_t$ which now use a richer information set including not only the history of dividends and returns, but also the history of the new observable. Interestingly, the predictive systems with and without BM_t have very similar ability to predict future returns. Thus, adding the book-to-market ratio to the system does not increase the predictive power of $\hat{\mu}^r_t$ indicating that most of the relevant information is already contained in the history of dividends and returns (recall that the correlation between BM_t and $\hat{\mu}^r_t$ constructed from dividends and returns only is 0.71). However, BM_t helps to disentangle expected dividend growth and expected returns and this improves the predictability of dividend growth.

The equity issuance variable S_t exhibits a different pattern. Similar to BM_t it has a positive coefficient before μ^r_t and a negative coefficient before μ^d_t, however, the absolute value of b is greater than a, so S_t is a surprisingly better proxy for expected dividend growth than for expected returns. Thus, given a positive correlation between μ^r_t and μ^d_t, when expected dividend growth is high expected returns are also high and S_t is low (b is negative). This induces negative correlation between expected return and the equity issuance variable, which is consistent with the negative slope coefficient in the predictive regression of returns on S_t.

Interestingly, the OLS regression of future dividend growth on S_t also demonstrates some ability of S_t to predict Δd_{t+1}. In particular, the corresponding slope coefficient is negative and its $t-$statistic is -2.56. This result is consistent with high and negative coefficient b.

Unlike the case of BM_t, adding S_t significantly increases the predictive power of $\hat{\mu}^r_t$ and $\hat{\mu}^d_t$. The history of S_t allows to raise the R^2 statistics for $\hat{\mu}^r_t$ from 5% to 13% and the R^2

statistics for $\hat{\mu}_t^d$ from 3.5% to 11%. It means that S_t is likely to contain new information about future dividends and returns, which is not incorporated in the history of dividends and returns, and the filtering approach allows to merge efficiently these sources of information. This conclusion is also supported by low correlation between S_t and $\hat{\mu}_t^r$ constructed only from dividends and returns (cf. Table 1.2) and by the ability of S_t to predict dividends and returns in univariate regressions.

Higher R^2 statistics in the model (1.12) relative to the benchmark case deserve an additional comment. One might think that the improvement in the sample predictability almost mechanically results from the larger number of parameters. However, it is not the case. Recall that by construction, while maximizing the log-likelihood function, we maximize the fit of the model, but not the degree of predictability *per se*. Essentially, the optimization procedure matches statistical moments, and adding a new observable not only gives additional degrees of freedom in the parameter space, but also increases the number of moments needed to be matched. Hence, it is not clear a priori that new information will help to boost the predictive ability of the model, thus the fact that it helps is a non-trivial result.

1.7 Conclusion

In this chapter, I suggest a new approach to analysis of predictability of aggregate stock returns. This approach is more robust to structural breaks and allows dividend growth to be predictable. Overall, although many other approaches fail to provide statistical evidence for predictability, I demonstrate that the suggested robust method rejects the null of constant expected returns at a high confidence level.

In my research, I mostly focus on the predictability of stock returns and improvements that can be achieved by allowing for time varying expected dividend growth and relaxing the standard no-bubble constraint. However, the same approach enables us to examine the predictability of dividend growth. Although the price-dividend ratio lacks the power to predict future dividends, the history of returns and dividends allows one to construct a new variable that can uncover time variation in expected dividend growth.

I concentrate on the analysis of aggregate stock returns and aggregate dividend growth. However, the suggested approach has much broader applicability. It is a general method

of extracting expected returns and expected dividend growth from the realized dividends and returns, which is applicable to a wide range of portfolios and trading strategies. For instance, it can be used for analysis of industry portfolios, growth and value portfolios, and many others. Because of its advantages, the filtering approach would be especially valuable in the cases where dividend growth has a predictable component. As demonstrated above, the ability to distinguish innovations to expected returns and expected dividend growth can significantly improve the quality of forecasters. Since the filtering approach produces its own forecasting variable, it can be very useful when there are no other economically motivated exogenous variables that are expected to predict dividends and returns. Also, due to its relatively flexible assumptions the filtering approach is particularly advantageous when the valuation ratio is mildly non-stationary or can suffer from structural breaks. When it is not clear *ex-ante* whether the structural break will occur or not, the robustness of the filtering approach to such breaks is appealing.

One of the most promising portfolio applications of the suggested approach is the study of time variation of the value premium. Since the filtering approach does not need exogenous variables that are supposed to predict returns, it can be used for analysis of value and growth portfolios separately. This sort of analysis can give new insights about sources of value premium variation, as well as clarify the relation between value premium and business cycles. Understanding time variation of value premium will help to support or refute the existing theoretical explanations of why value stocks earn higher returns relative to growth stocks. Also, the filtering approach allows us to uncover innovations in expected returns and expected dividend growth for value and growth portfolios, and compute their covariations with innovations in market expected returns and market expected dividend growth. Doing this will allow one to reexamine discount rate betas and cash flows betas in the line of Campbell and Vuolteenaho (2004) and Campbell, Polk, and Vuolteenaho (2005), without relying on specific exogenous proxies for expected returns and avoiding the critique of Chen and Zhao (2006).

The applications of the filtering approach can go even further than a system of returns and cash flows. In general, it can be extended to any system with unobservable expectations and a present value relation type constraint. For example, one can filter out expected asset returns and expected consumption growth given data on asset returns and consumption and imposing a linearized budget constraint, as in Campbell (1993). Although predictive OLS

regressions give only mixed evidence for predictability of aggregate consumption growth, the filtering approach can shed new light on this question. This is another interesting avenue for future research.

1.8 Appendix A. Present value relation and generalized Campbell - Shiller linearization

In this Appendix I provide details on the generalized Campbell-Shiller linearization of the present value relation when expected returns and expected dividend growth are time varying. I call the considered linearization generalized because in contrast to its original version I do not impose the no-bubble condition.

The starting point is the well-known present value relation which in logs is

$$r_{t+1} = \Delta d_{t+1} + dpr_t + \log(1 + \exp(-dpr_{t+1})), \tag{1.13}$$

where $dpr_t = \log(D_t/P_t)$. This is an identity which holds for every period, so for all $\rho \in (0,1)$

$$dpr_t = E_t \sum_{i=1}^{\infty} \rho^{i-1}(r_{t+i} - \Delta d_{t+i}) + B_t, \tag{1.14}$$

where the last term B_t is

$$B_t = -E_t \sum_{i=1}^{\infty} \rho^{i-1} \left(\rho dpr_{t+i} + \log(1 + \exp(-dpr_{t+i}))\right).$$

It is important that the expectation operator in (1.14) conditions on the time t information available to market participants, but not to the econometrician. Plugging (1.2) into (1.14) we get

$$dpr_t = E_t \sum_{i=1}^{\infty} \rho^{i-1} \left(\mu^r_{t+i-1} - \mu^d_{t+i-1}\right) + B_t = E_t \sum_{i=0}^{\infty} \rho^{i} \left(\mu^r_{t+i} - \mu^d_{t+i}\right) + B_t.$$

The dynamics of expectations stated in Eq. (1.1) yield:

$$E_t(\mu_{t+i} - \bar{\mu}) = \Phi^i(\mu_t - \bar{\mu})$$

Consequently,

$$dpr_t = -\sum_{i=0}^{\infty} e_{12}\rho^i\Phi^i(\mu_t - \bar{\mu}) - \sum_{i=0}^{\infty}\rho^i e_{12}\bar{\mu} + B_t$$

$$= -\frac{e_{12}\bar{\mu}}{1-\rho} - e_{12}(1-\rho\Phi)^{-1}(\mu_t - \bar{\mu}) + B_t,$$

where $e_{12} = (-1, 1, 0, ..., 0)$. Introducing the deviations from averages $\widetilde{\mu}_t = \mu_t - \bar{\mu}$, we obtain the following representation of the price-dividend ratio in terms of expectations:

$$dpr_t = -e_{12}(1-\rho\Phi)^{-1}\widetilde{\mu}_t - \frac{e_{12}\bar{\mu}}{1-\rho} + B_t. \tag{1.15}$$

The linearization of the present value relation (1.13) originally derived in Campbell and Shiller (1988) is

$$r_{t+1} \approx -k + dpr_t - \rho dpr_{t+1} + \Delta d_{t+1} \tag{1.16}$$

where $\rho = 1/(1+\exp(\overline{dpr}))$, $k = \log(\rho) + \overline{dpr}(1-\rho)$ and the Taylor expansion is taken around a specified point $\overline{dpr}$. Note that at this stage only proximity of dpr_{t+1} to $\overline{dpr}$ is used. Substitution of (1.15) into (1.16) yields

$$r_{t+1} \approx -k + dpr_t - \rho dpr_{t+1} + \Delta d_{t+1} = -k + dpr_t + \rho\left(e_{12}(1-\rho\Phi)^{-1}\widetilde{\mu_{t+1}} + \frac{e_{12}\bar{\mu}}{1-\rho} - B_{t+1}\right) + \mu_t^d + \varepsilon_{t+1}^d$$

$$= -k - e_{12}(1-\rho\Phi)^{-1}\widetilde{\mu}_t - \frac{e_{12}\bar{\mu}}{1-\rho} + B_t + \rho\left(e_{12}(1-\rho\Phi)^{-1}(\Phi\widetilde{\mu}_t + \varepsilon_{t+1}) + \frac{e_{12}\bar{\mu}}{1-\rho} - B_{t+1}\right)$$

$$+\mu_t^d + \varepsilon_{t+1}^d = -k - e_{12}\bar{\mu} - e_{12}\widetilde{\mu}_t + \rho e_{12}(1-\rho\Phi)^{-1}\varepsilon_{t+1} + \mu_t^d + \varepsilon_{t+1}^d + B_t - \rho B_{t+1}$$

$$= -k + \mu_t^r + \rho e_{12}(1-\rho\Phi)^{-1}\varepsilon_{t+1} + \varepsilon_{t+1}^d + B_t - \rho B_{t+1}.$$

As a result, in the linear approximation we get

$$\varepsilon_{t+1}^r = \rho e_{12}(1-\rho\Phi)^{-1}\varepsilon_{t+1} + \varepsilon_{t+1}^d + B_t - \rho B_{t+1} - k. \tag{1.17}$$

Using the definition of B_t and applying the linearization again we get

$$B_t - \rho B_{t+1} - k = (E_{t+1} - E_t)\sum_{i=2}^{\infty}\rho^{i-1}\left(\rho dpr_{t+i} + \log(1+\exp(-dpr_{t+i}))\right).$$

As a result, unexpected returns can be presented in the following form

$$\varepsilon_{t+1}^{r} = Q\varepsilon_{t+1} + \varepsilon_{t+1}^{d} + (E_{t+1} - E_t)\sum_{i=2}^{\infty}\rho^{i-1}\left(\rho dpr_{t+i} + \log(1+\exp(-dpr_{t+i})\right),$$

where $Q = \rho e_{12}(1-\rho\Phi)^{-1}$.

1.9 Appendix B. Proof of Proposition 1

Let x_t be a joint vector combining past state variables μ_{t-1} with current shocks $\varepsilon_t^{\mu r}$, $\varepsilon_t^{\mu d}$, and ε_t^{d}: $x_t = (\mu_{t-1}, \varepsilon_t^{\mu r}, \varepsilon_t^{\mu d}, \varepsilon_t^{d})'$. Note, that no one component of x_t is observable at time t. With this notation, the state space system can be put into canonical form:

$$x_{t+1} = Fx_t + \Gamma\varepsilon_{t+1}^{x},$$

where

$$F = \left(\begin{array}{c|ccc} & 1 & 0 & 0 \\ & 0 & 1 & 0 \\ \Phi & 0 & 0 & 0 \\ & & \cdots & \\ & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad \Gamma = \begin{pmatrix} 0 & 0 & 0 \\ & \cdots & \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \varepsilon_t^{x} = \begin{pmatrix} \varepsilon_t^{\mu r} \\ \varepsilon_t^{\mu d} \\ \varepsilon_t^{d} \end{pmatrix}.$$

The observables $y_t = (r_t, \Delta d_t)$ are

$$y_t = Mx_t,$$

$$M = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & Q_1 & Q_2 & 1 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 0 & 1 \end{pmatrix}$$

and $Q = \rho e_{12}(1-\rho\Phi)^{-1}$. Expected returns and expected dividend growth are the first two

components of $\hat{\mu}_t$, which can be obtained from $\widehat{x}_t$ as

$$\hat{\mu}_t = \left(\begin{array}{c|ccc} & 1 & 0 & 0 \\ & 0 & 1 & 0 \\ \Phi & 0 & 0 & 0 \\ & & \cdots & \\ & 0 & 0 & 0 \end{array} \right) \widehat{x}_t,$$

Thus, the problem of constructing $\hat{\mu}_t^r$ and $\hat{\mu}_t^d$ reduces to obtaining $\widehat{x}_t = E[x_t | y_\tau : \tau \leq t]$. Applying the Kalman filter[34] we get a recursive equation for $\widehat{x}_t$:

$$\hat{x}_t = (I - KM)F\hat{x}_{t-1} + Ky_t,$$

where the Kalman gain matrix K along with the error covariance matrix $U = E[(x_t - \hat{x}_t)(x_t - \hat{x}_t)' | y_\tau : \tau \leq t]$ are determined from the set of matrix equations

$$U = (I - KM)(FUF' + \Gamma\Sigma\Gamma'),$$

$$K = (FUF' + \Gamma\Sigma\Gamma')M'[M(FUF' + \Gamma\Sigma\Gamma')M']^{-1},$$

where I is the identity matrix.

1.10 Appendix C. Proof of Proposition 2

To examine the identification of the model (1.6) - (1.8) with $AR(1)$ processes I look for combinations of parameters that can be uniquely determined given all population moments of observables. First, the parameters ϕ_r and ϕ_d are identifiable. Indeed,

$$cov(r_t - \phi_r r_{t-1}, r_{t-2}) = 0, \qquad cov(\Delta d_t - \phi_d \Delta d_{t-1}, \Delta d_{t-2}) = 0.$$

Consequently,

$$\phi_r = \frac{cov(r_t, r_{t-2})}{cov(r_t, r_{t-1})}, \quad \phi_d = \frac{cov(\Delta d_t, \Delta d_{t-2})}{cov(\Delta d_t, \Delta d_{t-1})}.$$

[34] Jazwinski (1970) provides a textbook discussion of linear filtering theory.

Next, notice that the observables $y_t = (r_t, \Delta d_t)$ follow the VARMA(1,1) process:

$$\begin{pmatrix} 1-\phi_r L & 0 \\ 0 & 1-\phi_d L \end{pmatrix} \begin{pmatrix} r_t \\ \Delta d_t \end{pmatrix} = \begin{pmatrix} \frac{L-\rho}{1-\rho\phi_r} & \frac{\rho}{1-\rho\phi_d}(1-\phi_r L) & 1-\phi_r L \\ 0 & L & 1-\phi_d L \end{pmatrix} \begin{pmatrix} \varepsilon_t^{\mu r} \\ \varepsilon_t^{\mu d} \\ \varepsilon_t^{d} \end{pmatrix}, \tag{1.18}$$

where L is a lag operator. Indeed, applying the operators $(1-\phi_r L)$ and $(1-\phi_d L)$ to (1.8) and (1.7) and using that $(1-\phi_r L)\mu_t^r = \varepsilon_t^{\mu r}$ and $(1-\phi_d L)\mu_t^d = \varepsilon_t^{\mu d}$ we get (1.18).

Making use of this representation, I show that the covariance matrix Σ of shocks $\varepsilon_t^{\mu r}$, $\varepsilon_t^{\mu d}$ and ε_t^d is identifiable up to one parameter. Note that it is not the case for a general VARMA model with a two-dimensional vector of observed variables and a three-dimensional vector of shocks. It is convenient to introduce the following modification of observables:

$$\widetilde{y}_t = \begin{pmatrix} \widetilde{r}_t \\ \widetilde{\Delta d}_t \end{pmatrix} = \begin{pmatrix} 1-\phi_r L & 0 \\ 0 & 1-\phi_d L \end{pmatrix} \begin{pmatrix} r_t \\ \Delta d_t \end{pmatrix}.$$

Here I explicitly use that ϕ_r and ϕ_d are identifiable. Hence,

$$\widetilde{y}_t = \begin{pmatrix} \frac{L-\rho}{1-\rho\phi_r} & \frac{\rho}{1-\rho\phi_d}(1-\phi_r L) & 1-\phi_r L \\ 0 & L & 1-\phi_d L \end{pmatrix} \begin{pmatrix} \varepsilon_t^{\mu r} \\ \varepsilon_t^{\mu d} \\ \varepsilon_t^{d} \end{pmatrix} = (A + BL) \begin{pmatrix} \varepsilon_t^{\mu r} \\ \varepsilon_t^{\mu d} \\ \varepsilon_t^{d} \end{pmatrix},$$

where

$$A = \begin{pmatrix} \frac{-\rho}{1-\rho\phi_r} & \frac{\rho}{1-\rho\phi_d} & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} \frac{1}{1-\rho\phi_r} & -\frac{\rho\phi_r}{1-\rho\phi_d} & -\phi_r \\ 0 & 1 & -\phi_d \end{pmatrix}.$$

Provided the observations $\widetilde{y}_t$ the only non-trivial moments I can construct are the following ones:

$$var(\widetilde{y}_t) = \begin{pmatrix} var(\widetilde{r}_t) & cov(\widetilde{r}_t, \widetilde{\Delta d}_t) \\ cov(\widetilde{r}_t, \widetilde{\Delta d}_t) & var(\widetilde{\Delta d}_t) \end{pmatrix} = A\Sigma A' + B\Sigma B', \tag{1.19}$$

$$cov(\widetilde{y}_t, \widetilde{y}_{t-1}) = \begin{pmatrix} cov(\widetilde{r}_t, \widetilde{r}_{t-1}) & cov(\widetilde{r}_t, \widetilde{\Delta d}_{t-1}) \\ cov(\widetilde{r}_{t-1}, \widetilde{\Delta d}_t) & cov(\widetilde{\Delta d}_t, \widetilde{\Delta d}_{t-1}) \end{pmatrix} = B\Sigma A'. \tag{1.20}$$

In total, $var(\widetilde{y}_t)$ and $cov(\widetilde{y}_t, \widetilde{y}_{t-1})$ contain 7 different statistics, so (1.19) and (1.20) give 7 linear equations for 6 unknowns $\sigma^2_{\mu r}, \sigma^2_{\mu d}, \sigma^2_d, \sigma_{\mu r \mu d}, \sigma_{\mu r d}, \sigma_{\mu d d}$. The matrix of this system is

$$M = \begin{pmatrix} \frac{\rho^2+1}{(1-\rho\,\phi_r)^2} & \frac{\rho^2\left(1+\phi_r^2\right)}{(1-\rho\,\phi_d)^2} & 1+\phi_r^2 & -\frac{2\rho\,(\rho+\phi_r)}{(1-\rho\,\phi_r)(1-\rho\,\phi_d)} & -\frac{2(\rho+\phi_r)}{1-\rho\,\phi_r} & \frac{2\rho\left(1+\phi_r^2\right)}{1-\rho\,\phi_d} \\ 0 & 1 & 1+\phi_d^2 & 0 & 0 & -2\,\phi_d \\ 0 & \frac{\rho\phi_r}{-1+\rho\,\phi_d} & 1+\phi_r\,\phi_d & \frac{1}{1-\rho\,\phi_r} & \frac{\rho+\phi_d}{-1+\rho\,\phi_r} & \frac{\rho-\phi_r+2\,\phi_d\rho\,\phi_r}{1-\rho\,\phi_d} \\ -\frac{\rho}{(1-\rho\,\phi_r)^2} & -\frac{\rho^2\phi_r}{(1-\rho\,\phi_d)^2} & -\phi_r & \frac{\rho\,(\rho\,\phi_r+1)}{(1-\rho\,\phi_r)(1-\rho\,\phi_d)} & -\frac{\rho\,\phi_r+1}{-1+\rho\,\phi_r} & \frac{2\rho\,\phi_r}{-1+\rho\,\phi_d} \\ 0 & 0 & -\phi_d & 0 & 0 & 1 \\ 0 & 0 & -\phi_r & 0 & \frac{1}{1-\rho\,\phi_r} & -\frac{\rho\,\phi_r}{1-\rho\,\phi_d} \\ 0 & \frac{\rho}{1-\rho\,\phi_d} & -\phi_d & -\frac{\rho}{1-\rho\,\phi_r} & \frac{\rho\,\phi_d}{1-\rho\,\phi_r} & \frac{2\,\rho\,\phi_d-1}{-1+\rho\,\phi_d} \end{pmatrix}$$

A tedious calculation shows that the rank of this matrix is 5. It means that the linear system is degenerate and only 5 equations are linearly independent. Therefore, the matrix M has a non-trivial kernel z: $Mz = 0$. This kernel is

$$z = \begin{pmatrix} -\frac{(1-\phi_r^2)(1-\rho\phi_r)^2}{(1-\rho\phi_d)^2} & -1+\phi_d^2 & 1 & -\frac{(1-\phi_r\phi_d)(1-\rho\phi_r)}{1-\rho\phi_d} & -\frac{\phi_r\,(1-\rho\,\phi_r)}{-1+\rho\,\phi_d} & \phi_d \end{pmatrix}'.$$

It corresponds to the matrix

$$\Omega = \begin{pmatrix} -\frac{(1-\phi_r^2)(1-\rho\phi_r)^2}{(1-\rho\phi_d)^2} & -\frac{(1-\rho\phi_r)(1-\phi_d\phi_r)}{1-\rho\phi_d} & \frac{\phi_r(1-\rho\phi_r)}{1-\rho\phi_d} \\ -\frac{(1-\rho\phi_r)(1-\phi_d\phi_r)}{1-\rho\phi_d} & -(1-\phi_d^2) & \phi_d \\ \frac{\phi_r(1-\rho\phi_r)}{1-\rho\phi_d} & \phi_d & 1 \end{pmatrix}$$

such that $A\Omega A' + B\Omega B = 0$, $B\Omega A' = 0$. It means that this matrix being appropriately rescaled can be added to Σ without any changes in the observable statistics. This completes the proof.

