Working Paper

Linking Cross-Sectional and Aggregate Expected Returns

Serhiy Kozak
Stephen M. Ross School of Business
University of Michigan

Shrihari Santosh
Booth School of Business
University of Chicago

Ross School of Business Working Paper
Working Paper No. 1257
February 2015

Linking Cross-Sectional and Aggregate Expected Returns*

Serhiy Kozak[†], **Shrihari Santosh**[‡]

February 8, 2015

Abstract

We propose a one-state-variable ICAPM which rationalizes a large set of stock return anomalies, including size, value, and momentum. Differential covariance with news about future market discount rates drives observed cross-sectional patterns in expected returns. In response to discount-rate shocks, large, growth, and recent loser stocks outperform small, value, and winner stocks, respectively. Our interpretation is that increases in discount rates represent "bad" news, increasing investors' marginal utility of wealth. Ignoring this state variable causes drastic underestimation of the equilibrium price of "level risk" in bond returns. The model augmented with a "level" factor jointly prices stocks and bonds.

*We thank John Cochrane, Eugene Fama, Stefano Giglio, Valentin Haddad, Lars Peter Hansen, John Heaton, Bryan Kelly, Ralph Koijen, Stefan Nagel, Diogo Palhares, Jay Shanken, and seminar participants at Chicago, Emory, Maryland, Michigan, UIUC, Wisconsin for helpful comments and suggestions.

[†]Ross School of Business, University of Michigan. Email: sekozak@umich.edu.

[‡]R.H. Smith School of Business, University of Maryland. Email: shrihari@umd.edu.

1 Introduction

The logic of Merton (1973) suggests there exists a representation for the stochastic discount factor which is a linear combination of the return on the aggregate wealth portfolio and state variables which capture changes in the investment opportunity set[1]. We propose and test an equilibrium model in which shocks to risk aversion (sentiment) generate time-varying expected equity returns. We find support for the model's prediction that such shocks represent "bad" news for the representative investor and manifest in the cross-section of expected asset returns.

Empirically, we augment the one-factor CAPM[2] with a single additional factor, the shock to expected future excess market returns. This single additional factor is sufficient to rationalize a large cross-section of expected returns, including size, book-to-market ratio, recent performance (momentum), and anomaly portfolios studied in Novy-Marx and Velikov (2014). Since the CAPM completely fails in fitting the cross-section of assets we analyze, the success of our model comes entirely from the expected market return factor. This result is similar to Campbell and Vuolteenaho (2004) but with two important differences. First, we specify a general equilibrium model in which time variation in expected returns is endogenous. Second, instead of the typical vector auto-regression (VAR) approach, we use future realized returns to proxy for expected future returns (see Section 2.4). This provides a consistent estimate of expected returns under *any* information set. Our methodology produces very different cross-sectional patterns in factor loadings compared to the VAR approach. We find that growth stocks and large firm stocks outperform value stocks and small firm stocks, respectively, in response to an increase in expected market returns.

The pattern in loadings we find results in an opposite conclusion about the compensation an investor requires for bearing the risk of time-varying expected returns. We conclude that an increase in the expected market return corresponds to a drop in the investor's utility and hence an increase in his marginal utility of wealth. This implies the investor is willing to pay in order to eliminate this risk from his portfolio. We solve a general equilibrium model that

[1]Under loose restrictions, even with trading constraints, the absence of arbitrage implies the existence of a stochastic discount factor, M_{t+1}, such that the relationship $1 = E_t\left[M_{t+1}R_{i,t+1}\right]$ holds for all assets Skiadas (2009, Chapter 1). The SDF is also called a state-price density. Equivalently, no arbitrage implies the existence of a present-value function and of an equivalent martingale measure. If some wealth is not fully tradable then shocks to that wealth generates hedging demands in the space of tradable securities.

[2]See Sharpe (1964), Treynor (1961)

delivers exactly this prediction. In the model, expected market returns are high when the representative investor's risk aversion is high. These states of the world correspond to times of high marginal utility, consistent with our estimated price of risk. Further evidence of the relation between marginal utility and the market risk premium comes from bonds. Long-term bonds have higher covariance with innovations to market discount rates than do short-term bonds. Overall, these patterns are consistent with a "flight-to-quality" interpretation where effective investor risk aversion rises, stock prices fall, bond prices rise, and "good" companies outperform "bad" companies.

