Testing Asset Pricing Models with Long-Run Expected Returns

Lars A. Lochstoer and Paul C. Tetlock*
Columbia Business School

November 2015

Abstract

We introduce a framework for testing asset pricing models based on their implications for book-to-market ratios. We focus on the performance of beta pricing models, such as the Fama-French five-factor model. Our tests exploit the fact that book-to-market ratios represent expectations of long-run cash flows and stock returns. Imposing this relation and a given asset pricing model, we jointly estimate cash flow and discount rate dynamics. We estimate factors’ risk premiums from the relation between firms’ discount rates and risk exposures (betas). Our estimates of risk prices are robust to transitory fluctuations in prices that can exert substantial influence on standard tests based on the cross-section of short-term returns. Because they do not depend heavily on short-term returns, our tests are more powerful than standard tests. Our methodology can accommodate time-varying betas and risk premiums, as well as irrational cash flow expectations.

*Comments welcome, including letting us know if we have omitted related papers from our citations. Contact information: Lochstoer: 405B Uris Hall, Columbia Business School, 3022 Broadway, New York, NY 10027. E-mail: LL2609@gsb.columbia.edu. Tetlock: 811 Uris Hall, Columbia Business School, 3022 Broadway, New York, NY 10027. E-mail: paul.tetlock@columbia.edu.
1 Introduction

Standard asset pricing tests rely on short-term returns that are sensitive to transitory price fluctuations. Our contribution is to introduce a tractable framework for testing asset pricing models’ predictions of asset price levels. Our tests focus on a stationary measure of stock prices: the ratio of a stock’s book value to its market value. A stock’s book-to-market ratio represents the difference between expected long-run returns and earnings. We estimate both components of book-to-market by modeling the dynamics of firms’ earnings and imposing that discount rates come from the asset pricing model being tested. We estimate factors’ risk premiums from the relation between firms’ discount rates and risk exposures (betas).

Our framework adopts a long-run perspective on asset pricing that complements the existing paradigm based on short-term returns. Our estimates and tests are based on long-term expected returns. Short-term stock price fluctuations, such as stock price momentum, price reversal, and price pressure, have little direct impact on earnings and book-to-market ratios. Consequently, short-term returns do not materially affect our estimates of risk prices when they arise from transitory factors. Such robustness is important if the sources of short-term and long-term expected returns differ and long-term expected returns are of primary interest, as they are capital budgeting applications. Imprecision in our estimates arises primarily from imperfect forecasts of firm earnings, as opposed to volatility in realized stock returns. Because expected earnings are less volatile than returns, our method provides more precise risk premium estimates and more powerful tests of models than standard methods.

Although our testing framework is new, it relies on familiar assumptions. We employ a standard log-linear approximation to firms’ book-to-market ratios to obtain the present value equation with log earnings and log returns. We assume joint lognormality of returns and the pricing kernel to relate expected log returns to the stochastic discount factor. We assume that log earnings and beta dynamics are well-approximated by first-order autoregressive processes, allowing for temporary shocks to all processes and fixed effects in firms’ earnings. Taking these dynamics as given, we employ an efficient Generalized Method of Moments (GMM) estimator that solves a set of linear equations to infer risk premiums and fixed effects in earnings from panel data on book-to-market ratios and earnings.

Our estimates deliver two key insights. First, standard beta pricing models explain very little
of the variation in book-to-market ratios, even though many of these models explain cross-sectional differences in short-term realized returns. For example, the Fama-French five-factor model, which includes market, size, value, profitability, and investment factors, explains less than 2% of variation in book-to-market ratios. Further analysis attributes a significant portion of the residual variation in book-to-market ratios to variation in long-run discount rates arising from unknown sources. This finding should provide fertile ground for future research. Second, our estimate of the market risk premium is positive and significant in economic and statistical terms. In a test of the CAPM, the GMM estimate of the market risk premium is 5% with a standard error of roughly 1%. In contrast, a Fama-MacBeth cross-sectional regression of realized returns on (predicted) market betas yields a risk premium estimate of -2% with a standard error over 3%. This finding could help explain why 73% of chief financial officers use the CAPM to estimate the cost of equity capital (Graham and Harvey (2001)), even though the model makes inaccurate predictions of short-term returns (Fama and French (1992)).

Few papers in the literature try to explain variation in firms’ book-to-market ratios. Cohen, Polk, and Vuolteenaho (2003) apply the clean surplus relation of Ohlson (1995) and log-linear techniques to relate book-to-market ratios to future cash flows and discount rates. They analyze the dynamics of the value spread—the cross-sectional spread of book-to-market ratios—and argue that most of the spread comes from differences in expected cash flows. Lyle and Wang (2015) in recent work use the same framework to estimate the expected cash flow component of firms’ book-to-market ratios and infer the discount rate component as a residual. They show that discount rate components are volatile and forecast long-run returns. Lu (2015) adopts a similar approach to Lyle and Wang (2015) and argues that firms’ long-run returns are not strongly related to standard factor betas. In contrast to our approach, Lu’s approach assumes no correlation in the cash flow and discount rate components of book-to-market, and there is no formal link to an asset pricing model. Asness, Frazzini, and Pedersen (2014) argue that "quality" is important for the cross-section of valuation ratios. In contrast to these studies, we develop a consistent estimation framework based on a given stochastic discount factor (the null hypothesis) and model the dynamics of risk prices and factor loadings as required by theory to impose the present-value relation between prices, cash flows, and returns. Koijen and van Binsbergen (2010) and Kelly and Pruitt (2013) also use present-value relations to provide estimates of expected returns and cash flows, but do not focus
on particular model tests and long-run dynamics.

2 General model

In this section, we present our general model, introducing notation and the dynamics of relevant state variables. For analytical tractability, we apply a standard log-linearization to relate log book-to-market ratios \((bm)\) to future stock returns and earnings:

\[
bm_{i,t} = DR_{i,t}^{bm} - CF_{i,t}^{bm}
\]

\[
DR_{i,t}^{bm} = \sum_{j=1}^{\infty} \kappa^{j-1} E_t [r_{i,t+j}]
\]

\[
CF_{i,t}^{bm} = \sum_{j=1}^{\infty} \kappa^{j-1} E_t [roe_{i,t+j}]
\]

following Ohlson (1995) and Cohen, Polk, and Vuolteenaho (2003). Here \(DR_{i,t}^{bm}\) and \(CF_{i,t}^{bm}\) represent the discount rate and cash flow components of book-to-market, respectively, \(\kappa\) is a log-linearization constant, \(r_{i,t}\) is the log stock return, and \(roe_{i,t}\) is the log return on equity for firm \(i\) in time \(t\), where:

\[
roe_{i,t} \equiv \ln (1 + ROE_{i,t})
\]

\[
= \ln \left(1 + \frac{E_{i,t}}{BE_{i,t}}\right).
\]

\(E\) is earnings and \(BE\) is book equity. The book-to-market ratio increases with expected long-run discount rates and decreases with expected long-run cash flows.

Turning first to the discount rate component, \(DR_{i,t}^{bm}\), assume that returns and the pricing kernel \((M_{t+1})\) are jointly lognormal. In this case, expected excess log returns satisfy:

\[
E_t [r_{i,t+1} - r_{f,t+1}] = -\frac{1}{2} \sigma_t^2 (r_{i,t+1}) - cov_t (r_{i,t+1}, m_{t+1}),
\]

where \(m_{t+1} \equiv \ln M_{t+1}\) is the log stochastic discount factor (SDF) and \(\sigma_t^2 (r_{i,t+1})\) denotes log return
variance expected at time $t$. Applying the law of iterated expectations and substituting, we obtain:

$$
DR_{i,t}^{bm} = \sum_{j=1}^{\infty} \kappa_{t+j}^{-1} E_t \left[ E_{t+j-1} \left[ r_{t+j} \right] \right] \\
= \sum_{j=1}^{\infty} \kappa_{t+j}^{-1} E_t \left[ r_{f,t+j} - \frac{1}{2} \sigma_{t+j-1}^2 (r_{i,t+j}) - \text{cov}_{t+j-1} (r_{i,t+j}, m_{t+j}) \right]. \quad (7)
$$

We consider the following class of SDFs:

$$
m_{t+1} = -r_{f,t+1} - \frac{1}{2} b'_t \text{var}_t (f_{t+1}) b_t - b'_t (f_{t+1} - E_t [f_{t+1}]), \quad (9)
$$

where $f_{t+1}$ is a conditionally normal vector of $K$ log factor realizations, $b_t$ is a vector of $K$ coefficients that determine factor premiums, and the log risk-free rate, $r_{f,t+1}$, is known at time $t$. It follows that:

$$
DR_{i,t}^{bm} = \sum_{j=1}^{\infty} \kappa_{t+j}^{-1} E_t \left[ r_{f,t+j} - \frac{1}{2} \sigma_{t+j-1}^2 (r_{i,t+j}) + \text{cov}_{t+j-1} (r_{i,t+j}, b'_{t+j-1} f_{t+j}) \right]. \quad (10)
$$

Let $\Omega_t = \text{var}_t (f_{t+1})$ and define $\lambda'_t \equiv b'_t \Omega_t$. Then the discount rate component becomes:

$$
DR_{i,t}^{bm} = \sum_{j=1}^{\infty} \kappa_{t+j}^{-1} E_t \left[ r_{f,t+j} - \frac{1}{2} \sigma_{t+j-1}^2 (r_{i,t+j}) + \lambda'_{t+j-1} \beta_{i,t+j-1} \right] \\
= \sum_{j=1}^{\infty} \kappa_{t+j}^{-1} \left\{ E_t \left[ r_{f,t+j} - \frac{1}{2} \sigma_{t+j-1}^2 (r_{i,t+j}) \right] + \lambda'_{t+j-1} \beta_{i,t+j-1} \right\}, \quad (11)
$$

where $\beta_{i,t+j-1}$ is a vector of $K$ coefficients from regressing $r_{t+j}$ on $f_{t+j}$ and an intercept at time $t+j$. The discount rate component of the book-to-market ratio depends on the term structure of expected risk-free rates, risk prices, factor loadings, and return variances (because of the Jensen’s inequality term) and the covariance between risk prices and factor loadings.

To relate factor exposures to long-run discount rates, we specify processes for cash flows, the risk-free rate, return variance, betas, and prices of risk. We choose simple time-series processes that we later show accurately characterize these variables’ dynamics. In particular, we assume that the risk-free rate, prices of risk, betas, and return variances ($\nu_{i,t} \equiv \sigma_{t,i}^2 (r_{i,t+1})$) follow first-order...
autoregressive (AR(1)) processes:

\[ r_{f,t+1} = \tau_f + \rho_{r_f}(r_{f,t} - \tau_f) + \sigma_{r_f}^2 \varepsilon_{f,t}^r, \]  
\[ \lambda_{t+1}^{(k)} = \bar{\lambda}^{(k)} + \rho_{\lambda^{(k)}}(\lambda_t^{(k)} - \bar{\lambda}^{(k)}) + \sigma_{\lambda^{(k)}}^2 \varepsilon_{t+1}^{\lambda^{(k)}}, \quad k = 1, \ldots, K, \]  
\[ \nu_{i,t+1} = \nu + \rho_{\nu}(\nu_{i,t} - \nu) + \sigma_{\nu}^2 \varepsilon_{i,t+1}^{\nu}, \]  
\[ \beta_{i,t+1}^{(k)} = \bar{\beta}^{(k)} + \rho_{\beta^{(k)}}(\beta_{i,t}^{(k)} - \bar{\beta}^{(k)}) + \sigma_{\beta^{(k)}}^2 \varepsilon_{i,t+1}^{\beta^{(k)}}, \quad k = 1, \ldots, K. \]

Further, we assume that \( \text{roe}_{i,t} \) has a persistent component, \( x_{i,t} \), with a fixed firm-specific mean \( (\mu_i) \):

\[ \text{roe}_{i,t} = x_{i,t} + \sigma_i \eta_i \]  
\[ x_{i,t} = \mu_i + \rho_x (x_{i,t-1} - \mu_i) + \sigma_x^2 \varepsilon_{i,t}^x. \]

We assume all error terms are jointly normally distributed. Shocks can be cross-sectionally correlated, but they are i.i.d. over time.