References

Andrews D. (1999) Estimation When a Parameter is on a Boundary, *Econometrica* 67, 1341-1383

Ang A. and G. Bekaert (2005) Stock Return Predictability: Is It There? Unpublished Manuscript, Columbia University

Baker, M. and J. Wurgler (2000) The Equity Share in New Issues and Aggregate Stock

Returns, *Journal of Finance* 55, 2219-2257

Bansal, R., V. Khatchatrian, and A. Yaron (2005) Interpretable Asset Markets? *European Economic Review* 49, 531-60

Bansal R. and A. Yaron (2004) Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance* 59, 1481 - 1509

Bansal R. and A. Yaron (2006) The Asset Pricing - Macro Nexus and Return - Cash Flow Predictability, Unpublished Manuscript

Barberis, N. (2000) Investing for the Long Run when Returns Are Predictable, *Journal of Finance* 55, 225-264

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts (2004) On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing, NBER Working Paper 10651

Boudoukh, J., M. Richardson and R. Whitelaw (2005) The Myth of Long-Horizon Predictability, Unpublished Manuscript

Brandt, M. W. and Q. Kang (2004) On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach, *Journal of Financial Economics* 72, 217-257

Burmeister, E. and K. Wall (1982) Kalman Filtering Estimation of Unobserved Rational Expectations with an Application to the German Hyperinflation, *Journal of Econometrics* 20, 255-284

Campbell J. (1990) Measuring the Persistence of Expected Returns, *American Economic Review Papers and Proceedings* 80, 43-47

Campbell J. (1991) A Variance Decomposition for Stock Returns, *Economic Journal* 101, 157-179

Campbell J. (1993) Intertemporal Asset Pricing without Consumption Data, *American Economic Review* 83, 487-512

Campbell J. and L. Viceira (1999) Consumption and Portfolio Decision when Expected Returns are Time Varying, *Quarterly Journal of Economics* 114, 433-495

Campbell J. and R. Shiller (1988) The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195-227

Campbell J. and T. Vuolteenaho (2004) Bad Beta, Good Beta, *American Economic Review* 94, 1249 - 1275

Campbell J., C. Polk, and T. Vuolteenaho (2005) Growth or glamour? Fundamentals and Systematic Risk in Stock Returns, Unpublished Manuscript, Harvard University

Campbell J. and S. Thompson (2005) Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average? Unpublished Manuscript, Harvard University

Campbell J. and M. Yogo (2006) Efficient Tests of Stock Return Predictability, *Journal of Financial Economics* 81, 27-60

Chen L. and X. Zhao (2006) Return Decomposition, Unpublished Manuscript

Chernozhukov V., H. Hong, and E. Tamer (2004) Inference on Parameter Sets in Econometric Models, Unpublished Manuscript

Cochrane J. (1992) Explaining the Variance of Price-Dividend Ratios, *Review of Financial Studies* 5, 243-280

Cochrane J. (2005a) *Asset Pricing*, Princeton University Press

Cochrane J. (2005b) The Dog That Did Not Bark: A Defense of Return Predictability, Unpublished Manuscript, University of Chicago GSB

Conrad J. and G. Kaul (1988) Time-Variation in Expected Returns, *Journal of Business* 61, 409-425

Dufour, J.-M. (2005) Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics, Unpublished Manuscript, University of Montreal

Durbin J. and S. Koopman (2001) Time Series Analysis by State Space Methods, Oxford University Press

Fama E. and M. Gibbons (1982) Inflation, Real Returns and Capital Investment, *Journal of Monetary Economics* 9, 297-323

Fama E. and K. French (1988) Dividend Yields and Expected Stock Returns, *Journal of Financial Economics* 22, 3-25

Fama E. (1990) Stock Returns, Expected Returns, and Real Activity, *Journal of Finance* 45, 1089-1108

Ferson, W. E., S. Sarkissian, and T. Simin (2003) Spurious Regressions in Financial Economics? *Journal of Finance* 58, 1393-1413

Goetzmann W. and P. Jorion (1993) Testing the Predictive Power of Dividend Yields, *Journal of Finance* 48, 663-679

Goetzmann W. and P. Jorion (1995) A Longer Look at Dividend Yields, *Journal of Business* 68, 483-508

Goyal A. and I. Welch (2005) A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, Unpublished Manuscript

Grullon G. and R. Michaely (2002) Dividends, Share Repurchases, and the Substitution Hypothesis, *Journal of Finance* 57, 1649-1684

Hamilton J. (1985) Uncovering Financial Market Expectations of Inflation, *Journal of Political Economy* 93, 1224-1241

Hamilton J. (1994) Time Series Analysis, Princeton University Press

Hansen L., J. Heaton, and N. Li (2005) Consumption Strikes Back? Measuring Long-Run Risk, Unpublished Manuscript

Hodrick R. (1992) Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies* 5, 357-386

Jagannathan, M., C. P. Stephens, and M. S. Weisbach (2000) Financial Flexibility and the Choice Between Dividends and Stock Repurchases, *Journal of Financial Economics* 57, 355 -384

Jansson, M. and M. Moreira (2006), Optimal Inference in Regression Models with Nearly Integrated Regressors, Unpublished Manuscript, Harvard University

Jazwinski, A. H. (1970) Stochastic Processes and Filtering Theory, Academic Press

Kandel, S. and R. F. Stambaugh, On the Predictability of Stock Returns: An Asset-Allocation Perspective, *Journal of Finance* 51, 385-424

Keim D. B. and R. F. Stambaugh (1986) Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics* 17, 357-390

Kothari, S.P. and J. Shanken (1992) Stock Return Variation and Expected Dividends, *Journal of Financial Economics* 31, 177-210

Kothari, S.P. and J. Shanken (1997) Book-to-market, Dividend Yield, and Expected Market Returns: A Time-series analysis, *Journal of Financial Economics* 44, 169 - 203

Lanne, M. (2002) Testing the Predictability of Stock Returns, *Review of Economics and Statistics* 84, 407-415

Larrain B. and M. Yogo (2006) Does Firm Value Move Too Much to be Justified by Subsequent Changes in Cash Flow? Unpublished Manuscript

Lettau M. and S. Ludvigson (2001) Consumption, Aggregate Wealth and Expected Stock Returns, *Journal of Finance* 56, 815 - 849

Lettau M. and S. Ludvigson (2005) Expected Returns and Expected Dividend Growth, *Journal of Financial Economics* 76, 583-626

Lettau M. and S. Van Nieuwerburgh (2005) Reconciling the Return Predictability Evidence: In-Sample Forecasts, Out-of-Sample Forecasts, and Parameter Instability, Unpublished Manuscript, New York University

Lewellen, J. (2004) Predicting Returns with Financial Ratios, *Journal of Financial Economics* 74, 209-235

Mankiw, N. G. and M. D. Shapiro (1986) Do We Reject Too Often? *Economics Letters* 20, 139-145

Menzly L., T. Santos, and P. Veronesi (2004) Understanding Predictability, *Journal of Political Economy* 112, 1-47

Pástor Ľ. and R. F. Stambaugh (2006) Predictive Systems: Living with Imperfect Predictors, Unpublished Manuscript

Paye B. and A. Timmermann (2005) Instability of Return Prediction Models, Unpublished Manuscript, Rice University

Pontiff J. and L. D. Schall (1998) Book-to-market Ratios as Predictors of market returns, *Journal of Financial Economics* 49, 141 - 160

Ribeiro R. (2004) Predictable Dividends and Returns: Identifying the Effect of Future Dividends on Stock Prices, Unpublished Manuscript, Wharton School - University of Pennsylvania

Schweppe, F. (1965) Evaluation of Likelihood Functions for Gaussian Signals, *IEEE Transactions on Information Theory* 11, 61 - 70

Schwert W. (1990) Stock Returns and Real Activity: A Century of Evidence, *Journal of Finance* 45, 1237 - 1257

Stambaugh, R. F. (1999) Predictive Regressions, *Journal of Financial Economics* 54, 375-421

Torous, W., R. Valkanov, and S. Yan (2004) On Predicting Stock Returns with Nearly Integrated Explanatory Variables, *Journal of Business* 77, 937-966

Valkanov R. (2003) Long-horizon Regressions: Theoretical Results and Applications, *Journal of Financial Economics* 68, 201-232

Wang, J. (1993) A Model of Intertemporal Asset Prices Under Asymmetric Information, *Review of Economic Studies* 60, 249-282

Xia, Y. (2001) Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation, *Journal of Finance* 56, 205-245

Chapter 2

Expected Returns on Value, Growth, and HML

2.1 Introduction

The filtering technique developed in the first chapter is very general and applicable not only to aggregate expected returns, but also to returns on many other portfolios and even trading strategies. In this chapter, I study time variation of the value premium and explore predictability of returns and dividends on value and growth portfolios.

Understanding time variation of the value premium is very important from theoretical point of view. The origin of the value premium has attracted much attention in the asset pricing literature, and empirical analysis of value premium dynamics might shed new light on this problem. Indeed, alternative explanations of the value premium differ in their implications for dynamical properties of expected returns on value and growth portfolios. Hence, the examination of time variation of the value premium serves as an additional empirical test for those theories.

Rational risk-based theories predict countercyclical behavior of the value premium[1]. For example, Zhang (2005) argues that this pattern naturally results from costly disinvestment coupled with the countercyclical price of risk. In bad times, it is more difficult for value firms to scale down their capital than for growth firms. As a result, value firms are adversely affected to a greater extent, and this makes them more risky in bad times. To compensate

[1]See, for example, Gomes, Kogan, and Zhang (2003), Zhang (2005), Kiku (2006), and others.

investors for this risk, expected returns go up widening the value premium. In contrast, the theories explaining value premium by various irrationalities in the behavior of a typical investor do not predict any cyclicality in the value premium[2].

From the practitioners' point of view, understanding predictability of the value premium might help to improve the performance of the strategy known as style timing. An investor who follows such a strategy shifts his wealth between two investment styles (value and growth) on the basis of predictions of relative style performance. To implement this strategy it is crucial to forecast the difference in returns on value and growth portfolios, so much effort has been made by practitioners to identify the variables possessing such forecasting power. In particular, such variables as innovations in industrial production (Sorensen and Lazzara, 1995), the forecast of the spread in the price-earnings ratios for value and growth portfolios, earnings revisions, and style specific risk spread (Fan, 1995) were argued to have some ability to predict relative performance of value and growth. Kao and Shumaker (1999) also mention several macroeconomic factors such as yield-curve spread (term premium), real bond yield, and earnings yield gap (the difference between the earnings-to-price ratio of the S&P 500 Index and the long-term bond yield) as good proxies for future value premium. One of the most reliable predictors of the value premium is the value spread, which is the difference between book-to-market ratios of growth and value portfolios (Asness, Friedman, Krail, and Liew, 2000; Cohen, Polk, and Vuolteenaho, 2003)[3]. In particular, Cohen, Polk, and Vuolteenaho (2003) find that the value spread is a significant predictor of the return on the HML portfolio constructed by Fama and French (1993).

Another interesting question is predictability of style indexes per se. The classification of mutual fund strategies in the value-growth dimension is widely used in industry where style indexes serve as benchmarks for style investing. As a result, returns on mutual funds have a factor structure with the style benchmark as a dominant factor. Predictability of the benchmark is important to investors who follow portfolio strategies that invest in mutual funds (Avramov and Wermers, 2006). In particular, if benchmark returns are predictable, the optimal portfolio is very different from what investors would choose in the absence of predictability, and tilts toward actively managed funds.

[2]For arguments in favor of a behavioral explanation of the value premium see Lakonishok, Shleifer, and Vishny (1994), La Porta, Lakonishok, Shleifer, and Vishny (1997), Rozeff and Zaman (1998), and others.

[3]Asness, Krail, and Liew (2003) argue that this result can be extended to value and growth portfolios constructed from country equity indexes.

Below, I apply the filtering technique developed in the first chapter of this dissertation to analysis of expected returns on value and growth portfolios as well as the value premium. The obtained results can be summarized as follows. First, the filtering approach allows me to construct new proxies for expected dividends and expected returns for value and growth portfolios, making use only of the history of dividends and returns. I show that filtered expected returns indeed have significant forecasting power for future returns on growth portfolios, but their ability to predict returns on the value portfolio is rather poor.[4] The corresponding R^2 statistic for the growth portfolio is 5%, whereas it is only 1% for value stocks. Moreover, dividends on the growth portfolio also appear to be highly predictable and the constructed forecaster manages to explain 27% of its time variation.

Next, I modify the filtering methodology to make it applicable to the value premium. I assume that the difference between expected log returns on value and growth portfolios as well as the difference between their expected log dividend growths follow AR(1) processes. An analog of the relaxed no-bubble condition is now imposed on the difference in the dividend-price ratios of value and growth portfolios. This difference is much more stable than the aggregate dividend-price ratio examined in the first chapter. Hence, it would not be unreasonable to assume that this difference is stationary and the no-bubble condition in its canonical form is satisfied. If this stronger assumption is valid, it is more efficient to use the history of returns and the dividend-price ratio differences instead of the history of returns and the differences in dividend growths. For comparison, I examine two types of the filtered value premium which differ in the data they use. The predictor based on dividends and returns is denoted as $\widehat{VP}^1$, whereas the predictor constructed from returns and the dividend-price ratios is denoted as $\widehat{VP}^2$.

The constructed forecasters $\widehat{VP}^1$ and $\widehat{VP}^2$ indeed predict the value premium. For the sample period 1950 - 2005 $\widehat{VP}^1$ explains 2% of time variation of the value premium. $\widehat{VP}^2$ is more powerful and explains around 7%. This difference in the predictive ability illustrates the importance of economic assumptions behind each forecaster: $\widehat{VP}^2$ appears to be much more powerful because it incorporates a stronger assumption on the behavior of the difference in the dividend-price ratios.

Given new predictors for the value premium, I get evidence that the value premium is

[4]Using a completely different framework, Guo, Su, and Yang (2006) also find that growth stock portfolios are generally more predictable than value stock portfolios or stock market indexes.

countercyclical. To establish this fact, I run contemporaneous regressions of the filtered value premium on several countercyclical variables such as filtered expected aggregate stock returns, the default premium, the book-to-market ratio, and the NBER recession dummy. I show that for both $\widehat{VP}^1$ and $\widehat{VP}^2$ the slope coefficients in almost all cases are positive and in many cases significant.

One of the most successful variables predicting the value premium is the value spread (Cohen, Polk, and Vuolteenaho, 2003). It contains additional information relative to the history of prices and dividends and its incorporation into the filtering framework can further improve the quality of the forecast. I demonstrate how to exploit the flexibility of the filtering approach and add this variable as an additional observable. Simple OLS regression shows that the value spread by its own can explain about 4% of in-sample time variation of the value premium. However, the expected value premium filtered from the data on returns, the dividend-price ratios, and the value spread is a much better predictor giving the R^2-statistic of 11%.

Time variation of the value premium has attracted much attention in the empirical literature. Petkova and Zhang (2005) argue that the value premium varies countercyclically, since value betas have positive correlation with the expected market risk premium, but growth betas have negative correlations. These results were extended to international markets by Fujimoto and Watanabe (2005). Also, countercyclical variation of the value premium was confirmed by Chen, Petkova, and Zhang (2006) who estimated conditional expected return spread between value and growth portfolios as a sum of expected dividend-price ratio and expected long-term growth rate of dividends. Another evidence that the spread in expected returns on value and growth portfolios displays countercyclical variations is provided by Kiku (2006). She defines "bad" times as periods with high consumption uncertainty and constructs the value premium by projecting the realized spread in returns on value and growth portfolios on lagged price-dividend ratios and dividend growth rates of these portfolios. Santos and Veronesi (2006) argue that the value premium is counter-cyclical by comparing average excess return of the extreme value and growth portfolios in "good" and "bad" states, which are defined through the price-to-book ratio of the market portfolio. When the market-to-book ratio is low the economy is in the "bad" state with high average market excess returns. In these periods the realized value premium is high, and this serves as evidence of counter-cyclical behavior of the value premium.

The rest of the chapter is organized as follows. Section 2.2 describes the filtering approach to the analysis of time variation of expected returns focusing on specificities related to the HML portfolio. Section 2.3 collects the main empirical results on predictability of value and growth portfolios and time variation of the value premium. Section 2.4 describes an extension of the benchmark model, which augments the data on dividends and returns with the value spread. Section 2.5 concludes.

2.2 Filtering Approach

2.2.1 Benchmark model

This section briefly describes the filtering approach to analysis of time variation in expected dividends and expected returns. Assume that an econometrician is given time series of realized cash flows (dividends) and returns generated by some portfolio or trading strategy. The only restriction on this portfolio or trading strategy is that the cash flows are positive. The problem of the econometrician is to utilize the available data and to construct the most efficient forecasts of future cash flow growth and returns imposing realistic restrictions on the joint behavior of prices and cash flows. The basic idea is to model logs of (demeaned) expected returns μ_t^r and (demeaned) expected dividend growth μ_t^d as latent state variables which are known to market participants, but unobservable to the econometrician. In the benchmark model I assume that μ_t^r and μ_t^d follow AR(1) processes:

$$\mu_{t+1}^r = \phi_r \mu_t^r + \varepsilon_{t+1}^{\mu r}, \qquad \mu_{t+1}^d = \phi_d \mu_t^d + \varepsilon_{t+1}^{\mu d}. \tag{2.1}$$

The econometrician observes realized returns and dividend growth, which are noisy proxies for past market expectations:

$$r_{t+1} = \mu_t^r + \varepsilon_{t+1}^r, \qquad \Delta d_{t+1} = \mu_t^d + \varepsilon_{t+1}^d. \tag{2.2}$$

The innovations $\varepsilon_{t+1}^{\mu r}$, $\varepsilon_{t+1}^{\mu d}$, ε_{t+1}^r, ε_{t+1}^d are assumed to be normally distributed and independent across time. However, there are no restrictions on their contemporaneous correlation structure. It is reasonable to assume that the dividend-price ratio cannot blow up quickly, so a relaxed no-bubble condition holds. In the linearized form this assumption produces a

linear restriction on the innovations $\varepsilon^{\mu r}_{t+1}$, $\varepsilon^{\mu d}_{t+1}$, ε^{r}_{t+1}, ε^{d}_{t+1} which has the following form[5]:

$$\varepsilon^{r}_{t+1} = -\frac{\rho}{1-\rho\phi_r}\varepsilon^{\mu r}_{t+1} + \frac{\rho}{1-\rho\phi_d}\varepsilon^{\mu d}_{t+1} + \varepsilon^{d}_{t+1}. \tag{2.3}$$

Here ρ is a linearization parameter. The system (2.1) - (2.3) can be represented as a canonical state space system with unobservable state variables $(\mu^r_{t-1}, \mu^d_{t-1}, \varepsilon^{\mu r}_t, \varepsilon^{\mu d}_t, \varepsilon^d_t)$ and observables $(r_t, \Delta d_t)$. Hence, the best linear estimate of unobservable expectations $\hat{\mu}^r_t$ and $\hat{\mu}^d_t$ is provided by the Kalman filter. The details of the Kalman filter implementation for this particular state space system can be found in Chapter 1.

The condition (2.3) is valid even if the dividend-price ratio dpr_t is mildly non-stationary. However, in some cases it is reasonable to impose a more restrictive condition and assume that this ratio is stationary. In this case, it is efficient to use the (demeaned) log dividend-price ratio as an additional observable, which in the linear apporximation is related to unobservable expectations as

$$dpr_t = \frac{\mu^r_t}{1-\rho\phi_r} - \frac{\mu^d_t}{1-\rho\phi_d}.$$

Since given the dividend-price ratio and returns it is possible to reconstruct dividend growth, one of these series is redundunt. For example, it is enough to use (r_t, dpr_t) as observables.

The Kalman filter assumes that the model parameters are known. However, in practice they must be estimated and the question is whether they all can be estimated unambiguously. As demonstrated in Chapter 1, the persistence parameters ϕ_r and ϕ_d are identifiable, but the covariance matrix of innovations $\varepsilon^{\mu r}_{t+1}$, $\varepsilon^{\mu d}_{t+1}$, and ε^d_{t+1} can be reconstructed up to one parameter. Consequently, many interesting statistics of the unobservable processes μ^r_{t+1} and μ^d_{t+1} are also unidentifiable and we can only find intervals where these parameters lie. In such cases, I will report only maximum and minimum values reached by the parameter on the identified set.

2.2.2 Value premium

The described procedure can be directly applied to the growth and value portfolios. Furthermore, with some modifications, this approach is applicable to analysis of the value premium, which is by definition the difference between expected returns on growth and value portfo-

[5]See details in Chapter 1.

lios (Fama and French, 1993). Although it is possible to filter out these expected returns separately and then compute the filtered value premium as a difference between them, this approach is not efficient. Indeed, value and growth stocks comprise a large part of the whole stock market, and returns on these portfolios to a large extent are determined by aggregate market returns. Hence, expected returns on the value and growth portfolios quite closely follow expected returns on the market portfolio. Taking the difference we cancel out the dominant component and the residual variation is largely represented by noise. Moreover, this approach would assume that the valuation ratios of the value and growth portfolios satisfy the relaxed no-bubble condition only individually. This assumption is too weak for uncovering time variation in the value premium.

Another way to uncover the value premium would be the application of the filtering approach to HML, since effectively the value premium is expected return on this portfolio. However, the HML portfolio cannot be described by the state space system (2.1) and (2.2), since HML might have negative dividend payments making log dividend growth ill-defined.

To overcome this problem and make the procedure more efficient, I modify the filtering approach in several ways. First, slightly abusing notation I denote the *difference* between expected log returns on the value and growth portfolios as μ_t^r. I assume that this difference follows an AR(1) process. Similarly, μ_t^d is the *difference* in expected log dividend growth between growth and value portfolios. Since both portfolios pay positive dividends, this difference is well defined. Correspondingly, r_{t+1} is now realized return on the HML portfolio, and Δd_{t+1} is the difference in realized realized dividend growth on value and growth portfolios (not dividend growth on the HML portfolio!)

Next, to filter the value premium more efficiently, I impose a constraint identical to (2.3), where now ε_t^r is unexpected innovation in the value premium, $\varepsilon_t^{\mu r}$ is innovation in the expected value premium, $\varepsilon_t^{\mu d}$ and ε_t^d are innovations in expected and unexpected difference of dividend growths. Although this assumption is an analog to the relaxed no-bubble condition introduced in Chapter 1, it is different from the assumption that the valuation ratio of each portfolio does not grow too fast. As can be shown, Eq. (2.3) is satisfied in the linear approximation if the *difference* between dividend-price ratios of growth and value portfolios is stationary and the common component of these ratios cannot grow up too fast. This new assumption is quite reasonable and economically motivated. Indeed, because of price comovements the dividend-price ratios of various portfolios also tend to

comove and Figure 2-1 supports this observation. As mentioned above, the assumption that the dividend-price ratios of each portfolio individually satisfy the relaxed no-bubble condition is relatively weak and does not impose sufficient restrictions on the innovations.

With the described modifications and changes in notation, the state space model for the value premium is identical to the benchmark model. Hence, all identification results hold and the Kalman filter for expected returns has exactly the same form. To save the space, I do not reproduce these results again.

2.2.3 Adding other observables

The benchmark model assumes that only data on dividends and returns are available to the econometrician. However, it is quite likely that investors as well as the econometrician possess other information, which might be helpful for predicting future returns. In the simplest case, this information is summarized by an additional variable q_t.[6] In general, q_t is related both to future returns and future dividend growth. With the new observable q_t, the state space model (2.1) and (2.2) takes the following form

$$\begin{gathered}
\mu^r_{t+1} = \phi_r \mu^r_t + \varepsilon^{\mu r}_{t+1}, \qquad \mu^d_{t+1} = \phi_d \mu^d_t + \varepsilon^{\mu d}_{t+1}, \qquad \mu^q_{t+1} = \phi_q \mu^q_t + \varepsilon^{\mu q}_{t+1}, \\
r_{t+1} = \mu^r_t - \frac{\rho}{1-\rho\phi_r}\varepsilon^{\mu r}_{t+1} + \frac{\rho}{1-\rho\phi_d}\varepsilon^{\mu d}_{t+1} + \varepsilon^d_{t+1}, \qquad \Delta d_{t+1} = \mu^d_t + \varepsilon^d_{t+1}, \qquad (2.4) \\
q_{t+1} = a_1\mu^r_{t+1} + a_2\mu^d_{t+1} + a_3\varepsilon^{\mu r}_{t+1} + a_4\varepsilon^{\mu d}_{t+1} + a_5\varepsilon^d_{t+1} + \mu^q_{t+1}.
\end{gathered}$$

The way in which q_t is added to the system (2.4) is quite general and agnostic why q_t captures future returns. Overall, there are two channels through which innovations to expected returns enter into q_t. First, a non-zero coefficient a_1 means that q_t is a proxy for the *level* of expected returns and $\epsilon^{\mu r}_{t+1}$ affects q_{t+1} through μ^r_{t+1}. Second, a non-zero coefficient a_3 means that q_t is a direct proxy for *innovations* in expected returns and is affected by $\epsilon^{\mu r}_{t+1}$ directly. It is an empirical question which channel works for each particular observable q_t. I do not exclude the possibility that q_{t+1} also captures innovations in expected cash flows and allow μ^d_{t+1} and $\epsilon^{\mu d}_{t+1}$ to enter q_{t+1} in the same way as μ^r_{t+1} and $\epsilon^{\mu r}_{t+1}$.

To make the model sufficiently flexible, I assume that the error term μ^q_{t+1} also follows AR(1) process with the persistence coefficient ϕ_q. This assumption allows to break the link

[6] In the empirical work, I use the value spread as an additional proxy for the value premium.

between autocorrelation of q_t and autocorrelations of μ_t^r and μ_t^d. However, to reduce the number of parameters and make the model identifiable I assume that μ_{t+1}^q is uncorrelated with other shocks. It means that all other shocks in the system affect q_{t+1} only directly and the strength of their effect is controlled by coefficients a_1,..., a_5. The indirect channel through μ_{t+1}^q is switched off.

2.3 Main Empirical Results

2.3.1 Data

The main data set used in the analysis consists of dividends and returns on the value and growth portfolios. These portfolios are constructed from the standard two-by-three independent sort on size and book-to-market (Fama and French, 1993).[7] As in the case of the market portfolio explored in the first chapter, I work only with annual data. The major problem arising on shorter horizons is seasonality in the dividend growth, and there is no unambiguous way to eliminate it. The value premium is conventionally defined as return on the HML portfolio, which is the value portfolio minus the growth portfolio (Fama and French, 1993). For robustness check, I also examine value and growth portfolios from the five-by-five sort on size and book-to-market.

Although the data on value and growth portfolios are available from 1927, I choose the year 1950 as the starting point of the sample. This choice is motivated by anomalous behavior of value and growth stocks during Great Depression and WWII, which seem to be quite different from the subsequent period. This is clear from Figure 2-1, which displays the log dividend-price ratios for growth and value portfolios. During the period 1927-1949 the ratio is much more volatile than afterwards, especially for value stocks. In particular, in the mid 1930's the dividend-price ratio for value stocks exhibits a trough, but by 1940 reverted back to its average historical level. This anomalous pattern is mostly explained by huge volatility of dividend growth in the 1930's. The observed unusual swings strongly suggest that the description of dividends and returns before and after 1950 by the same dynamic system with the same coefficients might be unwarranted[8].

[7]This data is available on Kenneth R. French's website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[8]Indeed, a preliminary estimation shows that the simplest system (2.1) does a poor job matching statistical moments of dividends and returns for the whole sample 1927-2005.

