Learning About Beta: An Explanation of the Value Premium

Tobias Adrian and Francesco Franzoni*

JOB MARKET PAPER

November 26, 2002

Abstract

We develop an equilibrium model of learning about time-varying risk factor loadings. In the model, CAPM holds from investors' ex-ante perspective. However, positive mispricing can be observed when investors' expectations of beta are above ex-post realizations. This model is used to explain the value premium. In a learning framework, the fact that value stocks used to be more risky in the past leads to investors' expectations of beta that exceed the estimates from more recent samples. We propose an empirical methodology that takes investors' expectations of the factor loadings explicitly into account when estimating betas. With the adjusted estimates of beta, we can explain the cross-section of average returns on the ten book-to-market decile portfolios, and account for the value premium in the relevant sample.

*MIT and UPF. *Acknowledgments:* We are grateful to Olivier Blanchard, Xavier Gabaix, Jonathan Lewellen, Sendhil Mullainathan, Steve Ross, and Dimitri Vayanos for guidance and comments. We would also like to thank Daron Acemoglu, Manuel Amador, John Y. Campbell, Augustin Landier, Jiang Wang, and seminar participants at MIT and UPF for their helpful insights. We thank Ken French for the data. All the errors are the authors' sole responsibility. *Correspondence: Tobias Adrian*: Department of Economics, E52-262A; Massachusetts Institute of Technology; 50 Memorial Drive; Cambridge, MA02142; e-mail: tobias@mit.edu; tel: 1 617 4615138; http//:web.mit.edu/tobias/www; *Francesco Franzoni*: Department of Economics; Universidad Pompeu Fabra; c. R. Trias Fargas 25-27; Barcelona, 08005, Spain; e-mail: francesco.franzoni@econ.upf.es.

1 Introduction

Since its invention by Sharpe (1964) and Lintner (1965), the Capital Asset Pricing Model (CAPM) has experienced varying fortune. Although the early tests by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) established its empirical success, the CAPM anomalies were discovered soon after. Two of the main empirical failures - the value and small stock premia - were pointed out by Basu (1977) and Banz (1981) respectively. Fama and French (1992, 1993) present these failures of CAPM in the most dramatic way. They show that market risk, as measured by beta, does not explain the cross-section of average returns for portfolios sorted on past betas, and on size and book-to-market (B/M).

We provide a novel explanation for the value premium, and propose a new empirical methodology for testing asset pricing models. The motivation for our work comes from the observation in Franzoni (2002) that the beta of value and small stocks has decreased significantly over the past sixty years. In particular, the value stocks beta has dropped by about 77%, from 2.2 in the early forties to below 0.50 in the late nineties.

In an environment where the risk-factor loadings change dramatically, investors may not know the exact riskiness of the portfolios that they are going to hold. Consequently, they need to form beliefs about the betas, and these beliefs are affected by past levels of the loading. Moreover, as the riskiness of value stocks has been decreasing over time, it is likely that investors' expected beta is significantly higher than the actual beta. The implication is that while investors require an expected excess return that is proportional to the riskiness they perceive, the econometrician observes a premium in excess of the realized riskiness of these stocks, and the CAPM is rejected.

We develop this argument through an equilibrium model of learning, with unobservable and time-varying factor loadings, in the model CAPM holds from investors' ex-ante perspective. However, an econometrician, who looks at realized returns, observes positive mispricing relative to the CAPM, when expected beta differs from realized beta.

The model indicates new directions for asset pricing tests. If an asset pricing model holds under investor's probability distribution, the econometrician needs to assume investors' stand point, when testing the model. In particular, the unobservability and the variability of systematic risk should be explicitly taken into account.

In practice, we derive an empirical methodology in which we obtain a proxy for investors' expectations of beta by applying the Kalman filter to realized portfolio returns. This measure of beta is then used to explain the cross-section of average returns on a set of ten B/M sorted portfolios. While the standard tests reject CAPM on these data, our version of CAPM augmented with learning is not rejected. Therefore, we provide an account for the value premium. These findings are robust to modifications in the empirical strategy. In particular, they extend to portfolios sorted on size, as well as on B/M.

The driving force of our results is the assumption that both the current level as well as the long-run average of beta are unobserved by investors. Learning about these two elements causes investors' expectations to adjust slowly to the new realizations of the loading, and to stick more closely to historical values of beta.

The paper is organized as follows. Section 2 examines the relation of our work to other papers in the asset pricing literature. In Section 3, we develop a model in which CAPM is augmented with learning about factor loadings. Also, the performance of the model in explaining the value premium is assessed through simulations. Section 4 proposes an empirical strategy for testing asset pricing models, which is consistent with investors learning about the relevant parameters. Moreover, we present empirical evidence that our model explains the cross-section of average returns for portfolios sorted on B/M. Section 5 presents some robustness checks and extensions of our findings. Section 6 discusses the assumptions on the learning process that are needed to obtain our results. Finally, Section 7 draws the conclusions of our work.

2 Related literature

Our model incorporates insights from different strands of the empirical and theoretical asset pricing literature. Fama and French (1993, 1995, 1996) propose a risk-based explanation to resolve the value and size puzzles. By including size and B/M factors as proxies for some underlying distress risk, they show that they can account for a large fraction of the cross-sectional variation of stocks. Fama and French motivate their findings with multifactor asset pricing models, such as Merton's (1971) ICAPM, and Ross's (1976) APT. Vassalou and Xing (2002) examine the link between the SMB and HML factors and distress risk in detail. They conclude that "although SMB and HML contain default-related information, this is not the reason that the FF model can explain the cross-section. SMB and HML appear to contain important price information, unrelated to default risk." (Vassalou and Xing 2002). We provide an explanation for this finding. In our framework, the SMB and HML factors are detected in ex-post tests and stem from the fact that the systematic riskiness of value and small stocks has changed. We postulate that ex-ante only market risk is priced.

The behavioral finance literature suggests that characteristics, rather than risks, are priced in equilibrium. Daniel and Titman (1997), for example, argue that investors have an exaggerated perception of the riskiness of small and value stocks, and require a premium to hold them. LaPorta, Lakonishok Shleifer, and Vishny (1997) provide evidence that investors have biased expectations about the earnings of value and growth stocks. Our framework explains why the perception of risk for value stocks is understated by usual CAPM regressions.

More recently, Lettau and Ludvigson (2001) have produced favorable evidence for the Consumption CAPM of Breeden (1979) and Lucas (1978). In particular, they argue that the CCAPM holds under conditional probability distributions. Hence, the correct empirical implementation of the model requires the use of scaled factors, i.e. risk factors multiplied by appropriate state variables. Our theory has elements of the Lettau and Ludvigson's (2001) explanation, and more generally of the conditional CAPM literature, as it stresses the importance of studying conditional moments of the return distribution. Lettau and Ludvigson build on the important contribution of Jaganathan and Wang (1996), who estimate a conditional CAPM with labor income as conditioning variable.[1] Campbell and Cochrane (2000) explain the cross-section of returns with conditioning variables that result from a habit-formation model.[2]

A number of recent papers make the distinction between ex-ante expectations and ex-post measurements of returns. Fama and French (2002) extend work by Blanchard (1993) to estimate ex-ante expectations of the equity premium, pointing out that average equity returns are very likely to be above ex-ante expectations. Elton (1999) warns of the dangers of equating average returns with expected returns, and gives a number of examples where average realized returns are clearly different from expected returns. Brav, Lehavy, and Michaely (2002) examine the expectations of stock analysts by analyzing earnings forecasts. The authors find that CAPM cannot be rejected for these expectations. Furthermore, size and B/M factors are insignificant for stock analyst expectations. This evidence is in the spirit of our model, as we argue that expected returns are driven by CAPM, and the size and B/M factors proxy for the difference between ex-post measures and ex-ante expectations. Other recent evidence by Doukas, Kim and Pantzalis (2002) shows that expectations of future earnings of stock analysts are unbiased for value stocks and small stocks, implying that the abnormal returns of value stocks on earnings announcement days found by LaPorta, Lakonishok Shleifer, and Vishny (1997) are caused by a different mechanism. We are suggesting a possible mechanism: it might not be a surprise in the level in earnings, but the news about beta that has caused the excess returns of value stocks.

The reduced form of the general equilibrium pricing model that we derive is a Kalman filter. Some of the early proponents of applying Kalman filtering to economics are Engle and Watson (1987). The reduced form of our pricing model resembles a particular specification of a GARCH model, such as the one estimated by Bollerslev, Engle, and Wooldridge (1988). The resemblance to GARCH models is directly resulting from our focus on modelling time-varying betas explicitly.[3]

[1]See Santos and Veronesi (2001) for a more recent version of using labor income as conditioning varible.

[2]Habit formation models were treated by Constantinides (1990) and Sundaresan (1989), and more recently by Menzly, Santos and Veronesi (2002).

[3]Moskowitz (2002) estimates a GARCH model, including estimates of time-varying covariances of stocks with the value and size factors. Campbell and Hentschel (1992) is another example of a model that gives a

Learning in a symmetric information setting has recently attracted a considerable amount of attention in the academic research literature [4]. Brennan and Xia (2001) address the equity premium puzzle in a learning model. In their model, the drift of the aggregate dividend process is unobservable. Investors update their belief about the true drift of the dividend process, which leads to a learning premium in aggregate asset prices. In our model, we focus on cross-sectional asset pricing anomalies, and calibrate parameters, in particular risk aversion, to reproduce the observed equity premium. The aggregate equity premium is not the focus of our paper. Veronesi (1999) also examines the impact of latent variables on the aggregate stock market: unobserved switches from booms to recessions exacerbate the volatility of the stock market.

A paper that is closely related to ours is the one by Lewellen and Shanken (2002). The authors assume that the mean of the dividend process is unobserved. Their paper accounts for both predictability and excess variance of returns (Shiller, 1981). Ex-ante investors use all the available information, so there are no arbitrage opportunities. Predictability is observable ex-post, and is driven by the Bayesian updating of investors' beliefs about the unobservable mean of the dividend. The authors demonstrate that an econometrician will detect excess volatility as a consequence of learning, as new realizations of dividends cause investors to constantly revise their beliefs. Lewellen and Shanken explain cross-sectional anomalies as the result of learning on fundamentals. Our approach, instead, focuses on learning about riskiness, in an environment with time-varying factor loadings.

Other papers have focused on the portfolio allocation problem for investors with incomplete information. In Barberis (2000), investors are unsure whether returns are predictable or not. This uncertainty leads to an excessive allocation of wealth to stocks, and this allocation is larger, the longer the investment horizon. Barberis does not address cross-sectional implications of incomplete information, which is the focus of our paper. Finally, Pastor (2000) examines the Bayesian decision problem of investors who are unsure whether CAPM holds. One of the alternatives to the CAPM that Pastor considers is the multifactor model

GARCH specification in the reduced form.

[4] Earlier literature in learning include Genotte (1986) treats portfolio choice with parameter uncertainty, and Bawa, Brown and Klein (1979) examine the influence of estimation risk on portfolio choice.

proposed by Fama and French (1993). Pastor finds that even if an investor strongly believes that the market portfolio is mean variance efficient, he should invest a substantial amount of her wealth in value stocks. Unlike our paper, Pastor does not examine the reasons for the premium generated by the value portfolio, he takes this premium as given.

There are two recent papers that exploit the time-variation in beta explicitly. Campbell and Vuolteenaho (2002) decompose the time-variation in betas into variation driven by dividend news of individual stocks, and variation driven by expected return news by the market portfolio. This approach is complementary to ours, as we abstract completely from time-variation in expected market returns. In our approach, it is the time variation in expected returns of individual stocks that is driving the results.

The work of Ang and Chen (2002) is closely related to ours, as they take the time-variation in betas explicitly into account. Using Bayesian statistics, they estimate some of the parameters of the time-series process for betas. Then, through a bootstrapping procedure they show that the CAPM is not rejected on B/M sorted portfolios, once time variation in betas is taken into account. They make a purely statistical argument, as they do not fully derive an equilibrium asset pricing model. Although we also reach the conclusion that CAPM holds when tested on those assets, our approach is different to the extent that we give a learning based motivation for the reasons behind CAPM rejections. Furthermore, we provide a fully specified equilibrium model to support our conclusions.

3 The Learning-CAPM

Franzoni (2002) demonstrates that the systematic risk of stocks varies substantially over time. This finding suggests that investors may not know the true riskiness of assets precisely. In a world with uncertainty about relevant parameters, investors have to infer the factor loadings from the observable information.

Depending on how fast learning occurs, investors' beliefs can diverge more or less significantly from true factor loadings. Since investors' expectations of factor loadings determine the risk premium required to hold assets, it can be the case that the current value of the riskiness of a portfolio, does not entirely explain expected returns.

In this section, a CAPM model is set-up and analyzed in which investors must form expectations on the true level of systematic risk. The key result is that CAPM holds under investors' information set, but the econometrician who looks at realized returns can observe mispricing.

3.1 The set-up

There are N risky assets in the economy indexed by i, where $i = 1, .., N$. Each stock is paying dividends D_t^i that are assumed to be generated by the following factor structure:

$$D_{t+1}^i = \bar{D}^i + b_{t+1}^i x_{t+1} + \varepsilon_{t+1}^i \tag{1}$$

Dividends are determined by a common risk factor x_{t+1}, idiosyncratic risk ε_{t+1}^i, and the factor loading b_{t+1}^i. The factor loading is assumed to be time-varying, according to the following process:

$$b_{t+1}^i = B^i + F^i b_t^i + u_{t+1}^i \tag{2}$$

B^i and F^i are constants drawn at time 0. The shocks to factor loadings u_{t+1}^i are independent across stocks, *i.i.d.* through time, and normally distributed $u_{t+1}^i | x_t \sim N\left(0, \sigma_u^{i2}\right) \quad \forall i, \forall t$. Furthermore, ε_t^i and u_t^j are independent for all i, j. Idiosyncratic risk is also normally distributed with mean zero conditional on x_t, so that $\varepsilon_t^i | x_t \sim N\left(0, \sigma_\varepsilon^{i2}\right) \quad \forall i, \forall t$. Idiosyncratic risk is independent across stocks, and *i.i.d.* through time. The common factor x_t is assumed to be normally distributed $N\left(0, \sigma_x^2\right)$. Therefore, $\bar{D}^i$ is the unconditional mean of dividends. The cross-sectional distribution of $\bar{D}^i$ is assumed to have bounded expectation, $E\left[\bar{D}^m\right] < \infty$.

There is an infinite number of overlapping generations of representative investors in the economy. The investor of each generation t is working when young, receiving labor income y_t, and consuming c_{t+1} when old. Labor income at time t is denoted by y_t, and assumed to be growing deterministically at rate g. The only decision that the investor has to make is the portfolio choice, i.e. deciding how to save between young and old age. Investors solve:

$$\begin{aligned}
&\underset{\left\{\mathbf{a}_t^i\right\}_{i=1}^N}{Max} \quad E\left[-e^{-Ac_{t+1}}|\Im_t\right] && \text{(P)} \\
st.\ c_{t+1} &= (1+r)\,y_t + \sum_{i=1}^N \mathbf{a}_t^i \left(D_{t+1}^i + P_{t+1}^i - (1+r)\,P_t^i\right)
\end{aligned}$$

We thus assume that investors have constant absolute risk aversion A. Labor income is deterministic, and grows at rate $(1+g)$. The risk-free asset is assumed to be in infinite supply at rate r. The number of shares of asset i owned by the investor at time t is denoted by $\mathbf{a}_t^i$.

The assumption that there are overlapping generations of investors in the economy allows one to focus on learning as the main intertemporal linkage. In particular, the OLG assumption implies that each investor has a one-period horizon. Changes in the investment opportunity set due to the time-variation of b_t^i only affect demand via a discount rate effect. We abstract from more general effects that would arise if there was correlation between consumption and changes in systematic risk, and that would lead to a multifactor pricing model in the spirit of Merton (1971). We deliberately focus on a set-up where CAPM holds, and study the impact of learning in this world. It is straightforward to extend the framework to include state variable.