The discount rate component is then:

\[ DR_{i,t}^{\text{bm}} = \left( \frac{1}{1 - \kappa} - \frac{1}{1 - \rho_{r_f} \kappa} \right) \tau_f + \frac{1}{1 - \rho_{r_f} \kappa} r_{f,t+1} \]
\[ - \frac{1}{2} \left( \frac{1}{1 - \kappa} - \frac{1}{1 - \rho_i \kappa} \right) \nu - \frac{1}{2} \frac{1}{1 - \rho_i \kappa} \nu_{i,t} \cdots \]
\[ + \sum_{k=1}^{K} \left( \left( \frac{1}{1 - \kappa} - \frac{1}{1 - \rho_{\beta^{(k)} \kappa}} \right) \bar{\lambda}^{(k)} - \frac{1}{1 - \rho_{\beta^{(k)} \kappa}} \rho_{\lambda^{(k)}}^2 \right) \left( \lambda_{t}^{(k)} - \bar{\lambda}^{(k)} \right) \beta_{i,t}^{(k)} + \xi_{i,t}, \]

where

\[ \xi_{i,t} = \sum_{j=1}^{\infty} \kappa^{j-1} \sum_{k=1}^{K} \text{cov}_t \left( \lambda_{i+j-1}^{(k)}, \beta_{i+j-1}^{(k)} \right). \]

1 Modeling variance, \( \nu_{i,t} \), as a Gaussian process has the drawback that negative values can arise. In the data, however, this is never a problem and therefore we view this as a reasonable and tractable approximation to the true variance process.
Given the assumed dynamics for roe, the cash flow component of the book-to-market ratio is:

\[
CF_{i,t}^{bm} = \frac{1}{1 - \kappa} \mu_i + \frac{\rho_x}{1 - \kappa \rho_x} (x_{it} - \mu_i).
\]  

Our identification strategy relies on the simple structure of the cash flow and discount rate components of the book-to-market ratio.

3 Simple case: One-factor model, constant betas and risk price

We adopt a two-stage procedure that enables us to estimate parameters by solving systems of linear equations. For ease of exposition, we describe our procedure in a simple case: a one-factor model in which firms’ betas, firms’ return variances, the risk-free rate, and the price of risk are constant. This case provides insights into how asset pricing tests fit into our framework. The components of book-to-market simplify to become:

\[
DR_{i,t}^{bm} = \frac{1}{1 - \kappa} r_f - \frac{1}{2} \frac{1}{1 - \kappa} v_i + \frac{1}{1 - \kappa} \beta_i \lambda,
\]  

\[
CF_{i,t}^{bm} = \frac{1}{1 - \kappa} \mu_i + \frac{\rho_x}{1 - \kappa \rho_x} (x_{it} - \mu_i).
\]

Thus, we must estimate the following parameters: \( \lambda, \rho_x, \kappa, r_f \), which apply to all firms; and \( \mu_i, v_i, \) and \( \beta_i \) for each \( i \); and \( x_{it} \) for each \( i \) and \( t \). Following Cohen, Polk, and Vuolteenaho (2003), we set the log-linearization constant \( \kappa \) to 0.96. We show later that the main results are not particularly sensitive to this value. We estimate the remaining parameters using a two-stage approach. First, we obtain forecasts of \( v_i, \beta_i, \rho_x, \) and \( r_f \) without imposing the structure of the model. Even when we allow return variances and betas to vary over time, we show later that first-stage estimates are accurate. Second, we estimate \( \lambda, \mu_i, \) and \( x_{it} \) conditioning on the estimated parameter values from the first stage.

This two-stage procedure allows us to estimate the model parameters by solving systems of linear equations. The intuition for the identification is as follows. Because we impose the present value relation, identification of the cash flow component of book-to-market is crucial for obtaining the discount rate component. Imposing the model’s structure enables us to identify the numerous
earnings parameters (μ_i for all i and x_{it} for all i and t) even though we observe roe only once per firm-year. The key to identification is that we also observe each firm’s book-to-market ratio in each year. Knowing the decomposition of book-to-market into its components enables the estimation of risk prices (λ). Specifically, risk prices reflect the slope of the relation between the discount rate component of firms’ book-to-market ratios and their betas.

3.1 Second-stage estimation procedure

We use annual observations of roe_{it} and bm_{it} for each firm to estimate the cash flow parameters (μ_i and x_{it}) and the price of risk λ. As required by the model, we use only firms with available data in the current and previous year. First, we define a version of the book-to-market ratio (bm^ex_{i,t}) that excludes the variance and risk-free rate components as follows:

\[ bm^ex_{i,t} = bm_{i,t} - \frac{1}{1-\kappa} \rho_f + \frac{1}{2(1-\kappa)} v_i \]

\[ = \frac{1}{1-\kappa} \beta_i \lambda - \frac{1}{1-\kappa} \mu_i - \frac{\rho_x}{1-\kappa \rho_x} (x_{it} - \mu_i). \]  

We observe bm^ex_{i,t} conditional on the first-stage parameters. Thus, we can express x_{it} as a function of bm^ex_{i,t} and the parameters of the model:

\[ x_{it} = k_{x_{it}} + \mu_{\mu_i}, \]  

where \( k_{x_{it}} = (bm^ex_{i,t} + \frac{1}{1-\kappa} \beta_i \lambda) \frac{1-\kappa \rho_x}{\rho_x} \) and \( k_{\mu_i} = \frac{\rho_x^{-1}}{(1-\kappa) \rho_x} \). Using this equation to substitute for x_{it} in terms of λ and μ_i dramatically reduces the number of parameters to just K + N, keeping the estimation tractable.

We propose a Quasi-Maximum Likelihood Estimation (QMLE) procedure based on a log-likelihood function of roe where shocks are normal and independent across firms. This likelihood function approximates the true likelihood function but does not require specifying nuisance parameters representing correlations and deviations from normality. Because QMLE is a special case of GMM, we can compute GMM test statistics that account for cross-correlation, autocorrelation, and nonnormalities of the moments.
For firm \( i \), the approximating log-likelihood function is:

\[
l_i(\theta) = \text{const} + \sum_{t=t_i+1}^{t_i+T_i} \left\{ -\frac{1}{2} \omega_{i,t-1} \sigma_{\eta}^{-2} (\text{roe}_{i,t} - x_{i,t})^2 - \frac{1}{2} \omega_{i,t-1} \sigma_{\varepsilon}^{-2} (x_{it} - \rho_x x_{i,t-1} - (1 - \rho_x) \mu_i)^2 \right\} - \frac{1}{2} \omega_i \sigma_{\mu}^{-2} (\mu_i - \bar{\mu})^2,
\]

(28)

where \( \omega_{i,t} \) is a weighting parameter for firm \( i \) at time \( t \), \( t_i \) is the first observation for firm \( i \), and \( T_i \) is the total number of observations for firm \( i \). We estimate the model using both equal weights (EW) with \( \omega_{i,t} = 1 \) and value weights (VW) with \( \omega_{i,t} \) proportional to each firm’s market capitalization and normalized to sum to 1 across all sample firms in each year. Using VW is equivalent to assuming that the variance of each observation is inversely proportional to the size of the firm, implying that large firms have lower variance than small firms. We define the average weight as \( \omega_i = \frac{1}{T_i} \sum_{t=t_i+1}^{t_i+T_i} \omega_{i,t-1} \). The first line of Equation (28) corresponds to the log-likelihood of the roe observations for firm \( i \). The term on the second line of this equation represents the prior distribution of the firm-specific mean: \( \mu_i \sim N(\bar{\mu}, \sigma_{\mu}^2) \), where \( \bar{\mu} \) and \( \sigma_{\mu}^2 \) come from the first-stage, as discussed in the next section.

We now maximize the likelihood function with respect to each parameter. The first-order condition (FOC) with respect to \( \mu_i \), after substituting the above expression for \( x_{i,t} \) and simplifying, becomes:

\[
\mu_i = \frac{\sum_{t=t_i+1}^{t_i+T_i} \omega_{i,t-1} \left\{ \sigma_{\eta}^{-2} (\text{roe}_{i,t} - k_{x_{i,t}}) k_{\mu_i} - \sigma_{\varepsilon}^{-2} (k_{x_{i,t}} - \rho_x k_{x_{i,t-1}}) (k_{\mu_i} - 1) (1 - \rho_x) \right\} + \omega_i \sigma_{\mu}^{-2} \bar{\mu}}{T_i \omega_i \left[ \sigma_{\eta}^{-2} k_{\mu_i}^2 + \sigma_{\varepsilon}^{-2} (1 - \rho_x)^2 (k_{\mu_i} - 1)^2 \right] + \omega_i \sigma_{\mu}^{-2}}.
\]

(29)

Next we compute the FOC with respect to \( \lambda \), which appears in the likelihood function for all firms, yielding:

\[
\sum_{i=1}^{N} \frac{\partial J_i}{\partial \lambda} = \sum_{i=1}^{N} \sum_{t=t_i+1}^{t_i+T_i} \left\{ -\omega_{i,t-1} \sigma_{\varepsilon}^{-2} (x_{it} - \rho_x x_{i,t-1} - (1 - \rho_x) \mu_i) \left( \frac{\partial x_{it}}{\partial \lambda} - \rho_x \frac{\partial x_{i,t-1}}{\partial \lambda} \right) \right\},
\]

(30)

where \( \frac{\partial x_{it}}{\partial \lambda} = \frac{1-\kappa \rho_x}{(1-\kappa) \rho_x} \beta_i \). Because the FOCs are linear in \( \lambda \) and \( \mu_i \), we can solve for the optimal parameters analytically.
3.1.1 Solving for $\mu_i$ and $\lambda$

To economize on algebra, we note the following relationships:

$$k_{x_{it}} = a_{it} + b_{it}\lambda,$$  \hspace{1cm} (31)

where $a_{it} = -bm_{it}^x - r f - v 1 - \kappa \rho_x$ and $b_{it} = \frac{\partial x_{it}}{\partial \lambda^{10^6}} = \frac{1 - \kappa \rho_x}{(1 - \kappa) \rho_x} \beta_{it}^2$. Now we can solve for $\mu_i$ as a linear function of $\lambda$:

$$\mu_i = a_{\mu_i} + b_{\mu_i}\lambda,$$  \hspace{1cm} (32)

where

$$a_{\mu_i} = \frac{\sum_{t=t_{i+1}}^{t_{i+T_i}} \{ \omega_{i,t-1}\sigma_{\eta}^{-2} (roe_{it} - x_{it}) k_{\mu_i} - \omega_{i,t-1}\sigma_{\varepsilon}^{-2} (a_{it} - \rho_x a_{it-1}) (k_{\mu_i} - 1) (1 - \rho_x) \} + \omega_i \sigma_{\mu}^{-2} \mu_{it}}{T_i \omega_i \left[ \sigma_{\eta}^{-2} k_{\mu_i}^2 + \sigma_{\varepsilon}^{-2} (1 - \rho_x)^2 (k_{\mu_i} - 1)^2 \right] + \omega_i \sigma_{\mu_i}^{-2}},$$

$$b_{\mu_i} = \frac{\sum_{t=t_{i+1}}^{t_{i+T_i}} \{ -\omega_{i,t-1}\sigma_{\eta}^{-2} b_{it} k_{\mu_i} - \omega_{i,t-1}\sigma_{\varepsilon}^{-2} (b_{it} - \rho_x b_{it-1}) (k_{\mu_i} - 1) (1 - \rho_x) \} + \omega_i \sigma_{\mu_i}^{-2}}{T_i \omega_i \left[ \sigma_{\eta}^{-2} k_{\mu_i}^2 + \sigma_{\varepsilon}^{-2} (1 - \rho_x)^2 (k_{\mu_i} - 1)^2 \right] + \omega_i \sigma_{\mu_i}^{-2}}.$$

Substituting the expressions for $\mu_i$ and $x_{it}$ into the FOC for $\lambda$, we obtain:

$$\sum_{i=1}^{N} \frac{\partial J_i}{\partial \lambda} = \sum_{i=1}^{N} \sum_{t=t_{i+1}}^{t_{i+T_i}} \left\{ \omega_{i,t-1}\sigma_{\eta}^{-2} (roe_{it} - x_{it}) \frac{\partial x_{it}}{\partial \lambda} + \omega_{i,t-1}\sigma_{\varepsilon}^{-2} (x_{it} - \rho_x x_{it-1} - (1 - \rho_x) \mu_i) \left( \frac{\partial x_{it}}{\partial \lambda} - \rho_x \frac{\partial x_{it-1}}{\partial \lambda} \right) \right\} \hspace{1cm} (33)$$

$$= k_{\lambda} \lambda + z = 0,$$  \hspace{1cm} (34)

where

$$k_{\lambda} = -\sum_{i=1}^{N} \sum_{t=t_{i+1}}^{t_{i+T_i}} \left\{ \omega_{i,t-1}\sigma_{\eta}^{-2} (b_{it} + k_{\mu_i} b_{it}) b_{it} \right\}, \hspace{1cm} (35)$$

$$z = -\sum_{i=1}^{N} \sum_{t=t_{i+1}}^{t_{i+T_i}} \left\{ -\omega_{i,t-1}\sigma_{\eta}^{-2} (roe_{it} - (a_{it} + k_{\mu_i} a_{it}) b_{it}^{(k)}) \right\}.$$  \hspace{1cm} (36)

\footnote{The $t$ subscript is superfluous here but will be important when we allow betas to vary over time.}
Thus, we conclude that the risk price estimates are:

\[ \lambda = -k_{X}^{-1}z, \]  

(37)

The \( \mu_i \) for all \( i \) can then be computed from Equation (29) and the value of \( \lambda \) from (37).