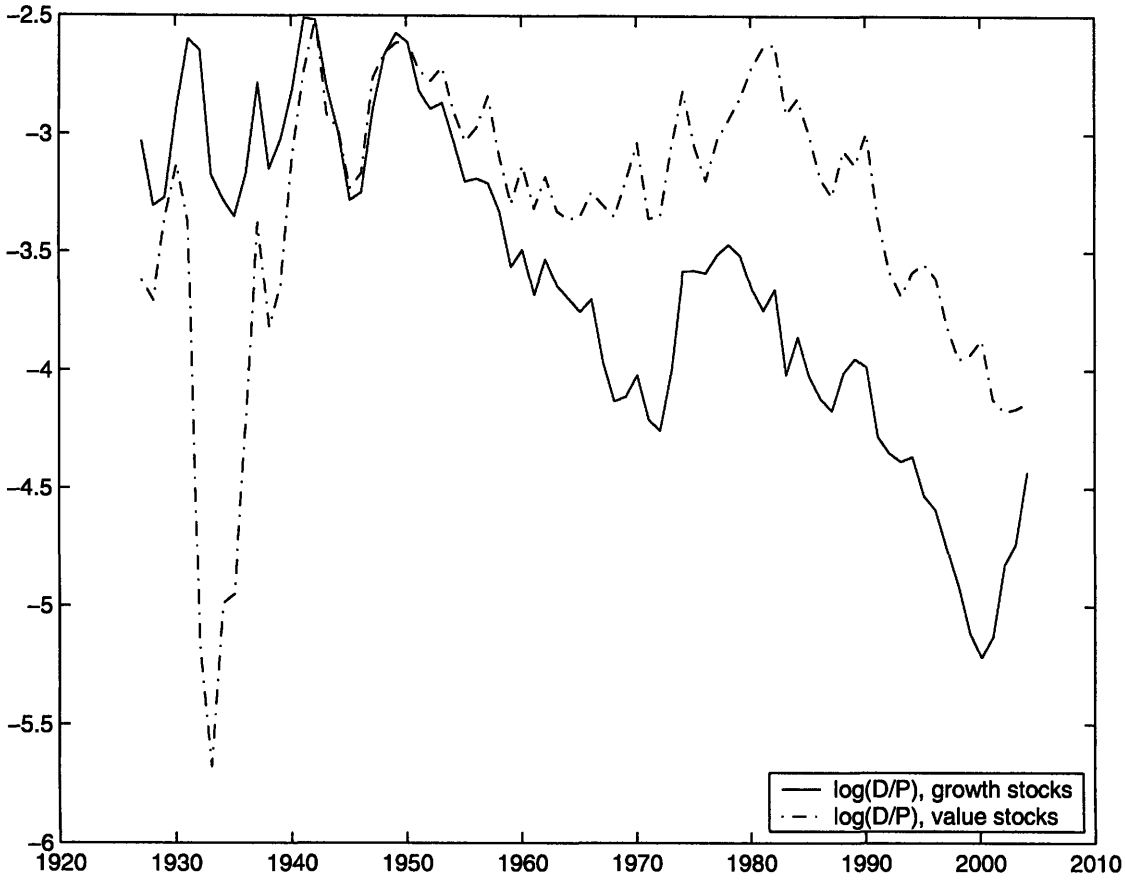


Figure 2-1: Log dividend-price ratio of value and growth portfolios.

Besides dividends and returns on growth and value stocks, this study exploits several other variables such as default premium, aggregate book-to-market ratio, equity share in total new and debt issues, which have been argued to predict aggregate stock returns or exhibit pronounced countercyclical pattern. Default premium is defined as the yield spread between Moody's Baa and Aaa corporate bonds with bond yields provided by Global Financial Data. The equity share in total new and debt issues is suggested by Baker and Wurgler (2000). The corresponding data can be found on Jeffrey Wurgler's website http://pages.stern.nyu.edu/~jwurgler/.

2.3.2 Value and growth portfolios

Before examining the value premium, I study time variation in expected returns and expected dividend growth for value and growth portfolios. I follow the standard definition of growth and value (Fama and French, 1993) based on six portfolios formed from sorts of stocks on market value and the book-to-market ratio. Value stocks are those with top 30% values of the book-to-market ratio, and growth stocks are in the bottom 30%. In the selected period from 1950 to 2005 both portfolios paid positive dividends, so it is possible to apply the filtering procedure described in Section 2.2.1 directly. Now μ_t^r and μ_t^d are expected returns and expected dividend growth on these portfolios.

Table 2.1 provides several estimated statistics of unobservable expected dividends and expected returns on value and growth portfolios. As discussed in the previous section, not all parameters of the model are point identifiable. For statistics that are not constant on the identified set I report the minimum (row 1) and maximum (row 2) values. Table 2.1 indicates that the obtained statistics are quite similar to their counterparts estimated for the market portfolio in Chapter 1. Again, expected returns are quite volatile, highly persistent, and correlated with expected dividend growth. Variance decomposition of unexpected returns indicates that "news about future returns" gives larger contribution than "news about future dividend growth". This similarity is not striking, since expected returns on growth and value portfolios are to a large extent driven by expected returns on the market portfolio.

Panel A: Growth Stocks											
$\sigma(\mu_t^r)$	$\sigma(\mu_t^d)$	$\rho(\mu_t^r, \mu_{t-1}^r)$	$\rho(\mu_t^d, \mu_{t-1}^d)$	$\rho(\mu_t^r, \mu_t^d)$	$\sigma(\varepsilon_t^r)$	$\rho(\varepsilon_t^r, \varepsilon_t^{\mu r})$	$\sigma^2(\eta_t^r)$	$\sigma^2(\eta_t^d)$	$-2\rho(\eta_t^r, \eta_t^d)$	R_r^2	R_d^2
0.055	0.07	0.95	0.53	0.45	0.207	-0.99	1.58	0.14	-0.78	0.05	0.27
0.056	0.10			0.47	0.207	-0.96	1.64	0.15	-0.74		
Panel B: Value Stocks											
$\sigma(\mu_t^r)$	$\sigma(\mu_t^r)$	$\rho(\mu_t^r, \mu_{t-1}^r)$	$\rho(\mu_t^d, \mu_{t-1}^d)$	$\rho(\mu_t^r, \mu_t^d)$	$\sigma(\varepsilon_t^r)$	$\rho(\varepsilon_t^r, \varepsilon_t^{\mu r})$	$\sigma^2(\eta_t^r)$	$\sigma^2(\eta_t^d)$	$-2\rho(\eta_t^r, \eta_t^d)$	R_r^2	R_d^2
0.052	0.03	0.89	-0.45	0.21	0.211	-0.78	0.85	0.42	-0.30	0.01	0.05
0.053	0.11			0.26	0.211	-0.77	0.88	0.43	-0.28		

Table 2.1: Statistics of expected returns and expected dividends for value stocks and growth stocks.

This table collects various statistics of expected returns and expected dividend growth for the value and growth portfolios. $\sigma(\mu_t^r)$ and $\sigma(\mu_t^d)$ are standard deviations of expected returns and expected dividend growth, $\rho(\mu_t^r, \mu_t^d)$ is a the correlation between them. $\rho(\mu_t^r, \mu_{t-1}^r)$ and $\rho(\mu_t^d, \mu_{t-1}^d)$ are autocorrelations. $\sigma(\varepsilon_t^r)$ is the standard deviation of unexpected stock return; $\rho(\varepsilon_t^r, \varepsilon_t^{\mu r})$ is the correlation between unexpected returns and the innovation in expected returns. $\sigma^2(\eta_t^r)$, $\sigma^2(\eta_t^d)$, and $-2\rho(\eta_t^r, \eta_t^d)$ represent Campbell (1991) decomposition of unexpected stock returns into "news about future returns", "news about future dividend growth", and the correlation term. R_r^2 and R_d^2 are in-sample R^2 statistics for filtered expected returns and expected dividend growth. For non-identified statistics the top and bottom numbers give the maximum and minimum values on the identified set.

An interesting result coming from Table 2.1 is that returns on growth portfolio exhibit more predictability than returns on the value portfolio and this is consistent with the results of Guo, Su, and Yang (2006). Although the volatility of expected returns on growth stocks is only slightly higher than the volatility of value stocks, the constructed predictors can explain 5% of time variation in expected returns on growth stocks, whereas the the corresponding R^2 statistics for value stocks is only 1%. Interestingly, dividend growth is also more predictable for the growth portfolio with the striking R^2 statistics of 27%. This observation admits the following interpretation. In the case with more predictable dividends, the filtering approach is more capable to disentangle expected dividends and expected returns, thus providing a better forecast for future returns.

To further evaluate the predictability of returns on value and growth portfolios, I run OLS regressions of realized returns on various forecasters. The results are displayed in Table 2.2. First, the filtered expected return is a significant forecaster for returns on growth stocks delivering the R^2 statistics of 8%, which is the largest among other considered predictors. Note that the regression-based R^2 statistics is larger than the R^2 statistics computed for the filtered expected returns. This is very natural, since the OLS regression adjusts the slope in front of the predictor producing better in-sample fit relative to the unscaled forecaster.

Second, the dividend-price ratio of the growth portfolio also possesses some forecasting power: the corresponding slope coefficient is significant and the R^2 statistics is around 5%. Third, returns on the growth portfolio can be predicted by some of the variables that predict market returns. Table 2.2 shows that filtered expected aggregate stock returns, the aggregate dividend-price ratio dpr_t, and the aggregate book-to-market ratio BM_t are statistically significant forecasters with the R^2 statistics around 5%.

The results are different for the value portfolio. Again, comparing the R^2 statistics we can conclude that returns on the growth portfolio are more predictable than returns on the value portfolio. Indeed, the filtered expected return has no forecasting power and the adjusted R^2 statistics for it is even negative. Although the aggregate expected returns, default premium, and the book-to-market ratio seem to have some forecasting power, the R^2-statistics are lower than provided by predictors for growth stocks.

The right hand side variables in regressions reported in Table 2.2 are quite persistent. Thus, there might be concern that the slope coefficients are upward biased in a finite sample and the distributions of t-statistics are non-standard making the reported results

	Growth							Value						
$\widehat{\mu}$	3.22							0.65						
	(2.40)							(1.57)						
$\widehat{\mu}^{aggr}$		1.05							0.87					
		(3.48)							(3.56)					
dpr			0.08							0.08				
			(2.71)							(1.75)				
dpr^{aggr}				0.13							0.09			
				(2.60)							(1.87)			
DEF					6.60							9.93		
					(1.57)							(2.50)		
BM						0.14							0.14	
						(2.24)							(2.27)	
S							−0.74							−0.41
							(−1.70)							(−0.97)
Adj. R^2	0.08	0.06	0.05	0.05	0.00	0.01	0.07	−0.01	0.04	0.01	0.02	0.04	0.02	0.01
N	55	55	55	55	55	55	55	55	55	55	55	55	55	55

Table 2.2: OLS regression of realized returns on value and growth portfolios on various predicting variables.

This table reports estimates from OLS regressions of realized returns on value and growth portfolios on filtered expected returns on these portfolios and several other variables. The sample is from 1950 to 2005. $\hat{\mu}$ is filtered expected return on the corresponding portfolio; $\hat{\mu}^{aggr}$ is filtered expected return on the market portfolio; dpr is the log dividend-price ratio of the corresponding portfolio; dpr^{aggr} is the log dividend-price ratio of the aggregate stock market; DEF is the default premium, defined as the yield spread between Moody's Baa and Aaa corporate bonds; BM is the aggregate book-to-market ratio; S is the equity share in total debt and equity issues. The t-statistics based on the Newey-West standard errors are reported in parentheses.

unreliable (Mankiw and Shapiro, 1986; Stambaugh, 1999). To alleviate this concern, I use the testing procedure developed by Campbell and Yogo (2006), which assesses the severity of the problem. I find, that in almost all regressions the correlation between innovations to returns and the predictor variable is sufficiently small and the inference based on the t-test with conventional quantiles is reliable.

To check the robustness of obtained results, I examine an alternative definition of the value and growth portfolios. Instead of a two-by-three sort on size and book-to-market, I take five-by-five sort and pick the portfolios with the highest and lowest values of the book-to-market ratio. Overall, the reported conclusions are not sensitive to this modification and even quantitatively the results are very similar to the benchmark case. To save the space, I do not report them.

2.3.3 Expected value premium

In this section I study the time series properties of the value premium with the special focus on its relation to business cycles. As described in Section 2.2, I estimate the benchmark model (2.1), (2.2) for *differences* in expected log returns on value and growth portfolios (value premium) and *differences* in expected log dividend growth. I examine two versions of the filtered value premium, which I denote as $\widehat{VP}^1$ and $\widehat{VP}^2$. The first one is constructed from the observables $(r_t^v - r_t^g, \Delta d_t^v - \Delta d_t^g)$, where indexes v and g stand for value and growth. The second variant of the filtered value premium $\widehat{VP}^2$ is built using $(r_t^v - r_t^g, dpr_t^v - dpr_t^g)$. As pointed out in Section 2.2, if the difference of the log dividend-price ratios $dpr_t^v - dpr_t^g$ is stationary, the estimate $\widehat{VP}^2$ is more efficient.

Table 2.3 reports several estimated statistics pertaining to the value premium. The obtained estimates indicate that the filtered value premium is very persistent with the autocorrelation coefficient about 0.93, which is comparable with the autocorrelations of expected returns on value and growth portfolios. However, the value premium is less volatile with the standard deviation around 4%. This is quite intuitive, since expected returns on value and growth portfolios to some degree are driven by expected returns on the market portfolio. In the difference this dominant component cancels out making expected value premium less volatile. Table 2.3 also displays the R^2 statistic for the filtered value premium. Although it is much more difficult to predict the value premium relative to returns on value and growth portfolios, both $\widehat{VP}^1$ and $\widehat{VP}^2$ have some ability to predict future value

$\sigma(VP_t)$	$\rho(VP_t, VP_{t-1})$	R^2_r
Panel A: Value Premium $\widehat{VP}^1$		
0.039 0.039	0.92	0.019
Panel B: Value Premium $\widehat{VP}^2$		
0.042 0.044	0.93	0.067

Table 2.3: Statistics of the filtered value premium.

This table displays statistics for the unobservable value premium. $\sigma(VP_t)$ is the standard deviation of the value premium, $\rho(VP_t, VP_{t-1})$ is its autocorrelation. R^2_r is the in-sample R^2 statistics for the filtered value premium.

	I	II	III	IV	V
VS	0.14			0.14	0.12
	(2.82)			(4.23)	(4.59)
$\widehat{VP}^1$		1.41		1.56	
		(2.83)		(2.39)	
$\widehat{VP}^2$			1.05		0.97
			(3.72)		(3.17)
Adj. R^2	0.04	0.02	0.05	0.06	0.08
N	55	55	55	55	55

Table 2.4: OLS regression of the realized value premium on the filtered value premium and the value spread.

This table reports coefficient estimates from OLS regression of the realized value premium on the filtered value premiums $\widehat{VP}^1$, $\widehat{VP}^2$, and the value spread VS. The sample is from 1950 to 2005. The value premium is defined as expected returns on HML portfolio. The value spread VS is defined as the difference in the logs of the book-to-market ratios for value and growth portfolios. The t-statistics based on the Newey-West standard errors are reported in parentheses.

premium. The first predictor has the R^2 statistics 2% whereas the second predictor manages to explain almost 7% of future time variation of the value premium. It is not surprising, since $\widehat{VP}^2$ uses more restrictive assumption on the dividend-price ratios of the value and growth portfolios and, as a result, is more efficient.

To get more evidence about the forecasting power of the constructed variable, I run predictive regression of realized value premium on the obtained predictors. The results are presented in Table 2.4. As a benchmark, I also consider the value spread VS, which also has some forecasting power for the value premium (Cohen, Polk, and Vuolteenaho, 2003).[9] The juxtaposition of the value spread and the filtered value premium allows me to show

[9]See more details on the value spread in Section 2.4.

how information contained in one predictor complements the information from the other.

The first regression in Table 2.4 shows that in my sample VS indeed appears to be a statistically significant predictor for the value premium with the R^2 statistics of 4%. In general, this confirms the results of Cohen, Polk, and Vuolteenaho (2003), but the R^2 statistics is a bit lower than they find for the sample period 1938-1997.

Predictive OLS regressions of the realized value premium on $\widehat{VP}^1$ and $\widehat{VP}^2$ confirm that these variables can forecast the value premium. The slope coefficients in both of them are positive and statistically significant. The adjusted R^2s are 2% and 5%, correspondingly, being very similar to those reported in Table 2.3. Notably, the coefficient in front of $\widehat{VP}^2$ is very close to 1 indicating that there is no need to rescale this variable to get better in-sample fit.

It is remarkable that the information contained in the value spread and the filtered value premium is different and these variables capture different components of the value premium. The correlation between VS and $\widehat{VP}^1$ is only -0.07, whereas the correlation between VS and $\widehat{VP}^2$ is 0.08 implying that these forecasters are almost orthogonal to each other. As a result, the combination of these predictors can provide a better forecast than each predictor alone. This is confirmed by Table 2.4. In the joint OLS regression both slope coefficients are positive and significant and the adjusted R^2 statistics are higher than in the individual regressions. The novelty of information brought into the system by the value spread will be further exploited in Section 2.4.

The most interesting question is the cyclical behavior of value premium. Rational theories providing risk-based explanation for value premium predict counter-cyclical variation of expected returns on the HML portfolio (e.g. Zhang, 2005; Kiku, 2006; Santos and Veronesi, 2006). To examine the cyclical behavior of the value premium, I run contemporaneous regressions of filtered expected HML returns on several business cycle variables. Since it is widely recognized that expected returns are countercyclical, I include as regressors proxies for aggregate expected returns. [10] One of such variables is filtered expected return on the market portfolio $\hat{\mu}^{aggr}$, constructed in Chapter 1. Also, I examine several other variables that have been argued to predict aggregate stock returns. In particular, I take the default premium DEF_t defined as the yield spread between Moody's Baa and Aaa corpo-

[10] Petkova and Zhang (2005) emphasize importance of using expected returns as opposed to realized returns for characterization of economic conditions.

	$\widehat{VP}^1$				$\widehat{VP}^2$			
	I	II	II	IV	I	II	III	IV
$\hat{\mu}^{aggr}$	0.22				0.80			
	(1.05)				(1.87)			
DEF		1.75				3.26		
		(5.79)				(2.64)		
BM			0.02				−0.02	
			(3.21)				(−0.87)	
$NBER$				0.01				0.005
				(2.99)				(0.87)
Adj. R^2	0.04	0.24	0.06	0.03	0.17	0.23	0.01	−0.01
N	56	56	56	56	56	56	56	56

Table 2.5: Regression of the filtered value premium on business cycle variables.

This table reports coefficient estimates from OLS regression of filtered value premium $\widehat{VP}^1$ and $\widehat{VP}^2$ on expected market returns and several business cycle variables. The sample covers the period from 1950 to 2005. The value premium is defined as expected returns on the HML portfolio. $\hat{\mu}^{aggr}$ is filtered expected aggregate stock returns; DEF is the default premium, defined as the yield spread between Moody's Baa and Aaa corporate bonds; $NBER$ is the NBER recession is dummy. BM is the aggregate book-to-market ratio. The t-statistics based on the Newey-West standard errors are reported in parentheses.

rate bonds (Fama and French, 1989) and the aggregate book-to-market ratio BM_t (Kothari and Shanken, 1997; Pontiff and Schall, 1997). All these variables exhibit a counter-cyclical variation. Finally, I add the NBER recession dummy which equals to 1 if December the particular year belongs to the NBER recession period.

The regression results are reported in Table 2.5. Although not all slope coefficients are significant, most of them make a good sense. Indeed, excepting the regression of $\widehat{VP}^2$ on BM, all coefficients are positive indicating positive correlation between $\widehat{VP}^1$ and $\widehat{VP}^2$ on one hand and the selected countercyclical variables on the other. Moreover, many of these coefficients are statistically significant. The strongest relation is observed between the value premium and the default premium. In both regressions with DEF on the right hand side the slopes the t-statistics are far above 2 and the R^2s demonstrate that the default premium and the value premium have almost 25% of common variation.

To visualize the commonality in the value premium and the selected countercyclical variables I plot them on the same graph. The result is displayed on Figure 2-2. The solid line and the dashed line represent the filtered value premiums $\widehat{VP}^1$ and $\widehat{VP}^2$, correspondingly. It is worth to note, that although $\widehat{VP}^1$ and $\widehat{VP}^2$ are built from different data, they appear to be highly correlated and track each other.[11] Also, from the graph we observe that both

[11]The correlation between them is 0.76.

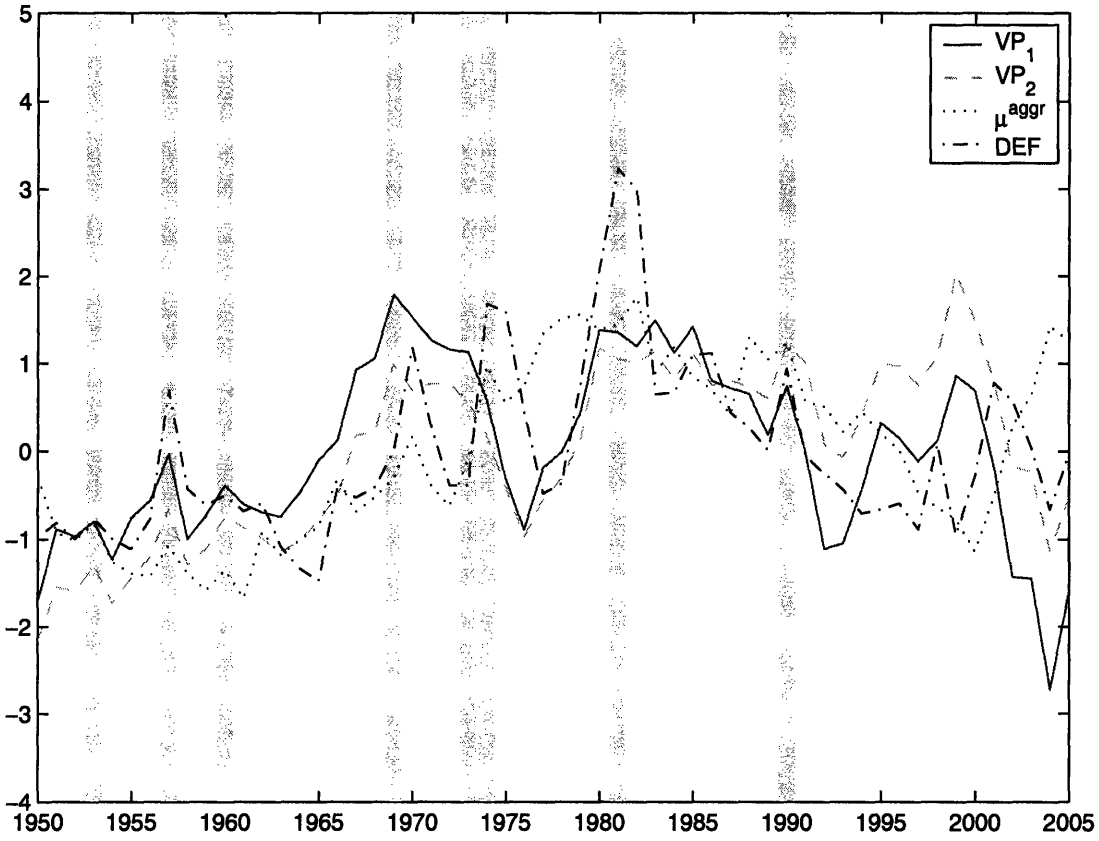


Figure 2-2: The filtered value premium and business cycles.

This figure plots the filtered value premium $\widehat{VP}^1$ and $\widehat{VP}^2$ along with several business cycle variables. $\hat{\mu}^{aggr}$ is filtered expected aggregate stock returns; DEF is the default premium. All variables are standardized to unit variance. The shaded areas indicate NBER recession dates.

estimates of the expected value premium jump up during the NBER recession periods, indicated by shaded bars. This is even true for the short recession in 2000, even though this year does not formally qualify for my definition of a recession year (the recession was over by December). Exactly the same behavior is demonstrated by the default premium, which is represented by the dash-dot line. Moreover, even beyond the recession period we observe some comovement between DEF and the filtered value premiums. The countercyclical pattern and comovement with $\widehat{VP}^1$ and $\widehat{VP}^2$ are less obvious for filtered expected aggregate stock returns $\hat{\mu}^{aggr}$. Although, as established in Chapter 1, $\hat{\mu}^{aggr}$ evolves countercyclically, it is a noisy proxy for business cycles. Hence, its comovement with the expected value premium might not be pronounced.

2.4 Value Spread

Forecasting the value premium is a daunting task, which is much more difficult than forecasting aggregate stock returns. Although there is a dozen of variables which have been argued to track aggregate expected returns, the existing literature suggests only few exogenous variables predicting the value premium. One of the most successful of them is the value spread (Asness, Friedman, Krail, and Liew, 2000; Cohen, Polk, and Vuolteenaho, 2003) defined as the difference between book-to-market ratios of the low- and high-B/M portfolios.