The agents information set evolves according to the following filtration $\Im_t$:

$$\Im_t = \left\{\Im_0, D_s^i, x_s, y_s \text{ for } s \leqslant t, \forall i\right\}$$

where $\Im_0 = \left\{x_0, y_0, \bar{D}^i, F^i, \bar{\mathbf{a}}^i \ \ \forall i\right\}$. The supply of shares outstanding denoted $\bar{\mathbf{a}}^i$, is thus assumed to be drawn at time 0 and known to the investor. The number of stocks is assumed to be large, so that the law of large number can be applied cross-sectionally. This assumption simplifies the learning process that is introduced in the next section.

3.2 Learning

In the specification of the investor's filtration, it has been assumed that neither the systematic risk factor loadings b_t^i, nor B^i are observable to the investor. Investors must therefore form expectations about true factor loadings b_t^i, as well as the long-run behavior of factor loadings that is governed by B^i. Investors are assumed to behave rationally, and forecast changes in systematic risk according to Bayes rule. As the systematic risk of factor loadings changes constantly, learning occurs according to the Kalman filter:

$$b_{t+1|t}^{ie} = B_{t-1}^{ie} + F^i b_{t|t-1}^{ie} + k_t^i \left(D_t^i - \bar{D}^i - b_{t|t-1}^{ie} x_t \right) \tag{3}$$

where $B_{t-1}^{ie} = E\left[B^i | \Im_{t-1}\right]$, and $b_{t+1|t}^{ie} = E\left[b_{t+1|t}^i | \Im_t\right]$. The details of the Kalman filter are given in the appendix. The optimal rule is to use the unexpected part of the current dividend realization to update the previous period's estimate of systematic risk. The function k_t^i is the "gain" and can be interpreted as a regression coefficient. Exact expressions are given in the appendix. It can be seen from the filter that $E\left[b_{t+1|t}^i | \Im_{t-1}\right] = B_{t-1}^{ie} + F^i b_{t|t-1}^{ie}$, i.e. the one-period ahead forecast of systematic risk is a combination of the long-run behavior of b_t^i, as captured by B^i, and the current estimate of the level of risk. The updating equation for expectations about B^i is:

$$B_t^{ie} = B_{t-1}^{ie} + K_t^i \left(D_t^i - \bar{D}^i - b_{t|t-1}^{ie} x_t \right) \tag{4}$$

Note that it follows from this equation that B_t^{ie} is a martingale under the investors information set: $B_t^{ie} = E\left[B_{t+1}^{ie} | \Im_t\right]$. The gain matrix for the filter of B^i, which is interpreted as a time-varying regression coefficient, is defined in the appendix.

A key assumption of the model is that dividend news occurs in discrete time. In fact, if news arrived in continuous time, investors could learn beta without error. We think that discrete time is a realistic assumption with regards to information that is relevant to form expectations on beta. For example, one can think that estimates of beta are updated on a quarterly basis, when earnings announcements are released. More generally, our implicit assumption is that very high frequency information is likely to be more qualitative, and difficult to translate into covariances.

3.3 The equilibrium pricing function

The equilibrium in the economy has to fulfill three standard criteria:

Definition 1 *An equilibrium is defined such that:*

1. *Investors of each generation* $t = 0, .., \infty$ *solve the maximization program* (P)

2. *The good market clears:*

$$c_t = y_t + \sum_{i=1}^{N} \bar{\mathbf{a}}^i D_t^i \quad \forall t$$

3. *Asset markets clear:*

$$\mathbf{a}_t^i = \bar{\mathbf{a}}^i \quad \forall i, \forall t$$

This definition of equilibrium is standard. As investors are assumed to form expectations rationally, equilibrium implies that they employ the Kalman filter developed in Section 3.2.

We can show that there exists a linear pricing function:

Proposition 1 *There exists a linear equilibrium pricing function:*

$$P_t^i = \frac{\bar{D}^i}{r} - \theta^i \left(b_{t+1|t}^{ei} + \frac{B_t^{ei}}{r} \right) \tag{5}$$

$\theta^i = \sigma_x^2 A / \left(1 + r - F^i\right)$

The proof is in the appendix. The price is the average discounted expected dividend, less a risk premium. The risk premium consists of the expectation of tomorrow's factor loading - $b_{t+1|t}^{ei}$- plus the discounted expected long-term average of the factor loading - $\frac{1}{r} B_t^{ei}$. The result that prices only vary due to changes in factor loadings arises from the stylized set-up of the model. It is straightforward to extend the model to cases where other shocks cause movements in prices.[5]

3.4 Returns, Expected Returns, and CAPM

Returns are defined as absolute excess returns:

$$R_{t+1}^i = D_{t+1}^i + P_{t+1}^i - (1 + r) P_t^i$$

[5] If we change the dividend process from equation (1) to $D_{t+1}^i = D_t^i + b_{t+1}^i x_{t+1} + \varepsilon_{t+1}^i$, we can show that the equilibrium pricing function is $P_t^i = \frac{D_t^i}{r} - \frac{(1+r)^2 \sigma_x^2 A}{r^2 (1+r-F^i)} \left(b_{t+1|t}^{ei} + \frac{B_t^{ei}}{r} \right)$.

The reason to study absolute excess returns is analytical tractability, and is standard in a CARA-normal set-up. At the stage of calibration, we will transform the relevant quantities into relative returns. Using the pricing function, we can solve for returns:

$$R^i_{t+1} = b^i_{t+1}x_{t+1} + \varepsilon^i_{t+1} + \frac{\sigma^2_x A}{1+r-F^i}\left(b^{ei}_{t+1|t} - b^{ei}_{t+2|t+1} + \left(B^{ei}_t - B^{ei}_{t+1}\right)/r + r\left(b^{ei}_{t+1|t} + B^{ei}_t/r\right)\right) \tag{6}$$

There are three terms to returns. First, $b^i_{t+1}x_{t+1} + \varepsilon^i_{t+1}$ constitutes the unexpected realization of dividends, consisting of systematic and idiosyncratic risk. Second, returns are an increasing function in downward-revisions of factor loadings, $b^{ei}_{t+1|t} - b^{ei}_{t+2|t+1}$, and of the long-run factor loading $\left(B^{ei}_t - B^{ei}_{t+1}\right)$. The intuition is straightforward: when expected factor loadings decline unexpectedly from one period to the next, prices increase as investors require a lower risk premium to hold the stock. The third term is driven by the short-run and long-run components of factor loadings, $b^{ei}_{t+1|t}$ and B^{ei}_t. This is the usual CAPM pricing effect: stocks with higher systematic risk have higher returns, as investors need to be compensated to hold the asset.

The effect of learning on returns can be seen by replacing for the filtering equations (3) and (4):

$$R^i_{t+1} = b^i_{t+1}x_{t+1} + \varepsilon^i_{t+1} + b^{ei}_{t+1|t}\sigma^2_x A + \frac{\sigma^2_x A\left(k^i_{t+1} + K^i_{t+1}/r\right)}{1+r-F^i}\left(x_{t+1}\left(b^{ei}_{t+1|t} - b^i_{t+1}\right) + \varepsilon^i_{t+1}\right) \tag{7}$$

Returns are determined by three parts: dividends news $b^i_{t+1}x_{t+1} + \varepsilon^i_{t+1}$, a risk premium that is determined by the one-period ahead expectation of systematic risk, $b^{ei}_{t+1|t}\sigma^2_x A$, and the returns that are driven by surprises in the evolution of systematic risk, $\left(b^{ei}_{t+1|t} - b^i_{t+1}\right)$. Note that the one-period ahead expectation of last term in the excess returns is zero from the investors point of view, and uncorrelated with x_t. It is uncorrelated with x_t because we assumed that expectations are formed rationally, and the evolution of systematic risk is uncorrelated with x_t per assumption.

Using equation (7), the market return can be derived:[6]

[6] In this proof, we use a number of technical assumptions that ensure that the variation in beta averages out in the cross-section. The cross-sectional distribution of $\bar{\mathbf{a}}^i$, is independent of the shocks ε^i_t, u^i_t, and x_t,

Proposition 2 *The absolute market return in excess of the risk-free rate is:*

$$R_{t+1}^{m} = x_{t+1} + \sigma_x^2 A$$

The market return is solely driven by the common risk factor x_t. This comes from the assumption that innovations to risk factors b_t^i are idiosyncratic, and average out for the market as a whole. This means that both idiosyncratic and expected factor loadings average to 1 across stocks, and the updating part of returns averages to zero across stocks. From proposition 2, the expected market excess return - or equity premium- is $E\left[R_{t+1}^{m}|\Im_t\right] = \sigma_x^2 A$. The model thus abstracts from time-variation in expected market returns. This is a deliberate choice, as the focus of the paper is the impact of time-varying betas on cross-sectional pricing anomalies. For individual stocks, we can show that a form of CAPM holds:

Proposition 3 *Denote* $\beta_t = Cov\left(R_{t+1}^{i}, R_{t+1}^{m}|\Im_t\right)/Var_t\left(R_{t+1}^{m}|\Im_t\right)$. *CAPM holds:*

$$E\left[R_{t+1}^{i}|\Im_t\right] = \beta_t E\left[R_{t+1}^{m}|\Im_t\right]$$

Furthermore:

$$\beta_t = b_{t+1|t}^{ei}$$

Under the investor's information set, the Capital Asset Pricing Model proposed by Sharpe (1964) and Lintner (1965) holds period by period: expected returns in period t are the asset's beta times the expected returns on the market portfolio. The key difference to the static set-up is the time-variation in the asset's betas. In our model, the beta for each asset equals the expected factor loading of the dividends of that asset in period $t+1$, conditional on information in period t.

and independent of the cross-sectional distribution of F^i. It is assumed that $\bar{\mathbf{a}}^i = O\left(N^{-1}\right)$, so that every stock is small relative to the market as the number of stocks gets large. We distinguish two cases, $F^i = 1$ and $F^i < 1$. In the case when $F^i < 1$, it is assumed that there exists a transformation of B^i, such that $\tilde{B}^i\left(1 - F^i\right) = B^i$, and $\sum_{i=1}^{N} \bar{\mathbf{a}}^i \tilde{B}^i = 1$. Furthermore, the economy is in the steady state at time 0, so that $b_0^i = \tilde{B}^i \quad \forall i$. In the case $F^i = 1 \ \ \forall i$, it is assumed that $\sum_{i=1}^{N} \bar{\mathbf{a}}^i B^i = 0$ and initially, $\sum_{i=1}^{N} \bar{\mathbf{a}}^i b_0^i = 1$. These assumptions are used to prove proposition 2.

3.5 Unobserved b, and ex-post measures of alpha

The model so far implied that CAPM holds under the investors information set, i.e. the beta that is determining returns corresponds to investors' estimate of the factor loading. In this section, we will contrast investors' expectations with the ex-post measures of risk by the econometrician. The key is now what we assume about the econometrician's information set. This subsection starts with a general result, that is then used to derive empirical implementations of the model. To start off, assume that the econometrician observes the history of returns:

$$\Im_t^E = \left\{ R_s^i, R_s^m \text{ for } t_0 \leqslant s \leqslant t, \forall i \right\}$$

The majority of tests of CAPM are performed using only return data. When betas are not time-varying, returns are a sufficient statistic for the riskiness that is expected by market participants.

Proposition 4 *Under the econometrician's information set, we find that estimated beta is:*

$$\hat{\beta}_t^{iE} = \frac{Cov\left(R_{t+1}^i, R_{t+1}^m | \Im_t^E\right)}{Var\left(R_{t+1}^m | \Im_t^E\right)} = E\left[b_{t+1}^i | \Im_t^E\right]$$

The proposition says that the econometrician's estimate of systematic risk, $\hat{\beta}_t^{iE}$, is the econometrician's conditional expectation of the true factor loading, $E\left[b_{t+1}^i | \Im_t^E\right]$. This result stems from the assumption that market returns are uncorrelated with changes in systematic risk, and therefore the Kalman filter updates are uncorrelated with changes in the market return.

Proposition 5

$$E\left[R_{t+1}^i | \Im_t^E\right] = \hat{\alpha}_t^{iE} + \hat{\beta}_t^{iE} E\left[R_{t+1}^m | \Im_t^E\right]$$

where

$$\hat{\alpha}_t^E = E\left[b_{t+1|t}^{ei} - b_{t+1}^i | \Im_t^E\right] \left(1 + G_t\right) \sigma_x^2 A \tag{8}$$

and

$$G_t = \frac{E\left[x_{t+1}\left(k_{t+1}^i + K_{t+1}^i / r\right) | \Im_t^E\right]}{1 + r - F^i} > 0$$

Proposition 5 is the key result of the model. If ex-post estimation of factor loadings differ from estimation of ex-ante expectations of factor loadings, the model predicts that the econometrician will observe positive mispricing. This corresponds to the term $E\left[b^{ei}_{t+1|t} - b^{i}_{t+1}|\Im^{E}_{t}\right]$.

The proposition also shows the role of learning. If there is a wedge in the measurement of ex-ante expectations and ex-post realizations of factors, such that $E\left[b^{ei}_{t+1|t} - b^{i}_{t+1}|\Im^{E}_{t}\right] > 0$, learning amplifies this wedge, through the term G_t, which is always positive. The learning process of investors is underlying G_t: the first term in the numerator, $E\left[x_{t+1}k^{i}_{t+1}|\Im^{E}_{t}\right]$, is the covariation of learning on b^{i}_{t+1} with the market factor, x_{t+1}. The second term in the numerator, $E\left[x_{t+1}K^{i}_{t+1}/r|\Im^{E}_{t}\right]$, is the (discounted) covariation of learning on B^i with the market factor x_{t+1}. These positive covariations of the learning process with the market factor are a direct consequence of the dividend process in equation (1): higher dividends can be either generated by higher x_{t+1} or a higher factor loading b^{i}_{t+1}. The role of learning on the long-run mean is particularly important: when B^i is known, $K^{i}_{t+1} = 0$.

The effect of learning and econometric misspecification is multiplicative: we will show in the next paragraphs that it is misspecification of the econometric model that gives rise to $E\left[b^{ei}_{t+1|t} - b^{i}_{t+1}|\Im^{E}_{t}\right] \neq 0$. This is multiplied with $(1 + G_t)$, and particularly with the effect of learning about the long-run mean.

In simulations in subsection 3.6, we examine the extent to which learning matters for measured mispricing with respect to CAPM. Equation (8) is directly tested in simulated data. The effects of learning about b^{i}_{t} and B^i are disentangled from the effects of surprise moves in factor loadings. A measure of the difference between the ex-post estimation and the ex-ante expectation of factor loadings can be seen in Figure 1, that is discussed in more detail in Section 4.

One of the implications of equation (8) is that there should be no mispricing if the econometric model is properly specified, and the time-variation of b^{i}_{t+1} is properly taken into account. This observation will be tested in section 4. In particular, we will show that the model implies that returns should be estimated with a Kalman filter, and that no mispricing with respect to CAPM should be observed once this is done. This prediction is tested cross-sectionally using estimated betas that result from Kalman filtering returns.

Our theory links the time-variation in beta to the ex-post measurement of *alpha*. Performing OLS regressions implicitly assume that betas are constant, at least over some time-interval. The OLS slope coefficient, from the regression that uses the whole series of data, is the correct estimator when the data generating process is:

$$H^{ols} : R_t^i = \beta^{i,ols} R_t^m + \varepsilon_t^i$$

In this section, we will examine more closely what the OLS assumption does to ex-post CAPM regressions. Let us assume now that the econometrician's information set consists of the history of returns, and in addition the null hypothesis H^{ols} that we labeled OLS:

$$\Im^{ols} = \left\{ R_s^i, R_s^m, \forall s \leq t; \forall i; H^{ols} \right\}$$

Let $\hat{\beta}_\tau^{i,ols} = E\left[b_\tau^i | \Im_t^{ols}\right]$ and $\hat{b}_t^{ei,ols} = E\left[b_\tau^{ei} | \Im_t^{ols}\right]$. Due to the misspecification, we find from proposition 5 that:

$$E\left[R_{t+1}^i | \Im^{ols}\right] = \hat{\alpha}_t^{i,ols} + \hat{\beta}_t^{i,ols} E\left[R_{t+1}^m | \Im_t^E\right]$$

where

$$\hat{\alpha}_t^{i,ols} = \left(\hat{b}_t^{ei,ols} - \hat{\beta}_t^{i,ols}\right)(1 + G_t)\,\sigma_x^2 A \tag{9}$$

The estimate of the intercept in a CAPM regression is thus proportional to the bias that arises from differences in the estimates between realized and expected returns.