We get standard errors by writing our QMLE procedure in a Generalized Method of Moments setting, using Newey-West with 3 lags so standard errors account for cross-correlation and autocorrelation in the moments. See the Appendix for details.

In summary, we obtain second-stage estimates as the unique solution to a set of linear equations, resulting in efficient estimates of risk prices as well as fixed and persistent effects in \( \text{roei}_t \). The standard errors are robust, though admittedly they do not take into account uncertainty in the first-stage parameter estimates. It is possible to estimate all parameters jointly at the significant cost of tractability: high-dimensional nonlinear numerical methods would be required.

4 Baseline estimation: K-factor model with time-varying betas

Here we present our baseline estimation where we allow for time variation in betas, variances, and the risk-free rate but assume that risk premiums are constant. We consider three well-known K-factor models: the CAPM, the Fama-French three-factor model, and the Fama-French five-factor model. We will compare the QMLE/GMM approach to standard cross-sectional return regressions used in prior studies. Our four main findings are: 1) standard factor models explain a tiny fraction of the total variation in book-to-market ratios; 2) the market risk premium is the largest among the five that we estimate; 3) imposing the present-value relation reduces the standard errors of risk premiums, increasing the power of asset pricing tests; and 4) risk premium estimates from our GMM approach are far less sensitive to transitory stock price fluctuations than standard methods.

With \( K \) factors and time-varying betas, the discount rate component of book-to-market ratio
becomes:

\[
D R_{i,t}^{pm} = \frac{1}{1 - \kappa} \bar{r}_f + \frac{1}{1 - \rho_{r_i} \kappa} (r_{f,t+1} - \bar{r}_f)
\]

\[
- \frac{1}{2} \frac{1}{1 - \kappa} \bar{v} - \frac{1}{2} \frac{1}{1 - \rho_v \kappa} (v_{i,t} - \bar{v})...
\]

\[
+ \sum_{k=1}^{K} \lambda^{(k)} \left\{ \frac{1}{1 - \kappa} \beta^{(k)} + \frac{1}{1 - \rho_{\beta^{(k)}} \kappa} \left( \beta_{i,t}^{(k)} - \beta^{(k)} \right) \right\}.
\]  

The appendix provides analytical solutions for model parameters and standard errors. The derivations are similar in spirit to those for the simple case discussed above. Next, we describe our data sample and explain how we estimate the first-stage parameters, including expected betas, return variances, and risk-free rates.

4.1 Data and sample

Our primary analysis requires panel data on firms’ book values, market values, earnings, and returns, as well as time series data on asset pricing factor returns and risk-free rate. We obtain all accounting data from Compustat, though we augment our book data with that from Davis, Fama, and French (2000). We obtain stock pricing data, including prices, returns, and shares outstanding from the Center for Research on Securities Prices (CRSP). Our main sample includes Compustat and CRSP data from 1963 through 2014. Because computations of variables such as factor loadings and return variance require up to five years of historical information, our model estimation focuses on the period from 1968 through 2014.

When computing a firm’s book-to-market ratio, we adopt the standard convention of dividing its book equity by its market equity at the end of the calendar year in which the firm’s fiscal year ended. We compute book equity using Compustat data when available and supplementing it with the hand-collected data from the Davis, Fama, and French (2000) study. We follow the Fama and French (1992) procedure for computing book equity. Market equity is equal to shares outstanding times stock price per share. We sum market equity across firms that have more than one share class of stock—i.e., multiple CRSP identifiers per Compustat identifier.

We measure earnings using return on equity as defined by net income divided by book equity. We use income before extraordinary items if net income is unavailable. For comparison to prior
tests of asset pricing models, we also compute several firm characteristics that predict short-term stock returns in historical samples. Following Novy-Marx (2013) and Fama and French (2015), we compute profitability as annual revenues minus costs of goods sold, interest expense, and selling, general, and administrative expenses, all divided by book equity from the same fiscal year. Following Cooper, Gulen, and Schill (2008) and Fama and French (2015), we compute investment as the annual percentage growth in total assets.

We obtain the monthly returns for the market, size, value, profitability, and investment factors in the Fama-French (2015) five-factor model from Kenneth French’s website. We obtain one-month and one-year risk-free rate data from one-month and one-year yields of US Treasury Bills, which are available on Kenneth French’s website and the Fama Files in the Monthly CRSP US Treasury Database, respectively.

We impose sample restrictions to ensure the availability of high-quality accounting and stock price information. We exclude firms with negative book values as we cannot compute the logarithm of their book-to-market ratios. Similarly, firms included in our sample all have nonmissing market equity data at the end of the most recent calendar year. They also have nonmissing stock return data for at least 225 days in the past year and at least 500 days over the past five years, which is necessary for the accurate estimation of stock return variance and risk exposures discussed below. Lastly, firms must have nonmissing earnings data in their most recent fiscal year to meet the requirements of our model.

### 4.2 Computing betas and return variance

Model estimation requires accurate measures of firms’ betas and return variances, which we estimate using the following procedure. We first compute simple one-year and five-year measures of betas and variance. Both are based on daily excess returns, which are daily stock returns minus the daily return from the one-month risk-free rate as of the beginning of the month.

A firm’s variance for each period is simply the annualized average value of its squared daily excess returns during the period. We do not subtract each firm’s mean squared excess return to minimize estimation error in this calculation. A firm’s beta with respect to a factor in an asset pricing model is the coefficient on the factor from a multivariate linear regression that includes all factors in the asset pricing model. Thus, a firm’s market beta depends on which factors other than
the market, such as size and value, are included in the model. We estimate betas for all factors in each of three asset pricing models: the CAPM, which includes only the market factor; the Fama-French (1993) three-factor model (FF3 hereafter), which also includes size and value factors; and the Fama-French (2015) five-factor model (FF5), which also includes profitability and investment factors. We consider stochastic discount factors that are log-linear in the factors. Standard linear-beta pricing models imply stochastic discount factors that are linear in the factors. This distinction has a negligible impact on the beta estimates as they are based on daily data.

4.3 First-stage parameter estimation

To facilitate estimation of the model, we estimate several parameters before implementing our GMM procedure. These parameters are less central to our results and including them in a simultaneous estimation would result in nonlinear estimation equations. Accordingly, we model the dynamics of earnings, betas, and variance, and the risk-free rate as first-order autoregressive processes, consistent with the model. We allow for temporary estimation error in betas and variance and temporary shocks to earnings, as well as fixed effects in earnings. We identify the persistence and variance parameters solely from the autocovariances of these processes. Each process conforms to the general model of a process, \( z_{i,t} \), that is the sum of a temporary component \( \eta_{it} \) and an AR(1) process \( x_{it} \) that includes a firm fixed effect \( \mu_i \):

\[
\begin{align*}
    z_{i,t} &= x_{i,t} + \sigma_\eta \eta_{it} \\
    x_{i,t} &= \mu_i + \rho (x_{i,t-1} - \mu_i) + \sigma_\varepsilon \varepsilon_{i,t}.
\end{align*}
\]

Shocks can be cross-sectionally correlated, but they are assumed to be i.i.d. over time. Let the cross-sectional variance of the fixed effect be \( \sigma_\mu^2 \). The autocovariances of \( z \) are highly informative about the persistence and three variance parameters, as shown by the following equations:

\[
\begin{align*}
    \text{Cov}(z_{i,t}, z_{i,t-k}) &= \sigma_\mu^2 + \frac{\rho^k}{1 - \rho^2} \sigma_\varepsilon^2 \text{ for } k > 0 \\
    \text{Var}(z_{i,t}) &= \sigma_\mu^2 + \frac{1}{1 - \rho^2} \sigma_\varepsilon^2 + \sigma_\eta^2
\end{align*}
\]

For each process, we employ a method of moments algorithm to pick the persistence and variance parameters that best fit the empirical variance and 20 autocovariances of each process.
As shown in Figures 1 to 4, the autocovariance functions of each process closely match the predictions from the augmented AR(1) model. The abrupt jump in the autocovariance function at zero represents the variance of the temporary shock ($\sigma^2_t$) and demonstrates how we identify the magnitude of that shock. The rate of decay of the autocovariances from the first lag through the last enables us to identify the persistence parameter ($\rho$). The level of the autocovariance at intermediate lags identifies the variance of the persistent component ($\sigma^2_p$).

In the \textit{roe} process, the level to which the autocovariance function appears to converge at long lags identifies the variance in the permanent (fixed) component ($\sigma^2_p$). Extrapolating the other graphs to lags beyond 20 years suggests that the autocovariances will eventually converge to zero, consistent with no fixed effect. Yet, even in a sample of 50 years, it is difficult to distinguish a highly persistent process without a fixed effect from a slightly less persistent process with a fixed effect. Our initial exploration of alternative specifications in which we allow for fixed effects in firms’ betas suggests that our main results are not sensitive to this modeling choice.

Table 1 shows the parameters for the \textit{roe}, $v$ (variance), and $\beta$ processes, where the latter are based on estimated one-year betas from the FF5 model. The table also shows the average values of each process, along with the average value and persistence of the risk-free rate. Throughout the analysis, we focus on value-weighted (VW) estimates to emphasize economic importance, though we also present equal-weighted (EW) results to show how results differ for smaller firms.

In Table 1, the estimated VW persistence parameters for the market, size, and value betas are quite high at 0.94 to 0.95, reflecting these processes significant autocovariances even after 20 lags. In contrast, the estimated persistence of \textit{roe} is lower at 0.75, which is necessary to match the rapid decline in autocovariances in lags two through five. The 4.5% volatility of the fixed effect in \textit{roe} matches the long-run autocovariance of the process reasonably well. The volatility of the temporary and persistent components in \textit{roe} are substantial at 10.4% and 5.9%, respectively. The volatility of the temporary components in the beta are quite large, ranging from 0.3 to 0.7, as a consequence of estimation error in betas.

In addition to the parameters above, we estimate the long-run means of the risk-free rate, return variance, and beta processes as the value-weighted averages of these variables across all observations in the sample. Similarly, we estimate the average fixed effect in earnings ($\hat{\mu}$) as the value-weighted average \textit{roe} across all firm-years.
4.4 Forecasting betas and return variance

Including estimated betas and return variance in our GMM procedure would cause attenuation bias in risk prices. To minimize this issue, we forecast estimated betas and variance and use these predicted values for $\beta_{it}$ and $\nu_{it}$ in our procedure. Although the resulting forecasts still suffer from some measurement error, we can quantify and bound the impact of this error on our findings.

We employ a simple linear model for forecasting a firm’s one-year estimated beta based on three sets of known betas: its adjusted one-year and five-year betas and the average one-year betas of firms with similar betas. We use all betas from a given asset pricing model to forecast each beta. For example, when testing the FF3 model, we use a firm’s market, size, and value betas to forecast each of its three betas. To minimize estimation error in the one-year and five-year firm betas used as regressors, we apply Bayesian shrinkage to each beta assuming a normal prior distribution of betas and normally distributed estimation error. The third predictor of beta is the average beta of firms ranked in the same beta decile as the firm. We include portfolio betas as predictors because they effectively perform a nonparametric form of shrinkage.

We evaluate the performance of the resulting beta forecasts using estimates from the augmented AR(1) model in the previous section as a benchmark. Ideal forecasts would have explanatory power ($R^2_{\text{forecast}}$) equal to the fraction of variation in betas explained by the latent persistent component in betas $R^2_x$. However, the actual relation between the $R^2$ values will satisfy:

$$R^2_{\text{forecast}} < R^2_x = \frac{1}{1-\rho^2} \frac{\sigma^2_{\eta}}{\sigma^2_{\theta} + \sigma^2_{\eta}}.$$  \hspace{1cm} (43)

The empirical forecasts from these linear models perform reasonably well, as measured by the ratio of the two $R^2$ values ($R^2_{\text{forecast}}/R^2_x$). In the CAPM, forecasts of market beta explain 75% as much variation as the latent persistent component in market betas ($R^2_{\text{forecast}} = 55.3\%$ vs. $R^2_x = 73.7\%$). In the FF3 model, the forecasts of market, size, and value betas explain 72%, 79%, and 87% as much variation as the latent persistent components in the respective betas. In the FF5 model, the forecasts of the profitability and investment betas are somewhat less accurate, explaining 48% and 44% of the possible variation.
4.5 GMM estimation

We employ efficient GMM estimation of risk prices ($\lambda$) and the components of earnings ($\mu_i$ and $x_{it}$), conditioning on the estimated parameter values from the first stage. In the moment equations, we substitute the first-stage values of $\beta$, $v$, and $r_f$, along with annual data on $roe$ and $bm$. Before estimation, we simplify the problem by using the present-value relation to express $x_{it}$ in terms of observable data and the $\mu_i$ and $\lambda$ parameters. The GMM point estimates of $\mu_i$ and $\lambda$ solve the equations for these parameters shown earlier. We compute GMM standard errors that are robust to heteroskedasticity and autocorrelation across three lags ($q = 3$) as described earlier.