As demonstrated in Chapter 1, the filtering approach is quite flexible and allows to add other available proxies for expected returns as additional observables. In this section, I estimate the augmented system (2.4) with the value spread as an additional predictor. I show that the constructed predictor optimally incorporates all available information and outperformes the value spread and the filtered value premium taken individually.

The value spread used in most of this section is constructed in the following way. All CRSP stocks are sorted into 3 groups on the basis of their book-to-market ratio (with 30% of stocks in the portfolio with low book-to-market ratio, 40% in the middle portfolio, and 30% in the portfolio with high ratio). For the high- and low-B/M portfolios equal-weighted averages of the book-to-market ratios are computed and the value spread VS_t is defined as the difference between logs of these averages[12].

[12]The portfolios that I use are constructed according to Fama and French (1993). For each year t, they are formed at the end of June of year t. B/M for year t is the book equity for the fiscal year $t-1$ divided

	I	II	III	IV	V
VS	0.14			−0.12	−0.01
	(2.82)			(−1.93)	(−0.48)
$\widehat{VP}^{q1}$		1.04		1.72	
		(4.08)		(4.80)	
$\widehat{VP}^{q2}$			1.69		1.74
			(7.09)		(6.22)
Adj. R^2	0.04	0.08	0.17	0.08	0.15
N	55	55	55	55	55

Table 2.6: OLS regression of the realized value premium on the filtered value premium from the augmented system (2.4) and the value spread.

This table reports coefficient estimates from OLS regression of the realized value premium on the filtered value premiums $\widehat{VP}^{q1}$, $\widehat{VP}^{q2}$, and the value spread VS. The sample is from 1950 to 2005. The value premium is defined as expected returns on HML portfolio. The value spread VS is defined as the difference in the logs of the book-to-market ratios for value and growth portfolios. The t-statistics based on the Newey-West standard errors are reported in parentheses.

As in the previous section, I consider two types of the filtered value premium, which I denote as $\widehat{VP}^{q1}$ and $\widehat{VP}^{q2}$. The first one is constructed from the history of realized differences in log returns $r_t^v - r_t^g$, differences in log dividend growth $\Delta d_t^v - \Delta d_t^g$, and the value spread VS_t. The second one uses the assumption that the difference in the log dividend-price ratios $dpr_t^v - dpr_t^g$ is stationary and it also contains some information about future value premium. Hence, $\widehat{VP}^{q2}$ exploits $r_t^v - r_t^g$, $dpr_t^v - dpr_t^g$, and VS_t.

To evaluate the forecasting ability of obtained predictors relative to the value spread, I run OLS regressions of the realized value premium on VS and the forecasters $\widehat{VP}^{q1}$, $\widehat{VP}^{q2}$. The results are displayed in Table 2.6.

Regressions II and III confirm that $\widehat{VP}^{q1}$ and $\widehat{VP}^{q2}$ are good forecasters for future value premium. The slope coefficients are positive and highly statistically significant. The R^2 statistics for $\widehat{VP}^{q1}$ is 8%, which is much higher than the R^2 statistics for VS (4%) and $\widehat{VP}^{1}$ (2%). It is even higher than in the multivariate OLS regression on VS and $\widehat{VP}^{1}$ (6%). Exactly the same pattern is observed for $\widehat{VP}^{q2}$, which alone manages to explain 17% of variation in the value premium.

Regressions IV and V demonstrate, that $\widehat{VP}^{q1}$ and $\widehat{VP}^{q2}$ absorb the information contained in the value spread. When the value premium is regressed on both the filtered value premium and the value spread, the slope coefficient in front of VS is insignificant. Correspondingly, the adjusted R^2 statistics does not increase and even go down in regression

by market equity for December of year $t-1$. This methodology slightly differs from the methodology of Cohen, Polk, and Vuolteenaho (2003) who use the market equity value recorded in May of year t.

V.

To ensure that my results are not specific to the particular definition of the value spread, I also examine alternative ways to construct it. In particular, I also consider the value spread based on sorting stocks into 5 equal groups and taking the difference between the log book-to-market ratios of the low- and high-BM portfolios. The obtained results are qualitatively very similar and I do not report them.

2.5 Concluding Remarks

In this chapter I apply the filtering methodology developed in Chapter 1 to analysis of time variation of value premium. Making use of the history of the value premium and the differences in the log dividend-price ratios of the value and growth portfolios, I construct a new predictor which can forecast 5% of time variation in the future value premium. This predictor evolves countercyclically, providing evidence in favor of rational explanations of the value premium. Augmented with the value spread, the state space system produces the forecasting variable that explains over 17% of the value premium variance.

In my analysis, I mostly focus on time series properties of the value premium. However, the filtering approach allows to reconstruct not only unobservable expectations, but also to disentangle the contributions of "news about future returns" and "news about future cash flows". Hence, it might serve as an alternative to the VAR methodology, developed by Campbell and Shiller (1988) and Campbell (1991). Given innovations to expected returns and expected dividend growth on the book-to-market sorted portfolios, it would be interesting to compare their cash flow and discount rate betas and check whether the value premium can be explained by higher cash flow betas of value stocks (Campbell and Vuolteenaho, 2004).

References

Asness, C., J. Friedman, R. Krail, and J. Liew (2000) Style Timing: Value versus Growth, *Journal of Portfolio Management* 26, 50-60

Asness, C., R. Krail, and J. Liew (2003) Country-Level Equity Style Timing, in T. Coggin and F. Fabozzi, eds., *The Handbook of Equity Style Management*, John Wiley & Sons, Inc., Hoboken

Avramov, D. and R. Wermers (2006) Investing in Mutual Funds when Returns are Predictable, *Journal of Financial Economics* 81, 339-377

Baker, M. and J. Wurgler (2000) The Equity Share in New Issues and Aggregate Stock Returns, *Journal of Finance* 55, 2219-2257

Campbell J. (1991) A Variance Decomposition for Stock Returns, *Economic Journal* 101, 157-179

Campbell J. and R. Shiller (1988) The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195-227

Campbell J. and T. Vuolteenaho (2004) Bad Beta, Good Beta, *American Economic Review* 94, 1249 - 1275

Campbell J. and M. Yogo (2006) Efficient Tests of Stock Return Predictability, *Journal of Financial Economics* 81, 27-60

Chen, L., R. Petkova, and L. Zhang (2006) The Expected Value Premium, Unpublished Manuscript

Cohen, R. B., C. Polk, and T. Vuolteenaho (2003) The Value Spread, *Journal of Finance* 58, 609-641

Fama E. and K. French (1988) Dividend Yields and Expected Stock Returns, *Journal of Financial Economics* 22, 3-25

Fama, E. and K. French (1989) Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics* 25, 23-49

Fama, E. and K. French (1993) Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56

Fan, S. (1995) Equity Style Timing and Allocation, in R. Klein and J. Lederman, eds., *Equity Style Management*, Irwin, Chicago

Fujimoto, A. and M. Watanabe (2005) Value Risk in International Equity Markets, Unpublished Manuscript

Gomes, J., L. Kogan, and L. Zhang (2003) Equilibrium cross section of returns, *Journal of Political Economy* 111, 693-732

Guo, H., X. Su, and J. Yang (2006) Are Growth and Value More Predictable than the Market? Unpublished Manuscript

Hamilton, J. (1994) Time Series Analysis, Princeton University Press

Kao, D. and R. Shumaker (1999) Equity Style Timing, *Financial Analysts Journal* 55, 37-48

Kiku, D. (2006) Is the Value Premium a Puzzle? Unpublished Manuscript

Kothari, S.P. and J. Shanken (1992) Stock Return Variation and Expected Dividends, *Journal of Financial Economics* 31, 177-210

Lakonishok, J., A. Shleifer, and R. Vishny (1994) Contrarian Investment, Extrapolation, and Risk, *Journal of Finance* 49, 1541-1578

La Porta, R., J. Lakonishok, A. Shleifer, and R. Vishny (1997) Good News for Value Stocks: Further Evidence on Market Efficiency, *Journal of Finance* 52, 859-874

Lettau M. and S. Ludvigson (2001) Consumption, Aggregate Wealth and Expected Stock Returns, *Journal of Finance* 56, 815 - 849

Liu, N. and L. Zhang (2005) The Value Spread Is Not a Useful Predictor of Returns, Unpublished Manuscript

Mankiw, N. G. and M. D. Shapiro (1986) Do We Reject Too Often? *Economics Letters* 20, 139-145

Petkova, R. and L. Zhang (2005) Is value riskier than growth? *Journal of Financial Economics* 78, 187-202

Pontiff J. and L. D. Schall (1998) Book-to-market Ratios as Predictors of market returns, *Journal of Financial Economics* 49, 141 - 160

Rozeff, M. and M. A. Zaman (1998) Overreaction and Insider Trading: Evidence from Growth and Value Portfolios, *Journal of Finance* 53, 701-716

Santos, T. and P. Veronesi (2006) Habit Formation, the Cross Section of Stock Returns and the Cash-Flow Risk Puzzle, Unpublished Manuscript

Sorensen, E. and C. Lazzara (1995) Equity Style Management: The Case of Growth and Value, in R. Klein and J. Lederman, eds., *Equity Style Management*, Irwin, Chicago

Stambaugh, R. F. (1999) Predictive Regressions, *Journal of Financial Economics* 54, 375-421

Zhang, L. (2005) The value premium, *Journal of Finance* 60, 67 - 103

Chapter 3

Forecasting the Forecasts of Others: Implications for Asset Pricing

3.1 Introduction

This chapter studies the properties of linear rational expectation equilibria with imperfectly informed agents. Since the 1970s the concept of rational expectation equilibrium (REE) has become central to both macroeconomics and finance, where agents' expectations are of paramount importance. In financial markets information is distributed unevenly across different agents. As a result, prices reflect expectations of various market participants and are, therefore, themselves essential sources of information. While making investment decisions, an agent, then, needs to worry not only about her own expectation, but also about expectations of other agents, or, in Townsend's (1983a) words, agents forecast the forecasts of others. Iterating this logic forward, prices must depend on the whole hierarchy of investors' beliefs. Many economists have seen this as an important feature of financial markets and a possible source of business cycle fluctuations. However, the formal analysis of dynamic models with heterogeneous information has proven to be very difficult. The reason is that in most cases the successive forecasts of the forecasts of others differ from one another, which leads to an infinite number of expectations to be accounted for. So unless some recursive representation is available, one needs an infinite number of state variables

to describe all of them. As a result, the model becomes not only analytically, but also numerically, involved.

A lot of effort has been made to identify some special cases where this hierarchy can be summarized by a few cleverly chosen state variables, and nearly all existing models study exactly these cases. Previous research found three assumptions that help to preserve tractability. The first requires agents to be hierarchically informed, which means that agents can be ranked according to the quality of their information. A special example is when there are two classes of agents, informed and uninformed. In this case, the informed know the forecasting error of the uninformed and therefore do not need to forecast it, so higher order expectations collapse. The second assumption calls for fundamentals to be constant over time[1]. Finally, the third one demands the number of endogenous variables (prices) that agents can condition their forecasts upon to be greater or equal than the number of variables that agents seek to forecast[2]. Clearly, the above assumptions are quite restrictive and the insights obtained under them may not survive in a more general informational environment.

There is a continuing quest for other cases that admit tractable analysis. This is a challenging problem, since even when a model does have a finite dimensional state space, often it is very difficult to identify the state variables in which equilibrium dynamics takes a tractable form. One of the approaches is suggested by Marcet and Sargent (1989) and Sargent (1991). The key idea of these papers is to model agents' beliefs as an ARMA process and compute the rational expectations equilibrium as a fixed point of a map from the perceived law of motion to the real law of motion. As an example, the authors calculate the equilibrium in Townsend's (1983a) model.

While the complete characterization of tractable models is undoubtedly a daunting task, in this paper we provide a boundary example. In the example, we show that if (i) each agent lacks some information that is known to other agents, (ii) fundamentals evolve stochastically over time, and (iii) dimensionality of unknown shocks exceeds that of conditioning variables, then the infinite regress problem cannot be avoided and an infinite number of state variables is required to describe the price dynamics. The first condition guarantees that information held by other agents is relevant to each agent's payoff and, as a result, her beliefs about

[1]See, for example, He and Wang (1995), Allen, Morris, and Shin (2004).

[2]See analysis of Townsend (1983a) model in Sargent (1991), Kasa (2000), Pearlman and Sargent (2004).

other agents' beliefs affect her demand for the risky asset. We call this information setup differential and contrast it with the hierarchical setup, in which one agent is better informed than the other. The second condition forces agents to form new sets of higher order beliefs every period. Since no agent ever becomes fully informed, they all need to incorporate the entire history of prices into their predictions[3]. The third condition makes it impossible for agents to reconstruct the unknown shocks (or their observational equivalents) from observable signals[4].

The proof goes as follows. First, we show that if the price dynamics can be described with the finite number of variables then it must be an ARMA (auto-regressive, moving average) process of a finite order. This provides a justification for Sargent's (1991) methodology and implies that the agents' demand should be a finite order ARMA process as well. Using a one-to-one correspondence between rational functions in frequency domain and ARMA processes, we show that a solution to a closed system of equations obtained from the market clearing conditions cannot be represented by a rational function. So by contradiction, price dynamics must invoke an infinite number of state variables.

Having established this result, which leaves little hope for an analytical solution, we proceed with the numerical analysis. By comparing the equilibria supported by the same fundamentals but with different distributions of information among investors, we are able to isolate the effect of information dispersion on expectations. We find that mistakes that agents make in forming their expectations are much larger under differential information than under hierarchical information. Therefore, differential information gives rise to a larger deviation of prices from the benchmark case without information asymmetry. To better understand the dynamics of price, in addition to static contributions of expectations, we also study the joint dynamics of expectations and fundamentals. We show that agents' forecasting errors are much more persistent when information is differentially distributed among them, which is a direct consequence of the absence of superiorly informed agents who arbitrage these errors away.

Analysis of a more general informational setup allows us to evaluate the robustness of previous findings and get new insights. We find that under the differential information

[3]This observation suggests that price histories may play an important role in financial markets in which asymmetric information is ubiquitous, thus lending support to technical analysis, which is often employed in practice.

[4]See Pearlman and Sargent (2004) for more details.

setup the absorption of information into prices can be very slow. As a result, returns can be positively autocorrelated, which may be a step towards an explanation of momentum. The driving force behind this effect in our model is underreaction of agents to new information. It should be emphasized that we consider an equilibrium model in which the diffusion of information into prices is an endogenous process: it is an equilibrium outcome of agents' portfolio decisions and the resulting effect on the price. This distinguishes our explanation of momentum from a number of behavioral theories which appeal to different cognitive biases[5]. Our model predicts that momentum is stronger in stocks with little analyst coverage and higher analysts dispersion forecasts, and this is consistent with empirical evidence[6].

Furthermore, our model allows us to investigate the effects of information dispersion on trading volume, whose empirically observed high levels present a puzzle to financial economists. We differentiate between two types of trades: informational trades between informed agents and exogenous trades with liquidity traders. We show that under the hierarchical setup there is almost no trade between informed and uninformed agents. This is intuitive, since the uninformed are aware of their disadvantage and, therefore, averse to trade. As a result, in this framework, trading volume is almost exclusively determined by the properties of the exogenously assumed process for noise trader demands. In contrast, in a framework with differential information, trade between informed agents is high, since no one has a clear advantage. More importantly, we demonstrate that the contribution of informational trade to total volume can be significantly higher than that of exogenous volume.

The rest of the chapter is organized as follows. In Section 3.2 we discuss the related literature. Section 3.3 describes the model. In Section 3.4 we solve for the equilibria for benchmark cases of full and hierarchical information dispersion setups. In Section 3.5 we consider differential information. Section 3.6 presents details of the numerical algorithm used to solve the model. In Section 3.7 we analyze higher order expectations. In Section 3.8 we examine the impact of information dispersion on prices and returns. In particular, we demonstrate that differential information accompanied by evolving fundamentals can generate momentum in returns. Section 3.9 is devoted to analysis of trading volume generated in our model. Section 3.10 concludes. Technical details are presented in Appendices A, B,

[5]See Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999) among others.

[6]See Hong, Lim, and Stein (2000), Lee and Swaminathan (2000), Verardo (2005) among others.

C and D.

3.2 Related Literature

There is a vast literature related to dynamic asset pricing with asymmetric information. In most papers, however, higher order expectations play a very limited role and can be summarized by a finite number of variables. Grundy and McNichols (1989) and Brown and Jennings (1989) study two-period models, which are very restrictive for analysis of dynamic effects of differential information. Singleton (1987), and Grundy and Kim (2002) consider multiperiod models in which all private information is revealed after one or two periods. Although such models deliver predictions about dynamic properties of prices and returns, the investors' learning problem in these models is effectively static. This diminishes the impact of asymmetric information on expectations and prices, which could be more pronounced would the learning problem be also dynamic. As demonstrated below, enabling private information to be long-lived allows for non-trivial interplay between expectations and fundamentals, which sometimes reverts the conclusions provided by simplified models. For example, we show that in contrast to Grundy and Kim (2002), the volatility of returns under differential information may be lower than in an otherwise identical economy with no information asymmetry.

The dynamic nature of the learning problem is a necessary, but not a sufficient condition for getting a non-trivial contribution of all higher order expectations. To avoid the aforementioned infinite regress problem, Wang (1993, 1994) considers a multi-period economy in which agents are hierarchically informed. Although this immensely simplifies the analysis, the obtained conclusions might significantly rely on the information distribution assumption. An alternative approach to avoid the whole hierarchy of iterated expectations was developed in He and Wang (1995). In their model, agents have differential information about the unknown final payoff, which, however, does not change over time. Our setup naturally extends these models by allowing both a general information structure and stochastic evolution of fundamentals.

The above papers assume that asymmetrically informed agents behave competitively and do not have any market power. However, there is a vast literature exploring strategic behavior of asymmetrically informed agents and the resulting effect on prices of securities.

In multiperiod models with strategic traders there is also a room for non-trivial dynamics of higher order expectations but, similar to the case with price-taking agents, most existing models try to avoid the "forecasting the forecasts of others" problem. For example, Foster and Viswanathan (1996) and Back, Cao, and Willard (2000) use the Kyle (1985) framework, in which the liquidation value of the risky asset does not change over time. The only paper allowing the "forecasting the forecasts of others" problem is Bernhardt, Seiler, and Taub (2004). In this paper authors study price dynamics of an asset with stochastic fundamentals which is traded by heterogeneously informed strategic agents.

The theme of our research is also aligned with another strand of the literature, which explicitly analyzes higher order expectations. Allen, Morris, and Shin (2006) argue that under asymmetric information agents tend to underreact to private information, making the price biased towards the public signal. Bacchetta and Wincoop (2004) show that under asymmetric information price deviations from fundamentals can be large. Having a fully-fledged dynamic model enables us to provide a more thorough analysis of agents' expectations and their dynamics as well as to give specific predictions about their relationship to the behavior of prices, returns, and trading volume.

3.3 The Model

In this section, we present our model. Throughout the rest of the chapter, it is assumed that investors are fully rational and know the structure of the economy.

3.3.1 Financial Assets

There are two assets. The first asset is a riskless asset in perfectly elastic supply that generates a rate of return $1 + r$. The second asset is a claim on a hypothetical firm which pays no dividends[7] but has a chance of being liquidated every period. We assume that the probability of liquidation in period $t + 1$, given that the firm has survived until period t, is equal to λ. Upon liquidation the firm pays its equity holders a stochastic liquidation value V_t. This liquidation value can be decomposed into two components: $V_t = V_t^1 + V_t^2$, and

[7]We model the firm as not paying dividends for the sake of simplicity, since the current dividend would be an additional signal about future cash flows.

each component evolves according to a first-order autoregressive process:

$$V_{t+1}^j = aV_t^j + b_V \epsilon_{t+1}^j, \; j = 1, 2.$$

We assume that $\varepsilon_t^j \sim \mathcal{N}(0,1)$ are *i.i.d.* across time and components. For simplicity we take identical parameters a and b_V for the processes V_t^1 and V_t^2. The total amount of risky equity[8] available to rational agents is $1 + \theta_t$, where $\theta_t \equiv b_\Theta \varepsilon_t^\Theta$ and $\varepsilon_t^\Theta \sim \mathcal{N}(0,1)$.

3.3.2 Preferences

There is an infinite set of competitive rational investors indexed by i and uniformly distributed on a unit interval $[0,1]$. Each of them is endowed with some piece of information about the fundamentals V_t^1 and V_t^2. We assume that investors are mean-variance optimizers and each investor i submits the demand X^i which is proportional to his expectation of excess stock return Q_{t+1}:

$$X_t^i = \frac{1}{\alpha} \frac{E[Q_{t+1}|\mathcal{F}_t^i]}{\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^i]}, \qquad Q_{t+1} = \lambda V_{t+1} + (1-\lambda)P_{t+1} - (1+r)P_t. \tag{3.1}$$

Here $\mathcal{F}_t^i$ is the information set of investor i at time t. All investors are assumed to have the same coefficient of risk aversion α.

3.3.3 Properties of the model

Before we turn to analysis of equilibrium, it is worthwhile to make several comments about the model. First of all, we make the model very stylized, since we want to demonstrate and analyze the "forecasting the forecasts of others" problem in the simplest setting. In particular, we assume that all shocks are normally distributed and this property is inherited by other random variables in the model, leading to the linear form of conditional expectations and, therefore, to a linear equilibrium. Next, we consider a model with an infinite horizon and focus on stationary equilibria which enables us to use powerful methods from the theory of stationary Gaussian processes. Finally, a major simplification is achieved by assigning agents' mean-variance preferences. This assumption is similar to the assumption of logarithmic utility with log-normally distributed shocks under which the hedging

[8]This can be interpreted as supply of stock by noise traders. Following Grossman and Stiglitz (1980), we introduce stochastic amount of equity to prevent prices from being fully revealing.

demand is zero. Since calculation of hedging demand in the economy with infinite number of state variables is complicated by itself[9], sidestepping this problem allows us to preserve tractability of the model but still relate equilibrium price to agents' higher order beliefs and characterize their dynamics.

3.3.4 The rational expectation equilibrium

We focus on a rational expectation equilibrium of this model which is defined by two conditions:

1) all agents rationally form their demands according to (3.1);

2) market clearing condition holds: $\int X_t^i di = 1 + \theta_t$.

In the most general case, information sets of investors $\mathcal{F}_t^i$ are different, investors have to forecast the forecasts of others, and non-trivial higher order expectations appear. As a basis for our subsequent analysis, it is useful to represent the price in terms of fundamentals and expectations of agents. It is convenient to first define the *weighted average expectation operator* $\bar{E}_t^w[x]$ of agents as follows:

$$\bar{E}_t^w[x] = \int \frac{\omega_i}{\Omega} E[x|\mathcal{F}_t^i]di, \quad \Omega = \int \omega_i di, \quad \omega_i = \frac{1}{\alpha}\frac{1}{\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^i]}$$

Note that the weights ω_i are endogenous and determined by the conditional variances of excess returns given investors' information sets. The expectations of agents with better information get larger weights than those of the less informed. Using the market clearing condition we can derive a relation between the current price and the next period price:

$$P_t = -\frac{1+\theta_t}{\Omega(1+r)} + \frac{1}{1+r}\bar{E}_t^w[\lambda V_{t+1} + (1-\lambda)P_{t+1}].$$

Iterating this relation forward and imposing the no-bubble condition, we get

$$P_t = -\frac{1}{\Omega(r+\lambda)} - \frac{1}{\Omega(1+r)}\theta_t + \frac{a\lambda}{1+r}\sum_{s=0}^{\infty}\left(\frac{1-\lambda}{1+r}\right)^s \bar{E}_t^w \bar{E}_{t+1}^w .. \bar{E}_{t+s}^w V_{t+s}. \qquad (3.2)$$

This equation represents the price as a series over iterated weighted average expectations of future values of V_t: we have arrived at a mathematical formulation of forecasting the forecasts of others. It highlights two essential difficulties. The first is that the law of iterated

[9]See Schroder and Skiadas (1999) for some results in this case.

expectations need not hold because agents may have different information; this point was recently emphasized by Allen, Morris and Shin (2004). The second and even more significant obstacle is that the current price also depends on agents' future expectations which, in turn, depend on future prices. Consequently, in order to compute their expectations, we have to solve for the entire sequence of prices as a fixed point. Since this problem is quite complicated, before attempting to find a solution for the general case, let us first consider some special cases in which the solution is not as involved.

3.4 Benchmarks

3.4.1 Full information

As a starting point, we consider the full information setup. Full information means that all investors $i \in [0,1]$ observe both components V_t^1 and V_t^2 and their information sets are

$$\mathcal{F}_t^i = \{P_\tau, V_\tau^1, V_\tau^2 : \tau \leq t\}.$$

In this case we are back to the representative agent framework, and the law of iterated expectations holds: $\bar{E}_t^w \bar{E}_{t+1}^w .. \bar{E}_{t+s}^w V_{t+s} = E_t V_{t+s} = a^s V_t$. Now observing the price is sufficient to infer the demand of noise traders θ_t. We have the following proposition:

Proposition 1 *Suppose that*

1) all investors observe V_t;

2) $2\sqrt{2} b_V b_\Theta \frac{\lambda(1-\lambda)}{1+r-a(1-\lambda)} \leq \frac{1}{\alpha}$.