For value stocks, factor loadings have declined dramatically over the past decades. Performing OLS on the whole realization of the data implicitly assumes that investors expectations are well proxied by realized factor loadings. However, OLS on ex-post data tends to underestimate ex-ante risk. The intuition for this result is that the decrease in beta over the sample period causes investors' expectations to be above realized factor loadings, and the α for value stocks is picking up this wedge between expectations and factor loadings.

3.6 The Role of Learning: Simulations

This section presents the results of simulations of the model for value stocks. The simulations are done in order to disentangle the effects that arise from surprise moves in factor loadings from the effects of learning about time-varying beta. The main conclusion of these simulations is that the model can explain significant amounts of the value premium when learning occurs slowly enough. This is the case if investors have to learn on both the current level of the loading, b_t, and the long-run mean, B. Moreover, the autoregression coefficient F has to be sufficiently low, namely below one. The reader for whom these results are intuitive already at this stage, can skip directly to the empirical analysis in Section 4.

First, we consider the case when both the short-run and the long-run behavior of factor loadings - b and B - are known. In this case, the only uncertainty arises from the period-to-period change in b. Depending on the beliefs that investors have about the data generating process of factor loadings, varying degrees of the value premium can be explained.

Second, the effect of learning about the long-run behavior governed by B is studied, when the variation in factor loadings b is observable. Learning about B gives more weight to past observations, and increases the fraction of the value premium that can be explained. Lastly, it is shown that unobservability about the short-run and long-run systematic risk - b and B - can account for a large fraction of the value premium. In particular, we argue that the decline in the riskiness of value stocks leads to an ex-post bias to underestimate risk when OLS regressions of portfolio returns on the market return are performed.

In the simulations, the path for $\{b_t^i\}$ and $\{x_t\}$ are fixed. The only variables that are randomly generated are the paths of idiosyncratic risk, $\{\varepsilon_t^i\}$. The calibration and data description is in the appendix, in Section A.7. In the case when neither b_t nor B are observable to the investor, Kalman filtering equations discussed in Section 3.2 are used to compute investors' expectations. For each time t, the Kalman filter updates in expectations can be computed. In particular, equations (3) and (4) can be rewritten as:

$$\begin{aligned} b_{t+1|t}^{ie} &= B_{t-1}^{ie} + F^i b_{t|t-1}^{ie} + k_t^i \left(\varepsilon_t^i + \left(b_{t+1} - b_{t|t-1}^{ie} \right) x_t \right) \\ B_t^{ie} &= B_{t-1}^{ie} + K_t^i \left(\varepsilon_t^i + \left(b_{t+1} - b_{t|t-1}^{ie} \right) x_t \right) \end{aligned}$$

The updating of investor's beliefs from period $t-1$ to t is solely a function of the realizations of $(\varepsilon_t^i, b_{t+1}^i, x_t)$, and the parameters. In part A.1 of the appendix , the expression for k_t^i and K_t^i are given, and they depend on the same variables and parameters. Given a path of $\{b_t^i\}$, beliefs are fully determined. Once beliefs are determined, portfolio returns are computed from equation (6). These simulated returns are then regressed on the market return, and summary statistics are reported in Table 1.

Results from two different types of regressions are reported in this table. The first type of regressions are for 10-year rolling windows, and results are reported in the columns denoted (1) and (2). The second type of regressions are over the period 1963:7-2001:12. Each table reports summary statistics from 500 repetitions. Average alphas are estimated from the simulated data under different assumptions about the observability of b and B. The values for F and B are the assumed values for the time-series process of b, in equation (2).

There are a number of comparative static results in the case that B is observable. First, lower values of F lead to a higher fraction of the value premium that can be accounted for in the 1963-2001 sample, holding B fix. This can be seen in the columns denoted (3). The intuition is that lower F implies a less persistent process for b, leading systematic risk to have a stronger tendency to revert back to the long-run mean. Thus, the lower F, and the higher B, the higher the fraction of the value premium that is explained.

When B is unobservable, the values for B and $B/(1-F)$ correspond to initial values of the Kalman filter. It is clear from looking at the columns denoted (3) that these initial conditions do not matter much for the 1963-2001 sample as long as $F < 1$. Priors about B do not have a very persistent effect, unless the process has a unit root, i.e. $F = 1$. This is true regardless of whether b is observable or not.

A key insight of the simulations comes from looking at the effect of learning about the long-run mean. The unobservability of B leads to measures of α that are similar to assuming that B is observable and large. Learning about B makes the learning process slower as long as $F < 1$.

The effect of learning about the long-run mean is even more pronounced in the case of unobservable b. Take the case when $F = .97$. In this case, column (3) gives a fraction of the value premium explained in the 1963-2001 sample of .24 and .62 when B is assumed to be 1.2 and 1.5 respectively. In the case that B and b are unobservable and $F = .97$, the corresponding values from column (3) are .55 and .57. This shows that the prior about B does not matter much, and it shows that the unobservability of B pushes the fraction of the value premium that can be explained close to the level where $B = 1.5$ is assumed. In the case that $F = .97$, learning does drive the results. In particular, comparing the values of column (3) horizontally for the case $F = .97$ and $B/(1-F) = 1.2$ shows that the fraction of the value premium that the model can explain goes from 8% to 55%. Most of the increase in explanatory power comes from the interaction in learning about b and B.

For the case that $F = .99$, the increase in explanatory power comes from the assumption that b is unobservable. This is because mean reversion of b is the less important, the closer F is to 1. Finally, it might be worthwhile pointing out that the case $\{F = 1, B = .01, b$ unobserved and B observed$\}$ has high explanatory power (the fraction of the premium explained in the 1963:7-2001:12 sample is .81), but this is due to an unrealistic assumption: investors belief in this case that beta is increasing on average, even though it is decreasing in reality.

Our preferred estimates are the ones when $F = .97$ or $F = .99$ and both b and B are unobserved. In these cases, initial conditions do not matter much, and 29% or $55\% - 58\%$ of the value premium can respectively be explained. Furthermore,in the columns denoted (1) and (2) show that the α in the simulated data is significantly different from 0 in $92\% - 100\%$ of the cases when the α in the real data is significant.[7]

As a conclusion to this section, the simulations demonstrate that the model can account for a large fraction of the value premium when both b and B are unobservable. Learning about the long-run mean gives weight to realizations of b that are far back in the past, and thus lead to expectations about the level of systematic risk that are relatively high.

We conjecture that our results are robust to set-ups that include more general stochastic processes for b, as long as investors have to learn about the long-run properties of the process of b. This is really the key lesson of the simulation exercise: a large fraction of the value premium can be explained as long as observations about b from the past matter a lot. In a set-up where the long-run mean can switch between states, our result would go through as long as investors have to estimate what those states are. The implication that b is mean-reverting in cases when $F < 1$ is sensible, as it is hard to believe that factor loadings will diverge to infinity with certainty, as would be implied by a nonstationary process.

[7] As a comparison, we have also computed the corresponding values of columns $(1) - (3)$ in the case that investors learn using 5-year rolling window regressions. The simulated values we obtain are 18%, .20 and .06 for the three columns. This low explanatory power of the 5-year rolling window regression occurs because not enough information from the past is taken into account. On the contrary, when investors learn about the long-run mean, their expectations are affected by realizations of b that lie far back in the past.

4 Testing the Learning CAPM

4.1 The empirical predictions of the Learning CAPM

The model presented in Section 3 suggests that CAPM holds under investors' probability distribution. Hence, in order to test CAPM in a consistent way, the econometrician needs to fulfill two requirements. First, the information set that is used to compute betas has to be a subset of investors' information set. Second, all the available information has to be used in an 'optimal' way, meaning that the econometrician has to replicate the filtering process that investors undertake. We label the empirical model that follows from these prescriptions Learning CAPM (LCAPM).

Our argument can be made formally by referring to proposition 5. If the econometrician only uses information up to time t, her information set is a subset of the investors' one: $\Im_t^E \subseteq \Im_t$[8]. Moreover, if the econometrician 'optimally' uses this information to compute betas, the conditional expectation is the proper operator to express her forecast of factor loadings. Under these two conditions, the law of iterated expectations can be applied to proposition 5, and prove that mispricing relative to CAPM is zero also from the econometrician's point of view[9]:

$$\begin{aligned} \hat{\alpha}_t^E &= E\left[b_{t+1|t}^{ei} - b_{t+1}^i | \Im_t^E\right] (1+G_t)\,\sigma_x^2 A \\ &= E\left[b_{t+1|t}^{ei} - E\left[b_{t+1}^i | \Im_t\right] | \Im_t^E\right] (1+G_t)\,\sigma_x^2 A \\ &= 0 \end{aligned}$$

Among the implications of this result is the fact that CAPM is not expected to hold when betas are computed at time t with information not yet available to investors ($\Im_t^E \nsubseteq \Im_t$). This is particularly true if betas are time-varying. Let us abstract for a moment from the fact that

[8]It is a subset because the econometrician only observes returns, while investors' information set also includes dividends. We could expand the econometrician's information set to dividends, but while not altering the conclusions of our argument, it would unnecessarily force us to move from the space of returns to the space of dividends, which is less preferred in empirical tests.

[9]A further assumption subsumed by this result is that the econometrician has in mind the same process for the factor loadings as the one used by investors in their filtering mechanism.

investors have to learn about factor loadings , and focus only on the variability of beta over time. Suppose, for example, that the true underlying beta is decreasing from t to $t+T$, as in the case of value stocks. If the econometrician assumes it is constant, and beta is estimated using the whole interval of data, a biased estimate of the underlying β_τ for $\tau \in [t; t+T]$ is obtained. In particular, the estimate is biased downward for values of τ close to t, and biased upward for τ close to $t+T$. The argument extends for any variability of the underlying factor loading. As a consequence of estimating β_τ with a bias, the econometrician introduces measurement error in the right-hand side variable, and the tests of the CAPM lose power.

This problem is the more serious, the longer the interval of time over which a constant beta is estimated, and the larger the variability of the underlying factor loading. Therefore, standard cross-sectional tests of CAPM, that assume that beta is constant over almost forty years, and that use portfolios sorted on B/M and size, are biased against CAPM. The evidence in Franzoni (2002) that betas for B/M and size portfolios have been changing dramatically over the past six decades exacerbates the bias. The time-series tests are equally subject to error, as they also assume that beta is constant.

The implications of time-variation in betas for CAPM tests, which are evident from our model, are not new to the literature. Time variation in the underlying factor loadings has been taken into account in a number of ways. One alternative is the Fama-MacBeth methodology (1973), which provides more flexibility to the cross-sectional tests by shortening the window over which beta is estimated. In this approach beta is usually computed over the prior five years of data (Fama and French, 1992). A more elaborate alternative is to model the variation of conditional moments as an autoregressive-conditional heteroskedasticity in the mean (ARCH-M) model (Engle, Lilien, and Robbins, 1987). Another alternative is the conditional CAPM methodology, which uses state variables that are available to investors to predict beta in each period. One of the first studies in this strand of literature was Harvey (1989), and one of the latest examples is Lettau and Ludvigson (2001). More recently, Ang and Chen (2002) have proposed a bootstrapping methodology to adjust confidence intervals in order to account for time-variation in beta.

The original contribution of our model, however, is to point out that not only betas are time-varying, but they are also unobserved. This fact implies that investors have to learn about betas using the available information in an 'optimal' way. The simulation results in Section 3.6 show that the assumption of learning, combined with a decreasing factor loading, is able to account for a large part of the estimated value premium. As investors tie their expectations to the high past realizations of the loading, they require an expected return that is not justified by the betas that are measured ex-post, after the decrease has occurred.

While the empirical methodologies described above may account for time-variation in realized betas, they fail to capture investors' expectations of factor loadings, which is what matters for pricing, and for the tests of CAPM. Even the conditional CAPM approach, although including variables that are available to investors at time t to forecast beta at time $t+1$, does not include the history of betas among the conditioning variables. In this way it disregards information that has a relevant impact on today's expectation of the loading.

We propose a new empirical methodology for testing CAPM that fully takes into account the predictions of our model. Since CAPM holds from investors' point of view, the econometrician needs to estimate betas by replicating the filtering process, which investors undertake. In our approach the betas that explain the cross-section of returns at time $t+1$ are the result of Kalman filtering returns up to time t.

Therefore the first step in implementing a test of a Learning CAPM consists of obtaining estimates of the factor loading from the Kalman filter of returns.

4.2 Kalman filtering returns

When the econometrician's information set is a subset of investors' information, using equation (7) we can rewrite returns in a way that her expectation of the error term is zero:

$$R_{t+1}^{i} = b_{t+1}^{i} R_{t+1}^{m} + \eta_{t+1}^{i}$$

where

$$\eta_{t+1}^{i} = \left(1 + g_{t+1}^{i}\right)\left(x_{t+1}\sigma_x^2 A\left(b_{t+1|t}^{ei} - b_{t+1}^{i}\right) + \varepsilon_{t+1}^{i}\right)$$

and

$$g_{t+1} = \left(1 + r - F^{i}\right)^{-1}\sigma_x^2 A\left(K_{t+1}^{ib} + K_{t+1}^{iB}/r\right)$$

The result that $E\left[\eta_{t+1}^{i}|\Im_t\right] = 0$ follows from the properties of the Kalman filter. From here, by the law of iterated expectations, we prove that $E\left[\eta^i_{t+1}|\Im^E_t\right] = 0$. Furthermore, $E\left[\eta^i_{t+1}|\Im_t, x_{t+1}\right] = E\left[\eta^i_{t+1}|\Im^E_t, x_{t+1}\right] = 0$.

So, the correct specification of the state space system from the econometrician's point of view is very simple:

$$\begin{aligned} Observable &: R^i_{t+1} = \beta_{t+1} R^m_{t+1} + \eta^i_{t+1} \\ State\ Eq\ 1 &: b^{ie}_{t+1} = B^i + F^i b^i_t + u^i_{t+1} \\ State\ Eq\ 2 &: B^i \ \ constant \end{aligned} \tag{10}$$

These equations imply that the econometrician applies the Kalman filter directly to realized returns.

In practice, we use the Kalman filtering procedure for a model with time-varying coefficients that is exposed in Chapter 13 of Hamilton (1994). For more details on the filtering equations, we refere to Section A.1 of the appendix[10].

This methodology allows us to derive a filtered beta series for each of the ten B/M decile portfolios in the 1931:7-2001:12 sample. We use the data points in the 1926:7-1931:6 interval to compute OLS estimates that provide initial conditions for β and the forecast error on β in the Kalman filter. However, this choice of initial conditions is not crucial for our results.

Note that one limitation of this approach is that it is not possible to have investors, or the econometrician, learn about the autocorrelation F^i of factor loadings using the Kalman filter. The reason is that the Kalman filter is linear, except for the conditioning variable x_t. Learning about both the current state of the factor loading, and the autocorrelation F^i is inherently nonlinear.

However, we can directly estimate the parameters of the state equations in (10). In particular, the autocorrelation F^i, and the variance-covariance matrix of error terms is estimated using maximum likelihood on the whole history of portfolio and market returns. One of the identifying assumptions of our model is that innovations to u^i_t are uncorrelated with innovations to η^i_t, for each stock, and error terms are uncorrelated across stocks. We impose

[10]The equations in the appendix refer to the filter for investors' problem. To obtain the filtering system that is used in the empirical analysis, one just needs to replace D^i_{t+1} with R^i_{t+1}, and x_{t+1} with R^m_{t+1} in the Observable equation.

that the autocorrelation of factor loadings $F^i = F \; \forall i$[11]. The ML-estimate of factor loadings from the Kalman filter on the ten B/M portfolios is around .97 monthly.

The simulations in Section 3.6 show that a value for F of .97 delivers expectations of beta that do not adjust too quickly to new realizations. So, expected returns today are still affected by the history of beta. In particular, Table 1 tells us that in the simulated data when investors assume F to be equal to .97, we can account for almost 60% of the value premium.