Table 2 shows the GMM risk price estimates for VW and EW versions of the CAPM, FF3, and FF5 models. As a test of each model’s predicted average return, we include an intercept ($\lambda_0$), operationalized as an extra price of risk associated with a constant $\beta_0 = 1$. The estimated coefficient on the intercept represents deviations in the average firm’s long-term return from the average prediction of the model. Under the joint null hypothesis of the asset pricing model and the specified earnings dynamics, the intercept should be zero.

The VW CAPM estimates in Table 2 indicate that the risk premium associated with market beta ($\lambda_{mkt}$) is positive at 5.27% per year with a standard error of 0.98%. We can reject the hypothesis that the market risk premium is zero at conventional statistical significance levels. The estimated intercept is economically material at 2.89% but statistically insignificant based on the standard error of 2.22%. The uncertainty in the intercept arises because all stocks’ average returns are subject to high market-wide volatility. The EW CAPM estimate of the market premium is quite similar, though the intercept is higher.

The VW FF5 estimates in column five indicate a somewhat lower but still positive and significant market premium of 3.39%. The other four estimated risk premiums are significantly smaller in absolute value than the market premium. Only the value premium and the profitability premium are statistically significant. The value premium is positive at 1.26% with a standard error of just 0.40%. Surprisingly, the profitability premium is negative at -1.60% with a standard error of 0.57%. We cannot reject the hypothesis that the size and investment premiums are zero. The EW FF5 estimates in column six provide similar insights.

The final column in Table 2 shows how the VW FF5 results change when we use a lower value of $\kappa$, the log-linearization constant. Switching from $\kappa = 0.96$ to $\kappa = 0.94$ has little impact on
the results. The estimated risk premiums increase in absolute value slightly, while the estimated intercept declines and becomes insignificant.

### 4.6 Results discussion

The GMM estimates of the market premium range from 3% to 5%, whereas the estimates of the premiums associated with non-market risks are all within 2% of zero. Although the market premium estimates differ from those in prior studies, they are consistent with the high average equity premium of stocks relative to safe assets. We explicitly analyze the extent to which the risk premium estimates above conflict with the realized average returns of portfolios formed on betas in the next section.

The GMM estimates have implications for firms’ expected cash flows and returns. Table 3 reports decompositions of book-to-market ratios and estimates of short- and long-term cash flow and return expectations from the CAPM and FF5 models. Panel A shows means and standard deviations, and Panel B shows correlations among variables.

The decomposition of log book-to-market ratios ($bm$) shows that variation in discount rates explains a trivial amount of variation in $bm$ in both asset pricing models. For ease of interpretation, we exclude the impact of return variance and the risk-free rate when performing the decomposition of $bm$. By analogy to $bm^{ex}$, which excludes the impact of return variance and the risk-free rate on $bm$, we define $bm^{ex}_{DR}$ as the discount rate component that excludes these same components. With this definition, we can write $bm^{ex} = bm^{ex}_{DR} - bm_{CF}$. Table 3 shows that the discount rate component ($bm^{ex}_{DR}$) from the CAPM is actually negatively correlated with $bm^{ex}$, implying that discount rates explain none of the variation in book-to-market ratios. In the FF5 model, although $bm^{ex}_{DR}$ is positively correlated with $bm^{ex}$, its minimal variance of just $(0.1063)^2 = 0.0113$ can explain just 1.64% of variation in $bm^{ex}$.

The rows in Panel A showing the variation in annualized expected returns provide another perspective on this finding. The value of $\lambda \beta_{it}$ represents the conditional one-year expected excess return in each model. The cross-sectional standard deviation of one-year expected returns is just 0.94% in the CAPM and 1.65% in the FF5 model, suggesting that few firms’ short-term expected returns differ by more than 5%. The value of $bm^{ex}_{DR}(1 - \kappa)$ represents firms’ annualized long-term expected excess returns from each model. Because of mean reversion in firms’ betas, long-term
returns vary even less than short-term returns. The cross-sectional standard deviations of long-term returns are just 0.34% and 0.44% in the CAPM and FF5 models, respectively.

The bottom rows in Panel A show the estimates of the persistent and permanent components of earnings, $x_{it}$ and $\mu_i$. Because discount rate variation from both asset pricing models fails to explain variation in book-to-market ratios, the cash flow component of book-to-market must be extremely volatile and is approximately equal to total book-to-market, after excluding the variance and risk-free rate components. This logic explains why the cash flow parameter estimates are similar in both models. The estimated volatility in the permanent component is near 2%, which is somewhat smaller than the 4% estimate from the augmented AR(1) model of earnings. Partly as a result, the estimated volatility of the persistent component of earnings exceeds 22%, which is implausibly high compared to the total volatility of earnings ($\text{roeq}$) of 13%. Because total earnings includes volatility from the temporary component ($\sigma^2_x$), the total should be more volatile than the persistent component.

The implausible cash flow dynamics implied by both asset pricing models (CAPM and FF5) suggests that they are misspecified. In a correctly specified asset pricing model, discount rates would account for more variation in book-to-market ratios to be consistent with the modest variation in earnings data. Identifying and characterizing the source of this discount rate variation is a promising agenda for future research.

## 5 Relation to existing approaches

The leading alternative to our approach is to estimate risk premiums using Fama-MacBeth (1973) cross-sectional regressions of monthly or annual returns on betas. Here we present estimates from simple implementations of the alternative approach. To facilitate comparison to our estimates, we use annual returns as the dependent variable and forecasted betas from our first-stage procedure as regressors. To minimize measurement error in these betas, we rank firms into deciles according to their forecasted betas at each time $t$ and assign each firm the value-weighted forecasted beta of all firms in its decile. Table 4 shows the estimated coefficients, which represent time-series averages of the cross-sectional coefficients, along with standard errors. The lefthand columns show the estimated risk premiums from the CAPM, FF3, and FF5 models. The righthand columns show regressions using firm characteristics rather than betas as the regressors.
The first and foremost finding in Table 4 is that the standard errors on the risk prices range from 1.76% to 3.64% with the highest standard errors being associated with the market premium. In a related fact, one cannot reject the hypothesis that most of the risk premium estimates are zero. The only minor exception is the value premium estimate in the EW version of the FF3 model, which is significant at the 10% level. The range of the 95% confidence intervals on the market premium exceeds 12% in all specifications, implying that cross-sectional regressions are weak tests of models that include the market as a factor. In contrast, the width of the confidence interval around the market premium estimates in our GMM approach is 4% or smaller in all three VW specifications.

In the cross-sectional return regressions with characteristics as regressors, shown in the right-hand columns of Table 4, one can detect some return predictability, mainly in the EW specification. Only the coefficient on log investment is statistically significant at the 5% level in the VW specification. Overall, the results indicate that large standard errors are a problem for the standard approach.

The nonparametric analogue of the regression approach is to estimate risk premiums by sorting firms into portfolios and measuring these portfolios’ returns. Figures 5 to 9 show the annual returns of portfolios sorted by forecasted betas and characteristics held for one-year and 10-year horizons. We use the Jegadeesh and Titman (1993) methodology for creating long-horizon portfolios. For example, the 10-year return of the market beta decile 1 portfolio is the average of the returns from 10 cohort portfolios, each formed by selecting firms with the lowest market betas \(k\) years ago, where \(k\) ranges from 1 to 10. We value weight firms within each cohort portfolio. We rebalance portfolios annually in June following Fama and French (1992).

The beta sorts shown in Figures 5 to 9 result in only minimal differences in returns across beta deciles. At the 10-year horizon, the largest top-to-bottom decile spreads in returns are the +2% and -2% spreads from the SMB and RMW beta sorts, respectively. Both spreads are statistically insignificant at even the 10% level because the standard errors (not shown) are approximately 2%. At the one-year horizon, the return spreads from beta sorts are again statistically and economically insignificant. In a related fact, the magnitudes and signs of some of the return spreads change by large amounts depending on which horizon one examines. For example, the return spread from sorting on RMW beta changes from positive at the one-year horizon to negative at the 10-year horizon.
The characteristic sorts shown in the companion figures tell a similar story with some minor exceptions. The portfolios sorted by size and value produce consistently negative and positive return spreads, respectively, across all horizons. However, none of the 10-year return spreads from characteristic sorted portfolios is statistically significant at the 5% level, again because of the large standard errors.

There are two main advantages to estimating long-term expected returns using book-to-market ratios instead of standard cross-sectional regressions of short-term realized returns on betas: 1) transitory price fluctuations have little impact on risk prices estimated from long-term returns, whereas these fluctuations could dramatically affect the cross-section of monthly or even annual returns; and 2) long-term expected returns are less noisy than average realized returns, leading to more precise risk price estimates. We formalize these arguments below.

5.1 Robustness to transitory price fluctuations

Consider the case in which the econometrician estimates a one-factor model with a constant price of risk but potentially time-varying betas. Suppose that in reality there are two factors and that firms’ exposures to the second factor are transient. In particular, the single factor that the econometrician analyzes, $f_t$, actually consists of two independent and homoskedastic factors, $f_t = f_{1t} + f_{2t}$. In this case, the time-varying factor loading on the single factor satisfies:

$$
\beta_{it} = \frac{\text{cov}(r_{it+1}, f_{it+1})}{\text{var}(f_{it+1})} = \frac{\text{var}(f_{1t+1})}{\text{var}(f_{it+1})} \frac{\text{cov}(r_{it+1}, f_{1t+1})}{\text{var}(f_{1t+1})} + \frac{\text{var}(f_{2t+1})}{\text{var}(f_{it+1})} \frac{\text{cov}(r_{it+1}, f_{2t+1})}{\text{var}(f_{2t+1})}
$$

(44)

(45)

where $\beta_{1i}$ is the long-run (highly persistent) factor loading of firm $i$, and $\beta_{2it}$ is transitory

$$
\beta_{2it} = \sigma_\varepsilon \varepsilon_{it}.
$$

(46)

Assume the loading on the misspecified single factor, $\beta_{it}$, is known to the econometrician. For the purpose of the example, let the price of risk for $f_{1t}$ be $\lambda_1 > 0$ and the price of risk for $f_{2t}$ be $\lambda_2 < 0$. One can estimate $\lambda$ from the misspecified one-factor model using either standard cross-sectional return regressions or cross-sectional book-to-market regressions. The population estimate ($\hat{\lambda}_r$) from
the cross-sectional return regression is:

\[ \hat{\lambda}_r = \frac{\text{cov} (r_{i,t+1}, \beta_{it})}{\text{var} (\beta_{it})} \]
\[ = \frac{\text{cov} (\beta_{1i} \lambda_1 + \beta_{2it} \lambda_2, k_1 \beta_{1i} + k_2 \beta_{2it})}{\text{var} (\beta_{it})} \]
\[ = \frac{\text{var} (\beta_{1i}) k_1}{\text{var} (\beta_{it})} \lambda_1 + \frac{\text{var} (\beta_{2it}) k_2}{\text{var} (\beta_{it})} \lambda_2. \quad (47) \]

Next consider the book-to-market cross-sectional regression. The discount rate component in this case is:

\[ DR_{i,t}^{bm} = \frac{1}{1 - \kappa} r_f - \frac{1}{2} \frac{1}{1 - \kappa} \nu_i + \frac{1}{1 - \kappa} \beta_{1i} \lambda_1 + \beta_{2it} \lambda_2. \]
\[ (50) \]

Thus, a misspecified cross-sectional regression with a dependent variable \((y_{it})\) based on book-to-market satisfying

\[ y_{it} \equiv (1 - \kappa) bm_{it} - \left( r_f - \frac{1}{2} \nu_i \right) + (1 - \kappa) CF_{it}^{bm} = \beta_1 \lambda_{bm} + \alpha_{it}, \quad (51) \]

yields an estimate, \(\hat{\lambda}_{bm}\), that simplifies to

\[ \hat{\lambda}_{bm} = \frac{\text{cov} (y_{i,t+1}, \beta_{it})}{\text{var} (\beta_{it})} \]
\[ = \frac{\text{cov} (\beta_{1i} \lambda_1 + (1 - \kappa) \beta_{2it} \lambda_2, k_1 \beta_{1i} + k_2 \beta_{2it})}{\text{var} (\beta_{it})} \]
\[ = \frac{\text{var} (\beta_{1i}) k_1}{\text{var} (\beta_{it})} \lambda_1 + (1 - \kappa) \frac{\text{var} (\beta_{2it}) k_2}{\text{var} (\beta_{it})} \lambda_2. \quad (52) \]

Comparing the equations for \(\hat{\lambda}_r\) and \(\hat{\lambda}_{bm}\), we see that the book-to-market regression estimate depends much less on the transitory factor, by a factor of \((1 - \kappa)\), in fact. As an example, if \(\frac{\text{var} (\beta_{1i}) k_1}{\text{var} (\beta_{it})} \lambda_1 = 4\%\), \(\frac{\text{var} (\beta_{2it}) k_2}{\text{var} (\beta_{it})} \lambda_2 = -8\%\), and \(\kappa = 0.96\), the two approaches yield very different estimates of the price of risk: \(\hat{\lambda}_r = -4\%\) versus \(\hat{\lambda}_{bm} = 3.68\%\).