Then there exists a full information equilibrium in which the equilibrium price of the risky asset is given by

$$P_t = -\frac{1}{\Omega(r+\lambda)} + \frac{a\lambda}{1+r-a(1-\lambda)} V_t - \frac{1}{\Omega(1+r)} \theta_t. \tag{3.3}$$

$$\Omega = \frac{(1+r-a(1-\lambda))^2}{4b_V^2 \lambda^2 (1+r)^2} \left(\frac{1}{\alpha} + \sqrt{\frac{1}{\alpha^2} - \frac{8 b_V^2 b_\Theta^2 \lambda^2 (1-\lambda)^2}{(1+r-a(1-\lambda))^2}} \right) \tag{3.4}$$

Proof. See Appendix A.

The obtained price function has a structure which is common to linear rational expectations models[10]. The first term corresponds to a risk premium for uncertain payoffs. The second term is the value of expected future payoffs discounted at the risk-free rate adjusted for the probability of liquidation. The third term compensates the investors for noise trading related risk.

Formally, the equations determining equilibrium price admit two solutions. One of them is given in Proposition 1, and we take this solution as the full information benchmark in the future. The reason for discriminating between equilibria is that the other solution is unstable, meaning that minor errors in agents' behavior significantly impact prices and destabilize the economy. Having this in mind, we consider only the full information equilibrium which is most sensible from the economic point of view.

3.4.2 Hierarchical information

Now consider the equilibrium with hierarchical information[11], which means that investors can be ranked according to the amount of their information: some investors are better informed than others. Formally, the information sets of investors at time t are hierarchically embedded in each other and generate a filtration: $\mathcal{F}_t^i \subseteq \mathcal{F}_t^{i'} \subseteq \ldots$. We focus on the simplest case, and assume that there are only two types of investors which we denote as 1 and 2. Investors of type 1, which are indexed by $i \in [0, \gamma]$, are informed and observe both V_t^1 and V_t^2. Investors of type 2, with $i \in (\gamma, 1]$, are partially informed and observe V_t^2 only. We can write their information sets of informed and uninformed investors as

$$\mathcal{F}_t^1 = \{P_\tau, V_\tau^1, V_\tau^2 : \tau \leq t\}, \qquad \mathcal{F}_t^2 = \{P_\tau, V_\tau^2 : \tau \leq t\}.$$

There are several reasons why this informational structure is interesting. First of all, it is an intermediate setup between the full information and the differential information equilibria. Despite the investors having heterogeneous information, the infinite regress problem does not arise and we can find a closed-form solution. The intuition behind this result is simple and can be easily conveyed in terms of expectations. When trying to extract the unknown piece of information from the price, investors of type 2 form their expec-

[10] See Campbell and Kyle (1993), Wang (1993)

[11] The idea to analyze hierarchical information setup in order to avoid the infinite regress problem was suggested by Townsend (1983) and elaborated in the asset pricing context by Wang (1993, 1994).

tations $\hat{V}_t^1 = E[V_t^1|\mathcal{F}_t^2]$ about the current value of V_t^1. Since all agents of type 2 make an identical estimation error, $\hat{V}_t^1$ is a new state variable influencing the price of the asset. In their turn, the investors of type 1 need to form their own expectations about expectations of type 2 investors, $E[\hat{V}_t^1|\mathcal{F}_t^1]$, and in the general case of differential information, it would be represented by another state variable. However, since $\mathcal{F}_t^2 \subseteq \mathcal{F}_t^1$ we get $E[\hat{V}_t^1|\mathcal{F}_t^1] = E[E[V_t^1|\mathcal{F}_t^2]|\mathcal{F}_t^1] = \hat{V}_t^1$ and the infinite regress problem does not arise. Basically, since the type 1 agents have all the information, they can, without mistake, deduce the mistake of type 2 agents, thus their prediction of the price is accurate. So the hierarchical information case illustrates how iterated expectations collapse and the state space of the model remains finite dimensional. The hierarchical information equilibrium in our model is characterized by Proposition 2.

Proposition 2 *If investors of type 1, with $i \in [0, \gamma]$, observe V_t^1 and V_t^2 and investors of type 2, with $i \in (\gamma, 1]$, observe only V_t^2 the equilibrium price of the risky asset is given by*

$$P_t = -\frac{1}{\Omega(r+\lambda)} + p_V V_t - \frac{1}{\Omega(1+r)}\theta_t + p_\Delta(\hat{V}_t^1 - V_t^1), \tag{3.5}$$

where p_V, p_Δ and Ω are constants which solve a system of nonlinear equations given in Appendix B.

Proof. See Appendix B.

3.5 Differential information equilibrium

Now consider the informational structure in which all agents are endowed only with a piece of relevant information and the rest of the information is never revealed. Again, assume that there are two types of agents, $j = 1, 2$ with $i \in [0, \gamma]$ and $i \in (\gamma, 1]$ respectively, such that their information sets are given by

$$\mathcal{F}_t^1 = \{P_\tau, V_\tau^1 : \tau \leq t\}, \qquad \mathcal{F}_t^2 = \{P_\tau, V_\tau^2 : \tau \leq t\}. \tag{3.6}$$

3.5.1 Forecasting the forecasts of others

In means that the agents of type j can observe only V^j and the history of prices. Let us show how the problem of "forecasting the forecasts of others" arises in this case. First

of all, due to the presence of noise traders, the price is not fully revealing, i.e. knowing the price and their own component of information V^j, the agents cannot infer the other component V^{-j}. However, the information about V^{-j} is relevant to agent j, since it helps him predict his own future payoff and, consequently, to form his demand for the asset. Moreover, due to the market clearing condition, the information of each investor is partially incorporated in the price, each agent has an incentive to extract the missing information of the other type from the price. Therefore, an agent will form his own expectations about the unknown piece of information. For example, agent 1 forms his expectations about agent 2's information. These expectations of agent 1 affect his demand and, subsequently, the price. So the inference problem of agent 2 is not only to extract the information of agent 1, but also the expectations of agent 1 about the information of agent 2. Agent 1, in turn, faces a similar problem; we can see how the infinite regress starts to appear.

The above reasoning might seem to be quite general, however, it does not always produce an infinite set of different higher order expectations. He and Wang (1995) provide an example how the higher order expectations can be reduced to first-order expectations even when investors have differential information. They consider a similar setup but assume that the firm is liquidated with probability one at some future time T and that the liquidation value does not evolve over time. In this situation, investors also try to predict the weighted average of investors' expectations $\widehat{V}$ of V. The paper demonstrates that $\widehat{V}$ can be written as a weighted average of V conditional on public information (price) and the true value of V. Given this, investor i's expectation of $\widehat{V}$ is a weighed average of his first-order expectations, conditional on price and on his private signals. Averaging them, one can show that second-order expectations of V can be again expressed as weighted average of V conditional on price and the true value of V. As will be shown later, this logic breaks down when V evolves stochastically over time.

It is necessary to distinguish between the cases with finite vs. infinite dimensional state space because they are conceptually different and call for different solution techniques. In the former case, the major problem is to find appropriate state space variables. In the latter, the search for a finite set of state variables that can capture the dynamics is worthless by default, and the solution of such models presents a greater challenge.

3.5.2 Markovian dynamics

To provide the ground for rigorous treatment of the "forecasting the forecasts of others", we introduce the concept of Markovian dynamics. Let $(\Omega, \mathcal{F}_t, \mu)$, $t \in \mathbb{Z}$ be a complete probability space equipped with a filtration $\mathcal{F}_t$. In what follows, all the processes are assumed to be defined on this space.

Definition. Let X_t be an adaptive random process. We say that X_t *admits Markovian dynamics* if there exists a collection of $n < \infty$ adaptive random processes $\bar{Y}_t = \{Y_t^i\}$, $i = 1..n$, such that the joint process $(X_t, \bar{Y}_t)$ is Markov, that is

$$Prob\left(X_t \leq x, \bar{Y}_t \leq y | X_\tau, \bar{Y}_\tau : \tau \leq t-1\right) = Prob\left(X_t \leq x, \bar{Y}_t \leq y | X_{t-1}, \bar{Y}_{t-1}\right).$$

Obviously any Markov process admits Markovian dynamics. The next example will further help to clarify the ideas.

Example. Let ε_t, $t \in \mathbb{Z}$ be *i.i.d.* standard normal random variables. Define $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, an MA(1) process. X_t is not a Markov process, or even an n-Markov process: $Prob\left(X_t | X_\tau : \tau \leq t-1\right) \neq Prob\left(X_t | X_{t-1}, \ldots, X_{t-n}\right)$ for any n. However, X_t can be easily extended to a Markov process if one augments it with ε_t.

An important consequence of X_t admitting Markovian dynamics is that the filtered process $\hat{X}_t$ then also admits Markovian dynamics, provided that signals obey the Markov property. As a result, all relevant information is summarized by a finite number of variables.

Applying the concept of Markovian dynamics to our model we get the following result.

Proposition 3 *Let $\mathcal{F}_t = \sigma(V_s^1, V_s^2, \theta_s,\ s \leq t)$. Suppose agents' information sets are given by*

$$\mathcal{F}_t^j = \{P_\tau, V_\tau^j : \tau \leq t\}, \qquad j = 1, 2.$$

Then in the linear equilibrium of the described economy the system $\{V^1, V^2, \theta, P\}$ does not admit Markovian dynamics.

Proof. See Appendix C.

Although we give a detailed proof in Appendix C, it is useful to make some comments on it here. The idea behind the proof is to use the following result from the theory of stationary Gaussian processes: if the process admits Markovian dynamics, then it is described by a rational function in the frequency domain. We start with the assumption that the price

admits Markovian dynamics. The main part of the proof is to show that it is impossible to satisfy the market clearing condition and to simultaneously solve the optimal filtering problem of each agent working only with rational functions. This contradiction proves that the equilibrium price does not admit Markovian dynamics and the infinite regress problem is there.

To highlight the significance of this result from the theoretical standpoint, we refer to the paper by Townsend (1983), which inspired the study of the infinite regress problem and coined the term "forecasting the forecasts of others". Townsend attempted to create a setup in which traders would have to estimate the beliefs of others in order to solve their own forecasting problems. However, Sargent (1991) and Kasa (2000) show how to reduce all higher order expectations in his model to just a small number of cleverly chosen low order expectations. Since then, a lot of effort has been made to state the necessary and sufficient conditions for the infinite regress problem to exist. We demonstrate that our setup is, in a sense, a minimal model where this phenomenon appears. We know from the result of He and Wang (1995) that if the value of the payoff remains constant over time, it is possible to reduce higher order expectations to first order expectations. In our model, we relax just this condition. It is still interesting to search for other cases, in which solution can take a simple form. Our result, however, severely restricts the set of possible candidates. It suggests that the infinite regress problem is almost unavoidable if one is willing to consider a situation more general than the ones previously studied.

The result also provides support for technical analysis. The simplicity of the fundamentals in our model leads to a straightforward solution in the case of complete information. However, asymmetric information results in highly non-trivial price dynamics. Now, to be as efficient as possible, agents have to use the entire price history in their predictions: as stated in Proposition 3, they cannot choose a finite number of state variables to summarize the price dynamics. This suggests that in financial markets, where fundamentals are not as simple and asymmetric information is commonplace, price history may be informative for investors.

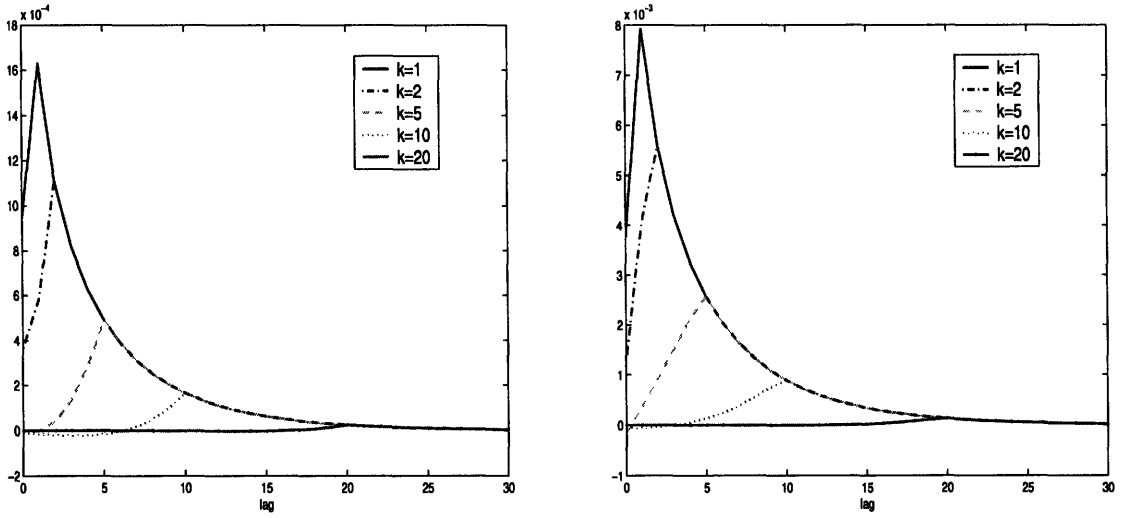
3.6 Numerical procedure

Systems with an infinite number of state variables are very difficult to analyze and, in general, they do not admit analytical solutions. In order to evaluate the implications of information distribution, we construct numerical approximation to the solution. Instead of the initial setup, we consider the k-lag revelation approximation, in which all information is revealed to all investors after k periods[12]. In this case, the state space of the model is finite dimensional and the equations determining the equilibrium can be solved numerically. We relegate all computational details of the k-lag revelation approximation to Appendix D.

To demonstrate the properties of the numerical approximation, we compute the different information equilibria in the k-lag revelation approximation for particular numerical values. In setting parameter values for our numerical solution, we assume that the length of one period is a month. It is reasonable to set the probability of liquidation λ to 5% annually, so that the expected life of a firm is 20 years. We make risk-free rate r equal to 1% annually. We let α, the coefficient of risk aversion, equal 3, which is a commonly chosen value, for example, as in Campbell, Grossman, and Wang (1993). We set the mean-reversion parameter a to 0.85. We make the size of supply shocks b_Θ equal to 15%. In Section 3.9 we relate this parameter to volume turnover. In order for risk premium and return volatility to roughly match their empirical counterparts, we set b_V equal to 1.2. Although the parameters are chosen somewhat arbitrarily, they match several observable statistics. In particular, with these parameters we get risk premium equal to slightly more than 7% and return volatility of 15%. Most of our reported results are computed for this combination of parameters, but we also examined a wide range of them and found that our conclusions are not driven by this particular choice.

k-lag revelation approximation is a workhorse of our computations. Hence, before delving into numerical analysis we have to establish the precision of this procedure. Below we demonstrate that the k-lag revelation approximation is sufficiently precise and converges to the exact differential information equilibrium. Moreover, the rate of this convergence is quite high and even small number of lags provide very good precision. It means that we can legitimately use this approximation for numerical analysis of analytically untractable equilibria.

[12]This approximation was initially suggested by Townsend (1983).



(a) (b)

Figure 3-1: Precision of the k-revelation approximation.

These graphs depict errors in the price decomposition over lagged shocks ϵ^i (Panel (a)) and ϵ^θ (Panel (b)) in the differential information equilibrium relative to the benchmark $k = 50$. Information is revealed to both types of agents after 1, 2, 5, 10, and 20 lags correspondingly. The model parameters are as follows: $a = 0.85$, $b_V = 1.2$, $b_\Theta = 0.15$, $\lambda = 0.05/12$, $r = 0.01/12$, $\alpha = 3$, $\gamma = 0.5$. Since the shocks ϵ^1 and ϵ^2 enter the price function symmetrically, the corresponding coefficients are identical and we report them only once.

In order to analyze precision of the k-lag revelation approximation we compute the differential information equilibrium with various number of lags k after which information is fully revealed. In particular, we take $k = 1$, $k = 2$, $k = 5$, $k = 10$, $k = 20$ and $k = 50$. It is reasonable to think that if the computed equilibria for sufficiently large k are not significantly different from each other and the price coefficient for each lag has a well defined limit then the approximation is quite good. Indeed, in this case it is not likely that for some greater k the coefficients jump or have irregular behavior. To demonstrate the quality of the k-lag revelation approximation, we decompose prices over shocks ϵ_t^1 and ϵ_t^Θ for different values of k and plot the differences in the corresponding coefficients for the given k and $k = 50$. The results are presented in Figure 3-1.

Inspecting Figure 3-1 we conclude that as k goes up the coefficients for each lag tend to the value corresponding to $k = 50$. Moreover, even 10 lags give a good approximation of the first five coefficients. Starting from the 6th coefficient the approximation works poorer because the 10th revealed lag introduces substantial distortions. However, the price

coefficients quickly decrease with the lag, so the error introduced by revealed information gets very small for sufficient number of lags. Indeed, the coefficients calculated for 20 and 50 unknown lags are virtually indistinguishable. This observation suggests that the k-lag revelation procedure gives very good approximation to exact differential information equilibrium even for 20 lags. Thus, the k-lag revelation approximation is quite reliable and produces meaningful results giving justification for its use in our analysis. To be on a safe side, we use 50 unknown lags in our computations.[13]

3.7 Analysis of higher order expectations

When information is dispersed among agents, higher order expectations play an important role in price formation. Moreover, higher order expectations determine not only the wedge between the price and the fundamental value of the firm, but also the statistical properties of prices and returns. Thus, analysis of higher order expectations and especially of their dynamical properties can shed new light on the impact of information distribution on prices and returns.

It is convenient to decompose the price as given in equation (3.2) into the part determined by fundamentals and the correction Δ_t arising as a consequence of heterogenous expectations:

$$P_t = -\frac{1}{\Omega(r+\lambda)} + \frac{a\lambda}{1+r-a(1-\lambda)}V_t - \frac{1}{\Omega(1+r)}\theta_t + \Delta_t, \tag{3.7}$$

where

$$\Delta_t = \frac{a\lambda}{1+r}\sum_{s=0}^{\infty}\left(\frac{1-\lambda}{1+r}\right)^s (\bar{E}_t^w \bar{E}_{t+1}^w .. \bar{E}_{t+s}^w - E_t)V_{t+s}. \tag{3.8}$$

E_t is the expectation operator with respect to full information. The differences $\Delta_t^s = (\bar{E}_t^w \bar{E}_{t+1}^w .. \bar{E}_{t+s}^w - E_t)V_{t+s}$ represent pure effects of asymmetric information. Obviously, if investors are fully informed $\Delta_t = 0$. The price decomposition (3.7) is valid for any information distribution. Thus, we can apply it to both the hierarchical and differential information cases and compare contributions of higher order expectations in different information setups.

In the hierarchical information case, all terms in the infinite series of expectations can

[13] As an additional check, we computed distances between sequential equilibria corresponding to $k = 1, 2, ...$ in the Hilbert space of price processes. These distances decrease as $k \longrightarrow \infty$, thus the equilibria form a Cauchy sequence. Provided the Hilbert space of price processes is complete, the sequence of computed equilibria converges to the equilibrium in the model without information revelation.

	Hierarchical	Differential
$\sigma(\Delta_t)$	$9.5 \cdot 10^{-4}$	$1.8 \cdot 10^{-2}$
$\rho(V_t, \Delta_t)$	-0.02	-0.27
$\rho(\theta_t, \Delta_t)$	-0.99	-0.73
$\rho(\Delta_{t-1}, \Delta_t)$	0.004	0.64

Table 3.1: Statistics of the informational term Δ_t in the hierarchical and differential information equilibria.
$\sigma(\Delta_t)$ is standard deviation; $\rho(V_t, \Delta_t)$ and $\rho(\theta_t, \Delta_t)$ are correlations with V_t and θ_t; $\rho(\Delta_{t-1}, \Delta_t)$ is autocorrelation coefficient.

be calculated explicitly. In particular, a simple calculation yields

$$\Delta_t^s = a^s \frac{\omega_2}{\Omega} \frac{1 - \left(c\frac{\omega_1}{\Omega}\right)^{s+1}}{1 - c\frac{\omega_1}{\Omega}} \left(\hat{V}_t^1 - V_t^1\right).$$

Here c, ω_1 and ω_2 are constants defined in Appendix B. As expected, all higher order expectations terms are proportional to the estimation error $\hat{V}_t^1 - V_t^1$. It means that all terms in the infinite series are perfectly correlated and the series collapses to one term $\Delta_t = p_\Delta(\hat{V}_t^1 - V_t^1)$, greatly simplifying the analysis.

We have already shown that with differential information all higher order expectations are different so we can evaluate the effect of Δ_t on prices only numerically. To do this, we compute volatility and autocorrelation of Δ_t as well as its correlations with V_t and θ_t. The results are presented in Table 3.1. The standard deviation of Δ_t are significantly greater in the differential information case, as opposed to the hierarchical case indicating a greater effect of information asymmetry when information is differentially distributed.

Also Table 3.1 reports that Δ_t is negatively correlated with contemporaneous values of both V_t and θ_t. Since Δ_t^s are highly correlated with each other the intuition behind the correlations can be easily conveyed by Δ_t^0. Let us start with the negative correlation with V_t. When V_t is high, the difference $\Delta_t^0 = \bar{E}_t V_t - V_t$ is low since at least some investors do not know exactly the value of V_t and their average estimation is biased towards the mean value of V_t, which is 0. The intuition behind negative correlation of Δ_t^0 and θ_t is also straightforward. If there is a positive supply shock, the price of the asset goes down. However, some investors cannot perfectly distinguish this shock from a negative shock to V_t, and therefore their estimation of V_t is low again.

This intuition is valid for both hierarchical and differential information, but the numbers

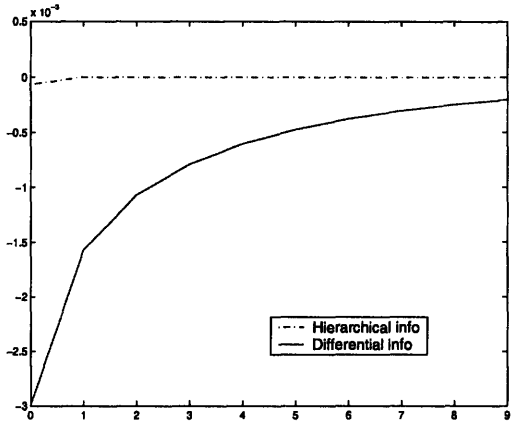


Figure 3-2: Impulse response of Δ_t to the innovation in fundamentals in the models with different informational structures.

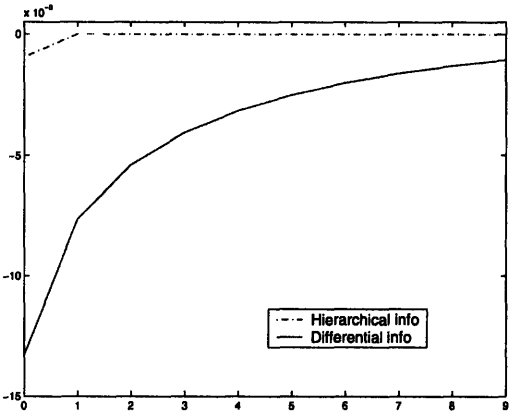


Figure 3-3: Impulse response of Δ_t to the noise trading shock in the models with different informational structures.

in these cases are significantly different. In the case of hierarchical information, informed investors take most of the fundamental risk, but leave some of the liquidity risk to the uninformed. With differential information, no agents are perfectly informed. It translates into a greater average mistake about the fundamentals.

It is also interesting to examine the autocorrelation of Δ_t. Table 3.1 shows that there is almost no autocorrelation in the hierarchical case, whereas Δ_t is highly persistent under differential information with autocorrelation coefficient 0.64. Indeed, under hierarchical information, fully informed investors can take advantage of the mistakes of the uninformed and therefore arbitrage them away. When investors are differentially informed they all make errors. Moreover, the errors made by one type depend not only on fundamentals but also the errors made by another type of investors. Without fully informed arbitrageurs, mistakes are much more persistent in comparison with hierarchical information case since it takes much longer to correct them.

We conclude our study of Δ_t by depicting how it depends on a particular shock represented by its coefficients in the decomposition over the current and past shocks under different information structures.

Figures 3-2 and 3-3 provide more support for the above results. We observe that in the differential information case, Δ_t not only has much higher negative loadings on both fundamentals and supply shocks than in the hierarchical one, but also its response to shocks declines significantly more slowly. In the former case it takes more time for information

to penetrate into price making it less responsive to innovations in fundamentals. This has important implications for statistical properties of prices and returns.