The graphs in Figure 1 provide a visual impression of the speed of adjustment of the filtered beta to new information. Each graph reports the series of beta coming out of the Kalman filter (with $F = .97$) for B/M deciles 1, 4, 7, and 10 respectively, along with the series of beta resulting from five-year rolling window regressions. It is evident how the filtered series gives more weight to past information. In particular, when the rolling window beta is increasing, as in the case of lower B/M deciles, the filtered series is below the estimated series. On the other hand, for higher B/M deciles the filtered beta lies above the estimated one, because there is learning about a decreasing series. For example, in the case of the 10th B/M decile the beta estimated on the last five years of data (1:1997-12:2001) is .51, while the filtered beta in December 2001 is 1.03, more than twice as much[12]. More information on the behavior of the filtered beta series is provided in Tables 2 and 3. Again, notice that for the higher B/M deciles, the filtered betas tend to be above the OLS estimates, as the filter gives more weight than OLS to the past, and the beta of value stocks has been decreasing. The opposite occurs for the lower deciles.

The wedge between the Kalman filtered betas and the betas from simple rolling window regressions captures the discrepancy that exists in our model between investors' expectations of the loading and the current value of the loading. The key to interpret our empirical strategy is to realize that the econometrician has to take this wedge into account, because

[11] It can be proved that this assumption has a justification on theoretical grounds using the constraint that the weighted sum of betas for a partion of the market is equal to one.

[12] The relative behavior of estimated and filtered betas for the remaining B/M deciles resembles that for the deciles presented in Figure 1. In particular, for the lower deciles the estimated series is above the filtered series, as they are both increasing; for the higher deciles the estimated beta is below the filtered beta, as the loading is decreasing.

investors' expectations of the loadings are what matters for pricing.

4.3 Empirical Implementation

Our focus for now is on the value premium. Hence, the empirical analysis of this Section uses the ten B/M decile portfolios as test assets. Davis, Fama, and French (2002) provide details on the construction of 25 B/M and size sorted portfolios that span the same sample period. Our test assets are formed in a similar fashion, the only difference is that the sorting is performed only along the B/M dimension. Tables 2 and 3 report summary statistics for these portfolios. Notice in particular that while value stocks have higher excess returns than growth stocks in each sample, their beta is lower in the second subsample. This phenomenon is at the basis of the value premium, as we will argue later.

The strategy we choose for estimating the learning CAPM resembles the cross-sectional tests of CAPM. These tests involve two stages. First, beta for each portfolio is estimated on the whole history of data with time-series regressions. Then, portfolio average returns are regressed on the estimated betas. The main implication of CAPM in his most general version is that beta is a significant predictor of the cross-section of average returns. In particular, the coefficient on beta should be equal to the market risk premium. In the Sharpe (1964) and Lintner (1965) version with risk free rate the constant in the regression should be equal to the risk free rate, or to zero if excess returns are used.

The two-stage methodology that we propose entails some important modifications of the one just described. In the first stage we obtain a series of Kalman filtered betas for each portfolio, according to the procedure that was exposed in the previous subsection, and using a value for F of 0.97.

In the second stage we relate the cross-section of average returns to the filtered betas. Unlike the standard tests, we have a different beta for each time period. So, in order to derive a cross-sectional test, we take unconditional expectations of the conditional version of CAPM. If CAPM holds conditionally, as it assumed by our model, it is the case that:

$$E_t\left[R^i_{t+1}\right] = \beta^i_t E_t\left[R^m_{t+1}\right] \tag{11}$$

where R_{t+1} denotes excess returns.

By the law of iterated expectations, taking the unconditional expectation of each side yields[13]:

$$E\left[R_{t+1}^{i}\right]=E\left[\beta_{t}^{i}R_{t+1}^{m}\right] \tag{12}$$

Equation (12) is the basis for the cross-sectional test. For each portfolio, we multiply the Kalman filtered beta series by the market return one period ahead. Then, we regress average portfolio returns on the time-series mean of these new series. In addition to the standard CAPM predictions, this methodology produces the theoretical restriction that the coefficient on the interaction between filtered betas and the market return has to equal one.

Hence, defining M_t^i as $(\hat{\beta}_t^{i,kf}\cdot R_{t+1}^m)$, where $\hat{\beta}_t^{i,kf}$ is the filtered beta series, the regression that we run in the case of LCAPM is

$$\bar{R}^{i}=g_{0}+g_{1}\bar{M}^{i}+e^{i} \tag{13}$$

where the bar above a variable denotes the time-series mean.

We compare the results of our tests with the standard cross-sectional tests of CAPM, for which the estimated regression is:

$$\bar{R}^{i}=g_{0}+g_{1}\hat{\beta}^{i,ols}+e^{i} \tag{14}$$

where $\hat{\beta}^{i\,ols}$ is the OLS estimate of portfolios beta on the whole sample.

At this point it is important to remark that the g_1 coefficients in equations (13) and (14) have different interpretations. The estimate of g_1 in the CAPM regression (14) is expected to equal the market premium. Instead, the g_1 in the LCAPM equation (13) is constrained by the theory to be equal to one, as the variable $\bar{M}^i$ already includes the market return.

From Table 4 we can assess the performance of our LCAPM against the standard CAPM tests in the 1931:7-2001:12 sample. In these data the traditional version of CAPM does fairly well in terms of explaining the cross-section of average returns, as it captures about 83% of the variation. However, the constant in the regression is significantly negative, and the estimated market premium is a preposterous 1.21% per month (14.5% annually). Comparatively, our LCAPM performs well on all fronts. The adjusted R^2 is about 90%, and the constant is not

[13]See Cochrane (2001) for a detailed discussion of the link between conditional and unconditional asset pricing equations.

significantly different from zero. Moreover, the theoretical restriction that g_1 has to equal one is not rejected by the data[14]. Figure 2 provides a graphical impression of these results. The graphs plot portfolio average returns against the predictions of each of the two models. If a model predicted 100% of the variation of average returns, all the points would fall on the 45-degree line.

It is an established fact that the performance of the CAPM varies across subsamples. Therefore, Table 5 and Figures 3 and 4 report results by subsamples.

In the early data (Panel A of Table 5) the standard CAPM does a fair job in explaining the cross-section of returns. The restriction on the constant is satisfied as well. Our LCAPM slightly outperforms it by achieving an R^2 of almost 87%.

In the second subsample (Panel B of Table 5), the rejection of CAPM by the standard tests is well known, and it constitutes the so-called 'value puzzle'. This fact can be briefly explained as follows. Although the average returns on value stocks are significantly higher than those on growth stocks, the ranking on betas is inverted: high beta for growth and low beta for value. For example, from Panel B of Table 3 we learn that the average return on the 10th B/M decile portfolio exceeds the one on the 1st B/M decile portfolio by 6.2% per annum in the 1963:7-2001:12 interval. On the other hand, the estimated beta of the 10th B/M decile portfolio in the same sample is .94, while the beta for the 1st decile portfolio is 1.10. This fact explains why in Panel B of Table 5 the coefficient on beta for the CAPM column is negative, in strong violation of the theoretical prediction[15].

In comparison to the rejection of CAPM by the standard tests, our LCAPM achieves its most striking success in the second subsample. The LCAPM can explain almost 60% of the cross-sectional variation of returns. Moreover, the constant is insignificantly different from zero at the 5% confidence level. Finally, the theoretical restriction that g_1 is equal to one is not rejected at the 99% confidence level. Consistent with these results, Figure 4 shows that for the LCAPM the dots are aligned around the 45-degree line, while for the CAPM they are very dispersed.

[14] We have preliminary evidence that our results are confirmed when we use a wild bootstrapping methodology to compute confidence intervals.

[15] The R^2 of 35% is meaningless as a summary measure of the performance of the model, as it is achieved with a negative slope coefficient.

In conclusion, we would like to provide an economic intuition for the empirical success of the LCAPM. By incorporating into the asset pricing test a proxy for investors' expectations of beta, we get around the inverse ranking between average returns and estimated betas for B/M portfolios. Because the beta of value stocks has decreased over time, it is possible that investors, who have to learn about the true level of the loading, expect it to be somewhere in between the high level of the past and today's low level. Given that a symmetric development characterizes growth stocks, it follows that the ranking in expected betas can reflect the ranking in average returns. The evidence that the Kalman filtered beta for the 10th B/M portfolio is above the corresponding series for the 1st decile portfolio over most the 1963:7-2001:12 period is supportive of this argument (see Figure 1).

The trending behavior of the betas of B/M sorted portfolios make our methodology particularly suitable to test CAPM on this set of assets. However, we believe this approach extends to other assets as well, and the next section presents some evidence in this direction.

More generally, we think our results remark the importance of modeling investors' expectations of unobservable parameters, and the necessity of incorporating them into the tests, especially in the presence of high parameter variability. This contribution is relevant for the tests of other asset pricing models as well.

5 Robustness and Extensions

There are several dimensions along which we can explore the robustness of the results presented in Section 4.

First, one might think that using a Fama-MacBeth (1973) procedure to test CAPM would account for the variability of beta, which plays a large role in the rejection of the model. In Section 4 we argued that this procedure is not enough to capture investors' expectations of the factor loading. In brief, while the Fama-MacBeth procedure accounts for variability of beta over time by estimating it with rolling windows, it does not incorporate the history of the loading, which matters for investors' expectations. Here we present some results for the ten B/M decile portfolios that corroborate our intuition.

For the modern subsample (1963:7-2001:12), we find that the time-series average of the slope coefficient from cross-sectional CAPM regressions, in which the betas estimated on the prior five years of data are the explanatory variable, is significantly negative (-0.02 percent monthly). Hence, even allowing the estimates of beta to change every month does not prevent the rejection of CAPM.

However, when we use the Kalman-filtered beta series as an explanatory variable, the time-series average of the slope coefficients is a significant 0.70 percent monthly. This estimate implies an 8% risk premium on an annual basis, which is not too far from the observed equity premium [16].

The conclusion that we draw from this analysis is that allowing for time variation in beta is not enough to account for the better performance of the Kalman-filtering methodology. In fact, the LCAPM is also successful because it provides a proxy for investors' expectations of the factor loadings.

It is interesting to assess the LCAPM vis-a-vis the Fama-French three-factor model. On this set of assets the Fama-French model naturally performs well, as the factors replicate the variation in the test portfolios. Consequently, the R^2 for this model are above 90% in all samples. However, when our $\bar{M}^i$ variable is included in the cross-sectional regressions along with the loadings on HML and SMB, the latter two are no longer significant in the whole sample, as well as in the early subsample [17]. We take this result as evidence of collinearity between the Fama-French factors and our learning based explanation.

A relevant element in the empirical implementation of the LCAMP is the choice of the value for the autoregression coefficient F, which is relevant in the filtering process. We based our choice of F (0.97) on maximum likelihood estimation performed on the whole sample. However, it is not necessarily the case that investors have in mind the same value for F, especially because, unlike the econometrician, the whole sample of data is not available to them.

Therefore, we test the robustness of our results to different specifications of F. As one might expect in the light of the simulation results, higher values for F reduce the

[16] As reported in Tables 2 and 3, the average excess market return is 8.2% annually in the 1931:7-2001:12 sample. It is 11.2% in the 1931:7-1963:6 subsample, and 5.6% in the 1963:7-2001:12 subsample.

[17] In the modern subsample HML remains significant even when we include our $\bar{M}^i$ variable.

wedge between filtered betas and estimated betas, by increasing the responsiveness to new information. On the other hand, a lower F makes updating less quick, and history matters more. For F equal to 0.98 the R^2 in the whole sample is still a remarkable 90%. In the second subsample it falls to 33%, but the coefficient on g_1 is still significantly positive, and the constant is insignificant. Therefore, even for F equal to 0.98 the LCAPM performs well. When F takes on values larger than 0.98, the LCAPM still performs well in the whole sample (R^2 of about 90%), but it has no explanatory power in the modern sample. On the other hand, F equal to 0.95 greatly magnifies the ability of the LCAPM to explain the cross-section of average returns. In the longer sample the R^2 is about 91%, while in the more recent interval it jumps up to 77%. These results highlight the role that the speed of learning plays in our explanation, and they are in line with the simulation results presented in Section 3.6.

Finally, an important extension of the results in the previous section consists of testing the LCAPM on a different set of assets. Another famous anomaly that emerged in the traditional CAPM literature besides the 'value premium' is the 'size effect' (see, e.g., Fama and French, 1992), namely the fact that small stocks pay a significantly positive premium relative the expected return predicted by CAPM. Although this anomaly disappeared after its first publication in the eighties (Banz, 1981), it represents a challenge for whoever wants to argue that CAPM is a valid pricing model. Therefore, we consider whether our learning version of CAPM can price portfolios that are sorted along the size, as well as the B/M dimension.

First, we analyze the performance of the LCAPM on 24 of the 25 portfolios constructed by Davis, Fama, and French (2000). In this set of results, we leave out the small-growth portfolio for comparability with other studies (Campbell and Vuolteenaho, 2002), and because there is evidence that the returns on this portfolio are affected by market imperfections for which our model does not account[18].

The cross-sectional tests for the whole sample are presented in Table 6, while Table 7

[18]Recent evidence by Lamont and Thaler (2001), Mitchell, Pulvino, and Stafford (2002), D'Avolio (2002), and others suggests that the equilibrium expected returns on the small-growth portfolio are heavily affected by short-sale constraints and other limits to arbitrage.

covers the 1931:7-1963:6 and 1963:7-2001:12 intervals.

As in the case of the ten B/M portfolios, the traditional CAPM does not perform poorly in the whole sample, and its R^2 is 45%. Still, our approach improves the R^2 by about 10%, while respecting the restrictions imposed by the theory.

Consistent with the evidence in Section 4, the most relevant contribution of our methodology occurs in the modern subsample (1963:7-2001:12). While the traditional version of CAPM has no explanatory power, our LCAPM can account for about 34% of the cross-sectional variation in average returns. Again, the predictions that the theory imposes on this model are not rejected by the data. The constant is insignificantly different from zero, and we do not reject the hypothesis that g_1 equals one.

The results from the whole set of 25 B/M and size sorted portfolios are less strong, but they confirm the success of the LCAPM, as it appears from Tables 8 and 9. In the long sample the R^2 for LCAPM is 33% (14% for CAPM), the constant is insignificantly different from zero at the 5% level, and g_1 is not different from one. The same occurs in the early subsample, where the R^2 is 43% (39% for CAPM). In the modern subsample, the R^2 is 7.3%, which is definitely an improvement over CAPM, for which a significantly negative slope coefficient explains the cross-section of average returns. Furthermore, the constant in the LCAPM regression is not significantly different from zero, and the constraint on g_1 is not rejected at the 95% confidence level. In this sample, g_1 is only significant with 90% confidence.

The successful extension of our results to a somewhat different set of assets suggests that the failure to account for investors' expectations of beta may be playing an important role in the rejections of CAPM.

In conclusion, we would like to point out that we do not consider our methodology as an alternative to other asset pricing models, which also perform well on this sample, such as Lettau and Ludvigson's (2001) CCAPM, or Campbell and Vuolteenaho's (2002) ICAPM. In fact, we have other versions of our model, and preliminary empirical results, that incorporate time-varying expected market returns, and state variables which track the evolution of beta. These additions have a complementary explanatory power to the one provided by investors' expectations of factor loading.

6 Discussion: learning in 20 years Vs. 5 years

The main conclusion of our empirical analysis is that a model in which CAPM is augmented to account for learning on unobservable parameters, can explain the value premium. To a certain extent, the result applies to the size effect as well.

Let us summarize our intuition for this evidence. The traditional failure of CAPM in explaining the cross-section of average returns on B/M portfolios depends on the fact that by estimating a constant beta over the whole testing period, the econometrician averages out different realization of betas. In particular, because beta has been decreasing for value stock, today's low realizations are averaged with high past realizations. This process not only causes measurement error in beta, but in a learning framework it also underplays the role that past realizations of the loading have on investors' expectations today. By using a Kalman filtered beta series in the tests, we restate the importance of historical betas for today's expected returns. As a consequence, we establish a link between the high average returns of value stocks and the high levels of beta that they experienced in the past. Said differently, by providing a proxy for investors' expectations of the factor loading, we increase the power of the asset pricing tests, and the alternatives to CAPM can be more easily rejected.

The analysis points out that the LCAPM has high explanatory power when learning occurs slowly enough, precisely for values of the autocorrelation coefficient of the factor loading below 0.99. This fact implies that the model explains the value premium, if realizations of beta from at least twenty years in the past have a non-negligible weight in today's expectations of the factor loading.