In summary, relative to the book-to-market based regressions, the cross-sectional return regressions emphasize short-term fluctuations in expected returns as opposed to persistent differences in discount rates. The latter are what mainly determine firm market values and what is important for buy-and-hold investors and firm investment, while the former is more important for investors that rebalance frequently. The reasoning above is relevant in practice unless all factors are known.
and the econometrician has the correct model—conditions that are rarely satisfied.

### 5.2 Increased statistical power

For ease of exposition, consider the simple constant beta case of the model discussion in Section 3, and assume that betas are known. In this case, our GMM approach effectively estimates the price of risk through the following panel regression:

\[
(1 - \kappa) bm_{it} - \left( r_f - \frac{1}{2} \hat{\sigma}_i \right) + (1 - \kappa) CF^{bm}_{it} = \beta_i \lambda + \alpha_{it},
\]

where \( \alpha_{it} \) is the pricing error of firm \( i \) and time \( t \). This pricing error arises, under the null hypothesis of the asset pricing model, from noise in estimates of \( \hat{\sigma}_i \) and \( CF^{bm}_{it} \). Such noise is likely to be relatively small because variance and cash flows are highly predictable relative to stock returns. In contrast, the standard cross-sectional regression relies on the following relation:

\[
r_{it}^e = \beta_i \lambda + \tilde{\alpha}_{it},
\]

where, under the null hypothesis, \( \tilde{\alpha}_{it} \) arises from the difference between average realized returns and expected return. The estimate of \( \lambda \) from the book-to-market regression is more efficient if the variance of \( \alpha_{it} \) is lower than the variance of \( \tilde{\alpha}_{it} \).\(^3\) This condition appears to hold based on the fact that standard errors from our approach are considerably smaller than those from standard regressions.

### 6 Robustness and extensions of the baseline model

Here we discuss the effects of model misspecification on GMM estimates of risk prices. We also discuss two extensions of the baseline framework that accommodate models in which risk prices vary over time and those in which some investors have irrational expectations of cash flows.

---

\(^3\) To ensure unbiased estimates, the errors in cash flow and return variance forecasts must be uncorrelated with \( \beta \).
6.1 Effects of model misspecification

We first consider the impact of model misspecification. Suppose the risk model is misspecified in that it omits a factor that is priced and is orthogonal to the other factors. Consider the constant beta model of Section 3 and assume that the first-stage parameters are known, as model misspecification does not affect these estimates. The parameters that remain to be estimated are the price of risk ($\lambda$) and fixed effects in earnings ($\mu_i$). As before, we substitute for the values of $x_{it}$ using the model.

In this case, the book-to-market ratio that excludes the variance and risk-free rate components is:

$$bm_{it}^{ex} = \frac{1}{1 - \kappa} \lambda \beta_i + \frac{1}{1 - \kappa} \lambda^o \beta_{i,t}^o - \frac{1}{1 - \kappa} \mu_i - \frac{\rho_x}{1 - \kappa \rho_x} (x_{it} - \mu_i), \quad (57)$$

where the superscript $o$ on $\lambda$ refers to the omitted factor. We allow the factor loading on the omitted factor to vary over time. We express firms’ betas on the omitted factor as $\beta_{i,t}^o = \beta^o + \frac{(1 - \kappa) \rho_x}{1 - \kappa \rho_x} \beta_{i,t}^o$, where $\beta^o = E[\beta_{i,t}^o]$. Then the solution for $x_{it}$ is:

$$x_{it} = \left( \frac{1}{1 - \kappa} \lambda^o \beta^o + \frac{1}{1 - \kappa} \lambda \beta_i - bm_{it}^{ex} \right) \frac{1 - \kappa \rho_x}{\rho_x} + \frac{\rho_x - 1}{(1 - \kappa) \rho_x} \mu_i + \beta_{i,t}^o \lambda^o. \quad (58)$$

For convenience, define $k_{x_{it}} = \left( \frac{1}{1 - \kappa} \lambda^o \beta^o + \frac{1}{1 - \kappa} \lambda \beta_i - bm_{it}^{ex} \right) \frac{1 - \kappa \rho_x}{\rho_x}$, $k_{\mu_i} = \frac{\rho_x - 1}{(1 - \kappa) \rho_x}$, such that $x_{it} = k_{x_{it}} + k_{\mu_i} \mu_i + \beta_{i,t}^o \lambda^o$. Here we note two facts. First, it is clear that the $x_{i,t}$ we extract from the misspecified model does not correspond to conditional expected $roe$, but in fact has two components related to discount rates. The first, $\frac{1}{1 - \kappa} \lambda^o \beta^o$, is based on the average discount rate component due to the omitted factor across firms. One can control for this constant component by adding an intercept ($\lambda_0$) in the price of risk specification, which is what we do in our empirical section. The second discount rate component, $\beta_{i,t}^o \lambda^o$, represents a mean zero error term in the extracted $x_{it}$. Thus, the estimated $x_{it}$ from the model will be too volatile, which is exactly what we find.

A related implication is that model misspecification biases the decomposition of book-to-market, resulting in too little (much) variation in the discount rate (cash flow) component, again consistent with our findings. However, if the factor loadings on the omitted factor are uncorrelated with the loadings on included factors ($Cov[\beta_{i,t}^o, \beta_i] = 0$), the price of risk estimates on included factors ($\lambda$) are unaffected by the omitted factor. On the other hand, if the factor loadings are correlated, our price of risk estimates would be biased. This latter possibility is a general problem caused by correlated omitted factors and is not specific to our approach.
6.2 Time-varying risk prices

We now extend the GMM framework to allow risk prices to vary over time. Suppose there are \( K \) factors with time-varying risk prices, \( \lambda_t^{(k)} \), that evolve according to:

\[
\lambda_{t+1}^{(k)} = \lambda_t^{(k)} + \rho^{(k)}(\lambda_t^{(k)} - \lambda_t^{(k)}) + \sigma_\lambda^{(k)} \varepsilon_{t+1}^{(k)}, \quad k = 1, \ldots, K. \tag{59}
\]

In this case, the discount rate component is:

\[
DR_{i,t}^{bm} = \left( \frac{1}{1 - \kappa} - \frac{1}{1 - \rho_f \kappa} \right) r_f + \frac{1}{1 - \rho_f \kappa} r_{f,t+1} - \frac{1}{2} \left( \frac{1}{1 - \kappa} - \frac{1}{1 - \rho_p \kappa} \right) \nu_i - \frac{1}{2} \frac{1}{1 - \rho_p \kappa} \nu_{i,t} \ldots + \sum_{k=1}^{K} \left( \frac{1}{1 - \kappa} \frac{1}{1 - \rho^{(k)} \kappa} \lambda_t^{(k)} - 1 \frac{1}{1 - \rho^{(k)} \rho^{(k)} \kappa} \lambda_t^{(k)} \lambda_t^{(k)} - \lambda_t^{(k)} \right) \beta_{i,t}^{(k)} + \xi_i, \tag{60}
\]

\[
\xi_i = \sum_{j=1}^{\infty} \kappa^{j-1} \sum_{k=1}^{K} \text{cov}_t \left( \lambda_{t+j-1}^{(k)}, \beta_{i,t+j-1}^{(k)} \right) + \sum_{k=1}^{K} \frac{\rho_{i,k} \sigma^{(k)} \sigma^{(k)}}{1 - \rho^{(k)} \rho^{(k)} \kappa}, \tag{61}
\]

where

\[
\xi_i = \sum_{j=1}^{\infty} \kappa^{j-1} \sum_{k=1}^{K} \text{cov}_t \left( \lambda_{t+j-1}^{(k)}, \beta_{i,t+j-1}^{(k)} \right) \tag{62}
\]

and where \( \rho_{i,k} \) is the correlation between firm \( i \)'s beta and factor \( k \). The numerator is measurable as the covariance of the shocks to estimated betas and lambdas because \( \rho_{i,k} \sigma^{(k)} \sigma^{(k)} = E \left[ \left( \lambda_{t+j}^{(k)} - \rho^{(k)} \lambda_t^{(k)} - (1 - \rho^{(k)}) \lambda_t^{(k)} \right) \left( \beta_{i,t+j}^{(k)} - \rho^{(k)} \beta_{i,t+j}^{(k)} - (1 - \rho^{(k)}) \beta_{i,t+j}^{(k)} \right) \right] \). Furthermore, linking these covariance terms to the realized covariance of betas and risk prices mitigates overfitting concerns—see Lewellen, Nagel, and Shanken (2010) for a critique of standard tests of conditional asset pricing models.

The first stage of our procedure remains the same as before, and we again condition on the first-stage parameters in the second stage of our procedure. Thus, the \( N + T + 2K \) parameters to be estimated are \( \{ \mu_i \}_{i=1}^{N}, \{ \lambda_t \}_{t=1}^{T}, \bar{\lambda}, \) and \( \rho_\lambda \) where \( \lambda_t, \bar{\lambda}, \rho_\lambda \) all are \( K \times 1 \) vectors. The presence of the \( \rho_\lambda \) parameters makes the system nonlinear. However, conditional on \( \rho_\lambda \), the system of equations
is still linear. We exploit this feature by solving for \( \{\mu_i\}_{i=1}^{N}, \{\lambda_t\}_{t=1}^{T}, \bar{\lambda} \) as a function of \( \rho_\lambda \), which we can do analytically. Then we find the \( K \) elements of \( \rho_\lambda \) numerically by maximizing the quasi-maximum likelihood function over \( \rho_\lambda \), where we substitute the \( \rho_\lambda \)-dependent expressions for the other parameters.

In summary, using our proposed methodology, one can test models with time-varying betas and risk prices in which realized covariances between betas and risk prices are explicitly linked to the corresponding conditional covariances.

### 6.3 Irrational expectations

Because our tests are based on the likelihood function of realized \( roe \), the results above therefore reflect rational expectations. However, if survey or analyst data about expected future \( roe \) is available for, say, annual horizons up to five years, one can use these expectations in our tests instead of realized \( roe \). Making this substitution in the first stage of our procedure would yield an estimate of the persistence of beliefs (\( \rho_x \)); and in the second stage it would yield estimates of agents’ beliefs about firm cash flows (\( \mu_i \) and \( x_{it} \)). Thus, by substituting survey/analyst forecasts into our procedure, one can directly test models in which investors’ exhibit biases in expectations of cash flows.

### 7 Conclusion

This study introduces a novel framework for testing asset pricing models’ predictions of long-run expected returns. We model the joint dynamics of earnings, betas, and return variance and impose the structure of the present value relation to estimate long-run expected returns and risk premiums. This approach has the benefits of being robust to short-term price fluctuations and providing precise estimates of risk premiums.

The main cost of imposing structure is that rejections of the model can occur because of misspecification in either the asset pricing model or the dynamics of earnings, betas, and return variance. That is, the null hypothesis is a joint hypothesis of the asset pricing model and the dynamics of these processes. Fortunately, simple econometric techniques enable reliable forecasts of earnings, betas, and return variance, mitigating this joint hypothesis problem. Thus, rejections
of the model are likely attributable to omitted asset pricing factors.

In the last 50 years, researchers in financial economics have learned much about the nature of short-term returns from conducting standard cross-sectional regressions. At this point, however, the marginal benefit of further short-term return regressions is low relative to the benefit of exploring the determinants of long-term returns. This study points to several issues that warrant further investigation.

Reconciling the observed differences in short-term and long-term expected returns is a natural starting point. For example, the estimated long-term market risk premium is significantly positive in sharp contrast to the negative, albeit statistically insignificant, estimates of the market premium from short-term tests. A next step is to understand why the betas from existing asset pricing models explain such a small fraction of the variation in long-term expected returns. Future research could identify and characterize the omitted asset pricing factors. Within our proposed framework, one could also analyze how risk premiums vary over time. Finally, one could test the extent to which asset pricing factors are related to systematic errors in investors’ expectations. We intend to analyze these issues in future work.
8 References


9 Appendix

9.1 Standard errors for the constant beta case in Section 3

Recognizing that the likelihood function is misspecified, we compute robust standard errors using GMM techniques. In other words, the standard errors are not directly linked to the Hessian matrix of the likelihood function. They allow for heteroskedasticity as well as cross- and auto-correlation of the moments. The moments are the QMLE first-order conditions.