3.8 Implications for asset prices

In this section we analyze how the underlying information structure affects stock prices, returns, and their basic statistical properties. Propositions 1 and 2 give equilibrium prices in economies with full and hierarchical information, respectively. To compute the statistics of interest in the economy with differential information, we use the k-lag revelation approximation described above. All computational details are collected in Appendix D.

3.8.1 Stock prices

Panel (a) of Figure 3-4 shows the decomposition of the equilibrium price with respect to fundamental shocks ε_t^1. As we move from full to hierarchical and then to differential information, it takes longer for fundamental shocks to be impounded into the price (cf. Figure 3-2). The quantitative effect is much more pronounced in the case of differential information. The reason for this is that under hierarchical information, fully informed investors know perfectly the states of the economy: mistakes of the uninformed and demand of the liquidity traders. Competition forces them to arbitrage the mistakes of the uninformed quite aggressively, and by the second lag the price reflects the underlying value almost perfectly. When investors are differentially informed they all make errors in valuations. Moreover, the errors made by one type depend not only on fundamentals, but also the errors made by the other type of investors. Without fully informed arbitrageurs, it takes much longer to correct mispricing: in the figure we see that it takes up to 20 lags for the price to reveal the true value.

Panel (b) of Figure 3-4 shows the decomposition of the price with respect to supply shocks ε_t^Θ. Here we observe the opposite effect: as we move from full to hierarchical and then to differential information the equilibrium price becomes more and more sensitive to noise trading. Investors with perfect information trade against liquidity traders. On the other hand, investors who do not have full information confuse supply and fundamental shocks and therefore require higher compensation to absorb supply shocks. The price is much more affected by supply shocks under differential information, since in this case there

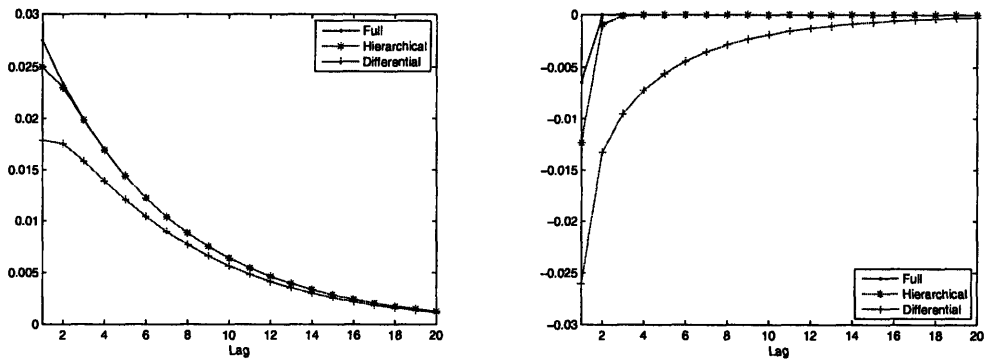


Figure 3-4: Impulse response of the price to underlying shocks.

Panel (a) plots impulse response of the price to shocks ε_t^1 for economies with full, hierarchical, and differential information respectively. Panel (b) plots impulse response of the price to shocks ε_t^Θ for economies with full, hierarchical, and differential information respectively. The following parameter values are used: $\lambda = 0.05/12$, $r = 0.01/12$, $a = 0.85$, $\alpha = 3$, $\gamma = 1/2$, $b_V = 1.2$, $b_\Theta = 1$.

is much more uncertainty about the true value of the firm.

3.8.2 Volatility of prices and returns

Table 3.2 collects standard deviations of prices and returns in equilibria with different information dispersions. The price volatility is almost identical in the full information and hierarchical information cases, but goes down in the differential information setup. This finding can be interpreted as an effect of higher order expectations studied above. From our previous analysis (cf. Table 3.1), we know that the higher order expectations have two opposite effects on price volatility. On one hand, they represent additional volatile state variables the inclusion of which increases total price volatility. On the other hand, these state variables are negatively correlated with V_t, and this correlation is higher for differential information. This leads to decrease in volatility. The overall effect depends on which of the two effects dominates. For the given choice of parameters, these effects almost cancel each other in the hierarchical information case, and the second effect dominates in the case of differential information, in which price volatility goes down.

Let us consider how the information dispersion setup affects the volatility of returns. Table 3.2 reports that volatility is lowest under differential information. This observation

	Full (%)	Hierarchical (%)	Differential (%)
$E(Q_t)$	7.56	7.57	7.16
$Std(Q_t)$	15.88	15.89	15.47
$Std(P_t)$	25.6	25.6	24.7

Table 3.2: Risk premium and volatilities of price and excess return in models with different informational structures.
The model parameters are as follows: $a = 0.85$, $b_V = 1.2$, $b_\Theta = 0.15$, $\lambda = 0.05/12$, $r = 0.01/12$, $\alpha = 3$, $\gamma = 0.5$.

contradicts the conclusion of Grundy and Kim (2002), who assert that differential information causes returns to be more volatile than in the benchmark case with no information asymmetry. The cause for this discrepancy is that in Grundy and Kim's model private information is short lived, so investors can only trade on their information for one period, and therefore trade more aggressively. If information is not revealed every period investors have plenty of time to trade on their information. As a result, it takes a long time for shocks to be impounded into prices, making returns less volatile.

Also, we can see that in the hierarchical information case volatility is slightly larger than under full information. This result is consistent with findings in Wang (1993) who considers a similar model. In this case it also takes longer for shocks to fundamentals to be impounded into price compared to the full information setup. However, another effect is also at work: the uninformed investors face the risk of being taken advantage of by the informed investors. As a result, they are afraid of trading and taking large positions against liquidity traders, which causes returns to be more volatile. The overall result depends on the interaction of these two effects. In our simulations we could not find a region where the first effect is stronger than the second one. It is interesting to notice that under the differential information the opposite is true: the first effect dominates the second one. These results provide another example in which introduction of fully informed arbitragers makes returns more volatile[14].

Because we assigned our agents a mean-variance demand over a one period horizon, the volatility of one period returns has a direct effect on their perception of risk, producing an inverse relation between expected returns and volatility. This is a result of our simplifying assumptions, and a more thorough modelling of agents' preferences, for example as in Wang

[14]See also Stein (1987) and Wang (1993).

(1993) would be required for rigorous analysis of the effect of asymmetric information on risk premium.

3.8.3 Serial correlation in returns

In a seminal paper by Jegadeesh and Titman (1993) it is shown that buying past winners, shorting past losers, and holding the position over 3 to 12 months generates high abnormal profits. Since its discovery, momentum has been one of the most resilient anomalies that challenge the market efficiency hypothesis. Despite a vast empirical literature about momentum[15], there are few theories that try to explain it. These theories are traditionally classified into rational and behavioral.

The rational theories provide risk-based explanations of momentum relating momentum to systematic risk of cash flows. In Berk, Green, and Naik (1999) momentum results from slow evolution of the project portfolio of the firm. Johnson (2002) demonstrates that momentum can arise in a fully rational and complete information setting with stochastic expected dividend growth rates. However, analysis in both papers is conducted in the partial equilibrium framework with an exogenously specified pricing kernel.

Other researchers have turned to behavioral models, which generally attribute momentum to underreaction or delayed overreaction, caused by cognitive biases. In Barberis, Shleifer and Vishny (1998) investors, due to the conservative bias, tend to underweight new information when they update their priors. In Daniel, Hirshleifer and Subrahmanyam (1998) investors are overconfident and overestimate the precision of their signals. As a result, they overreact to private information, but not to public information. Hong and Stein (1999) assume that information is slowly revealed to "news-watchers," who observe future payoff relevant signals but do not use price as a source of information.

In general, there can be three possible sources of momentum profits[16]. First, winners might be stocks with high unconditional expected returns. Second, if one assumes that factors are positively autocorrelated, then winners could be stocks with high loadings to these factors. Finally, it might come from positive autocorrelation of idiosyncratic returns. In all but the first explanation,[17] some components of stock returns should be positively autocorrelated. Therefore, any theory aimed to explain momentum should be able to generate

[15] Jegadeesh and Titman (2005) is a recent review.
[16] See Lo and MacKinlay (1990).
[17] Jegadeesh and Titman (2002) present evidence against this explanation.

positive autocorrelations in returns.

In this section we consider the correlation of Q_{t+1} with the realized excess return $\Delta P_t^e = P_t - (1+r)P_{t-1}$. Although most models with asymmetric information put severe restrictions on possible sign of this correlation, we demonstrate that differential information can make it positive.We use ΔP_t^e instead of Q_t because in the model investors do not observe Q_t, but rather the history of prices. We will say that there is momentum if ΔP_t^e is positively correlated with Q_{t+1}.

Proposition 4 *Define the sequences* a_k, $k = 0, 1, ...$ *and* p_k^θ, $k = 0, 1, ...$ *as* $a_k = E(\varepsilon_{t-k}^\Theta \theta_t)$ *and* $p_k^\theta = E(\varepsilon_{t-k}^\Theta P_t)$. *Then*

$$Cov(Q_{t+1}, \Delta P_t^e) = \frac{1}{\Omega} \sum_{k=0}^{\infty} p_k^\theta (a_k - (1+r)a_{k+1}). \tag{3.9}$$

Proof. Using market clearing condition $\int X_t^i di = 1 + \theta_t$ and the law of iterated expectations we get

$$\mathrm{Cov}(\theta_t, \Delta P_t^e) = \int \mathrm{Cov}(X_t^i, \Delta P_t^e) di = \Omega \mathrm{Cov}\left(Q_{t+1}, \Delta P_t^e\right). \tag{3.10}$$

From the definition of a_k and p_k^θ:

$$\mathrm{Cov}\left(\theta_t \Delta P_t^e\right) = a_0 p_0^\theta + \sum_{k=1}^{\infty} a_k (p_k^\theta - (1+r)p_{k-1}^\theta) = \sum_{k=0}^{\infty} p_k^\theta (a_k - (1+r)a_{k+1}). \tag{3.11}$$

Combining 3.10 and 3.11 we get 3.9. ■

In deriving this result, we use the fact that agents have myopic preferences. In general, there might also be a hedging demand. Note, however, that if the hedging demand results solely from information asymmetry, then it is a linear combination of agents' forecasting mistakes, and therefore is orthogonal to the public information set. Since everyone observes the price, the covariance of the hedging demand with ΔP_t^e is zero, and Eq. 3.10 holds. As a result, the distribution of information between agents can change the magnitude of the correlation but not the sign.

Proposition 4 allows us to study the possibility to observe positive serial correlation of returns for different specifications of θ_t.

Example 1. If θ_t are *i.i.d.* then $\mathrm{Cov}\left(Q_{t+1}, \Delta P_t^e\right) = b_\Theta p_0^\theta / \Omega$.

It means that the sign of the correlation $\mathrm{Cov}\,(Q_{t+1}, \Delta P_t^e)$ is determined by the sign of p_0^θ. This coefficient is negative, because a positive supply shock normally leads to lower price since risk averse agents require compensation for holding additional amount of risky equity. In other words, if θ_t are *i.i.d.* the incentive to follow contrarian strategies is very strong and momentum cannot arise. The logic suggests that if we reduce this incentive by modifying the process for θ_t it might be possible to make the correlation of Q_t with ΔP_t^e positive.

Example 2. If θ_t follows an $AR(1)$ process $\theta_t = b_\Theta(1 - a_\Theta L)^{-1}\varepsilon_t^\Theta$ then

$$\mathrm{Cov}\,(Q_{t+1}, \Delta P_t^e) = \frac{b_\Theta(1 - a_\Theta(1+r))}{\Omega} \sum_{k=0}^{\infty} p_k^\theta a_\Theta^k. \tag{3.12}$$

The sum $\sum_{k=0}^{\infty} p_k^\theta a_\Theta^k$ is likely to be negative in most models, since the price is negatively affected by shocks ε_t^Θ. Therefore, the sign of $\mathrm{Cov}\,(Q_{t+1}, \Delta P_t^e)$ depends only on the sign of $(1 - a_\Theta(1+r))$ and not on the information dispersion or any other model parameters[18]. If θ_t is sufficiently persistent then $(1 - a_\Theta(1+r))$ is negative and momentum arises. It is worthwhile to compare this result with that of Brown and Jennings (1989), who are able to generate positive autocorrelation of returns for a wide range of parameters in a two period, but otherwise similar model. This difference underscores the importance of considering a stationary economy where the initial conditions have little effect on properties of equilibrium.

Next, if we allow θ_t to have more general dynamics, information dispersion has a qualitative effect on serial correlation of returns. To keep the model parsimonious, we consider the case in which θ_t follows an $AR(2)$ process.

Example 3. If θ_t follows an $AR(2)$ process with non-coinciding real roots $a_{1\Theta}$ and $a_{2\Theta}$: $\theta_t = b_\Theta(1 - a_{1\Theta}L)^{-1}(1 - a_{2\Theta}L)^{-1}\varepsilon_t^\Theta$ then

$$\mathrm{Cov}\,(Q_{t+1}, \Delta P_t^e) = \frac{b_\Theta}{\Omega(a_{1\Theta} - a_{2\Theta})} \sum_{k=0}^{\infty} p_k^\theta \left((1 - a_{1\Theta}(1+r))a_{1\Theta}^{k+1} - (1 - a_{2\Theta}(1+r))a_{2\Theta}^{k+1} \right). \tag{3.13}$$

Eq. (3.13) shows that in this case the sign of $\mathrm{Cov}\,(Q_{t+1}, \Delta P_t^e)$ is not so obvious and, in general, depends on particular values of $a_{1\Theta}$ and $a_{2\Theta}$. For illustrative purpose, we set $a_{1\Theta}$ and $a_{2\Theta}$ to 0.54 and 0.89, respectively. This choice is somewhat arbitrary. It guarantees that

[18] Wang (1993) illustrates this observation.

	Full (%)	Hierarchical (%)	Differential (%)
$\text{Corr}(Q_{t+1}, \Delta P_t^e)$	-0.05	-0.02	0.01

Table 3.3: Correlation between Q_{t+1} and ΔP_t^e in models with different informational structures.

Stock supply by noise traders θ_t follows $AR(2)$ process with roots 0.89 and 0.54.

correlations in the full information case are negative at all lags, but at the same time, makes the incentive to trade against liquidity traders small enough. We verify that similar results can be attained with other parameter values as well. Table 3.3 presents the correlation between Q_{t+1} and ΔP_t^e in models with different informational structures. Remarkably, when investors are fully informed or information is dispersed hierarchically, the correlation $\text{Cov}\,(Q_{t+1}, \Delta P_t^e)$ is negative and prices exhibit mean reversion. However, in the differential information case, the correlation is positive and momentum arises. It is worth to emphasize that momentum is not a result of specific choice of fundamental parameters, but originates as a consequence of differential information. In the AR(2) case the mean reverting impact of liquidity traders is sufficiently reduced and the effect of slow diffusion of fundamental shocks dominates producing momentum. Diffusion of information in our model is an endogenous process which is consistent with demands of fully rational investors. It distinguishes our results from those of Hong and Stein (1999), who take the slow rate of information revelation as an assumption. Of course, since we are looking at just one stock, this is not the whole story about momentum: we do not take into account diversification at the limit, but this is beyond the scope of our analysis.

There are several empirical regularities which support the informational explanation of momentum and which are consistent with our model. Hong, Lim, and Stein (2000) show that momentum predominantly resides in small stocks, and that, controlling for size, momentum is greater for firms with little analyst coverage. These stocks are less informationally transparent, and if momentum is really due to slow diffusion of information into prices then exactly these stocks should exhibit the strongest momentum behavior. Verardo (2005) finds that momentum is more pronounced in stocks with high dispersion of analysts' forecasts. This observation is also consistent with the suggested information theory of momentum. Indeed, if analysts have diverse opinions on a particular stock it is likely that it is

more difficult to get objective and reliable information about the firm. Hence, it takes more time for news to be incorporated into prices. As a result, in accordance with our theory, this stock is more prone to momentum.

Although our model suggests that momentum arises due to slow diffusion of information, it lacks a well defined parameter controlling the precision of information that agents have. So, to get sharper predictions consistent with the empirical facts, we consider an extension of our model. We still assume that there are two types of investors, $j = 1, 2$ and investors of type j know V_τ^j, $\tau \leq t$. But now we introduce a third component, V_t^3, which is observed by both types of investors, so that the total value of the firm consists of three parts: $V_t = V_t^1 + V_t^2 + V_t^3$. Again, V_t^3 follows an $AR(1)$ process, $V_{t+1}^3 = aV_t^3 + b_V^* \epsilon_{t+1}$.

The third component allows us to control the magnitude of the information dispersion. To separate the impact of information from the effect of changing fundamentals we keep the variance of V_t constant. By increasing the contribution of V_t^3 to the total firm value V_t and decreasing that of V_t^1 and V_t^2, we decrease the information dispersion among agents. To make the results comparable across the sections we fix $Var(V_t)$ and control the contribution of V_t^3 by means of b_V^*. Thus, if b_V^* is close to zero the contribution of V_t^3 is negligible and we arrive at the differential information case with maximum information dispersion. On the contrary, if b_V^* is close to $\sqrt{2}b_V$ the third component dominates and we get the full information case with zero information dispersion. We measure how diverse are opinions among the agents as $\mathcal{D} = 1 - \sqrt{Var(V_t^*)/Var(V_t)}$.

To gain a better understanding of the relation between momentum and information dispersion we plot the correlation of Q_{t+1} with ΔP_t^e as a function of $\mathcal{D}$ in Figure 3-5. We see this correlation increases monotonically with information dispersion and eventually converges to the positive correlation observed under differential information. This observation is consistent with the results of Verardo (2005), thus providing support to our information-based theory of momentum.

3.9 Trading volume

In this section we examine the basic properties of trading volume under different information dispersion setups. This question has received a significant amount of attention in the past. For example, Wang (1994) conducts an extensive analysis of stock trading volume

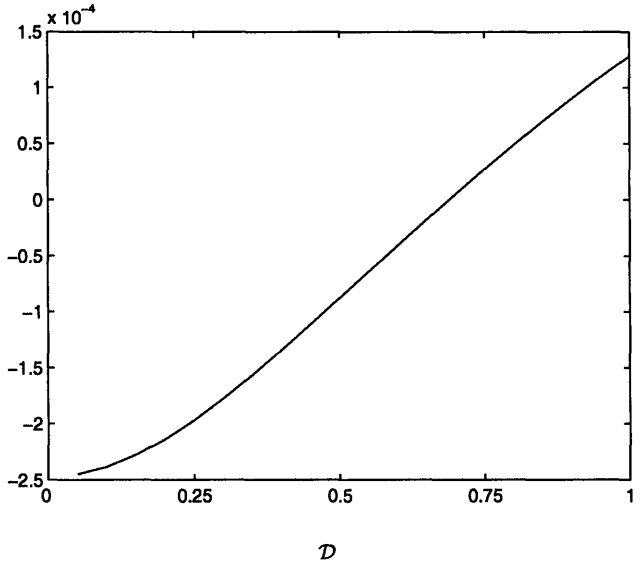


Figure 3-5: Correlation between Q_{t+1} and ΔP_t^e as a function of information dispersion measured as $\mathcal{D} = 1 - \sqrt{Var(V_t^*)/Var(V_t)}$.

under hierarchical information and He and Wang (1995) study it under differential information. However, as we have already pointed out, these papers employ various simplifying assumptions to avoid the infinite regress problem. As a result, there is no easy way to compare the findings among those models. In our work both hierarchical and differential information are nested within the same model, which makes it possible to conduct such analysis.

Let us first give a definition of volume in our model. Since the average number of shares in our model is equal to one, what we refer to as trading volume is actually the turnover. If in period t agent i holds X_t^i shares, but in period $t+1$ his holdings are X_{t+1}^i then his (unsigned) trading volume is $|X_{t+1}^i - X_t^i|$. We are interested in average trading volume $Vol^i = E|X_{t+1}^i - X_t^i|$ of each agent and the relation between volume and information the particular agent has. All liquidity shocks, produced by noise traders, are absorbed by rational investors, and their trading volume is mostly determined by these shocks. However, when investors have different information, they also trade with each other and, on average, this volume can be characterized as $Vol^{12} = E\left[\int |X_{t+1}^i - X_t^i| di - |\theta_{t+1} - \theta_t|\right]/2$. Because this volume is generated endogenously, we call it informational volume. On the other hand, the trading volume of noise traders $Vol^{NT} = E|\theta_{t+1} - \theta_t|$ is completely exogenous in the model. Total trading volume in the model is $Vol^{Tot} = Vol^{12} + Vol^{NT}$. The results are

	I(%)	II(%)	III(%)
Vol^1	16.9	33.7	37.8
Vol^2	16.9	1.2	37.8
Vol^{12}	0	0.2	10.5
Vol^{NT}	16.9	16.9	16.9
Vol^{Tot}	16.9	17.1	27.4
$\text{Corr}(\lvert\Delta P_t\rvert, \text{Vol}_t^{Tot})$	0.1	0.7	6.7

Table 3.4: Normalized trading volume in models with different informational structures. Vol^i – average trading volume of type i investors, Vol^{12} – trading volume between two classes of investors, Vol^{NT} – trading volume of noise traders, Vol^{Tot} – aggregate trading volume. I – full information equilibrium, II – hierarchical information equilibrium, III – differential information equilibrium.

collected in Table 3.4.

Under full information, the volume is completely exogenous: no trades occur between the informed agents. They simply absorb liquidity shocks, equally splitting the volume. Under hierarchical information, the informed agents absorb most of the trades, since the uninformed agents are aware of their disadvantage and, therefore, averse to trade. Their volume is not zero because they try to trade against the noise traders, but occasionally make mistakes and end up trading against informed investors. In the case of differential information, the situation is very different. Agents of different types are not afraid of trading against each other, and this leads to a high trading volume between them, as well as total volume.

It is interesting to consider the behavior of trading volume with respect to the amount of noise trading. Figures 3-6 and 3-7 show the ratio of the informational volume to the volume of noise traders and total volume, respectively, for both hierarchical and differential information.

We observe that total volume is increasing under both setups, since it is driven primarily by the increase in the exogenous volume of liquidity traders. With the normality assumption about the underlying shocks, we have

$$E|\theta_{t+1} - \theta_t| = \frac{2b_\Theta}{\sqrt{\pi}}, \tag{3.14}$$

which is linear in noise trading intensity b_Θ. Thus, it is more instructive to consider the behavior of the ratio $\text{Vol}^{12}/\text{Vol}^{NT}$. We can see that it displays a very different pattern.

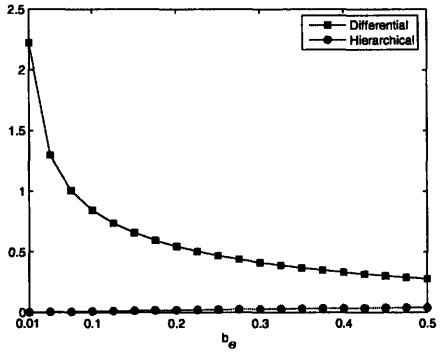


Figure 3-6: The ratio of informational volume in the hierarchical and differential information equilibria to the exogenous noise trading volume $\text{Vol}^{12}/\text{Vol}^{NT}$ as a function of noise trading intensity b_Θ.

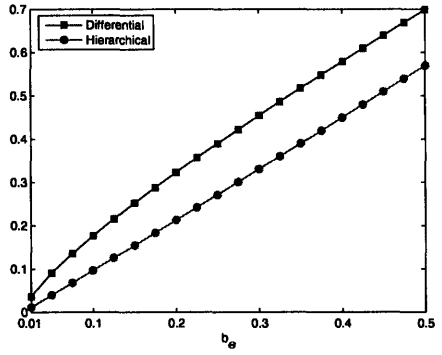


Figure 3-7: Total trading volume Vol^{Tot} in the hierarchical information and differential information equilibria as a function of noise trading intensity b_Θ.

In the case of hierarchical information, the ratio is increasing in b_Θ. The increase in the level of noise trading represents a better trading opportunity for the uninformed traders, so they start to trade more. However, the more they trade, the more often they trade against the informed investors. In the case of differential information the only obstacle to trade is the no-trade theorem. Price becomes more and more informative as the level of noise trading decreases. But in this case investors although trading less in absolute terms trade much more relative to liquidity traders. This result suggests that asymmetric, especially differential, information can help explain high trading volume levels observed in financial markets.

We can also notice that the model is capable of producing another stylized fact about volume: the positive correlation of trading volume with absolute price changes. Table 3.4 shows that correlation increases from full information to hierarchical information, and is strongest under differential information. In the case of full information, the price moves whenever any shock occurs. However, change in volume is only caused by supply shocks. As we move from full information to hierarchical and to differential, more and more trades come from shocks to fundamentals resulting in increased correlation.

3.10 Concluding remarks

This chapter presents a dynamic equilibrium model of asset pricing under different information dispersion setups. The model allows us to clarify the mechanics behind the infinite regress problem and explicitly demonstrate the effect of information distribution. By analyzing differential information coupled with time evolving fundamentals we are able to provide new insights about the behavior of prices and returns.