In particular, when we compute the Kalman filtered beta series with $F = .97$, the expected factor loading for the 10th B/M portfolio is 1.03 in December 2001 (see Figure 1). Perhaps more importantly, the expectation of the long run mean of the loading ($\frac{B}{1-F}$), on which investors also learn, is 1.26. In terms of betas from rolling window regressions, which provide a more up-to-date version of the underlying factor loading, one needs to go back to December 1980 to obtain an estimate as high as 1.26. So, investors' learning process is such that they believe realizations of the loadings from twenty years in the past still matter today in determining the long run level towards which betas will revert.

One objection that can be moved to our theory, is that a learning model in which investors give weight to data that are so far back in the past is not realistic. In fact, the practice in the industry is to update very frequently the expectations of the riskiness of a stock. For example, Bloomberg reports estimates for company betas that are computed using the most recent five years of data.

The way we think our results capture the real process of expectation formation, is related to the distinction between the riskiness of a single stock and the riskiness of an entire asset class. While it is the case that the assessment of the riskiness of a company is updated very frequently (e.g., using five-year rolling windows), we believe that the expectation of the risk-factor loading of an entire class of assets is formed over a longer time horizon. This can be true for two reason. First, the series of return data for portfolios is very long, unlike for individual companies, which can be very young. Hence, Bayesian investors are able to use a larger data set to form their expectations, and prefer to do so. Second, company betas are much more volatile than portfolio betas and can change along the life cycle of the firm. On the contrary, an asset class is defined around a characteristic whose riskiness is stable over the long run, or at least investors believe it is stable. Therefore, as the test portfolios are representative of entire asset classes (i.e., value Vs. growth, small Vs. large), we consider it plausible that the horizon of investors' expectations spans a few decades, rather than a few years.

Some readers might still be skeptical about the importance of learning on beta for an account of the value premium. At the same time they may acknowledge the stylized fact that betas from the past explain today's average returns. To these readers, we propose a different key to interpret the results of our work. Franzoni (2002) shows that in the case of a naive investor, who sets his expectation of beta for the second subsample (1926:7-1963:6) equal to the realized beta in the first subsample (1963:7-2000:12), 80% of the premium to value stocks (10th B/M decile) can be explained. The alternative interpretation of our paper emerges as a complement to this evidence. Within a fully rational framework, we ask the question of which assumptions one needs to make in order to explain the observed relation between past betas and current average returns. Then, in the light of our results the reader can decide whether a rational framework is a good account for the stylized fact.

7 Conclusions

The assumption that investors know the variance-covariance structure of asset returns is implausible if the parameters of the model vary over time. Dramatic changes in factor loadings, like the ones documented by Franzoni (2002) for value and small stocks, cause investors to continuously revise their expectations of the riskiness of assets. Depending on the speed of learning, and on the amount of noise in returns, these expectations can diverge significantly from the true level.

We have developed an equilibrium model of learning about risk factor loadings. In the model, CAPM holds from investors' point of view, but the econometrician observes mispricing when the expected factor loadings diverge from the actual level of systematic risk.

The simulations of the model show that in order to produce levels of mispricing that are close to the value premium, the (optimal) learning process of investors needs to be such that they put sufficient weight on past observations. This is the case when factor loadings follow a mean reverting process, and when investors do not know the long run mean of the loading.

The first condition assures that expectations of the loading are fairly stable in the face of new information, as they tend to adhere to the long run mean. The second condition causes the belief of the loading to decrease slowly, while investors learn about the long run mean as well.

We propose an approach for testing the CAPM that is consistent with the implications of the model, and we label the resulting empirical model Learning CAPM (LCAPM).

The idea behind this testing strategy is that investors' expectations of the factor loading determine equilibrium expected return. Therefore the econometrician who is testing a pricing model, needs to replicate the filtering process that investors undertake. In particular, the econometrician should not use more information than the one available to investors. Also, this information needs to be processed in an 'optimal' way.

We translate these prescriptions into practice by obtaining series of portfolio betas from Kalman filtering realized returns. Then, we use these filtered series to test the CAPM prediction that beta explains the cross-section of expected returns.

The fundamental result of the paper is that when we apply this methodology to the ten B/M decile portfolios, the learning augmented version of CAPM is not rejected by the data. This conclusion also holds in the last forty years of data, in which the standard tests systematically reject CAPM. Therefore, we conclude that learning provides an account for the value premium.

We have presented a few extensions of these results. In particular, the LCAPM also performs well in explaining the cross-section of average returns on portfolios sorted on size, as well as on B/M.

Our empirical results have the same characterizations as the simulations results. The high explanatory power of LCAPM relies on learning occurring slowly enough, namely for values of the autocorrelation coefficient of the factor loading below 0.99.

There are a few extensions of our research that we intend to pursue in the future. Maximum likelihood estimation of the autoregression parameter F suggests that its level is around 0.97. However, we cannot be sure that this is the level that investors have in mind. Hence, a direction for future research consists of letting investors and the econometrician learn about the parameters in the time-series process for beta as well. This non-parametric approach will lead us far away from the nice linearity provided by the Kalman filter.

Other extensions of our research will include testing the LCAPM on other sets of assets. The failures of the CAPM are not limited to the 'value premium' and the 'size effect', which have been addressed in this paper. In fact, in order to circumvent Daniel and Titman's (1997) critique that characteristics, rather than risk, may be explaining the cross-section of average returns, one needs to test an asset pricing model on risk-sorted portfolios.

In conclusion, we believe this work has made two important contributions to the literature on asset pricing tests. The positive contribution is that we have provided an account for the value premium within the traditional mean-variance analysis, by augmenting CAPM to incorporate investors' expectation of unobservable betas. The normative contribution is that with an intuitive model, and with some clear empirical evidence, we have remarked the importance of including investors' expectations of unobservable time-varying parameters into asset pricing tests. In the process, we have provided an innovative empirical methodology that can be easily replicated for the tests of other models. In this sense, we do not consider our research as alternative to other well performing asset pricing models, but complementary to them.

A Appendix

A.1 Derivation of the Kalman-Filter

Let us introduce the following notation:

$$\xi_t^i = \begin{pmatrix} B^i \\ b_t^i \end{pmatrix} \quad \tilde{F}^i = \begin{pmatrix} 1 & 0 \\ 1 & F^i \end{pmatrix} \quad H_t = \begin{pmatrix} 0 \\ x_t \end{pmatrix} \quad U_{t+1}^i = \begin{pmatrix} 0 \\ u_{t+1}^i \end{pmatrix}$$

Then equations (1) and (2) can be written as:

$$\begin{aligned} \xi_{t+1}^i &= \tilde{F}\xi_t^i + U_{t+1}^i \quad \forall i \\ D_t^i - \bar{D}^i &= H_t'\xi_t^i + \varepsilon_t^i \quad \forall i \end{aligned}$$

Furthermore denote the variance-covariance matrix of the forecast error as follows:

$$\Gamma_{t+1|t}^i = E\left[\left(\xi_{t+1}^i - E\left[\xi_{t+1}^i|\Im_t\right]\right)\left(\xi_{t+1}^i - E\left[\xi_{t+1}^i|\Im_t\right]\right)'|\Im_t\right]$$

With this notation, the Kalman-Filter from Hamilton (1994, chap. 13) or Liptser and Shiryaev (2000, chap. 14) can be directly applied:

$$\begin{aligned} E\left[\xi_{t+1}^i|\Im_t\right] &= \tilde{F}^i E\left[\xi_t^i|\Im_{t-1}\right] + \kappa_t^i\left(D_t^i - \bar{D}^i - H_t' E\left[\xi_t^i|\Im_{t-1}\right]\right) \\ \kappa_t^i &= \tilde{F}^i\Gamma_{t|t-1}^i H_t\left(H_t'\Gamma_{t|t-1}^i H_t + \sigma_\varepsilon^{i2}\right)^{-1} \end{aligned} \tag{15}$$

Now, using the notation introduced in Section 3.2, $b_{t+1|t}^{ie} = E\left[b_{t+1}^i|\Im_t\right]$ and $B_t^{ie} = E\left[B^i|\Im_t\right]$. Using this in 15, we obtain:

$$\begin{bmatrix} B_t^{ie} \\ b_{t+1|t}^{ie} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & F^i \end{bmatrix} \begin{bmatrix} B_{t-1}^{ie} \\ b_{t|t-1}^{ie} \end{bmatrix} + \kappa_t^i\left(D_t^i - \bar{D}^i - x_t b_{t|t-1}^{ie}\right)$$

where

$$\kappa_t^i = \begin{bmatrix} 1 & 0 \\ 1 & F^i \end{bmatrix} \Gamma_{t|t-1}^i \begin{bmatrix} 0 \\ x_t \end{bmatrix} \left(\begin{bmatrix} 0 \\ x_t \end{bmatrix}' \Gamma_{t|t-1}^i \begin{bmatrix} 0 \\ x_t \end{bmatrix} + \sigma_\varepsilon^{i2}\right)^{-1}$$

Or, writing each updating equation separately:

$$B_t^{ie} = B_{t-1}^{ie} + K_t^i\left(D_t^i - \bar{D}^i - x_t b_{t|t-1}^{ie}\right) \tag{16}$$

$$b_{t+1|t}^{e} = B_{t-1}^{ie} + F^i b_{t|t-1}^{ie} + k_t^i\left(D_t^i - \bar{D}^i - x_t b_{t|t-1}^{ie}\right) \tag{17}$$

where

$$K_t^i = \frac{\gamma_{t|t-1}^{i(1,2)}}{\gamma_{t|t-1}^{i(2,2)} x_t^2 + \sigma_\varepsilon^{i2}} x_t \qquad k_{t+1}^i = \frac{\gamma_{t|t-1}^{i(1,2)} + \gamma_{t|t-1}^{i(2,2)} F^i}{\gamma_{t|t-1}^{i(2,2)} x_t^2 + \sigma_\varepsilon^{i2}} x_t$$

and $\gamma_{t|t-1}^{i(q,r)}$ is are the (q-th, r-th) element of the matrix $\Gamma_{t|t-1}^i$ and: $\kappa_t^i = \begin{pmatrix} K_t^i & k_t^i \end{pmatrix}'$. From Hamilton (1994), we find the evolution of the forecast error:

$$\Gamma_{t+1|t}^i = \left(\tilde{F}^i - \kappa_t^i H_t^{i\prime}\right) \Gamma_{t|t-1}^i \left(\tilde{F}^{i\prime} - H_t^i \kappa_t^{i\prime}\right) + \sigma_\varepsilon^2 \kappa_t^i \kappa_t^{i\prime} + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}$$

A.2 Proof of Proposition 1:

Before the proof of Proposition 1, we need the following Lemma:

Lemma 1 *Under the distributional assumptions of sections 2.1.-2.3.,*

$$c_{t+1} \to \bar{D}^m + x_{t+1} + y_{t+1} \quad \text{as } N \uparrow \infty$$

where $\bar{D}^m = \sum_{i=1}^N \bar{\mathbf{a}}^i \bar{D}^i$

Proof. Lemma 1

Solving the process for b recursively gives:

$$b_t^i = \left(F^i\right)^t b_0^i + \sum_{s=1}^t \left[\left(F^i\right)^{s-1} \left(B^i + u_{t-s+1}^i\right)\right]$$

Let us first consider the case $F^i = 1 \;\; \forall i$. Taking the market-weighted sum gives:

$$\sum_{i=1}^N \bar{\mathbf{a}}^i b_t^i = \sum_{i=1}^N \bar{\mathbf{a}}^i b_0^i + \sum_{s=1}^t \left[\left(\sum_{i=1}^N \bar{\mathbf{a}}^i B^i + \sum_{i=1}^N \bar{\mathbf{a}}^i u_{t-s+1}^i\right)\right]$$

It is assumed that $\bar{\mathbf{a}}^i = O\left(N^{-1}\right)$, which implies that we can define $\tilde{\mathbf{a}}^i = \bar{\mathbf{a}}^i N$ with the property that

$$\sum_{i=1}^N \tilde{\mathbf{a}}^i < C < \infty \;\; \text{as } N \uparrow \infty$$

for some constant C. Furthermore, it has been assumed that the distribution of shares outstanding, $\bar{\mathbf{a}}^i$, is independent of the shocks ε_t^i, u_t^i, and x_t for all t and all i. The cross-sectional expectation of $\bar{\mathbf{a}}^i$ is finite: $E[\tilde{\mathbf{a}}^i] < \infty$. Initially, $\sum_{i=1}^N \bar{\mathbf{a}}^i b_0^i = 1$ and $\sum_{i=1}^N \tilde{\mathbf{a}}^i B^i = 0$ when $F^i = 1$. Using these assumptions, we get:

$$\begin{aligned}\sum_{i=1}^N \bar{\mathbf{a}}^i b_t^i &= 1 + \sum_{s=1}^t \left[\left(\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{a}}^i B^i + \frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{a}}^i u_{t-s+1}^i \right) \right] \\ &\overset{N\uparrow\infty}{\longrightarrow} 1 + \sum_{s=1}^t E\left[\tilde{\mathbf{a}}^i u_{t-s+1}^i\right] = 1 + \sum_{s=1}^t E\left[\tilde{\mathbf{a}}^i\right] E\left[u_{t-s+1}^i\right] = 1\end{aligned}$$

The case when $F < 1$ is similar. Recall from Section 2.1. that it is assumed that there exists a transformation of B^i, such that $\tilde{B}^i(1 - F^i) = B^i$, and $\sum_{i=1}^N \bar{\mathbf{a}}^i \tilde{B}^i = 1$. We can therefore write:

$$\begin{aligned}b_t^i &= \left(F^i\right)^t b_0^i + \sum_{s=1}^t \left[\left(F^i\right)^{s-1} \left(\tilde{B}^i \left(1 - F^i\right) + u_{t-s+1}^i \right) \right] \\ &= \tilde{B}^i + \left(F^i\right)^t \left(b_0^i - \tilde{B}^i \right) + \sum_{s=1}^t \left(F^i\right)^{s-1} u_{t-s+1}^i\end{aligned}$$

Recall that it is assumed that the economy is in the steady state at time 0, so that $b_0^i = \tilde{B}^i$. Taking the market weighted sum gives:

$$\sum_{i=1}^N \bar{\mathbf{a}}^i b_t^i = \sum_{i=1}^N \bar{\mathbf{a}}^i \tilde{B}^i + \sum_{s=1}^t \left(\sum_{i=1}^N \bar{\mathbf{a}}^i \left(F^i\right)^{s-1} u_{t-s+1}^i \right) \overset{N\uparrow\infty}{\longrightarrow} 1$$

Therefore we conclude that: $\sum_{i=1}^N \bar{\mathbf{a}}^i b_t^i \to 1 \quad \text{as } N \uparrow \infty \ \ \forall t.$

From this, it follows immediately that the market dividend, D_{t+1}^m is:

$$D_{t+1}^m = \sum_{i=0}^N \bar{\mathbf{a}}^i D_{t+1}^i = \sum_{i=0}^N \bar{\mathbf{a}}^i \left(\bar{D}^i + b_t^i x_t + \varepsilon_t^i \right) \overset{N\uparrow\infty}{\longrightarrow} \bar{D}^m + x_{t+1}$$

as $\sum_{i=0}^N \bar{\mathbf{a}}^i \varepsilon_t^i = \frac{1}{N} \sum_{i=0}^N \tilde{\mathbf{a}}^i \varepsilon_t^i \to E[\tilde{\mathbf{a}}^i \varepsilon_t^i] = E[\tilde{\mathbf{a}}^i] E[\varepsilon_t^i] = 0$.