One can view these $N + K$ moments as the basis for an exactly identified GMM system for estimating the $N + K$ parameters. Define an indicator variable that is one if firm $i$ has data for year $t$ and $t - 1$ and zero otherwise as $I_{it}$. We can express the derivatives of the log-likelihood function at time $t$ as:

$$\frac{\partial l_t}{\partial \lambda} = \sum_{i=1}^{N} I_{it}\omega_{i,t-1} \left\{ \sigma^{-2}_\eta (\text{roe}_{it} - x_{it}) b_{it} - \sigma^{-2}_\varepsilon (x_{it} - \rho x x_{it-1} - (1 - \rho) \mu_i) (b_{it} - \rho x b_{it-1}) \right\} , \quad (66)$$

and

$$\frac{\partial l_t}{\partial \mu_i} = I_{it}\omega_{i,t-1} \left\{ \frac{\sigma^{-2}_\eta (\text{roe}_{it} - x_{it}) k_{\mu_i}}{-\sigma^{-2}_\varepsilon (k_{\mu_i} - 1) (1 - \rho_x) [x_{it} - \rho x x_{it-1} - (1 - \rho_x) \mu_i] - \sigma^{-2}_\mu (\mu_i - \bar{\mu}) / T_i} \right\} . \quad (67)$$
Define the time $t$ moment condition as the score of the time $t$ likelihood function:

$$f_t(\lambda, \mu) = \begin{bmatrix} \frac{\partial l_t(\theta)}{\partial \lambda} \\ \frac{\partial l_t(\theta)}{\partial \mu} \end{bmatrix}.$$  \hspace{1cm} (68)

To obtain standard errors, we need the derivatives and spectral density of the moment conditions. The derivative of the average of the moment condition over the sample is:

$$d_T(\lambda, \mu) = \frac{\partial}{\partial (\lambda, \mu)} \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} \frac{\partial f_t}{\partial \lambda} \\ \frac{\partial f_t}{\partial \mu} \end{bmatrix},$$  \hspace{1cm} (69)

where

$$\frac{\partial^2 l_t}{\partial \lambda^2} = -\sum_{i=1}^{N} I_t \omega_{i,t-1} \left\{ \sigma_n^{-2} b_{it}^2 + \sigma_\epsilon^{-2} (b_{it} - \rho_\epsilon b_{it-1})^2 \right\},$$  \hspace{1cm} (70)

$$\frac{\partial^2 l_t}{\partial \lambda \partial \mu_i} = -I_t \omega_{i,t-1} \left\{ \sigma_n^{-2} k_{\mu_i} b_{it} + \sigma_\epsilon^{-2} (1 - \rho_\epsilon) \left( k_{\mu_i} - 1 \right) (b_{it} - \rho_\epsilon b_{it-1}) \right\},$$  \hspace{1cm} (71)

$$\frac{\partial^2 l_t}{\partial \mu_i \partial \mu_j} = -I_t \omega_{i,t-1} \left\{ \sigma_n^{-2} k_{\mu_i}^2 + \sigma_\epsilon^{-2} \left( k_{\mu_i} - 1 \right)^2 (1 - \rho_x)^2 + \sigma_\mu^{-2} / T \right\} \text{ if } i = j, \text{ zero otherwise.}$$  \hspace{1cm} (72)

We estimate the spectral density matrix via Newey-West with $q$ lags. First, define

$$R_T(v; \lambda, \mu) = \frac{1}{T} \sum_{t=1+u}^{T} f_t(\lambda, \mu) f_{t-u}(\lambda, \mu)'$$  \hspace{1cm} (73)

The estimate of the spectral density matrix is then

$$S_T = R_T(0; \lambda, \mu) + \sum_{v=1}^{q} \frac{q+1-v}{q+1} (R_T(v; \lambda, \mu) + R_T(v; \lambda, \mu)').$$  \hspace{1cm} (74)

Let $\hat{\theta} = (\hat{\lambda}, \hat{\mu})$ be our parameter estimates; and let $\theta_0$ be the true values of the parameters. Because the GMM system is exactly identified and the sample is reasonably long, we use efficient GMM standard errors. The estimated parameters are asymptotically distributed as:

$$\sqrt{T} \left( \hat{\theta} - \theta_0 \right) \sim N \left( 0, \left[ d_T \right]^{-1} S \left[ d_T \right]^{-1} \right).$$  \hspace{1cm} (75)
9.2 Testing a K-factor model with time-varying betas

Here we derive parameter estimates and test statistics for the model in which we allow for betas that vary over time for each of $K$ factors.

9.2.1 Estimating the model using QMLE/GMM

The first stage of our procedure remains unchanged. That is, we still estimate $\rho$, $\sigma^2_{\eta}$, and $\sigma^2_\varepsilon$ for the earnings, return variance, and beta processes based on the autocovariance functions of these processes and prior research (in the case of $\kappa$). In the second-stage of our procedure, we condition on the first-stage parameter estimates when estimating the remaining parameters, which include risk prices ($\lambda$), fixed effects in earnings ($\mu_i$), and persistent components of earnings ($x_{it}$). As before, we can use the present value relation to infer the value of $x_{it}$ as a function of parameters and the book-to-market ratio. Define a version of the book-to-market ratio ($bm_{it}^{ex}$) that excludes the variance and risk-free rate components as

$$bm_{it}^{ex} = \sum_{k=1}^{K} \lambda^{(k)} \left\{ \frac{1}{1 - \kappa} \beta_i^{(k)} + \frac{1}{1 - \rho_{\beta^{(k)}} \kappa} \left( \beta_{i,t}^{(k)} - \beta_i^{(k)} \right) \right\}$$

$$- \frac{1}{1 - \kappa} \mu_i - \frac{\rho_x}{1 - \kappa \rho_x} (x_{it} - \mu_i).$$

Then we can express $x_{it}$ as

$$x_{it} = \left( \sum_{k=1}^{K} \lambda^{(k)} \left\{ \frac{1}{1 - \kappa} \beta_i^{(k)} + \frac{1}{1 - \rho_{\beta^{(k)}} \kappa} \left( \beta_{i,t}^{(k)} - \beta_i^{(k)} \right) \right\} - \frac{1}{1 - \kappa} \mu_i - bm_{it}^{ex} \right) \frac{1 - \kappa \rho_x}{\rho_x} + \mu_i$$

$$= \left( \sum_{k=1}^{K} \lambda^{(k)} \left\{ \frac{1}{1 - \kappa} \beta_i^{(k)} + \frac{1}{1 - \rho_{\beta^{(k)}} \kappa} \left( \beta_{i,t}^{(k)} - \beta_i^{(k)} \right) \right\} - bm_{it}^{ex} \right) \frac{1 - \kappa \rho_x}{\rho_x} + \frac{\rho_x - 1}{(1 - \kappa) \rho_x} \mu_i.$$

For convenience, define $x_{it} = k_{x_{it}} + k_{\mu_i} \mu_i$, where

$$k_{x_{it}} = \left( \sum_{k=1}^{K} \lambda^{(k)} \left\{ \frac{1}{1 - \kappa} \beta_i^{(k)} + \frac{1}{1 - \rho_{\beta^{(k)}} \kappa} \left( \beta_{i,t}^{(k)} - \beta_i^{(k)} \right) \right\} - bm_{it}^{ex} \right) \frac{1 - \kappa \rho_x}{\rho_x},$$

$$k_{\mu_i} = \frac{\rho_x - 1}{(1 - \kappa) \rho_x}.$$

We propose a QMILE procedure based on a log-likelihood function of $roe$ where shocks are
normal and independent across firms. For firm $i$, the approximating log-likelihood function is:

$$l_i(\theta) = \text{const} + \sum_{t=t_i+1}^{t_i+T_i} \left\{ -\frac{1}{2} \omega_{i,t-1} \sigma_{\eta}^{-2} (\text{roe}_{i,t} - x_{i,t})^2 - \frac{1}{2} \omega_{i,t-1} \sigma_{\varepsilon}^{-2} (x_{it} - \rho_x x_{it-1} - (1 - \rho_x) \mu_i)^2 \right\}$$

$$-\frac{1}{2} \omega_i \sigma_{\mu}^{-2} (\mu_i - \bar{\mu})^2,$$

where the definitions of the number of observations for each firm, $T_i$, and the weights, $\omega_{i,t}$ and $\omega_i = \frac{1}{T_i} \sum_{t=t_i+1}^{t_i+T_i} \omega_{i,t-1}$, are the same as in the main text.

We now maximize the likelihood with respect to each parameter. Because the FOCs for $\mu_i$ and $\bar{\lambda}^{(k)}$ are linear, we proceed with solving for the optimal parameters exactly as in the main text.

Evaluating the FOC with respect to $\mu_i$, substituting for $x_{it}$, and solving for $\mu_i$ yields:

$$\mu_i = \frac{\sum_{t=t_i+1}^{t_i+T_i} \omega_{i,t-1} \left\{ \sigma_{\eta}^{-2} (\text{roe}_{i,t} - k_{x_{it}}) k_{\mu_i} - \sigma_{\varepsilon}^{-2} (k_{x_{it}} - \rho_x k_{x_{it-1}}) (k_{\mu_i} - 1) (1 - \rho_x) \right\} + \omega_i \sigma_{\mu}^{-2} \bar{\mu} / T_i \omega_i \left[ \sigma_{\eta}^{-2} k_{\mu_i}^2 + \sigma_{\varepsilon}^{-2} (1 - \rho_x)^2 (k_{\mu_i} - 1)^2 \right] + \omega_i \sigma_{\mu}^{-2}}{T_i \omega_i \left[ \sigma_{\eta}^{-2} k_{\mu_i}^2 + \sigma_{\varepsilon}^{-2} (1 - \rho_x)^2 (k_{\mu_i} - 1)^2 \right] + \omega_i \sigma_{\mu}^{-2}}$$

(80)

Evaluating the FOC of $\bar{\lambda}^{(k)}$, we obtain:

$$\sum_{i=1}^{N} \frac{\partial J_i}{\partial \bar{\lambda}^{(k)}} = \sum_{i=1}^{N} \sum_{t=t_i+1}^{t_i+T_i} \left\{ -\omega_{i,t-1} \sigma_{\varepsilon}^{-2} (x_{it} - \rho_x x_{it-1} - (1 - \rho_x) \mu_i) \left( \frac{\partial x_{it}}{\partial \bar{\lambda}^{(k)}} - \rho_x \frac{\partial x_{it-1}}{\partial \bar{\lambda}^{(k)}} \right) \right\}.$$

(82)

As before, we substitute the solution for $\mu_i$ and solve for $\bar{\lambda}$ to obtain the linear equation:

$$k_{\bar{\lambda}^{(k)}}^t \bar{\lambda} + z^{(k)} = 0,$$

where

$$k_{\bar{\lambda}^{(k)}} = -\sum_{i=1}^{N} \sum_{t=t_i+1}^{t_i+T_i} \left\{ \frac{\omega_{i,t-1} \sigma_{\varepsilon}^{-2} (b_{it}^t + k_{\mu_i} b_{it}^{(k)}) b_{it}^{(k)}}{+\omega_{i,t-1} \sigma_{\varepsilon}^{-2} (b_{it}^t - \rho_x b_{it-1}^t - (1 - \rho_x) (1 - k_{\mu_i}) b_{it}^{(k)}) (b_{it}^{(k)} - \rho_x b_{it-1}^{(k)})} \right\}.$$

(83)

$$z^{(k)} = -\sum_{i=1}^{N} \sum_{t=t_i+1}^{t_i+T_i} \left\{ \frac{-\omega_{i,t-1} \sigma_{\varepsilon}^{-2} (\text{roe}_{i,t} - (a_{it} + k_{\mu_i} a_{\mu_i})) b_{it}^{(k)}}{\omega_{i,t-1} \sigma_{\varepsilon}^{-2} (a_{it} - \rho_x a_{it-1} - (1 - \rho_x) (1 - k_{\mu_i}) a_{\mu_i}) (b_{it}^{(k)} - \rho_x b_{it-1}^{(k)})} \right\}.$$

(84)
Lastly, we substitute the solution to solve for the linear system for the $k$ risk prices: 

$$\tilde{\lambda} = -k_{\lambda}^{-1}z.$$ \hspace{1cm} (85)

Lastly, we substitute the $\tilde{\lambda}$ solution to solve for $\mu_i$:

$$\mu_i = \frac{\sum_{t=t_i+1}^{t_i+T_i} \{ \omega_{i,t-1} \sigma_\eta^{-2} (\text{roe}_{it} - k_{x_{it}}) k_{\mu_i} - \omega_{i,t-1} \sigma_\varepsilon^{-2} (k_{x_{it}} - \rho_x k_{x_{it-1}}) (k_{\mu_i} - 1) (1 - \rho_x) \} + \omega_i \sigma_\mu^{-2} \tilde{\mu}}{T_i \omega_i \left[ \sigma_\eta^{-2} k_{\mu_i}^2 + \sigma_\varepsilon^{-2} (1 - \rho_x)^2 (k_{\mu_i} - 1)^2 \right] + \omega_i \sigma_\mu^{-2}}.$$ \hspace{1cm} (86)