Due to the complexity of the problem, we made a number of simplifying assumptions. It is reasonable to believe that the intuition we gain from our analysis can be applied to more realistic models as well. There are several directions in which our research can be developed. First, it would be interesting to consider a setup with multiple stocks and analyze the effect of information distribution on cross-correlations of prices and returns[19]. Next, we consider myopic investors who do not have hedging demand. This significantly simplifies the model, since otherwise we would have to solve a dynamic program with an infinite dimensional space of state variables. The impact of hedging could be non-trivial and needs further research.

In our model the agents are exogenously endowed with their information and can neither buy new information, nor release their own information if they find this exchange profitable. It might be interesting to relax this assumption and to introduce the market for information. This direction was explored in a static setting by Verrecchia (1982), Admati and Pfleiderer (1986), and others but dynamic properties of the market for information are not thoroughly explored[20].

Although our analysis pertains mostly to asset pricing, the insights about various aspects of the "forecasting the forecasts of others" problem and iterated expectations, as well as the intuition behind our results, are much more general and also relevant for other fields. For example, higher order expectations naturally arise in different macroeconomic settings (Woodford (2002)), in the analysis of exchange rate dynamics (Bacchetta and Wincoop (2003)), in models of industrial organization where, for example, firms have to extract information about unknown cost structure of competitors (Vives (1988)). The application of our approach and analysis of higher order expectations in these fields might be fruitful

[19]See Admati (1985), Easley and O'Hara (2004), and Hughes, Liu, and Liu (2005), among others, for a static analysis.

[20]See Naik (1997b) for analysis of monopolistic information market in a dynamic framework.

and need further research.

3.11 Appendix A. Proof of Proposition 1

Our starting point is a representation of equilibrium price (3.2). If all investors know V_t^1 and V_t^2 then the infinite sum can be computed explicitly and we get

$$P_t = -\frac{1}{\Omega(r+\lambda)} + \frac{a\lambda}{1+r-a(1-\lambda)}V_t - \frac{1}{\Omega(1+r)}\theta_t.$$

So the only remaining problem is to calculate Ω which is endogenous and is determined by conditional variance of Q_{t+1}. A simple calculation yields

$$\mathrm{Var}(Q_{t+1}|\mathcal{F}_t^i) = \frac{2\lambda^2(1+r)^2 b_V^2}{(1+r-a(1-\lambda))^2} + \frac{(1-\lambda)^2 b_\Theta^2}{\Omega^2(1+r)^2}.$$

By definition of Ω

$$\Omega = \int \frac{1}{\alpha}\frac{di}{\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^i]} = \frac{1}{\alpha}\frac{1}{\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^i]},$$

which gives the following equation for Ω

$$\frac{1}{\alpha}\frac{1}{\Omega} = \frac{2\lambda^2(1+r)^2 b_V^2}{(1+r-a(1-\lambda))^2} + \frac{(1-\lambda)^2 b_\Theta^2}{\Omega^2(1+r)^2}$$

or, equivalently,

$$\frac{2\lambda^2(1+r)^2 b_V^2}{(1+r-a(1-\lambda))^2}\Omega^2 - \frac{\Omega}{\alpha} + \frac{(1-\lambda)^2 b_\Theta^2}{(1+r)^2} = 0.$$

This is a quadratic equation which has real solutions only if its discriminant is non-negative, or

$$2\sqrt{2}b_V b_\Theta \frac{\lambda(1-\lambda)}{1+r-a(1-\lambda)} \leqslant \frac{1}{\alpha}.$$

Under this condition there is a full information solution with Ω as given in Proposition 1.

3.12 Appendix B. Proof of Proposition 2

If investors are hierarchically informed the infinite sequence of iterated expectations collapses to one term $\hat{V}_t^1 = E[V_t^1|\mathcal{F}_t^2]$, which is a new state variable of the economy. So we

conjecture that the price is a linear function of state variables:

$$P_t = p_0 + p_{V^1} V_t^1 + p_{V^2} V_t^2 + p_\Theta \theta_t + p_\Delta (\hat{V}_t^1 - V_t^1), \tag{B1}$$

where p_0, p_{V^1}, p_{V^2}, p_Θ and p_Δ are constants. The dynamics of $\hat{V}_t^1$ can be found from the filtering problem of uninformed agents. To solve this problem we use the following theorem[21].

Theorem 1 *(Kalman - Bucy filter)*

Consider a discrete linear system of the form

$$x_t = \Phi x_{t-1} + \Gamma \epsilon_{x,t},$$
$$y_t = M x_t + \epsilon_{y,t},$$

where x_t is an n-vector of unobservable state variables at t, y_t is an m-vector of observations at t. Φ, Γ and M are $(n \times n)$, $(n \times r)$, and $(m \times n)$ constant matrices respectively. $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$ are r-vector and m-vector white Gaussian sequences: $\varepsilon_{x,t} \sim \mathcal{N}(0, Q)$, $\varepsilon_{y,t} \sim \mathcal{N}(0, R)$, $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$ are independent. Denote the optimal estimation of x_t at time t as $\hat{x}_t$:

$$\hat{x}_t = E[x_t | y_\tau : \tau \leq t]$$

and define

$$\Sigma = E[(x_t - \hat{x}_t)(x_t - \hat{x}_t)' | y_\tau : \tau \leq t].$$

Then

$$\hat{x}_t = (I_n - KM)\Phi \hat{x}_{t-1} + K y_t, \tag{B2}$$
$$\Sigma = (I_n - KM)(\Phi \Sigma \Phi' + \Gamma Q \Gamma'), \tag{B3}$$
$$K = (\Phi \Sigma \Phi' + \Gamma Q \Gamma') M' [M(\Phi \Sigma \Phi' + \Gamma Q \Gamma') M' + R]^{-1}, \tag{B4}$$

where I_n is the $(n \times n)$ identity matrix.

In our case the system of unobservable state variables is $V_{t+1}^1 = a V_t^1 + b_V \epsilon_{t+1}^1$. The partially informed investors effectively observe $Z_t = (p_{V^1} - p_\Delta) V_t^1 + p_\Theta \theta_t$. We have the

[21] See Jazwinski (1970) for textbook discussion of linear filtering theory.

following mapping:

$$x_t = V_t^1, \quad y_t = Z_t, \quad \Phi = a, \quad \Gamma = b_V,$$
$$M = p_{V^1} - p_\Delta, \quad R = (p_\Theta b_\Theta)^2, \quad Q = 1.$$

Applying the Kalman-Bucy filter we arrive at

$$\hat{V}_t^1 = a(1 - k(p_{V^1} - p_\Delta))\hat{V}_{t-1}^1 + k(p_{V^1} - p_\Delta)V_t^1 + kp_\Theta\theta_t, \tag{B5}$$

where k solves the quadratic equation

$$p_\Theta^2 b_\Theta^2 a^2 (p_{V^1} - p_\Delta)k^2 + (p_\Theta^2 b_\Theta^2(1 - a^2) + b_V^2(p_{V^1} - p_\Delta)^2)k - b_V^2(p_{V^1} - p_\Delta) = 0. \tag{B6}$$

Equation (B5) implies AR(1) dynamics of the estimation error:

$$\hat{V}_t^1 - V_t^1 = ac(\hat{V}_{t-1}^1 - V_{t-1}^1) - b_V c\epsilon_t^1 + kb_\Theta p_\Theta \epsilon_t^\Theta, \quad c = 1 - k(p_{V^1} - p_\Delta). \tag{B7}$$

Consider now the demand functions of investors and the market clearing condition. The aggregate demand of partially informed investors is

$$X_t^2 = (1 - \gamma)\frac{E[Q_{t+1}|\mathcal{F}_t^2]}{\alpha \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^2]}$$

Using our conjecture for the price function we can rewrite it as

$$\begin{aligned} X_t^2 &= \omega_2\Big((1-\lambda)p_0 + a(\lambda + (1-\lambda)p_{V^2})V_t^2 + a(\lambda + (1-\lambda)p_{V^1})\hat{V}_t^1 - (1+r)P_t\Big) \\ &= \omega_2\Big((1-\lambda)p_0 + a(\lambda + (1-\lambda)p_{V^1})V_t^1 + a(\lambda + (1-\lambda)p_{V^2})V_t^2 + \\ &\quad + a(\lambda + (1-\lambda)p_{V^1})(\hat{V}_t^1 - V_t^1) - (1+r)P_t\Big), \end{aligned}$$

where, by definition, $\omega_2 = (1-\gamma)/(\alpha \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^2])$. Similarly, the aggregate demand of informed investors is:

$$X_t^1 = \gamma\frac{E[Q_{t+1}|\mathcal{F}_t^1]}{\alpha \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1]} = \omega_1\Big(a\lambda V_t + (1-\lambda)E[P_{t+1}|\mathcal{F}_t^1] - (1+r)P_t\Big),$$
$$\omega_1 = \gamma/(\alpha \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1]).$$

Using (B7) we can rewrite it as

$$\omega_1\Big((1-\lambda)p_0+a(\lambda+(1-\lambda)p_{V^1})V_t^1+a(\lambda+(1-\lambda)p_{V^2})V_t^2+acp_\Delta(1-\lambda)(\hat{V}_t^1-V_t^1)-(1+r)P_t\Big).$$

The market clearing condition $X_t^1 + X_t^2 = 1 + \theta_t$ gives

$$P_t = \frac{\Omega(1-\lambda)p_0 - 1}{\Omega(1+r)} + \frac{a(\lambda+(1-\lambda)p_{V^1})}{1+r}V_t^1 + \frac{a(\lambda+(1-\lambda)p_{V^2})}{1+r}V_t^2 - \frac{1}{\Omega(1+r)}\theta_t + \frac{a(\omega_2\lambda+(1-\lambda)(\omega_2 p_{V^1}+\omega_1 cp_\Delta))}{\Omega(1+r)}(\hat{V}_t^1 - V_t^1), \quad \text{(B8)}$$

where $\Omega = \omega_1 + \omega_2$. Comparing (B8) with the conjectured expression for price we get a set of equations for the coefficients p_0, p_{V^1}, p_{V^2}, p_Θ and p_Δ:

$$p_0 = \frac{\Omega(1-\lambda)p_0 - 1}{\Omega(1+r)}, \qquad p_{V^1} = \frac{a(\lambda+(1-\lambda)p_{V^1})}{1+r},$$
$$p_{V^2} = \frac{a(\lambda+(1-\lambda)p_{V^2})}{1+r}, \qquad p_\Theta = -\frac{1}{\Omega(1+r)},$$
$$p_\Delta = \frac{a(\omega_2\lambda+(1-\lambda)(\omega_2 p_{V^1}+\omega_1 cp_\Delta))}{\Omega(1+r)}.$$

Solving these equations we obtain:

$$p_0 = -\frac{1}{\Omega(r+\lambda)}, \qquad p_{V^1} = p_{V^2} = \frac{a\lambda}{1+r-a(1-\lambda)},$$
$$p_\Theta = -\frac{1}{\Omega(1+r)}, \qquad p_\Delta = \frac{\omega_2\lambda a(1+r)}{(1+r-a(1-\lambda))(\Omega(1+r)-\omega_1 ac(1-\lambda))}.$$

Coefficients p_{V^1} and p_{V^2} are expressed in terms of exogenous parameters of the model. In order to get p_0, p_Θ, and p_Δ we have to compute $\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1]$ and $\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^2]$. We have:

$$\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1] = \\ = b_V^2\left[(\lambda+(1-\lambda)(p_{V^1}-cp_\Delta))^2+(\lambda+(1-\lambda)p_{V^2})^2\right]+b_\Theta^2(1-\lambda)^2p_\Theta^2(1+kp_\Delta)^2,$$

$$\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^2] = \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1] + a^2\left(\lambda+(1-\lambda)(p_{V^1}-cp_\Delta)\right)^2\mathrm{Var}[\hat{V}_t^1 - V_t^1|\mathcal{F}_t^2] = \\ = \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^2] = \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1] + (\lambda+(1-\lambda)(p_{V^1}-cp_\Delta))^2\frac{a^2c}{1-a^2c}b_V^2.$$

As a result, we have the following system of nonlinear equations for p_Θ, p_Δ, c, ω_1, ω_2 and Ω:

$$
\begin{aligned}
&p_\Theta = -\frac{1}{\Omega(1+r)},\\
&p_\Delta = \frac{\omega_2\lambda a(1+r)}{(1+r-a(1-\lambda))(\Omega(1+r)-\omega_1 ac(1-\lambda))},\\
&\omega_1\left(b_V^2\left[(\lambda+(1-\lambda)(p_{V^1}-cp_\Delta))^2+(\lambda+(1-\lambda)p_{V^2})^2\right]+b_\Theta^2(1-\lambda)^2p_\Theta^2(1+kp_\Delta)^2\right)=\gamma,\\
&\omega_2\left(b_V^2\left[\frac{1}{1-a^2c}(\lambda+(1-\lambda)(p_{V^1}-cp_\Delta))^2+(\lambda+(1-\lambda)p_{V^2})^2\right]+\right.\\
&\qquad\left.+b_\Theta^2(1-\lambda)^2p_\Theta^2(1+kp_\Delta)^2\right)=1-\gamma,\\
&p_\Theta^2b_\Theta^2a^2(1-c)^2+(p_\Theta^2b_\Theta^2(1-a^2)+b_V^2(p_{V^1}-p_\Delta)^2)(1-c)-b_V^2(p_{V^1}-p_\Delta)^2=0,\\
&\Omega=\omega_1+\omega_2.
\end{aligned}
$$

The solution to the above system then can be obtained numerically.

3.13 Appendix C. Proof of Proposition 3

To save space we give the proof for $\alpha=1$ and $\gamma=1/2$, and the components V_t^1 and V_t^2 are treated symmetrically. The proof for the general case follows the same logic but is more involved. Denote demeaned price by $\tilde{P}_t$. We assume that the model has a stationary linear equilibrium, i.e. $\tilde{P}_t$ is a stationary regular Gaussian process[22] which admits the following decomposition:

$$\tilde{P}_t = b_V\sum_{k=0}^{\infty}f_k\varepsilon_{t-k}^i + b_V\sum_{k=0}^{\infty}f_k\varepsilon_{t-k}^{-i} + b_\Theta\sum_{k=0}^{\infty}f_k^\Theta\varepsilon_{t-k}^\Theta, \tag{C1}$$

where

$$\sum_{k=0}^{\infty}\left(b_V^2f_k^2+b_V^2f_k^2+b_\Theta^2\left(f_k^\Theta\right)^2\right)<\infty. \tag{C2}$$

Instead of working with an infinite number of coefficients it is convenient to put the series in z-representation[23], i.e. introduce functions $f(z)$ and $f_\Theta(z)$ such that

$$f(z)=\sum_{k=0}^{\infty}f_kz^k, \qquad f_\Theta(z)=\sum_{k=0}^{\infty}f_k^\Theta z^k. \tag{C3}$$

[22]See all relevant definitions in Ibragimov and Rozanov (1978).

[23]For other applications of z-representation to analysis of rational expectation equilibrium see Futia (1981), Kasa (2000), Kasa, Walker, and Whiteman (2004).

Due to (C2) f and f_Θ are well-defined analytical functions in the unit disk $D_0 = \{z : |z| < 1\}$ in the complex plane $\mathbb{C}$. Let L be a shift operator defined as $L\varepsilon_t = \varepsilon_{t-1}$. Then using z-representation we can put the conjectured price function into the following form:

$$\tilde{P}_t = b_V f(L)\varepsilon_t^i + b_V f(L)\varepsilon_t^{-i} + b_\Theta f^\Theta(L)\varepsilon_t^\Theta. \tag{C4}$$

One can verify that if two random processes x_t and y_t are

$$x_t = b_V f_x^1(L)\varepsilon_t^i + b_V f_x^2(L)\varepsilon_t^{-i} + b_\Theta f_x^\Theta(L)\varepsilon_t^\Theta$$
$$y_t = b_V f_y^1(L)\varepsilon_t^i + b_V f_y^2(L)\varepsilon_t^{-i} + b_\Theta f_y^\Theta(L)\varepsilon_t^\Theta$$

then

$$E[x_t y_t] = \frac{1}{2\pi i}\oint \left\{ b_V^2 f_x^1(z) f_y^1\left(\frac{1}{z}\right) + b_V^2 f_x^2(z) f_y^2\left(\frac{1}{z}\right) + b_\Theta^2 f_x^\Theta(z) f_y^\Theta\left(\frac{1}{z}\right)\right\}\frac{dz}{z}.$$

It turns out that the notion of Markovian dynamics has a nice counterpart in the frequency domain. We will use extensively the following result from the theory of Gaussian stationary processes (see Doob (1944) for original results and Ibragimov and Rozanov (1978) for textbook treatment).

Theorem 2 *Let X_t be a regular Gaussian stationary process with discrete time defined on a complete probability space $(\Omega, \mathcal{F}, \mu)$. Let $\mathcal{F}_t$ be a natural filtration generated by X_t. The process X_t admits Markovian dynamics with a finite number of Gaussian state variables if and only if its spectral density is a rational function $e^{i\lambda}$.*

Remark. It is a well-known result then that a Gaussian process X_t with a rational spectral density is an ARMA(p,q) process, that is, it can be represented as

$$X_t - \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \tag{C5}$$

for some ϕ_i, $i = 1..p$, θ_i, $i = 1..q$, and ε_t, $t \in \mathbb{Z}$.

Let us reformulate the equilibrium conditions in terms of functions $f(z)$ and $f_\Theta(z)$. It is convenient to start from the filtering problem of each agent. When forming his demand each agent has to find the best estimate of $\lambda V_{t+1}^{-i} + (1-\lambda)P_{t+1}^i$ given his information

set $\mathcal{F}_t^i = \{V_s^i, P_s\}_{-\infty}^t$. Since some components of P_t are known to agent i, observation of $\mathcal{F}_t^i = \{V_s^i, P_s\}_{-\infty}^t$ is equivalent to observation of $\mathcal{F}_t^i = \{V_s^i, Z_s^i\}_{-\infty}^t$, where

$$Z_t^i = b_V f(L)\epsilon_t^{-i} + b_\Theta f^\Theta(L)\epsilon_t^\Theta. \tag{C6}$$

The filtering problem is equivalent to finding a projector G such that:

$$E[\lambda V_{t+1}^{-i} + (1-\lambda)Z_{t+1}^i | \mathcal{F}_t^i] = G(L)Z_t^i. \tag{C7}$$

By definition, $\lambda V_{t+1}^{-i} + (1-\lambda)Z_{t+1}^i - G(L)Z_t^i$ is orthogonal to all Z_s^i, $s \le t$:

$$E[\left(\lambda V_{t+1}^{-i} + (1-\lambda)Z_{t+1}^i - G(L)Z_t^i\right) Z_s^i] = 0. \tag{C8}$$

Calculating expectations we get

$$\begin{aligned}
E[V_{t+1}^{-i} Z_s^i] &= \frac{1}{2\pi i}\oint \frac{1}{z}\frac{a}{1-az}\frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) dz, \\
E[Z_{t+1}^i Z_s^i] &= \frac{1}{2\pi i}\oint \left\{ b_V^2 \frac{1}{z^2} f(z) \frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{z^2} f^\Theta(z) \frac{1}{z^{t-s}} f^\Theta\left(\frac{1}{z}\right)\right\}, \\
E[G(L)Z_t^i Z_s^i] &= \frac{1}{2\pi i}\oint \left\{ b_V^2 \frac{1}{z} G(z) f(z) \frac{1}{z^{t-s}} f\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{z} G(z) f^\Theta(z) \frac{1}{z^{t-s}} f^\Theta\left(\frac{1}{z}\right)\right\} dz.
\end{aligned} \tag{C9}$$

Collecting all terms the orthogonality condition (C8) takes the form

$$\frac{1}{2\pi i}\oint \frac{1}{z^k} U(z) = 0, \quad k = 1, 2, \ldots \tag{C10}$$

where the function $U(z)$ is

$$\begin{aligned}
U(z) = b_V^2 \frac{a\lambda}{1-az} f\left(\frac{1}{z}\right) + (1-\lambda)\left(b_V^2 \frac{1}{z} f(z) f\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{z} f_\Theta(z) f_\Theta\left(\frac{1}{z}\right)\right) - \\
- G(z)\left(b_V^2 f(z) f\left(\frac{1}{z}\right) + b_\Theta^2 f_\Theta(z) f_\Theta\left(\frac{1}{z}\right)\right).
\end{aligned} \tag{C11}$$

This means that $U(z)$ is analytic in $D_\infty = \{z : |z| > 1\}$ and $U(\infty) = 0$. In other words, the series expansion of $U(z)$ at $z = \infty$ doesn't have the terms $z^s, s \ge 0$. The demand function

of i' agent in z-representation can be written as

$$X_t^i = -(r+\lambda)p_0 + b_V\left(\frac{a\lambda}{1-aL} - (1+r)f(L) + (1-\lambda)\frac{f(L)-f(0)}{L}\right)\varepsilon_t^i$$
$$+ b_V\Big(-(1+r)+G(L)\Big)f(L)\varepsilon_t^{-i} + b_\Theta\Big(-(1+r)+G(L)\Big)f_\Theta(L)\varepsilon_t^\Theta. \quad \text{(C12)}$$

The market clearing condition $\omega_1 X_t^1 + \omega_2 X_t^2 = 1 + \theta_t$, where $\omega_1 = \omega_2 = \Omega/2$ should be valid for all realizations of shocks, which yields the following set of equations:

$$\frac{a\lambda}{1-az} - 2(1+r)f(z) + (1-\lambda)\frac{f(z)-f(0)}{z} + G(z)f(z) = 0, \quad \text{(C13)}$$

$$-\Omega(1+r)f_\Theta(z) + \Omega G(z)f_\Theta(z) = 1. \quad \text{(C14)}$$

Given these equations $U(z)$ can be rewritten as

$$U(z) = 2b_V^2(1+r)f(z)f\left(\frac{1}{z}\right) + b_V^2(1-\lambda)\frac{f(0)}{z}f\left(\frac{1}{z}\right) - 2b_V^2G(z)f(z)f\left(\frac{1}{z}\right) +$$
$$+ b_\Theta^2(1-\lambda)\frac{1}{z}f_\Theta(z)f_\Theta\left(\frac{1}{z}\right) - b_\Theta^2\left(\frac{1}{2} + (1+r)f_\Theta(z)\right)f_\Theta\left(\frac{1}{z}\right). \quad \text{(C15)}$$

Note that the term $b_V^2(1-\lambda)\frac{f(0)}{z}f\left(\frac{1}{z}\right)$ does not have terms with non-negative powers of z, so it can be discarded. Similarly, the term $-\frac{1}{2}b_\Theta^2 f_\Theta\left(\frac{1}{z}\right)$ contributes only the constant $-\frac{1}{2}b_\Theta^2 f_\Theta(0)$. So $U(z)$ takes an equivalent form:

$$U(z) = 2b_V^2\left((1+r) - G(z)\right)f(z)f\left(\frac{1}{z}\right) + b_\Theta^2\left((1-\lambda)\frac{1}{z} - (1+r)\right)\times$$
$$\times f_\Theta(z)f_\Theta\left(\frac{1}{z}\right) - \frac{1}{2}b_\Theta^2 f_\Theta(0). \quad \text{(C16)}$$

Let us introduce a function $g(z)$ such that $g(z) = G(z) - (1+r)$. Then equations (C13), (C14), and (C16) take the following forms:

$$f(z) = -\frac{a(\lambda + (1-\lambda)f(0))z - (1-\lambda)f(0)}{(1-az)(1-\lambda-(1+r)z + zg(z))}, \quad \text{(C17)}$$

$$f_\Theta(z) = \frac{1}{\Omega g(z)}, \quad \text{(C18)}$$

$$U(z) = -2b_V^2 g(z) f(z) f\left(\frac{1}{z}\right) + b_\Theta^2 \left(1 - \lambda - (1+r)z + zg(z)\right) \frac{1}{z} \times$$
$$\times f_\Theta(z) f_\Theta\left(\frac{1}{z}\right) - b_\Theta^2 (\frac{1}{2} + \frac{1}{\Omega}) f_\Theta(0). \quad \text{(C19)}$$

So the rational expectation equilibrium in our model is characterized by functions $f(z)$, $f_\Theta(z)$, $g(z)$ and $U(z)$ such that $f(z)$, $f_\Theta(z)$ and $g(z)$ are analytic inside the unit circle, $U(z)$ is analytic outside the unit circle, $U(\infty) = 0$ and equations (C17), (C18), and (C19) hold.

Now let us turn to the main part of the proof. By Theorem 2, if the system $\{V^1, V^2, \theta, P\}$ admits Markovian dynamics, then its joint spectral density should be rational, which, in turn, implies that function $g(z)$ has to be rational as well. Given relationships (C17) and (C18), functions $f(z)$ and $f_\Theta(z)$ should also be rational. So to prove that our model has non-Markovian dynamics we have to show that there do not exist rational functions $f(z)$ and $f_\Theta(z)$ solving equations (C17), (C18), (C19) and satisfying all conditions specified above.