Taking all these results together shows that total consumption is:

$$\begin{aligned}c_{t+1} &= y_{t+1} + \sum_{i=0}^N \bar{\mathbf{a}}^i D_{t+1}^i \\ &= y_{t+1} + \sum_{i=0}^N \bar{\mathbf{a}}^i \bar{D}^i + \sum_{i=0}^N \bar{\mathbf{a}}^i b_t^i x_t + \sum_{i=0}^N \bar{\mathbf{a}}^i \varepsilon_t^i \\ &\to y_{t+1} + \bar{D}^m + x_{t+1}\end{aligned}$$

■

With this lemma, the pricing function from proposition 1 can be derived:

Proof. Proposition 1:

The FOC for investor's optimization problem is:

$$E\left[e^{-Ac_{t+1}}\left(D_{t+1}+P_{t+1}-(1+r)P_t\right)|\Im_t\right]=0$$

Replacing for total consumption by application of Lemma 1 gives:

$$(1+r)P_t^iE\left[e^{-Ax_{t+1}-A\left(\bar{D}^m+y_{t+1}\right)}|\Im_t\right]=E\left[e^{-Ax_{t+1}-A\left(\bar{D}^m+y_{t+1}\right)}\left(D_{t+1}^i+P_{t+1}^i\right)|\Im_t\right]$$

As labor income y and $\bar{D}^m$ are independent of x, and both the dividend process and the variables in the pricing function are independent of labor income, this reduces to:

$$(1+r)P_t^iE\left[e^{-Ax_{t+1}}|\Im_t\right]=E\left[e^{-Ax_{t+1}}\left(D_{t+1}^i+P_{t+1}^i\right)|\Im_t\right]$$

A linear pricing function is:

$$P_t^i=\nu\bar{D}^i+\omega^ib_{t+1|t}^{ie}+\upsilon^iB_t^{ie}$$

Replacing for the guess of the pricing function gives:

$$E\left[e^{-Ax_{t+1}}\left(x_{t+1}b_{t+1}^i+\varepsilon_{t+1}^i+\bar{D}^i\left(1+\nu^i\right)+\omega^ib_{t+2|t+1}^{ei}+\upsilon^iB_{t+1}^{ei}\right)|\Im_t\right]=(1+r)P_t^iE\left[e^{-Ax_{t+1}}|\Im_t\right]$$

By the Law of Iterated Expectations and the fact that b_{t+1}^i and x_{t+1} are independent, it follows that:

$$E\left[e^{-Ax_{t+1}}x_{t+1}b_{t+1}^i|\Im_t\right]=E\left[e^{-Ax_{t+1}}x_{t+1}\right]E\left[b_{t+1}^i|\Im_t\right]=E\left[e^{-Ax_{t+1}}x_{t+1}|\Im_t\right]b_{t+1|t}^{ie}$$

Using this together with $E\left[\varepsilon_{t+1}^i|\Im_t,x_{t+1}\right]=0$, the pricing equation reduces to:

$$\begin{aligned}&(1+r)P_t^iE\left[e^{-Ax_{t+1}}|\Im_t\right]\\=\;&E\left[e^{-Ax_{t+1}}\left(x_{t+1}b_{t+1|t}^{ie}+\bar{D}^i\left(1+\nu^i\right)+\omega^ib_{t+2|t+1}^{ie}+\upsilon B_{t+1}^{ie}\right)\Im_t\right]\end{aligned}$$

Using the updating equations (16) and (17), we can replace for $E\left[B^{ie}_{t+1}|\Im_t\right] = B^{ie}_t$ and $E\left[b^{ie}_{t+2|t+1}|\Im_t\right] = B^{ie}_t + F^i b^{ie}_{t+1|t}$:

$$
\begin{aligned}
& E\left[e^{-Ax_{t+1}}\left(x_{t+1}b^{ie}_{t+1|t} + \bar{D}^i\left(1+\nu^i\right) + \omega^i\left(B^{ie}_t + F^i b^{ie}_{t+1|t}\right) + \upsilon^i B^{ie}_t\right)|\Im_t\right] \\
= \; & \left(1+r\right) P^i_t E\left[e^{-Ax_{t+1}}|\Im_t\right]
\end{aligned}
$$

Now we use again the iterated expectations, and the independence between b_{t+1} and x_{t+1}, this equation simplifies:

$$
\begin{aligned}
& \left\{\bar{D}^i\left(1+\nu^i\right) + \omega^i\left(B^{ie}_t + F^i b^{ie}_{t+1|t}\right) + \upsilon^i B^{ie}_t\right\} E\left[e^{-Ax_{t+1}}|\Im_t\right] + b^{ie}_{t+1|t} E\left[e^{-Ax_{t+1}}x_{t+1}|\Im_t\right] \\
= \; & \left(1+r\right) P^i_t E\left[e^{-Ax_{t+1}}|\Im_t\right]
\end{aligned}
$$

Computing the expectations shows:

$$E\left[e^{-Ax_{t+1}}|\Im_t\right] = e^{\frac{1}{2}A^2\sigma_x^2}$$

and

$$
\begin{aligned}
E\left[e^{-Ax_{t+1}}x_{t+1}|\Im_t\right] &= \frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{\infty} x_{t+1}\exp -Ax_{t+1}\exp\left(-\frac{x_{t+1}^2}{2\sigma_x^2}\right)dx_{t+1} \\
&= e^{\frac{1}{2}A^2\sigma_x^2}\frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{\infty} x_{t+1}\exp\left(-\frac{1}{2\sigma_x^2}\left(x_{t+1}^2 + 2\sigma_x^2 A x_{t+1} + \sigma_x^2 A^2\right)\right)dx_{t+1} \\
&= -e^{\frac{1}{2}A^2\sigma_x^2}\sigma_x^2 A
\end{aligned}
$$

and replacing into the pricing equation gives:

$$P^i_t = \bar{D}^i\frac{1+\nu^i}{1+r} + \frac{\left(\omega^i F^i - \sigma_x^2 A\right)}{1+r}b^{ie}_{t+1|t} + \frac{\omega^i + \upsilon^i}{1+r}B^{ie}_t$$

By the method of undetermined coefficients, we obtain three equations in three unknowns:

$$
\begin{aligned}
1+\nu^i &= \left(1+r\right)\nu^i \\
\omega^i F^i - \sigma_x^2 A &= \left(1+r\right)\omega^i \\
\omega + \upsilon^i &= \left(1+r\right)\upsilon^i
\end{aligned}
$$

Therefore the pricing function becomes:

$$P^i_t = \frac{\bar{D}^i}{r} - \frac{A\sigma^2}{1+r-F^i}\left(b^{ie}_{t+1|t} + \frac{B^{ie}_t}{r}\right)$$

■

A.3 Proof of Proposition 2

Rewriting equation (7) gives:

$$R^i_{t+1} = b^i_{t+1}\left(x_{t+1} + \sigma_x^2 A\right) + \left(1 + \frac{\left(k^i_{t+1} + K^i_{t+1}/r\right)}{1 + r - F^i}\right)\left(\sigma_x^2 A x_{t+1}\left(b^{ei}_{t+1|t} - b^i_{t+1}\right) + \varepsilon^i_{t+1}\right)$$

Summing the last term in this equation gives:

$$\frac{1}{N}\sum_{i=0}^{N}\tilde{\mathbf{a}}^i\left(b^{ei}_{t+1|t} - b^i_{t+1}\right) \longrightarrow 0 \quad \text{as} \quad N \uparrow \infty$$

This follows directly from the assumption that innovations in b^i are independent across assets. Furthermore, the assumptions about the distribution of F^i together with the properties of the Kalman filter give:

$$\begin{aligned}
\sum_{i=0}^{N}\bar{\mathbf{a}}^i R^i_{t+1} \quad &= \quad \frac{1}{N}\sum_{i=0}^{N}\tilde{\mathbf{a}}^i b^i_{t+1}x_{t+1} + \frac{1}{N}\sum_{i=0}^{N}\tilde{\mathbf{a}}^i\varepsilon^i_{t+1} \\
&\quad + \frac{1}{N}\sum_{i=0}^{N}\tilde{\mathbf{a}}^i\left(1 + \frac{\left(k^i_{t+1} + K^i_{t+1}/r\right)}{1 + r - F^i}\right)\left(\sigma_x^2 A x_{t+1}\left(b^{ei}_{t+1|t} - b^i_{t+1}\right) + \varepsilon^i_{t+1}\right) \\
&\quad \overset{N\uparrow\infty}{\longrightarrow} x_{t+1} + \sigma_x^2 A
\end{aligned}$$

where results from Lemma 1 were used.

■

A.4 Proof of Proposition 3:

Taking conditional expectations of the return equation (6) gives:

$$\begin{aligned}
E\left[R^i_{t+1}|\Im_t\right] \quad &= \quad E\left[b^i_{t+1}x_{t+1}|\Im_t\right] + E\left[\varepsilon^i_{t+1}|\Im_t\right] \\
&\quad + \frac{\sigma_x^2 A}{1 + r - F^i}\left(b^{ei}_{t+1|t} - E\left[b^{ei}_{t+2|t+1}|\Im_t\right] + \left(B^{ei}_t - E\left[B^{ei}_{t+1}|\Im_t\right]\right)/r + r b^{ei}_{t+1|t} + B^{ei}_t\right)
\end{aligned}$$

By the law of iterated expectation,

$$E\left[b^i_{t+1}x_{t+1}|\Im_t\right] = E\left[E\left(b^i_{t+1}|\Im_t, x_{t+1}\right)x_{t+1}|\Im_t\right] = b^{ie}_{t+1|t}E\left[x_{t+1}|\Im_t\right] = 0$$

Furthermore, we have that $E\left[\varepsilon^i_{t+1}|\Im_t\right] = 0$ per assumption, $E\left[B^{ei}_{t+1}|\Im_t\right] = B^{ei}_t$ from equation 16 and $E\left[b^{ei}_{t+2|t+1}|\Im_t\right] = B^{ei}_t + F^i b^{ei}_{t+1|t}$ from equation (17). Replacing back into the expected returns yields:

$$E\left[R^i_{t+1}|\Im_t\right] = \sigma^2_x A b^{ei}_{t+1|t}$$

Using proposition 2, we find

$$E\left[R^i_{t+1}|\Im_t\right] = b^{ei}_{t+1|t} E\left[R^m_{t+1}|\Im_t\right]$$

Now Let us compute the investor's beta:

$$\beta_t = \frac{Cov\left(R^i_{t+1}, R^m_{t+1}|\Im_t\right)}{Var\left(R^m_{t+1}|\Im_t\right)}$$

Taking into account the expression for market returns from proposition 2 this yields:

$$\beta_t = \frac{1}{\sigma^2_x} Cov\left(R^i_{t+1}, x_{t+1}|\Im_t\right)$$

Replacing for individual returns, using the fact that variables conditional on time t information drop out of the covariance, as they are not random conditional on information at time t, and using $E\left[\varepsilon^i_{t+1}|x_{t+1}\right] = 0$, gives:

$$\beta_t = \frac{1}{\sigma^2_x} Cov\left(b^i_{t+1}x_{t+1}, x_{t+1}|\Im_t\right) - \frac{1}{\sigma^2_x}\frac{\sigma^2_x A}{1+r-F^i} Cov\left(b^{ei}_{t+2|t+1} + B^{ei}_{t+1}/r, x_{t+1}|\Im_t\right) \tag{18}$$

Taking into account the Kalman-filtering equations (16) and (17), we find that the second term in equation (18) reduces to 0:

$$Cov\left(b^{ei}_{t+2|t+1} + B^{ei}_{t+1}/r, x_{t+1}|\Im_t\right) = 0$$

This is a consequence of the fact that innovations to factor loadings are assumed to be independent of x_t. As investor's are acting like Bayesian's, this implies that the filter for b and B is uncorrelated with x_t.

Taking this result into account, beta reduces to:

$$\begin{aligned}\beta_t &= \frac{1}{\sigma^2_x} Cov\left(b^i_{t+1}x_{t+1}, x_{t+1}|\Im_t\right) = \frac{1}{\sigma^2_x} E\left[b^i_{t+1}x^2_{t+1}|\Im_t\right] \\ &= \frac{1}{\sigma^2_x} E\left[E\left[b^i_{t+1}|\Im_t, x_{t+1}\right] x^2_{t+1}|\Im_t\right] = b^{ie}_{t+1|t}\frac{1}{\sigma^2_x} E\left[x^2_{t+1}|\Im_t\right] = b^{ie}_{t+1|t}\end{aligned}$$

which proofs the proposition.

■

A.5 Proof of Proposition 4

Recall the notation $b^{iE}_{t|t-1} = E\left[b^i_t|\Im^E_{t-1}\right]$

$$
\begin{aligned}
\hat{\beta}^E_{t-1} &= \frac{Cov\left[R^i_t, R^m_t|\Im^E_{t-1}\right]}{Var\left[R^m_t|\Im^E_{t-1}\right]} \\
&= E\left[b^i_t|\Im^E_{t-1}\right] - \frac{\theta^i}{\sigma^2_x} Cov\left[\left(b^i_t - b^{ie}_{t|t-1}\right)\left(k^i_t x_t + K^i_t x_t/r\right), x_t|\Im^E_{t-1}\right] \\
&= E\left[b^i_t|\Im^E_{t-1}\right] - \frac{\theta^i}{\sigma^2_x} E\left[b^i_t - b^{ie}_{t|t-1}|\Im^E_{t-1}\right] Cov\left[\left(k^i_t x_t + K^i_t x_t/r\right), x_t|\Im^E_{t-1}\right] \\
&= b^{iE}_{t|t-1} - \frac{\theta^i}{\sigma^2_x} E\left[b^i_t - b^{ie}_{t|t-1}|\Im^E_{t-1}\right]\left(\frac{1+r}{r}\gamma^{i(1,2)}_{t|t-1} + F^i \gamma^{i(2,2)}_{t|t-1}\right) Cov\left[\frac{x^2_t}{\gamma^{i(2,2)}_{t|t-1} x^2_t + \sigma^{i2}_\varepsilon}, x_t|\Im^E_{t-1}\right]
\end{aligned}
$$

Now note that the term $x^2_t\left(\gamma^{i(2,2)}_{t|t-1} x^2_t + \sigma^{i2}_\varepsilon\right)^{-1}$ is a symmetric function around 0. Furthermore, the distribution of x_t is symmetric around 0, as it is a normal with 0 mean. Therefore,

$$
Cov\left(\frac{x^2_t}{\gamma^{i(2,2)}_{t|t-1} x^2_t + \sigma^{i2}_\varepsilon}, x_t|\Im^E_{t-1}\right) = 0
$$

and we find that the econometrician's beta is $\hat{\beta}^E_{t-1} = b^{iE}_{t|t-1}$.

■

A.6 Proof of Proposition 5:

Taking expectations of individual returns from equation (7) gives under the econometricians information set:

$$
E\left[R^i_{t+1}|\Im^E_t\right] = (1+G_t)\, E\left[b^{ei}_{t+1|t}|\Im^E_t\right]\sigma^2_x A - G_t E\left[b^i_{t+1}|\Im^E_t\right]\sigma^2_x A
$$

where the following substitution was made:

$$
G_t = \frac{E\left[x_{t+1}\left(K^{bi}_{t+1} + K^{Bi}_{t+1}/r\right)|\Im^E_t\right]}{1 + r - F^i}
$$

Now recall from proposition 4 that the beta of the econometrician, $\hat{\beta}^E_t = E\left[b^i_{t+1}|\Im^E_t\right]$. We can therefore write:

$$
E\left[R^i_{t+1}|\Im^E_t\right] = E\left[b^{ei}_{t+1|t} - b^i_{t+1}|\Im^E_t\right](1+G_t)\,\sigma^2_x A + \hat{\beta}^E_t E\left[R^m_{t+1}|\Im^E_t\right]
$$

The proof of the proposition is concluded by recognizing that

$$\hat{\alpha}_t^E = E\left[b_{t+1|t}^{ei} - b_{t+1}^{i} | \Im_t^E\right](1 + G_t)\,\sigma_x^2 A$$

■

A.7 Calibration and Data Description

The calibration of the model is relatively straightforward, as there are few variables and parameters. We use monthly return data from July 1926 to December 2001.[19] Equation (2) shows that x_t is the market return, less the equity premium. The market return is computed as the value-weighted portfolio of the universe of stocks in CRSP. The data set also contains returns on B/M decile portfolios, formed as in Fama and French (1993). The simulations are done only for the 10th B/M decile, i.e. the value stocks that were also investigated by Franzoni (2002).

The model allows the price of the market portfolio to be normalized to one. This normalization implies that the relative market return is equal to the absolute market return, while relative portfolio returns are equal to the absolute returns times the weight of the portfolio in the market. σ_x^2 is set to equal the variance of the excess market return, which is 5.5% monthly in the data set. The constant of absolute risk aversion A is chosen to match the theoretical equity premium, $A\sigma_x^2$, with the average realized equity premium, which is 0.68% monthly. The risk free rate r equals the average realized value of 0.31%.