**Calculating standard errors** First, define an indicator variable that is one if firm $i$ has data for year $t$ and $t - 1$ and zero otherwise as $I_{it}$. The derivatives of the log-likelihood function are:

$$\frac{\partial l(\theta)}{\partial \lambda^{(k)}} = \sum_{t=1}^{T} \sum_{i=1}^{N} I_{it} \omega_{i,t-1} \left\{ \sigma_\eta^{-2} (\text{roe}_{it} - x_{it}) \left( \frac{\partial x_{it}}{\partial \lambda^{(k)}} \right) - \sigma_\varepsilon^{-2} (x_{it} - \rho_x x_{it-1} - (1 - \rho_x) \mu_i) \left( \frac{\partial x_{it}}{\partial \lambda^{(k)}} - \rho_x \frac{\partial x_{it-1}}{\partial \lambda^{(k)}} \right) \right\},$$ \hspace{1cm} (87)
where
\[
\frac{\partial x_{it}}{\partial \lambda^{(k)}} = \left\{ \frac{1}{1 - \kappa} \beta_{i}^{(k)} + \frac{1}{1 - \rho_{i,t}^{(k)}} \left( \beta_{i}^{(k)} - \beta_{i}^{(k)} \right) \right\} \frac{1 - \kappa \rho_{x}}{\rho_{x}}, \quad (88)
\]
\[
x_{it} = a_{it} + \sum_{k=1}^{K} b_{it}^{(k)} \lambda^{(k)} + c_{it}, \quad (89)
\]
and
\[
\frac{\partial l (\theta)}{\partial \mu_{i}} = \sum_{t=1}^{T} I_{it} \omega_{i,t-1} \left\{ \begin{array}{l}
\sigma_{\eta}^{-2} (\text{roe}_{it} - x_{it}) k_{\mu_{i}} \\
-\sigma_{\varepsilon}^{-2} (k_{\mu_{i}} - 1) (1 - \rho_{x}) [x_{it} - \rho_{x} x_{it-1} - (1 - \rho_{x}) \mu_{i}] - \sigma_{\mu}^{-2} (\mu_{i} - \bar{\mu}) / T_{i}
\end{array} \right\}, \quad (90)
\]
Define
\[
f_{t} (\lambda, \mu) = \left[ \begin{array}{c}
\frac{\partial l_{t}(\theta)}{\partial \lambda} \\
\frac{\partial l_{t}(\theta)}{\partial \mu}
\end{array} \right]_{(K+N) \times 1}, \quad (91)
\]
where
\[
\frac{\partial l_{t}(\theta)}{\partial \lambda^{(k)}} = \sum_{i=1}^{N} I_{it} \omega_{i,t-1} \left\{ \sigma_{\eta}^{-2} (\text{roe}_{it} - x_{it}) b_{it}^{(k)} - \sigma_{\varepsilon}^{-2} (x_{it} - \rho_{x} x_{it-1} - (1 - \rho_{x}) \mu_{i}) \left( b_{it}^{(k)} - \rho_{x} b_{it-1}^{(k)} \right) \right\} ; \quad (92)
\]
and
\[
\frac{\partial l_{t}(\theta)}{\partial \mu_{i}} = I_{it} \omega_{i,t-1} \left\{ \begin{array}{l}
\sigma_{\eta}^{-2} (\text{roe}_{it} - x_{it}) k_{\mu_{i}} \\
-\sigma_{\varepsilon}^{-2} (k_{\mu_{i}} - 1) (1 - \rho_{x}) [x_{it} - \rho_{x} x_{it-1} - (1 - \rho_{x}) \mu_{i}] - \sigma_{\mu}^{-2} (\mu_{i} - \bar{\mu}) / T_{i}
\end{array} \right\}. \quad (93)
\]
Next
\[
d_{T} (\lambda, \mu) = \frac{\partial}{\partial (\lambda, \mu)} \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial f_{t}}{\partial \lambda} \right]_{(K+N) \times (K+N)}, \quad (94)
\]
where
\[
\frac{\partial^{2} l_{t}(\theta)}{\partial \lambda^{(k)} \partial \lambda^{(j)}} = -\sum_{i=1}^{N} I_{it} \omega_{i,t-1} \left\{ \sigma_{\eta}^{-2} b_{it}^{(k)} b_{it}^{(j)} + \sigma_{\varepsilon}^{-2} \left( b_{it}^{(k)} - \rho_{x} b_{it-1}^{(k)} \right) \left( b_{it}^{(j)} - \rho_{x} b_{it-1}^{(j)} \right) \right\} ; \quad (95)
\]
\[
\frac{\partial^{2} l_{t}(\theta)}{\partial \lambda^{(k)} \partial \mu^{(i)}} = -I_{it} \omega_{i,t-1} \left\{ \sigma_{\eta}^{-2} k_{\mu_{i}} b_{it}^{(k)} + \sigma_{\varepsilon}^{-2} (1 - \rho_{x}) (k_{\mu_{i}} - 1) \left( b_{it}^{(k)} - \rho_{x} b_{it-1}^{(k)} \right) \right\} ; \quad (96)
\]
\[
\frac{\partial^{2} l_{t}(\theta)}{\partial \mu^{(i)} \partial \mu^{(j)}} = -I_{it} \omega_{i,t-1} \left\{ \sigma_{\eta}^{-2} k_{\mu_{i}} + \sigma_{\varepsilon}^{-2} (k_{\mu_{i}} - 1)^{2} (1 - \rho_{x})^{2} + \sigma_{\mu}^{-2} / T_{i} \right\} \text{ if } i = j, \quad (97)
\]
zero otherwise. \quad (98)
Next, estimate the spectral density matrix via Newey-West with \( q \) lags. First, define
\[
R_T (v; \lambda, \mu) = \frac{1}{T} \sum_{t=1+v}^{T} f_t (\lambda, \mu) f_{t-v} (\lambda, \mu)'
\]  
(99)

The estimate of the spectral density matrix is then
\[
S_T = R_T (0; \lambda, \mu) + \sum_{v=1}^{q} \frac{q + 1 - v}{q + 1} \left( R_T (v; \lambda, \mu) + R_T (v; \lambda, \mu)' \right).
\]  
(100)

Let \( \hat{\theta} = (\hat{\lambda}, \hat{\mu}) \) and let \( \theta_0 \) be the true values of the parameters. Then
\[
\sqrt{T} \left( \hat{\theta} - \theta_0 \right) \sim N \left( 0, \left[ d_T \right]^{-1} S \left[ d_T \right]^{-1} \right).
\]  
(101)
10 Tables

Table 1 - First-stage parameters

Table 1: This table shows the parameter estimates from the first-stage estimation. We model the dynamics of several variables, including \( \textit{roe}, \textit{r}_f, \textit{v}, \beta_{\text{mkt}}, \beta_{\text{smb}}, \beta_{\text{hml}}, \beta_{\text{rmw}}, \) and \( \beta_{\text{cma}} \), as AR(1) processes with measurement noise. For a generic variable \( z \) following such a process, we estimate the following parameters at an annual frequency: \( \mathbb{E}[z] \), mean; \( \rho[z] \), autocorrelation of the persistent component; \( \sigma_r[z] \), volatility of innovations in the persistent component; \( \sigma_\eta[z] \), volatility of measurement noise; and \( \sigma_\mu[z] \), standard deviation of heterogeneity in means across firms. The five betas are based on the Fama-French five-factor model. The left half of the table shows parameter estimates based on value weighting observations within each year, while the right half of the table shows parameter estimates based on equal weighting observations. The sample includes the years 1968 through 2014.

<table>
<thead>
<tr>
<th>( z )</th>
<th>( \mathbb{E}[z] )</th>
<th>( \rho[z] )</th>
<th>( \sigma_r[z] )</th>
<th>( \sigma_\eta[z] )</th>
<th>( \sigma_\mu[z] )</th>
<th>( \mathbb{E}[z] )</th>
<th>( \rho[z] )</th>
<th>( \sigma_r[z] )</th>
<th>( \sigma_\eta[z] )</th>
<th>( \sigma_\mu[z] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa )</td>
<td>0.960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.960</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \textit{roe} )</td>
<td>0.130</td>
<td>0.753</td>
<td>0.0587</td>
<td>0.104</td>
<td>0.0453</td>
<td>0.0365</td>
<td>0.670</td>
<td>0.155</td>
<td>0.252</td>
<td>0.0621</td>
</tr>
<tr>
<td>( \textit{r}_f )</td>
<td>0.0544</td>
<td>0.950</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0544</td>
<td>0.950</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \textit{v} )</td>
<td>0.105</td>
<td>0.883</td>
<td>0.0378</td>
<td>0.0926</td>
<td>0</td>
<td>0.341</td>
<td>0.885</td>
<td>0.138</td>
<td>0.260</td>
<td>0</td>
</tr>
<tr>
<td>( \beta_{\text{mkt}} )</td>
<td>1</td>
<td>0.938</td>
<td>0.0851</td>
<td>0.292</td>
<td>0</td>
<td>1</td>
<td>0.935</td>
<td>0.116</td>
<td>0.459</td>
<td>0</td>
</tr>
<tr>
<td>( \beta_{\text{smb}} )</td>
<td>0</td>
<td>0.953</td>
<td>0.137</td>
<td>0.372</td>
<td>0</td>
<td>0</td>
<td>0.978</td>
<td>0.0976</td>
<td>0.649</td>
<td>0</td>
</tr>
<tr>
<td>( \beta_{\text{hml}} )</td>
<td>0</td>
<td>0.953</td>
<td>0.162</td>
<td>0.685</td>
<td>0</td>
<td>0</td>
<td>0.958</td>
<td>0.129</td>
<td>0.955</td>
<td>0</td>
</tr>
<tr>
<td>( \beta_{\text{rmw}} )</td>
<td>0</td>
<td>0.853</td>
<td>0.295</td>
<td>0.687</td>
<td>0</td>
<td>0</td>
<td>0.915</td>
<td>0.195</td>
<td>1.071</td>
<td>0</td>
</tr>
<tr>
<td>( \beta_{\text{cma}} )</td>
<td>0</td>
<td>0.773</td>
<td>0.461</td>
<td>0.732</td>
<td>0</td>
<td>0</td>
<td>0.700</td>
<td>0.396</td>
<td>1.044</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2: This table shows quasi-maximum likelihood estimates (QMLE) of the prices of risk in the CAPM, Fama-French three-factor model, and Fama-French five-factor model for the case with constant risk prices. The sample includes the years 1968 through 2014. We report Newey-West standard errors with three lags in parentheses below the coefficient estimates. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>CAPM</th>
<th></th>
<th>FF3</th>
<th></th>
<th>FF5</th>
<th></th>
<th>FF5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VW</td>
<td>EW</td>
<td>VW</td>
<td>EW</td>
<td>VW</td>
<td>EW</td>
<td></td>
</tr>
<tr>
<td>(\lambda_0)</td>
<td>0.0289</td>
<td>0.0610***</td>
<td>0.0544***</td>
<td>0.0722***</td>
<td>0.0476**</td>
<td>0.0706***</td>
<td>0.0311</td>
</tr>
<tr>
<td></td>
<td>(0.0222)</td>
<td>(0.0191)</td>
<td>(0.0202)</td>
<td>(0.0160)</td>
<td>(0.0232)</td>
<td>(0.0168)</td>
<td>(0.0228)</td>
</tr>
<tr>
<td>(\lambda_{mkt})</td>
<td>0.0527***</td>
<td>0.0560***</td>
<td>0.0276***</td>
<td>0.0475***</td>
<td>0.0339***</td>
<td>0.0487***</td>
<td>0.0376***</td>
</tr>
<tr>
<td></td>
<td>(0.0098)</td>
<td>(0.0169)</td>
<td>(0.0076)</td>
<td>(0.0123)</td>
<td>(0.0107)</td>
<td>(0.0129)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>(\lambda_{smb})</td>
<td>-0.0032</td>
<td>-0.0087***</td>
<td>-0.0017</td>
<td>-0.0088***</td>
<td>-0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0048)</td>
<td>(0.0037)</td>
<td>(0.0042)</td>
<td>(0.0039)</td>
<td>(0.0049)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda_{hml})</td>
<td>0.0105***</td>
<td>0.0021</td>
<td>0.0126***</td>
<td>0.0024</td>
<td>0.0153***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0026)</td>
<td>(0.0068)</td>
<td>(0.0040)</td>
<td>(0.0074)</td>
<td>(0.0047)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda_{rmw})</td>
<td></td>
<td></td>
<td>-0.0160***</td>
<td>-0.0044</td>
<td>-0.0188***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0057)</td>
<td>(0.0039)</td>
<td>(0.0061)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda_{cma})</td>
<td>-0.0226</td>
<td>-0.0103</td>
<td>-0.0232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0165)</td>
<td>(0.0114)</td>
<td>(0.0165)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N)</td>
<td>2057</td>
<td>2057</td>
<td>2057</td>
<td>2057</td>
<td>2057</td>
<td>2057</td>
<td>2057</td>
</tr>
</tbody>
</table>
Table 3
Panel A - Analysis of Expected Returns and Earnings

Table 3: This table shows the decomposition of book-to-market ratios into expected cash flow and return components. Panel A shows the means and standard deviations of these variables, while Panel B shows the correlations among these variables. The sample includes the years 1968 through 2014.