The second contribution of our paper is decomposing the average return differential between long and short term government bonds into a large positive differential due to loadings on "level risk" (interest-rate risk) and a large negative spread due to loadings on our expected market return factor. These net to a slightly upward sloping term structure of expected bond returns. Koijen et al. (2010) find a similar decomposition of bond risk premia using time-varying expected bond returns instead of stock returns. Our results suggest that analyzing fixed income securities in isolation can lead to erroneous conclusions about bond risks and risk premia.

Finally, we document that returns on the Fama-French-Carhart factors SMB, HML, and UMD are useful forecasters of future market excess returns, significant both statistically and economically. In our sample from August 1963 to December 2013, the three factors combined greatly outperform the dividend price ratio, D/P, in terms of R^2. This confirms the spreads in covariances seen in Section 3 are economically relevant. Additionally, we show that the first principal component of anomaly portfolios has similar forecasting ability to the three Fama-French-Carhart factors combined in terms of R^2 and its statistical significance is substantially higher than that of SMB, HML, and UMD.

2 The Model

2.1 Specification

Following Duffie and Epstein (1992a), we define a stochastic differential utility by two primitive functions, $f(C_t, J_t) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ and $A(J_t) : \mathbb{R} \to \mathbb{R}$. For a given consumption process C, the utility process J is the unique Ito process that satisfies the stochastic differential equation,

$$dJ_t = \left[-f(C_t, J_t) - \frac{1}{2} A(J_t) \|\boldsymbol{\sigma_{J,t}}\|^2\right] dt + \boldsymbol{\sigma_{J,t}} d\boldsymbol{Z_t}, \tag{1}$$

where $\boldsymbol{\sigma_{J,t}}$ is an $\mathbb{R}^3$-valued square-integrable utility-"volatility" process, J_t is the continuation utility for C, conditional on current information at time t, $f(C_t, J_t)$ is the flow utility, $A(J_t)$ is a variance multiplier that penalizes the variance of the utility "volatility" $\|\boldsymbol{\sigma_{J,t}}\| \equiv \boldsymbol{\sigma_{J,t}} \boldsymbol{\sigma}'_{\boldsymbol{J,t}}$, and $d\boldsymbol{Z_t}$ is a vector of shocks. A pair (f, A) is called an aggregator. We use a Kreps-Porteus (Epstein-Zin-Weil) aggregator, defined as

$$\begin{aligned} f(C, J) &= \frac{\delta}{\rho} \frac{C^\rho - J^\rho}{J^{\rho-1}} = \frac{\delta}{\rho} J \left[\left(\frac{C}{J}\right)^\rho - 1\right] && (2) \\ A(J) &= -\frac{\alpha}{J}, && (3) \end{aligned}$$

where $\rho = 1 - \frac{1}{\psi}$ and ψ is the elasticity of intertemporal substitution; δ is a subjective discount factor, and α is the risk-aversion parameter.

We consider a simple endowment with a dividend process given by:

$$\frac{dD}{D_t} = g_t dt + \boldsymbol{\sigma_D} d\boldsymbol{Z} \tag{4}$$

where g_t is the mean consumption growth,

$$dg = \phi_g (\bar{g} - g_t) dt + \boldsymbol{\sigma_g} d\boldsymbol{Z}. \tag{5}$$

We extend the utility specification, the parameter α to be stochastic,

$$d\alpha = \phi_\alpha (\bar{\alpha} - \alpha_t) dt + \boldsymbol{\sigma_\alpha} d\boldsymbol{Z_\alpha}, \tag{6}$$

This parameter can be interpreted as either time-varying risk aversion (Campbell and Cochrane, 1999, Dew Becker, 2011, Kozak, 2013) or time-varying ambiguity aversion with respect to model specification (Drechsler, 2013, Hansen and Sargent, 2008)[3].