The variance of idiosyncratic risk σ_ε^2 represents the noise in the observation equation of the Kalman system. It is matched to the idiosyncratic risk of the 10th B/M portfolio. σ_ε^2 is set to a level such that the R^2 in a regression of the portfolio return is 70%. In the simulation, factor loadings b_t are proxied by estimates of five-year rolling window regressions with one-month increments of portfolio excess returns on the market excess return. σ_u^2 is the sample variance of fitting an $AR(1)$ on the estimated time-series of $\{b_t\}$.

A separate issue is the estimate of F, which is a crucial parameter in the learning process. Dickey-Fuller unit root tests do not reject a unit root for the time-series of betas from the rolling-window estimation. However, the rolling window procedure mechanically induces

[19] The data has been obtained from Ken French's website.

a unit root in the data, as 59 out 60 observations that are used for estimating beta are overlapping in two adjacent windows.

A more correct procedure to obtain an estimate of F is to use the Kalman filter on returns along with maximum likelihood estimation, as it is explained in Section 4. This strategy yields estimates of F that are between .94 and .97, and these estimates are statistically different from 1.

In order to understand the impact of different values of F on the learning process, we perform simulations for values of F ranging .95 to 1.

In the cases that B is observable, the drift (in cases when $F = 1$) or the level of the long-run mean of b (in cases when $F < 1$) must be calibrated. The ex-post, unconditional mean of b over the sample period is 1.2, and the average decline is -0.1 monthly. In order to study the impact of different beliefs about these parameters, simulation results for $B = \{-0.01, 0, 0.01\}$ when $F = 1$ and $B/(1-F) = \{1.5, 1.2, 1.0\}$ when $F < 1$ are reported.

The initial conditions for the variance-covariance matrix of the forecast error is estimated using the first 5 years of the estimates of betas from the first window of the sample. The initial condition for b_t^e is equal to the estimate of beta in that first window. Note that choice of the initial conditions turns out not to be crucial empirically. This fact is discussed in the text in some detail. The exact formulas for the evolution of the forecast error is given in Section A.1 of the appendix.

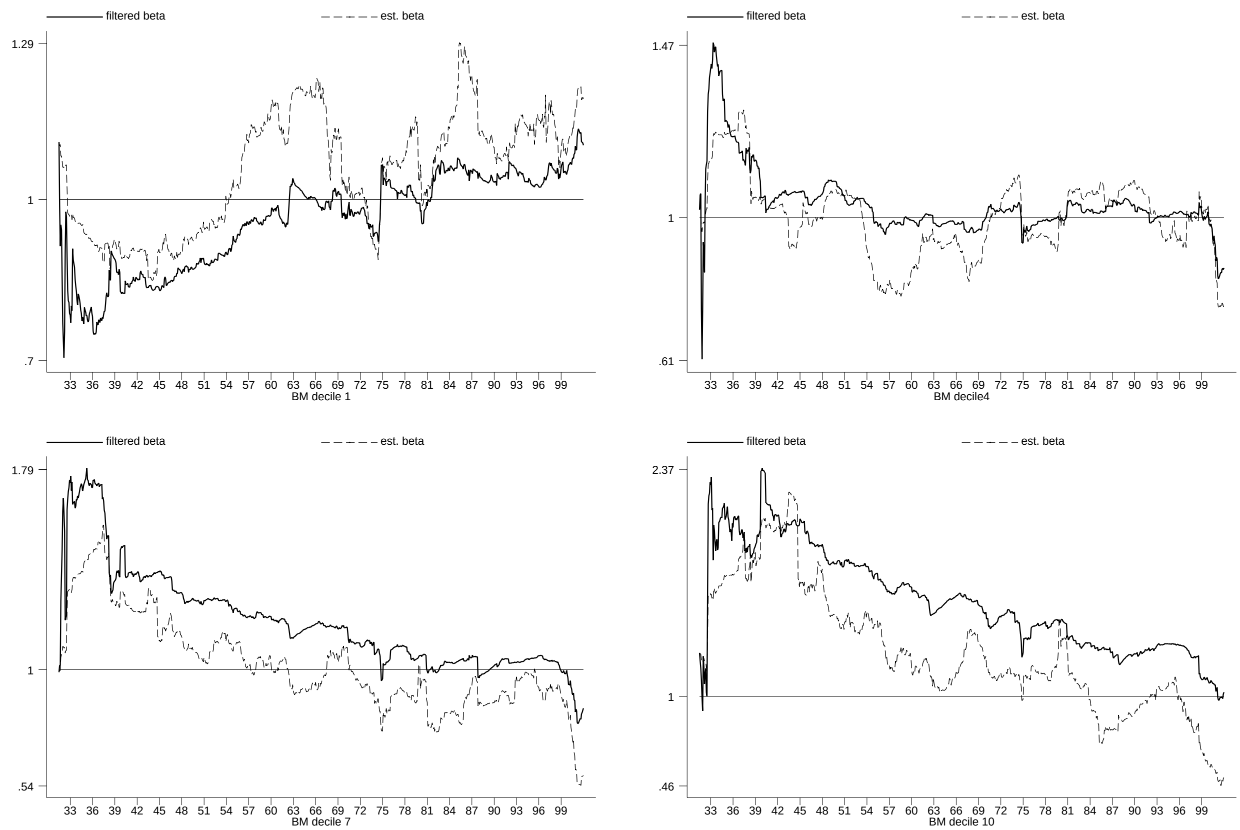

Figure 1: Estimated beta Vs. Filtered beta, by B/M deciles. Each of the graphs plots a time-series of estimated betas from five-year rolling window regression and a time-series of beta derived from Kalman filtering returns, with an autoregression coefficient $F = .97$. Time is on the horizontal axis, and the series have monthly frequency in the 1931:7-2001:12 interval. The return series used in the computation of the betas are for B/M decile portfolios. Deciles 1, 4, 7, and 10 (clockwise from the top-left graph) are reported.

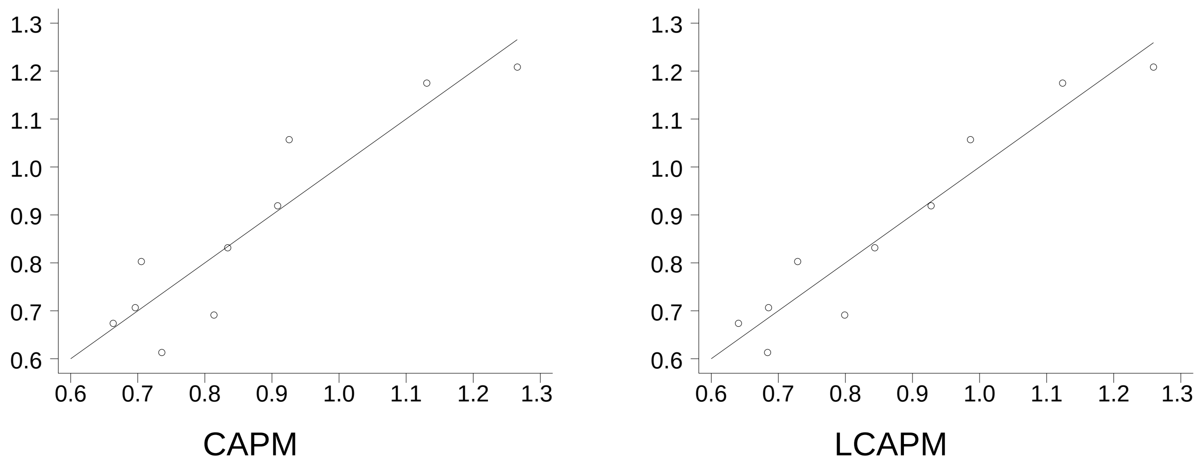

Figure 2: Performance of the CAPM and LCAPM, 10 B/M portfolios, 1931:7-2001:12. The diagram on the left corresponds to the CAPM, the one on the right to the LCAPM. The horizontal axes correpspond to the predicted average excess returns, and the vertical axes to the sample average excess returns, for the 10 B/M decile portfolios. The predicted average excess returns are based on the regressions in Table 4. The sample period is 1931:7-2001:12. The straight line is the 45-degree line.

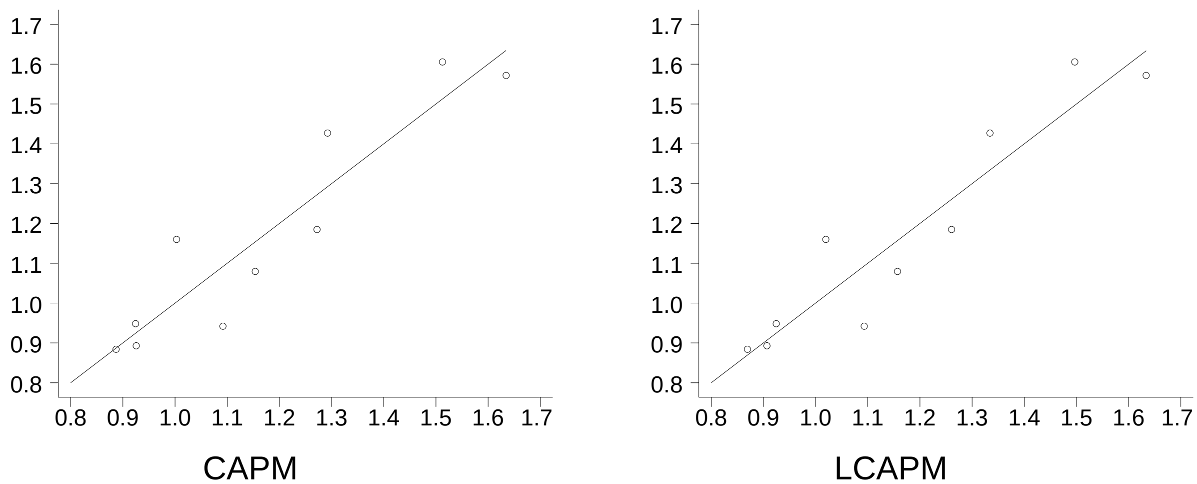

Figure 3: Performance of the CAPM and LCAPM, 10 B/M portfolios, 1931:7-1963:6. The diagram on the left corresponds to the CAPM, the one on the right to the LCAPM. The horizontal axes correpspond to the predicted average excess returns, and the vertical axes to the sample average excess returns, for the 10 B/M decile portfolios. The predicted average excess returns are based on the regressions in Table 5. The sample period is 1931:7-2001:12. The straight line is the 45-degree line.

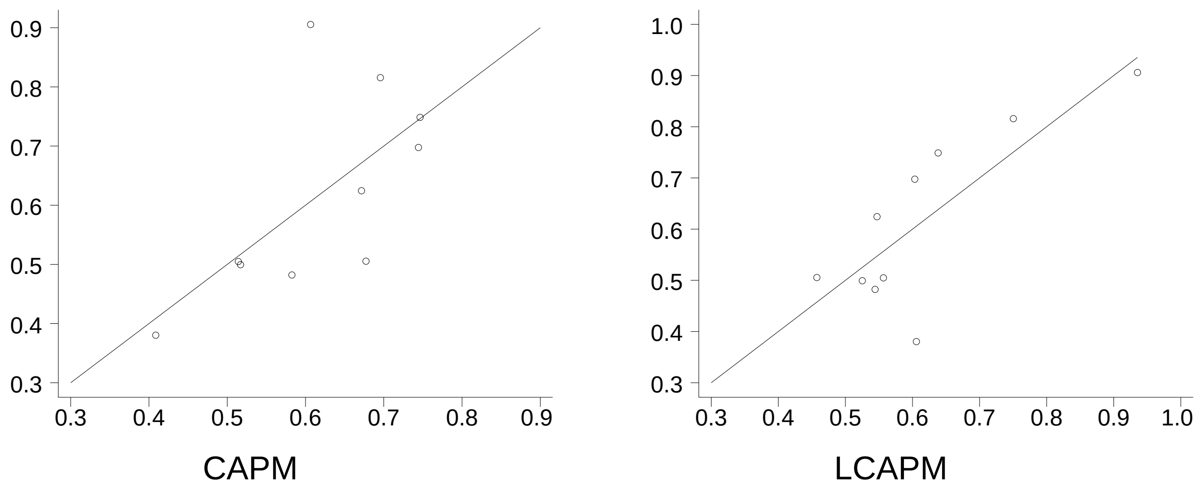

Figure 4: Performance of the CAPM and LCAPM, 10 B/M portfolios, 1963:7-2001:12. The diagram on the left corresponds to the CAPM, the one on the right to the LCAPM. The horizontal axes correpspond to the predicted average excess returns, and the vertical axes to the sample average excess returns, for the 10 B/M decile portfolios. The predicted average excess returns are based on the regressions in Table 5. The sample period is 1931:7-2001:12. The straight line is the 45-degree line.

Table 1: Simulation Results. Each cell in the table corresponds to averages accross 500 simulations. In each simulation, series of monthly returns for 1931:7 - 2001:12 are generated, investors' expectations are computed, and ex-post CAPM regressions are performed on 10-year rolling windows. The first two rows correspond to different assumptions about the observability of b and B. Different values for F correspond to different beliefs of investors. When B is unobservable, the values reported in the first column correspond to investors' priors. Columns denoted (1) report the percentage of simulated alphas that are significant in estimation windows when the alphas on real data are also sgnificant. The significance of the simulated alpha is computed using the standard deviation across repetitions. Columns denoted (2) report the average ratio of the simulated relative to the real alpha in windows when both are significant. Columns denoted (3) report the ratio of the simulated alpha to the real alpha obtained from a single regression in the 1963:7 - 2001:12 subsample. If the ratio is negative, zero is reported. In that sample, the real alpha is 0.45% monthly.

		Observable b Observable B			Observable b Unobservable B			Unobservable b Observable B			Unobservable b Unobservable B		
F	B	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	-.01	10	.18	.00	26	.14	.02	0	.00	.00	31	.14	.03
1	.00	21	.15	.01	10	.30	.02	29	.18	.09	30	.13	.07
1	.01	39	.14	.09	21	.15	.05	100	.51	.81	22	.17	.04
F	$\frac{B}{1-F}$	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
.99	1	14	.21	.01	13	.25	.02	12	.29	.00	94	.18	.29
.99	1.2	29	.15	.03	13	.20	.02	83	.14	.15	100	.21	.29
.99	1.5	47	.14	.09	11	.27	.05	100	.20	.34	92	.16	.29
.97	1	10	.28	.00	13	.23	.04	13	.29	.00	100	.37	.58
.97	1.2	44	.12	.08	49	.14	.09	87	.16	.24	100	.35	.55
.97	1.5	56	.15	.12	29	.15	.07	100	.41	.62	100	.37	.57
.95	1	10	.30	.00	65	.14	.17	12	.36	.02	100	.44	.69
.95	1.2	26	.16	.06	57	.14	.15	100	.20	.30	100	.47	.73
.95	1.5	34	.17	.15	52	.13	.12	100	.44	.70	100	.42	.69

Table 2: Summary statistics (1931:7-2001:12). The table reports summary statistics for the 10 B/M decile portfolios (constructed as in Davis, Fama, and French, 2000), for the market portfolio (CRSP value-weighted), and for the risk free rate (return on 3-month T-Bills). The relevant sample for this table is 1931:7-2001:12. For each B/M portfolio, and for the market the reported statistics are the average annualized return in excess of the risk free rate, with its standard error in parenthesis. For the risk free rate the simple annualized return is given. Also, for each decile portfolio the beta from an OLS time-series regression in the given sample is reported. For each B/M portfolio we report the average, the minimum, and the maximum of the Kalman filtered beta series in the given sample. The description of the estimation of Kalman filtered betas is given in Section 4.2.