<table>
<thead>
<tr>
<th></th>
<th>CAPM</th>
<th></th>
<th>FF5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E[z]$</td>
<td>$\sigma[z]$</td>
<td></td>
<td>$E[z]$</td>
</tr>
<tr>
<td>$bm$</td>
<td>$-0.7672$</td>
<td>$0.8110$</td>
<td></td>
<td>$-0.7672$</td>
</tr>
<tr>
<td>$bm^{ex}$</td>
<td>$-0.7794$</td>
<td>$0.8309$</td>
<td></td>
<td>$-0.7794$</td>
</tr>
<tr>
<td>$bm_{CF}$</td>
<td>$2.7671$</td>
<td>$0.8548$</td>
<td></td>
<td>$2.7607$</td>
</tr>
<tr>
<td>$bm_{DR}^{ex}$</td>
<td>$1.9878$</td>
<td>$0.0823$</td>
<td></td>
<td>$2.7184$</td>
</tr>
<tr>
<td>$\lambda \beta_{it}$</td>
<td>$0.0823$</td>
<td>$0.0094$</td>
<td></td>
<td>$0.0813$</td>
</tr>
<tr>
<td>$bm_{DR}^{ex}(1-\kappa)$</td>
<td>$0.0828$</td>
<td>$0.0034$</td>
<td></td>
<td>$0.0826$</td>
</tr>
<tr>
<td>$roe$</td>
<td>$0.1314$</td>
<td>$0.1287$</td>
<td></td>
<td>$0.1314$</td>
</tr>
<tr>
<td>$x_{it}$</td>
<td>$0.1460$</td>
<td>$0.2258$</td>
<td></td>
<td>$0.1459$</td>
</tr>
<tr>
<td>$\mu_{i}$</td>
<td>$0.1090$</td>
<td>$0.0212$</td>
<td></td>
<td>$0.1087$</td>
</tr>
</tbody>
</table>

Panel B - Correlations Among Components of Expected Returns and Earnings

Panel B shows the correlations among components of book-to-market ratios and earnings. Correlations based on the CAPM estimates appear below the diagonal, while correlations based on the FF5 estimates appear above the diagonal.

<table>
<thead>
<tr>
<th></th>
<th>CAPM</th>
<th></th>
<th>FF5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$bm$</td>
<td>$1.0000$</td>
<td>$-0.8274$</td>
<td></td>
<td>$-0.3958$</td>
</tr>
<tr>
<td>$bm^{ex}$</td>
<td>$-0.8274$</td>
<td>$1.0000$</td>
<td>$-0.9918$</td>
<td>$0.1685$</td>
</tr>
<tr>
<td>$bm_{CF}$</td>
<td>$-0.8307$</td>
<td>$-0.9956$</td>
<td>$1.0000$</td>
<td>$-0.0410$</td>
</tr>
<tr>
<td>$bm_{DR}^{ex}$</td>
<td>$-0.2757$</td>
<td>$-0.2459$</td>
<td>$-0.3353$</td>
<td>$1.0000$</td>
</tr>
<tr>
<td>$roe$</td>
<td>$-0.3958$</td>
<td>$-0.4387$</td>
<td></td>
<td>$0.4320$</td>
</tr>
<tr>
<td>$x_{it}$</td>
<td>$-0.6284$</td>
<td>$-0.8031$</td>
<td></td>
<td>$0.8039$</td>
</tr>
<tr>
<td>$\mu_{i}$</td>
<td>$-0.6134$</td>
<td>$-0.6737$</td>
<td></td>
<td>$0.6793$</td>
</tr>
</tbody>
</table>
Table 4 - Fama-MacBeth Regressions

Table 4: This table shows Fama-MacBeth regressions of firms’ annual excess returns on their forward-looking factor betas. The sample includes the years 1968 through 2014. In parentheses, we report White standard errors applied to the time-series of the Fama-MacBeth coefficients. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>CAPM</th>
<th></th>
<th>FF3</th>
<th></th>
<th>FF5</th>
<th></th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VW</td>
<td>EW</td>
<td>VW</td>
<td>EW</td>
<td>VW</td>
<td>EW</td>
<td>VW</td>
</tr>
<tr>
<td>intercept</td>
<td>0.0733***</td>
<td>0.1016***</td>
<td>0.0814***</td>
<td>0.0845**</td>
<td>0.0887***</td>
<td>0.0755**</td>
<td>intercept</td>
</tr>
<tr>
<td></td>
<td>(0.0253)</td>
<td>(0.0276)</td>
<td>(0.0277)</td>
<td>(0.0316)</td>
<td>(0.0282)</td>
<td>(0.0289)</td>
<td>(0.0861)</td>
</tr>
<tr>
<td>$\beta_{mkt}$</td>
<td>−0.0208</td>
<td>−0.0241</td>
<td>−0.0211</td>
<td>−0.0190</td>
<td>−0.0283</td>
<td>−0.0079</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0311)</td>
<td>(0.0322)</td>
<td>(0.0364)</td>
<td>(0.0351)</td>
<td>(0.0344)</td>
<td>(0.0344)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{smb}$</td>
<td>0.0260</td>
<td>0.0079</td>
<td>0.0318</td>
<td>0.0062</td>
<td>ln ME</td>
<td>−0.0067</td>
<td>−0.0078</td>
</tr>
<tr>
<td>$\beta_{hml}$</td>
<td>0.0129</td>
<td>0.0457*</td>
<td>0.0110</td>
<td>0.0318</td>
<td>ln B/M</td>
<td>0.0112</td>
<td>0.0274**</td>
</tr>
<tr>
<td>$\beta_{rmw}$</td>
<td>0.0075</td>
<td>0.0172</td>
<td>ln Prof</td>
<td>0.0895*</td>
<td>0.1234***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{ema}$</td>
<td>0.0209</td>
<td>0.0449</td>
<td>ln Inv</td>
<td>−0.0737**</td>
<td>−0.1098***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>2554</td>
<td>2554</td>
<td>2554</td>
<td>2554</td>
<td>2554</td>
<td>2554</td>
<td>2468</td>
</tr>
</tbody>
</table>
11 Figures

Figure 1 - Empirical and predicted autocovariance function of $roe_{i,t}$ and $\hat{\beta}_{i,t}^{mkt}$

Figure 1: The plots show empirical (solid lines) and predicted (dashed lines) autocovariance functions for $roe$ and one-year realized market betas across all firms and years. We estimate firms’ realized betas in each year using daily returns within the year. We compute the empirical autocovariance function of these realized betas. We predict autocovariances from a model in which estimated betas follow an AR(1) process with noise. We estimate the parameters, including persistence, variance of noise, and variance of innovations, that most closely match the autocovariances from this model of beta dynamics. We do the same for $roe$, which is the annual log return on equity as defined in the main text and reported in firms’ annual reports. The sample includes the years 1968 through 2014.
Figure 2: The plots show empirical (solid lines) and predicted (dashed lines) autocovariance functions for one-year realized HML and SMB betas across all firms and years. We estimate firms’ realized betas in each year using daily returns within the year. We compute the empirical autocovariance functions of these realized betas. We predict autocovariances from a model in which estimated betas follow an AR(1) process with noise. We estimate the parameters, including persistence, variance of noise, and variance of innovations, that most closely match the autocovariances from this model of beta dynamics. The sample includes the years 1968 through 2014.
Figure 3: The plots show empirical (solid lines) and predicted (dashed lines) autocovariance functions for one-year realized RMW and CMA betas across all firms and years. We estimate firms’ realized betas in each year using daily returns within the year. We compute the empirical autocovariance functions of these realized betas. We predict autocovariances from a model in which estimated betas follow an AR(1) process with noise. We estimate the parameters, including persistence, variance of noise, and variance of innovations, that most closely match the autocovariances from this model of beta dynamics. The sample includes the years 1968 through 2014.
Figure 4: The plots show empirical (solid lines) and predicted (dashed lines) autocovariance functions for one-year realized return variance across all firms and years. We estimate firm return variance in each year using daily returns within the year. We compute the empirical autocovariance function of these realized variances. We predict autocovariances from a model in which estimated variance follows an AR(1) process with noise. We estimate the parameters, including persistence, variance of noise, and variance of innovations, that most closely match the autocovariances from this model of variance dynamics. The sample includes the years 1968 through 2014.
Figure 5: The figure plots annual excess returns for 10 decile portfolios sorted on forward-looking market betas at different time horizons: one year ago (solid line) or one to ten years ago (dashed line). We estimate market betas from the Fama-French five-factor model. All portfolios are rebalanced annually and firms are value-weighted within each portfolio. The portfolio breakpoints are based on the distribution of betas for firms listed on the NYSE in each year. Each decile portfolio $j$ shown in the dashed line represents the average return of 10 cohort portfolios. Within each decile $j$, we construct each cohort portfolio $c$ from stocks that were ranked in beta decile $j$ as of $c$ years ago. The returns of the one-year portfolios span the years 1968 through 2014, while the 10-year portfolios span the years 1977 through 2014.
Figure 6: 1- and 10-year size and SMB beta decile portfolio returns

Figure 6: The bottom figure plots annual excess returns for 10 decile portfolios sorted on forward-looking SMB betas at different time horizons: one year ago (solid line) or one to ten years ago (dashed line). We estimate SMB betas from the Fama-French five-factor model. All portfolios are rebalanced annually and firms are value-weighted within each portfolio. The portfolio breakpoints are based on the distribution of betas for firms listed on the NYSE in each year. Each decile portfolio $j$ shown in the dashed line represents the average return of 10 cohort portfolios. Within each decile $j$, we construct each cohort portfolio $c$ from stocks that were ranked in beta decile $j$ as of $c$ years ago. The returns of the one-year portfolios span the years 1968 through 2014, while the 10-year portfolios span the years 1977 through 2014. The upper figure plots annual excess returns for 10 decile portfolios sorted on market equity using analogous methods.
Figure 7: The bottom figure plots annual excess returns for 10 decile portfolios sorted on forward-looking HML betas at different time horizons: one year ago (solid line) or one to ten years ago (dashed line). We estimate HML betas from the Fama-French five-factor model. All portfolios are rebalanced annually and firms are value-weighted within each portfolio. The portfolio breakpoints are based on the distribution of betas for firms listed on the NYSE in each year. Each decile portfolio $j$ shown in the dashed line represents the average return of 10 cohort portfolios. Within each decile $j$, we construct each cohort portfolio $c$ from stocks that were ranked in beta decile $j$ as of $c$ years ago. The returns of the one-year portfolios span the years 1968 through 2014, while the 10-year portfolios span the years 1977 through 2014. The upper figure plots annual excess returns for 10 decile portfolios sorted on book-to-market ratios using analogous methods.
Figure 8: 1- and 10-year profitability and RMW beta decile portfolio returns

Figure 8: The bottom figure plots annual excess returns for 10 decile portfolios sorted on forward-looking RMW betas at different time horizons: one year ago (solid line) or one to ten years ago (dashed line). We estimate RMW betas from the Fama-French five-factor model. All portfolios are rebalanced annually and firms are value-weighted within each portfolio. The portfolio breakpoints are based on the distribution of betas for firms listed on the NYSE in each year. Each decile portfolio $j$ shown in the dashed line represents the average return of 10 cohort portfolios. Within each decile $j$, we construct each cohort portfolio $c$ from stocks that were ranked in beta decile $j$ as of $c$ years ago. The returns of the one-year portfolios span the years 1968 through 2014, while the 10-year portfolios span the years 1977 through 2014. The upper figure plots annual excess returns for 10 decile portfolios sorted on profitability using analogous methods.
Figure 9: The bottom figure plots annual excess returns for 10 decile portfolios sorted on forward-looking CMA betas at different time horizons: one year ago (solid line) or one to ten years ago (dashed line). We estimate CMA betas from the Fama-French five-factor model. All portfolios are rebalanced annually and firms are value-weighted within each portfolio. The portfolio breakpoints are based on the distribution of betas for firms listed on the NYSE in each year. Each decile portfolio $j$ shown in the dashed line represents the average return of 10 cohort portfolios. Within each decile $j$, we construct each cohort portfolio $c$ from stocks that were ranked in beta decile $j$ as of $c$ years ago. The returns of the one-year portfolios span the years 1968 through 2014, while the 10-year portfolios span the years 1977 through 2014. The upper figure plots annual excess returns for 10 decile portfolios sorted on investment using analogous methods.