We construct the proof by contradiction. Suppose that function $g(z)$ is rational. For further convenience we introduce the function $H(z)$ such that

$$g(z) = (zg(z) + 1 - \lambda - (1+r)z) H(z) \quad \text{(C20)}$$

Consequently, in terms of $H(z)$, the function $g(z)$ is

$$g(z) = (1+r)\frac{z_0 - z}{H^{-1}(z) - z}, \quad z_0 = \frac{1-\lambda}{1+r} \quad \text{(C21)}$$

The following lemmas describe the properties of $H(z)$.

Lemma 5 *$H(z)$ is rational, $H(z) \neq 0$ for $z \in D_0$, and $H(z_0) = \frac{1}{z_0}$.*

Proof. Since $f_\Theta(z) = 1/(\Omega g(z))$, we have

$$f_\Theta(z) = \frac{1}{\Omega(1+r)} \frac{1 - zH(z)}{(z_0 - z)H(z)}. \quad \text{(C22)}$$

Statements of the lemma now follow from the fact that $f_\Theta(z)$ is rational and analytic in D_0.

Lemma 6 *$(z - z_1)H(z)$, where $z_1 = \frac{(1-\lambda)f(0)}{a(\lambda + (1-\lambda)f(0))}$ is analytic in D_0.*

Proof. Substituting (C21) into (C17) gives

$$f(z) = \frac{a(\lambda + (1-\lambda)f(0))}{(1+r)} \frac{z_1 - z}{(1-az)(z_0 - z)}(1 - zH(z)). \tag{C23}$$

The lemma now follows from analyticity of $f(z)$ in D_0.

Substitution of (C22) and (C23) into $U(z)$ results in

$$U(z) = -2b_V^2 a^2(\lambda + (1-\lambda)f(0))^2 \frac{(z-z_1)(\frac{1}{z} - z_1)}{(1-az)(1-\frac{a}{z})g\left(\frac{1}{z}\right)} \times$$
$$\times H(z)H\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{\Omega^2 z H(z) g\left(\frac{1}{z}\right)} - b_\Theta^2 \left(\frac{1}{2} + \frac{\Omega}{2}\right) f_\Theta(0). \tag{C24}$$

Also from (C22),

$$f_\Theta(0) = \frac{1}{\Omega(1-\lambda)H(0)}. \tag{C25}$$

Since $g(z)$ does not have poles in D_0 (and consequently $g\left(\frac{1}{z}\right)$ does not have poles in D_∞), analyticity of $U(z)$ in D_∞ implies analyticity of $U^g(z) = U(z)g\left(\frac{1}{z}\right)$ in D_∞. Using (C25) we see that

$$U^g(z) = -2b_V^2 a^2(\lambda + (1-\lambda)f(0))^2 \frac{(z-z_1)(\frac{1}{z} - z_1)}{(1-az)(1-\frac{a}{z})} \times$$
$$\times H(z)H\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{\Omega^2 z H(z)} - \left(\frac{1}{2} + \frac{1}{2\Omega}\right) b_\Theta^2 \tag{C26}$$

must be analytical in D_∞. This means that the pole $1/a$ in (C26) must be cancelled. It might happen only due to one of the following reasons:

1. $H(1/a) = 0$,
2. $H(a) = 0$,
3. $z_1 = a$, or, equivalently, $f(0) = \frac{a^2}{1-a^2}\frac{\lambda}{1-\lambda}$
4. $z_1 = 1/a$
5. The pole in the first term is cancelled by a pole in the second term.

It is easy to notice that the first reason does not work since in this case a pole in the second term appears. Similarly, the fifth possibility cannot realize. The equation $z_1 = 1/a$ is inconsistent unless $\lambda = 0$. The second option contradicts the condition that $H(z)$ does

not have zeros inside the unit circle. This leaves only the third possibility should realize and we can fix the value of $f(0)$. Consequently, we rewrite $U^g(z)$ as

$$U^g(z) = -2b_V^2 \frac{a^2\lambda^2}{(1-a^2)^2} H(z)H\left(\frac{1}{z}\right) + b_\Theta^2 \frac{1}{\Omega^2 z H(z)} - (\frac{1}{2} + \frac{1}{2\Omega})b_\Theta^2 \qquad \text{(C27)}$$

with the condition

$$H\left(\frac{1-\lambda}{1+r}\right) = \frac{1+r}{1-\lambda} \quad \text{and} \quad U^g(\infty) = 0. \qquad \text{(C28)}$$

Now we will show that there is no such rational function $H(z)$. Assume for now that $H(z)$ has a pole z_h. From Lemma 6, $z_h = a$ or $z_h \in D_\infty$. If $z_h \in D_\infty$ and $z_h \neq \infty$, then, for analyticity of $U(z)$ in D_∞, we have to have $H(1/z_h) = 0$, but it contradicts Lemma 5. If $z_h = a$ then $U(z)$ has a pole at $1/a$. Indeed, if a is a pole of $H(z)$ then $1/a$ is a pole of $H(1/z)$. The only possibility to cancel it in the first term of $U(z)$ is to have $H(1/a) = 0$. But in this case a pole in the second term arises. So $H(z)$ does not have poles in $\mathbb{C}$. As a result, the only possibility is $z_h = \infty$. This means that $H(z)$ is a polynomial. Let $w_0 \in \mathbb{C}$ be a zero of $H(z)$. Because of Lemma 5, w_0 can be only in D_∞. However, this means that, unless $H(1/z)$ or $H(z)$ have a pole at w_0, $U(z)$ is not analytic in D_∞. We know that $H(z)$ (and consequently $H(1/z)$) do not have poles in $\mathbb{C}$. Thus we can conclude that $H(z)$ does not have zeros. Hence by Liouville's theorem $H(z) = H = const$. We have two equations that this constant has to satisfy:

$$H = \frac{1+r}{1-\lambda}, \qquad -2b_V^2 \frac{a^2\lambda^2}{(1-a^2)^2} H^2 - \frac{1}{\Omega} b_\Theta^2 = 0.$$

Obviously, these conditions are inconsistent and this concludes the proof.

3.14 Appendix D. k-lag revelation approximation

In the k-lag revelation approximation all information is revealed to all investors after k periods, so the information set of investor i is $\mathcal{F}_t^i = \{V_\tau^i : \tau \leq t;\ V_\tau^{-i}, \theta_\tau : \tau \leq t-k\}$. It means that the state of this economy Ψ_t is characterized by the current values of V_t^1, V_t^2 and θ_t and by their k lags:

$$\Psi_t = (\psi_t, \psi_{t-1}, \ldots, \psi_{t-k}, \psi_{t-k})', \quad \text{where} \quad \psi_\tau = (V_\tau^1, V_\tau^2, \theta_\tau)'.$$

Demand of type i investors is

$$X_t^i = \omega_i E[Q_{t+1}|\mathcal{F}_t^i] = \omega_i(a\lambda V_t^i - (1+r)P_t + E[a\lambda V_t^{-i} + (1-\lambda)P_{t+1}|\mathcal{F}_t^i]),$$

where ω_i are endogenous constants given by

$$\omega_1 = \frac{\gamma}{\alpha \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^1]}, \qquad \omega_2 = \frac{1-\gamma}{\alpha \mathrm{Var}[Q_{t+1}|\mathcal{F}_t^2]}$$

We look for the equilibrium price process as a linear function of state variables, i.e. $P_t = p_0 + P\Psi_t$, where P is a $(1 \times 3(k+1))$ constant matrix. In the matrix form dynamics of ψ_t is:

$$\psi_{t+1} = a_\psi \psi_t + \epsilon_{t+1}^\psi, \quad \text{where} \quad a_\psi = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \epsilon_t^\psi = \begin{pmatrix} \epsilon_t^1 \\ \epsilon_t^1 \\ \epsilon_t^\Theta \end{pmatrix}, \quad \mathrm{Var}(\epsilon_t^\psi) = \begin{pmatrix} b_V^2 & 0 & 0 \\ 0 & b_V^2 & 0 \\ 0 & 0 & b_\Theta^2 \end{pmatrix}.$$

Consequently, dynamics of Ψ_t can be described as:

$$\Psi_{t+1} = A_\Psi \Psi_t + B_\Psi \epsilon_{t+1}^\psi, \quad \text{where} \quad A_\Psi = \begin{pmatrix} a_\psi & 0 & \dots & 0 & 0 \\ I_3 & 0 & \dots & 0 & 0 \\ 0 & I_3 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_3 & 0 \end{pmatrix}, \quad B_\Psi = \begin{pmatrix} I_3 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Here I_3 is a 3-dimensional unit matrix. Now demand can be rewritten as

$$X_t^i = \omega_i((1-\lambda)p_0 + a\lambda V_t^i - (1+r)P_t + E[a\lambda V_t^{-i} + (1-\lambda)PA_\Psi \Psi_t|\mathcal{F}_t^i]).$$

Introducing $(1 \times 3(k+1))$ constant matrices $V^1 = (1,0,0,\dots,0)$, $V^2 = (0,1,0,\dots,0)$ and $V = (1,1,0,\dots,0)$ we get

$$X_t^i = \omega_i((1-\lambda)p_0 - (1+r)P_t) + \omega_i(a\lambda V + (1-\lambda)PA_\Psi)E[\Psi_t|\mathcal{F}_t^i].$$

Thus, we have to calculate $E[\Psi_t|\mathcal{F}_t^i]$. Denoting time t observations of agent i as $y_t^i = (\tilde{P}_t, V_t^i)'$ we can gather all his relevant observations in one vector $Y_t^i = (y_t^i, y_{t-1}^i, \dots, y_{t-k+1}, \psi_{t-k})$.

It is also convenient to introduce a set of $\tilde{P}_\tau$, $\tau = t-k+1...t$ to separate the informative part of the price:

$$\tilde{P}_t = P_t - p_0,$$

$$\tilde{P}_{t-1} = P_{t-1} - p_0 - P^k \psi_{t-k-1},$$

$$\dots$$

$$\tilde{P}_{t-k+1} = P_{t-k+1} - p_0 - P^2 \psi_{t-k-1} - \dots - P^k \psi_{t-2k}.$$

Now we can put all observations in a matrix form:

$$Y_t^i = H^i \Psi_t, \quad \text{where} \quad H^i = \begin{pmatrix} h^i \\ h^i J \\ h^i J^2 \\ \vdots \\ h^i J^k \\ O_{3\times 3k} \quad I_3 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & I_3 & 0 & \dots & 0 \\ 0 & 0 & I_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_3 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad h^i = \begin{pmatrix} P \\ V^i \end{pmatrix}$$

We use the following well-known fact: if (Ψ, Y) are jointly normal with zero mean, then

$$E[\Psi|Y] = \beta' Y, \quad \text{where} \quad \beta = \text{Var}(Y)^{-1} E(Y\Psi'), \tag{C29}$$

$$\text{Var}[\Psi|Y] = \text{Var}(\Psi) - E(Y\Psi')' \text{Var}(Y)^{-1} E(Y\Psi'). \tag{C30}$$

In our particular case we have:

$$\text{Var}(Y_t^i) = H^i \text{Var}(\Psi_t) H^{i'}.$$

$$E(Y^i \Psi') = H^i \text{Var}(\Psi_t)$$

From the dynamic equation for Ψ_t we find that

$$\text{Var}(\Psi_t) = A_\Psi \text{Var}(\Psi_t) A_\Psi' + B_\Psi \text{Var}(\epsilon_t^\psi) B_\Psi'.$$

Iterating we get

$$\mathrm{Var}(\Psi_t) = \sum_{l=0}^{\infty} A_\Psi^l B_\Psi \mathrm{Var}(\epsilon_t^\psi) B_\Psi' A_\Psi^{l\,\prime}$$

Thus, the demand of agent i is

$$X_t^i = \omega_i((1-\lambda)p_0 - (1+r)P_t) + \omega_i(a\lambda V + (1-\lambda)PA_\Psi)\mathrm{Var}(\Psi_t)H^{i\prime}(H^i\mathrm{Var}(\Psi_t)H^{i\prime})^{-1}H^i\Psi_t.$$

The imposing of the market clearing condition gives

$$\begin{aligned}\Omega(1-\lambda)p_0 - \Omega(1+r)P_t \\ + \omega_1(a\lambda V + (1-\lambda)PA_\Psi)\mathrm{Var}(\Psi_t)H^{1\prime}(H^1\mathrm{Var}(\Psi_t)H^{1\prime})^{-1}H^1\Psi_t \\ + \omega_2(a\lambda V + (1-\lambda)PA_\Psi)\mathrm{Var}(\Psi_t)H^{2\prime}(H^2\mathrm{Var}(\Psi_t)H^{2\prime})^{-1}H^2\Psi_t = 1 + \Theta\Psi_t, \quad \text{(C31)}\end{aligned}$$

where $\Theta = (0,0,1,0,0,\ldots,0)$ and $\Omega = \omega_1 + \omega_2$. Rearranging terms we get:

$$\begin{aligned}P_t = \frac{\Omega(1-\lambda)p_0 - 1}{\Omega(1+r)} + \frac{1}{\Omega(1+r)}(a\lambda V + (1-\lambda)PA_\Psi)\mathrm{Var}(\Psi_t)\times \\ \times\left[\omega_1 H^{1\prime}(H^1\mathrm{Var}(\Psi_t)H^{1\prime})^{-1}H^1 + \omega_2 H^{2\prime}(H^2\mathrm{Var}(\Psi_t)H^{2\prime})^{-1}H^2\right]\Psi_t - \frac{1}{\Omega(1+r)}\Theta\Psi_t.\end{aligned} \quad \text{(C32)}$$

Comparing this equation with the price representation $P_t = p_0 + P\Psi_t$ we get a set of equations:

$$p_0 = \frac{\Omega(1-\lambda)p_0 - 1}{\Omega(1+r)}, \qquad \text{or} \qquad p_0 = -\frac{1}{\Omega(r+\lambda)},$$

$$\begin{aligned}\Omega(1+r)P = (a\lambda V + (1-\lambda)PA_\Psi)\mathrm{Var}(\Psi_t)\times \\ \times\left(\omega_1 H^{1\prime}(H^1\mathrm{Var}(\Psi_t)H^{1\prime})^{-1}H^1 + \omega_2 H^{2\prime}(H^2\mathrm{Var}(\Psi_t)H^{2\prime})^{-1}H^2\right) - \Theta. \quad \text{(C33)}\end{aligned}$$

This system of equations on matrix P shoud be supplemented by two equations determining ω_1 and ω_2. By definition, ω_i are determined by conditional variances $\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^i]$

$$\mathrm{Var}[Q_{t+1}|\mathcal{F}_t^i] = \mathrm{Var}[\lambda V_{t+1} + (1-\lambda)P_{t+1}|\mathcal{F}_t^i] = \mathrm{Var}[(\lambda V + (1-\lambda)P)\Psi_{t+1}|\mathcal{F}_t^i]$$

$$= (\lambda V + (1-\lambda)P)\mathrm{Var}[\Psi_{t+1}|\mathcal{F}_t^i](\lambda V + (1-\lambda)P)'.$$

From (C30) we get

$$\mathrm{Var}[\Psi_{t+1}|\mathcal{F}_t^i] = A_\Psi(\mathrm{Var}(\Psi_t) - \mathrm{Var}(\Psi_t)H^{i'}(H^i\mathrm{Var}(\Psi_t)H^{i'})^{-1}H^{i'}\mathrm{Var}(\Psi_t))A'_\Psi + B_\Psi\mathrm{Var}(\epsilon_t^\psi)B'_\Psi.$$

So the additional equations are

$$\frac{1}{\omega_i} = (\lambda V + (1-\lambda)P)(A_\Psi(\mathrm{Var}(\Psi_t) - \mathrm{Var}(\Psi_t)H^{i'}(H^i\mathrm{Var}(\Psi_t)H^{i'})^{-1}H^{i'}\mathrm{Var}(\Psi_t))A'_\Psi + B_\Psi\mathrm{Var}(\epsilon_t^\psi)B'_\Psi)(\lambda V + (1-\lambda)P)'. \quad \text{(C34)}$$

As a result, when all information is revealed after k lags the equilibrium condition transforms into a complicated system of non-linear equations(C33) and (C34) determining P, ω_1 and ω_2. Numerical solution to these equations give us an approximation to the original heterogeneous information equilibrium.

k-lag approximation allows us to calculate explicitly the decomposition of higher order expectations over the state variables Ψ_t. Indeed,

$$\begin{aligned}\bar{E}_t^w[V_t] &= \frac{1}{\Omega}\left(\omega_1 E[V_t|\mathcal{F}_t^1] + \omega_2 E[V_t|\mathcal{F}_t^2]\right) \\ &= \frac{1}{\Omega}\left(\omega_1 V E[\Psi_t|\mathcal{F}_t^1] + \omega_2 V E[\Psi_t|\mathcal{F}_t^2]\right) = \frac{1}{\Omega}V\left(\omega_1\Pi^1 + \omega_2\Pi^2\right)\Psi_t, \quad \text{(C35)}\end{aligned}$$

where

$$\Pi^i = \mathrm{Var}(\Psi_t)H^{i'}\left(H^i\mathrm{Var}(\Psi_t)H^{i'}\right)^{-1}H^i.$$

Iterating we get:

$$\begin{aligned}\bar{E}_t^w\bar{E}_{t+1}^w[V_{t+1}] &= \bar{E}_t^w[\frac{1}{\Omega}V\left(\omega_1\Pi^1 + \omega_2\Pi^2\right)\Psi_{t+1}] \\ &= \frac{1}{\Omega}V\left(\omega_1\Pi^1 + \omega_2\Pi^2\right)A_\Psi\frac{1}{\Omega}(\omega_1\Pi^1 + \omega_2\Pi^2)\Psi_t,\end{aligned}$$

...

$$\bar{E}_t^w\bar{E}_{t+1}^w...\bar{E}_{t+s}^w[V_{t+s}] = \frac{1}{\Omega}V\left(\omega_1\Pi^1 + \omega_2\Pi^2\right)\left[A_\Psi\frac{1}{\Omega}\left(\omega_1\Pi^1 + \omega_2\Pi^2\right)\right]^s\Psi_t$$

References

Admati A. R. (1985) A Noisy Rational Expectations Equilibrium for Multi-asset Securities Markets, *Econometrica* 53, 629-657

Allen F., S. Morris and H. S. Shin (2006) Beauty Contests and Iterated Expectations in Asset Markets, *Review of Financial Studies* 19, 719-752

Bacchetta, Ph., and E. van Wincoop (2003) Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle? *NBER WP.* 9498

Bacchetta, Ph., and E. van Wincoop (2004) Higher Order Expectations in Asset Pricing, *mimeo*

Back K., C. H. Cao and G. A. Willard (2000) Imperfect Competition among Informed Traders, *Journal of Finance* 55, 2117-2155

Barberis N., A. Shleifer and R. Vishny (1998) A Model of Investor Sentiment, *Journal of Financial Economics* 49, 307-343

Berk J. B., R. C. Green, and V. Naik (1999) Optimal Investment, Growth Options, and Security Returns, *Journal of Finance* 54, 1553-1607

Bernhardt, D., P. Seiler, and B. Taub (2004) Speculative dynamics, *Working Paper, University of Illinois*

Brown, D. P. and R. H. Jennings (1989) On Technical Analysis, *Review of Financial Studies* 2, 527-552

Brunnermeier, M. (2001) Asset Pricing Under Asymmetric Information, Oxford University Press

Campbell, J. Y. and S. J. Grossman, and J. Wang (1993) Trading Volume and Serial Correlation in Stock Returns, *Quarterly Journal Economics* 108, 905-939

Campbell, J. Y., and A. S. Kyle (1993) Smart Money, Noise Trading, and Stock Price Behavior, *Review of Economic Studies* 60, 1-34

Daniel K., D. Hirshleifer and A. Subrahmanyam (1998) Investor Psychology and Security Market under-and Overreaction, *Journal of Finance* 53, 1839-1885

Doob J. L. (1944) The elementary Gaussian processes, *Annals of Mathematical Statistics* 15, 229-282

Easley, D., and M. O'Hara (2004) Information and the cost of capital, *Journal of Finance* 59, 1553-1583

Foster F. D. and S. Viswanathan (1996) Strategic Trading When Agents Forecast the Forecasts of Others, *Journal of Finance* 51, 1437-1477

Futia, C. A. (1981) Rational Expectations in Stationary Linear Models, *Econometrica* 49, 171-192

Grossman, S. J. and J. E. Stiglitz (1980) On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, 393-408

Grundy, B. D. and M. McNichols (1989) Trade and Revelation of Information through Prices and Direct Disclosure, *Review of Financial Studies* 2, 495-526

Grundy, B. D. and Y. Kim (2002) Stock Market Volatility in a Heterogenous Information Economy, *Journal of Financial and Quantitative Analysis* 37, 1-27

He, H., and J. Wang (1995) Differential Information and Dynamic Behavior of Stock Trading Volume, *The Review of Financial Studies* 8, 919-972

Hong, H. and J. Stein (1999) A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance* 54, 2143-2184

Hong, H., T. Lim, and J. Stein (2000) Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies, *Journal of Finance* 55, 265-295

Hughes, J., J. Liu, and J. Liu (2005) Information, Diversification, and Cost of capital, *Working Paper*

Ibragimov I.A., and Y.A. Rozanov (1978) Gaussian Random Processes, Springer-Verlag

Jazwinski, A. H. (1970) Stochastic Processes and Filtering Theory, Academic Press

Jegadeesh N. and S. Titman (1993) Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65-91

Jegadeesh N. and S. Titman (2002) Cross-Sectional and Time-Series Determinants of Momentum Returns *Review of Financial Studies* 15, 143 - 157.

Jegadeesh N. and S. Titman (2005) Momentum, in R. H. Thaler (editor), *Advances in Behavioral Finance*, Vol. 2, Princeton University Press

Johnson, T. C. (2002) Rational Momentum Effects, *Journal of Finance* 57, 585-607

Kasa, K. (2000) Forecasting the Forecasts of Others in the Frequency Domain, *Review of Economic Dynamics* 3, 726 - 756

Kasa, K., T. B. Walker and C. H. Whiteman (2004) Asset Pricing with Heterogeneous Beliefs: A Frequency Domain Approach, *Mimeo*

Keynes, J. M. (1936) The General Theory of Employment, Interest and Money, MacMillan, London

Kyle, A. (1985) Continuous Auctions and Insider Trading, *Econometrica* 53, 1315-1335

Lo, A. W. and J. Wang (2000) Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory, *Review of Financial Studies* 13, 257-300

Lo, A. and A. C. MacKinlay (1990) When are Contrarian Profits Due to Stock Market Overreaction?, *Review of Financial Studies* 3, 175-208.

Lee, C. M. C. and B. Swaminathan (2000) Price Momentum and Trading Volume, *Journal of Finance* 55, 2017-2069

Naik, N. Y. (1997a) On aggregation of information in competitive markets: The dynamic case, *Journal of Economic Dynamics and Control* 21, 1199-1227

Naik, N. Y. (1997b) Multi-period information markets, *Journal of Economic Dynamics and Control* 21, 1229-1258

Pearlman, J. G. and T. J. Sargent (2004) Knowing the Forecasts of Others, *Mimeo, New York University*

Sargent, T. J. (1991) Equilibrium with Signal Extraction from Endogenous Variables, *Journal of Economic Dynamics and Control* 15, 245-273

Schroder M. and C. Skiadas (1999) Optimal Consumption and Portfolio Selection with Stochastic Differential Utility, *Journal of Economic Theory* 89, 68-126

Singleton K. J. (1987) Asset prices in a time-series model with disparately informed, competitive traders, in Barnett, W. A. and Singleton , K. J. (eds.), New Approaches to Monetary Economics, Proceedings of the Second International Symposium in Economic Theory and Econometrics, Cambridge University Press

Stein J. (1987) Informational Externalities and Welfare-reducing speculation, *Journal of Political Economy* 95, 1123-1145

Townsend, R. M. (1983a) Forecasting the Forecasts of Others, *Journal of Political Economy* 91, 546 - 588

Townsend, R. M. (1983b) Equilibrium theory with learning and disparate expectations: some issues and methods, in R. Frydman and E. Phelps (eds), Individual Forecasting and Aggregate Outcomes, Cambridge University Press

Verrecchia, R. E. (1982) Information Acquisition in a Noisy Rational Expectations Economy, *Econometrica* 50, 1415-1430

Verardo, M. (2005) Heterogeneous Beliefs and Momentum Profits, *Working Paper*

Vives, X. (1988) Aggregation of Information in Large Cournot Markets, *Econometrica* 56, 851-876.

Wang, J. (1993) A Model of Intertemporal Asset Prices Under Asymmetric Information, *Review of Economic Studies* 60, 249-282

Wang, J. (1994) A Model of Competitive Stock Trading Volume, *Journal of Political Economy* 102, 127-168

Woodford, M. (2002) Imperfect Common Knowledge and the Effects of Monetary Policy, in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (eds.), Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, Princeton Univ. Press.