	1931:7-2001:12				
	$\bar{R}^{e\,i}$	$\hat{\beta}^{ols}$	$\hat{\beta}^{kf}$		
			avg	min	max
dec. 1	7.17 [2.36]	1.00	.95	.70	1.13
dec. 2	8.31 [2.25]	.97	.96	.72	1.04
dec. 3	7.92 [2.20]	.94	.93	.70	1.03
dec. 4	9.05 [2.51]	1.06	1.03	.61	1.47
dec. 5	9.43 [2.32]	.97	.96	.79	1.29
dec. 6	9.75 [2.59]	1.08	1.10	.86	1.72
dec. 7	10.74 [2.80]	1.14	1.19	.78	1.79
dec. 8	12.37 [2.85]	1.15	1.24	.92	1.93
dec. 9	13.74 [3.36]	1.32	1.41	.85	2.23
dec. 10	14.15 [3.86]	1.43	1.57	.91	2.37
Mkt	8.20 [2.23]				
Rf	3.79 [0.10]				

Table 3: Summary statistics (1931:7-1963:6, and 1963:7-2001:12). The table reports summary statistics for the 10 B/M decile portfolios (constructed as in Davis, Fama, and French, 2000), for the market portfolio (CRSP value-weighted), and for the risk free rate (return on 3-month T-Bills). The relevant samples for this table are 1931:7-1963:6, and 1963:7-2001:12. For each B/M portfolio, and for the market the reported statistics are the average annualized excess return in excess of the risk free rate, with its standard error in parenthesis. For the risk free rate the simple annualized return is given. Also, for each decile portfolio the beta from an OLS time-series regression in the given sample is reported. For each B/M portfolio we report the average, the minimum, and the maximum of the Kalman filtered beta series in the given sample. The description of the estimation of Kalman filtered betas is given in Section 4.2.

	Panel A: 1931:7-1963:6					Panel B: 1963:7-2001:12				
	$\bar{R}^{e\,i}$	$\hat{\beta}^{ols}$	$\hat{\beta}^{kf}$			$\bar{R}^{e\,i}$	$\hat{\beta}^{ols}$	$\hat{\beta}^{kf}$		
			avg	min	max			avg	min	max
dec. 1	10.31 [3.78]	.94	.88	.70	1.10	4.56 [2.97]	1.10	1.02	.92	1.13
dec. 2	11.03 [3.78]	.94	.92	.72	1.04	6.05 [2.66]	1.02	1.00	.90	1.04
dec. 3	10.24 [3.72]	.89	.88	.70	1.00	5.99 [2.67]	1.02	.97	.89	1.03
dec. 4	10.78 [4.65]	1.12	1.07	.61	1.47	5.78 [2.62]	.96	1.00	.83	1.05
dec. 5	13.49 [4.19]	1.02	1.01	.82	1.29	6.06 [2.45]	.89	.93	.79	0.98
dec. 6	12.46 [4.89]	1.19	1.21	.93	1.72	7.49 [2.45]	.89	1.01	.86	1.08
dec. 7	13.60 [5.43]	1.32	1.35	.98	1.79	8.37 [2.42]	.83	1.05	.78	1.19
dec. 8	16.45 [5.56]	1.34	1.41	.95	1.93	8.98 [2.41]	.83	1.10	.82	1.25
dec. 9	18.50 [6.72]	1.58	1.65	1.11	2.23	9.79 [2.57]	.87	1.20	.85	1.39
dec. 10	18.11 [7.72]	1.72	1.84	.91	2.37	10.86 [2.98]	.94	1.35	.97	1.62
Mkt	11.27 [3.90]					5.64 [2.49]				
Rf	1.06 [0.05]					6.07 [0.11]				

Table 4: Cross-sectional Tests (10 B/M portfolios, whole sample). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the ten B/M decile portfolios. The estimation sample 1931:7-2001:12.

	1931:7-2001:12	
	CAPM	LCAPM
g_0	$-.48$	$-.04$
p-value	[.04]	[.64]
conf. int.	$(-.95; -.01)$	$(-.27; .18)$
g_1	1.21	1.12
p-value	[.00]	[.00]
conf. int.	$(.79; 1.63)$	$(.84; 1.39)$
adj. R^2	83.01%	90.70%

Table 5: Cross-sectional Tests (10 B/M porfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the ten B/M decile portfolios. The estimation samples are given in the table.

Panel A: 1931:7-1963:6		
	CAPM	LCAPM
g_0	.07	.28
p-value	[.65]	[.04]
conf. int.	$(-.28; .43)$	$(.01; .55)$
g_1	.90	.73
p-value	[.00]	[.00]
conf. int.	$(.61; 1.19)$	$(.51; .95)$
adj. R^2	84.74%	86.90%

Panel B: 1963:7-2001:12		
	CAPM	LCAPM
g_0	1.78	$-.75$
p-value	[.00]	[.07]
conf. int.	$(.67; 2.89)$	$(-1.58; .08)$
g_1	-1.24	2.74
p-value	[.04]	[.00]
conf. int.	$(-2.42; -.06)$	$(1.08; 4.40)$
adj. R^2	35.47%	59.96%

Table 6: Cross-sectional Tests (24 B/M and Size sorted portfolios, whole sample). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are 24 B/M and size sorted portfolios, i.e. the 25 portfolios of Davis, Fama, and French (2002), minus the small-growh portfolio. The estimation sample is 1931:7-2001:12.

	1931:7-2001:12	
	CAPM	LCAPM
g_0	$-.24$	$-.39$
p-value	[.40]	[.14]
conf. int.	$(-.82; .34)$	$(-.92; .14)$
g_1	1.02	1.47
p-value	[.00]	[.00]
conf. int.	$(.54; 1.50)$	$(.92; 2.03)$
adj. R^2	45.15%	55.91%

Table 7: Cross-sectional Tests (24 B/M and Size sorted portfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are 24 B/M and size sorted portfolios, i.e. the 25 portfolios of Davis, Fama, and French (2002), minus the small-growh portfolio. The estimation samples are given in the table.

Panel A: 1931:7-1963:6		
	CAPM	LCAPM
g_0	.13	$-.11$
p-value	[.52]	[.69]
conf. int.	$(-.30; .58)$	$(-.68; .45)$
g_1	.92	1.07
p-value	[.00]	[.00]
conf. int.	$(.59; 1.24)$	$(.66; 1.48)$
adj. R^2	58.88%	55.39%

Panel B: 1963:7-2001:12		
	CAPM	LCAPM
g_0	1.12	$-.59$
p-value	[.00]	[.12]
conf. int.	$(.55; 1.69)$	$(-1.36; .17)$
g_1	$-.38$	2.19
p-value	[.15]	[.00]
conf. int.	$(-.93; .15)$	$(.91; 3.46)$
adj. R^2	5.00%	33.83%

Table 8: Cross-sectional Tests (25 B/M and Size sorted portfolios, whole sample). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the 25 portfolios of Davis, Fama, and French (2002). The estimation samples is 1931:7-2001:12.

1931:7-2001:12		
	CAPM	LCAPM
g_0	.28	$-.10$
p-value	[.37]	[.73]
conf. int.	$(-.37; .95)$	$(-.75; .53)$
g_1	.57	1.15
p-value	[.03]	[.00]
conf. int.	$(.04; 1.09)$	$(.48; 1.82)$
adj. R^2	14.36	32.91

Table 9: Cross-sectional Tests (25 B/M and Size sorted portfolios, by subsamples). The table reports estimates g_1 from cross-sectional regressions of average excess returns on betas (for the CAPM), and on the time-series average of the product between the kalman-filtered beta and the excess market return (LCAPM). The regressions also include a constant (g_0). P-values and 5% confidence intervals are provided. The test assets are the 25 portfolios of Davis, Fama, and French (2002). The estimation samples are given in the table.

Panel A: 1931:7-1963:6		
	CAPM	LCAPM
g_0	.37	.05
p-value	[.13]	[.85]
conf. int.	$(-.13; .87)$	$(-.56; .67)$
g_1	.72	.94
p-value	[.00]	[.00]
conf. int.	$(.36; 1.09)$	$(.49; 1.38)$
adj. R^2	39.79	43.16

Panel B: 1963:7-2001:12		
	CAPM	LCAPM
g_0	1.26	$-.02$
p-value	[.00]	[.95]
conf. int.	$(.74; 1.79)$	$(-.91; .86)$
g_1	$-.53$	1.20
p-value	[.03]	[.10]
conf. int.	$(-1.03; -.04)$	$(-.25; 2.66)$
adj. R^2	14.64	7.32

References

[1] Ang, A., and J. Chen, 2002, "CAPM Over the Long-Run: 1926-2001", mimeo, University of Southern California and Columbia University.

[2] Banz, R., 1981, "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics 9*, pp. 3-18.

[3] Barberis, N., 2000, "Investing for the Long Run when Returns are Predictable", *Journal of Finance 55*, pp. 225-264.

[4] Basu, S., 1977, "The Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis", *Journal of Finance, 32*, pp. 663-682.

[5] Bawa, V. S., S. Brown, and R. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North Holland.

[6] Bekaert, G., and Liu, J. 2001, "Conditioning Information and Variance Bounds on Pricing Kernels", NBER Working Paper 6880.

[7] Black, F., M. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Tests", in M. Jensen ed., *Studies in the Theory of Capital Markets*, (Praeger, New York, NY).

[8] Blanchard, O. J., 1993, "Movements in the Equity Premium", *Brookings Papers on Economic Activity, 1993-2*, pp. 75-138.

[9] Bollerslev, T., R. Engle, and J. Wooldridge, 1988, "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy 96,* pp. 116-131.

[10] Brav, A., R. Lehavy, and R. Michaely, 2002, "Expected Returns and Asset Pricing", mimeo, Duke University, University of Michigan, and Cornell University.

[11] Breeden, D., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics, 7*, pp. 265-296.

[12] Brennan, M. J. and Y. Xia, 2001, “Stock Return Volatility and Equity Premium”, *Journal of Monetary Economics 47*, pp. 249-283.

[13] Campbell, J.Y., and J. Cochrane, 2000, ”Explaining the Poor Performance of Consumption-based Asset Pricing Models”, *Journal of Finance 55*, pp. 2863-2878.

[14] Campbell, J. Y., and L. Hentschel, 1992, ”No News is Good News: an Asymmetric Information Model of Changing Volatility in Stock Returns”, *Journal of Financial Economics 31*, 281-318.

[15] Campbell, J. Y., and R. Shiller, 1988, “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors”, *Review of Financial Studies 1*, pp. 195-227.

[16] Campbell, J. Y., and T. Vuolteenaho, 2002, ”Bad Beta, Good Beta”, mimeo, Harvard University.

[17] Chen, N.-F., and F. Zhang, 1998, ”Risk and Return of Value Stocks”, *Journal of Business 71*, 501-536.

[18] Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

[19] Cohen, R. B., C. Polk, and T. Vuolteenaho, 2002, ”The Value Spread”, *Journal of Finance*, forthcoming.

[20] Constantinides, G., 1990, ”Habit Formation: A Resolution of the Equity Premium Puzzle”, *Journal of Political Economy 98*, pp. 519-543.

[21] Daniel, K., and S. Titman, 1997, “Evidence on the characteristics of cross-sectional variation in stock returns”, *Journal of Finance 52*, pp. 1-33.

[22] Davis, J., E. Fama, and K. French, 2000, ”Characteristics, Covariances, and Average Returns: 1929 to 1997”, *Journal of Finance 55*, pp. 389-406.

[23] D’Avolio, G., 2002, ”The Market for Borrowing Stock”, *Journal of Financial Economics*, forthcoming.

[24] Doukas, J., C. F. Kim, and C. Pantzalis, 2002, "A test of the Errors-in-Expectations Explanation of the Value/Glamour Stock Returns Performance: Evidence from Analysts' Forecasts", *Journal of Finance 57*, pp.2143-2165.

[25] Engle, R., and M. W. Watson, 1987, "The Kalman Filter: Applications to Forecasting and Rational Expectations Models", in: *Advances in Econometrics Fifth World Congress*, Volume I, ed. T. Bewley, Cambridge University Press, pp. 245-283.

[26] Engle, R. D. Lilien, and R. Robins, 1987, "Estimating Time-varying Risk Premia in the Term Structure: the ARCH-M Model", *Econometrica 55*, pp. 391-407.

[27] Elton, E., 1999, "Expected Returns, Realized Return, and Asset Pricing Tests", *Journal of Finance 54,* pp.1199-1220.

[28] Fama, E., and K. French, 1992, "The Cross-section of Expected Stock Returns", *Journal of Finance 47*, pp. 427-465.

[29] Fama, E., and K. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics 33*, pp. 3-56.

[30] Fama, E., and K. French, 1995, "Size and Book-to-Market Factors in Earnings and Returns", *Journal of Finance 50*, pp. 131-155.

[31] Fama, E., and K. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies", *Journal of Finance 51*, pp. 55-84.

[32] Fama, E., and K. French, 2002, "The Equity Premium", *Journal of Finance 57*, pp. 637-659.

[33] Fama, E. and J. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests", *Journal of Political Economy, 81*, pp. 607-636.

[34] Franzoni, F., 2002, "Where is beta going? The Riskiness of Value and Small Stocks", Ph.D. Dissertation, MIT.

[35] Gennotte, G., 1986, "Optimal Portfolio Choice under Incomplete Information", *Journal of Finance 61*, pp. 733-749.

[36] Hansen, L. and Jagannathan, R., 1997, "Assessing Specification Errors in Stochastic Discount Factor Models" *Journal of Finance 52*, pp. 557–590.

[37] Hamilton, J. D., 1994, *Time Series Analysis*. Princeton, NJ: Princeton University Press.

[38] Harvey, C., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models", *Journal of Financial Economics 24*, pp. 289-317.

[39] Hodrick, R. and Zhang, X. 2001, "Evaluating the Specification Errors of Asset Pricing Models", *Journal of Financial Economics 62*, pp. 327-376.

[40] Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance 51*, pp. 3-53.

[41] 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*, pp. 859-874.

[42] Lakonishok, J., A. Shleifer, and R. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk", *Journal of Finance 50*, pp. 1541-1578.

[43] Lamont, O., and R. Thaler, 2001, "Can the Market add and substract? Mispricing in Tech-stock carve-outs", mimeo, GSB Chicago.

[44] Lettau, M., and S. Ludvigson, 2002, "Expected Returns and Expected Dividend growth", mimeo, NYU.

[45] Lettau, M., and S. Ludvigson, 2001, "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying", *Journal of Political Economy*, 109, pp. 1238-1287.

[46] Lewellen, J. and J. Shanken, 2002, "Learning, Asset-pricing Tests, and Market Efficiency", *Journal of Finance 57*, pp. 1113-1145.

[47] Lintner, J., 1965, "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", *Review of Economics and Statistics*, 47, pp. 13.-37.

[48] Liptser, R., and A. Shiryaev, 2000, *Statistics of Random Processes*, Vol. I & II, 2nd Edition, Berlin: Springer Verlag.

[49] Lucas, R., 1978, "Asset Prices in an Exchange Economy", *Econometrica 46*, pp.1429-45.

[50] MacKinley, A. C., 1995, "Multifactor Models do not Explain Deviations from CAPM", *Journal of Financial Economics 38*, pp. 3-28.

[51] Mehra, R. and E. Prescott, 1985, "The Equity Premium: A Puzzle", *Journal of Monetary Economics 15*, pp. 145-161.

[52] Merton R. C., 1973 "An Intertemporal Asset Pricing Model", *Econometrica 41*, pp. 867-887.

[53] Mitchell, M., T. Pulvino, and Erik Stafford, 2001, "Limited Arbitrage in Equity Markets", *Journal of Finance 57.* pp. 551-584.

[54] Moskowitz, T., 2002, "An Analysis of Covariance Risk and Pricing Anomalies", mimeo, University of Chicago.

[55] Pastor, L., 2000, "Portfolio Selection and Asset Pricing Models", *Journal of Finance 55*, pp. 179–223.

[56] Roll, R., 1977, "A critique of the Asset Pricing Theory's Tests: Part I", *Journal of Financial Economics 4*, pp.129-176.

[57] Ross, S., 1976, "The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory 13*, pp. 341-360.

[58] Santos, T., and P. Veronesi, 2001, "Labor Income and Predictable Stock Returns, mimeo, GSB Chicago.

[59] Shanken, J., 1990, "Intertemporal Asset Pricing: An Empirical Investigation", *Journal of Econometrics*, 45, pp. 99-120.

[60] Sharpe, W., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance 19*, pp. 425-442.

[61] Shiller, R. J., 1981 "Do Stock Prices Move too much to be Explained by Subsequent Changes in Dividends?", *American Economic Review 71*, pp. 421-436.

[62] Sundaresan, S., 1989, "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *Review of Financial Studies 3*, pp. 73-89.

[63] Veronesi, P., 1999, "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model", *Review of Financial Studies 12*, 975-1007.

[64] Vassalou, M., and Xing, Y., 2002, "Default Risk and Equity Returns", *Journal of Finance*, forthecoming.