Finally, agents' wealth evolves according to:

$$\frac{dW}{W_t} = (\theta_t \lambda_t + r_t - C_t)\, dt + \theta_t \boldsymbol{\sigma_R} d\boldsymbol{Z} \tag{7}$$

where θ_t is the share of wealth invested in the risky asset, λ_t is the expected excess return on the risky asset, and r_t is the risk-free rate.

2.2 Solution

Market clearing requires $C_t = D_t$, $\theta_t = 1$, and $W_t = P_t$. For simplicity, we further assume that the elasticity of intertemporal substitution is equal to unity. We solve the Hamilton-Jacobi-Bellman equation corresponding to the above problem in the Appendix A.

Theorem 1. *The equilibrium SDF corresponding to the problem in equations (1), (2), (3), and (6) is given by:*

$$\frac{d\Lambda}{\Lambda_t} = -r_t dt - \alpha_t dR_M - (\alpha_t - 1) \underbrace{a_\alpha}_{<0} \boldsymbol{\sigma_\alpha} d\boldsymbol{Z} - (\alpha_t - 1) \underbrace{a_g}_{>0} \boldsymbol{\sigma_g} d\boldsymbol{Z} \tag{8}$$

where R_M is the return on the market portfolio, a_α and a_g are constants provided in the Appendix A. When risk aversion is higher than 1, the price of risk-aversion risk is negative.

The equity risk premium is given by

$$\lambda_t = \lambda_0 + \alpha_t \times \boldsymbol{\sigma_D} \boldsymbol{\sigma}_{\boldsymbol{D}}' \tag{9}$$

[3] Variability in α generates time-varying prices of risk in the model. Our main results survive if we instead model time-varying quantity of risk (through time-varying volatility of consumption growth).

where λ_0 is a constant defined in Appendix A[4]*. Finally, the risk-free rate is:*

$$r_t = \delta + g_t - \lambda_t. \tag{10}$$

Proof. See Appendix A. □

Theorem 2. *The price of market discount rate risk is negative when investors' risk aversion is higher than unity. Investors dislike assets that pay off poorly in states when the discount rate is high, and require a higher risk premium on those assets.*

Proof. The result follows immediately from the fact that the coefficient a_α in (8) is negative, $a_\alpha < 0$ (see Appendix A). □

Theorem 2 is a result of endogenous variation of discount rates in the model (due to movements in risk aversion). This contrasts with the corresponding result for exogenous variation in expected returns as in Campbell (1996) and Campbell and Vuolteenaho (2004).

Theorem 3. *Given the results in Theorem 1, the three-factor model holds:*

$$\begin{aligned}
\mu_{i,t}^{R} &= \alpha_t \times cov\left(dR_{i,t}^{e}, dR_{M,t}^{e}\right) + a_g\left(\alpha_t - 1\right) \times cov\left(dR_{i,t}^{e}, dg_t\right) && (11)\\
&\quad + a_\alpha\left(\alpha_t - 1\right) \times cov_t\left(dR_{i,t}^{e}, d\lambda_t\right) \\
&\equiv \delta_{M,t} \times cov_t\left(R_{i,t+1}^{e}, R_{M,t+1}^{e}\right) + \delta_{\lambda,t} \times cov_t\left(R_{i,t+1}^{e}, \lambda_{t+1}\right) && (12)\\
&\quad + \delta_{g,t} \times cov_t\left(R_{i,t+1}^{e}, g_{t+1}\right)
\end{aligned}$$

where $\mu_{i,t}^{R}$ is the conditional excess return on asset i, $dR_{M,t}^{e}$ is the excess return on the market portfolio, $d\lambda_t$ are shocks to the equity risk premium, and dg_t are shocks to the growth rate.

Proof. See Appendix A. □

2.3 Unconditional Pricing

We discretize the model for empirical implementation and define $rx_{i,t+1}$ as the log return on an asset i over the time period $t \to t+1$. When instantaneous excess returns $dR_{i,t}^{e}$ are

[4] Given the dynamics of α_t this implies λ_t follows an AR(1) process. It allows us to capture persistence in the risk premium while keeping the model tractable. The assumption is similar to those made in the literature, which typically assume that a vector of state variables follows a VAR(1) process (see, for example, Campbell (1996), Campbell and Vuolteenaho (2004)).

normally distributed, corresponding discrete-time log excess returns $rx_{i,t+1}$ are as well.

The model above is specified conditionally with "shocks" as factors. In Appendix B.2 we derive an unconditional representation in terms of "levels". Theorem 4 below summarizes the result.

Assumption 1. *Consumption growth rate is constant,* $\boldsymbol{\sigma_g} = 0$.

Assumption 1 serves solely for expositional purposes in this section and will be relaxed in Section 4.

Theorem 4. *Given Assumption 1, and the conditional model in Equation 11, we obtain the following linear pricing relation:*

$$\forall i, t: \qquad E\left(rx_{i,t+1}\right) + \frac{V_i}{2} = cov\left(rx_{i,t+1}, rx_{M,t+1}\right) \times \hat{\delta}_M + cov\left(rx_{i,t+1}, \lambda_{t+1}\right) \times \hat{\delta}_\lambda \tag{13}$$

where $\hat{\delta}_M$ *and* $\hat{\delta}_\lambda$ *are two constant unconditional prices of market and expected returns risk, respectively. Appendix B.2 derives the link between these prices and the conditional ones in Appendix B, Equation 65.*

Proof. Refer to the proof of a more general Theorem 8 with an arbitrary number of factors and risk prices in Appendix B.2. □

Two aspects about this relation are worth emphasizing. First, all moments in the formula are unconditional and hence can be easily estimated using time-series regressions. Second, the formula involves the covariances with *levels* rather than shocks. This results from the time-series properties of α.

2.4 Expected Returns Factor

2.4.1 Using Future Realized Returns

The unconditional model in Equation 13 involves the expected market return λ_{t+1}, which is not directly observable by an econometrician. Typically, a predictive regression (or VAR) using macroeconomic and financial variables is employed to address this issue. A major limitation of such an approach is that it restricts the information set to a small number of variables. Since investors are presumed to condition on all available information, the forecasts from a predictive regression will not equal market expectations. This measurement error may bias estimates of covariance and risk premia.

We employ a novel methodology designed to circumvent the issue. We use future realized returns as an unbiased estimator of current risk premia required by investors. This is valid since for any information set $\mathcal{F}_{t+1}$ at time $t+1$,

$$rx_{M,t+2} = E_{t+1} rx_{M,t+2} + \varepsilon_{t+2} \tag{14}$$

$$E_{t+1}\left(\varepsilon_{t+2}\right) = E_{t+1}\left(\varepsilon_{t+2} | \mathcal{F}_{t+1}\right) = 0 \tag{15}$$

True (population) conditional covariances are also equal

$$cov_t\left(rx_{i,t+1}, rx_{M,t+2}\right) = cov_t\left(rx_{i,t+1}, E_{t+1} rx_{M,t+2}\right) \tag{16}$$

and thus a consistent estimator of the covariance on the RHS is also a consistent estimator of the covariance on the LHS. See Appendix B.3 for a more formal argument. Our empirical method allows us to test the model without directly estimating expected market returns[5]. Instead, our approach is based on investors', rather than the econometrician's, ability to forecast returns.

2.4.2 Moving Average of Future Realized Returns

An issue with using one period ahead realized returns is that the signal-to-noise ratio of such a proxy is low. Because expected returns are persistent, we can cumulate a longer series of future returns in order to improve the ratio. There is an obvious trade-off here: increasing the length of the cumulative sum improves the signal-to-noise ratio at a decreasing rate. We determine the horizon length empirically in later sections.

We now show that the cumulative sum of future realized returns can be used as a proxy for the market risk premium in Equation 13.

Theorem 5. *When the expected returns factor is measured over a long horizon, the unconditional relation in Equation 13 still holds, with the new prices of risk that are a linear transform of the ones in Equation 13:*

[5] Discount rates are often estimated using forecasting regressions

$$\forall i,t: \qquad E\left(rx_{i,t+1}\right)+\frac{V_i}{2} = cov\left(rx_{i,t+1},rx_{M,t+1}\right)\times\tilde{\delta}_M + cov\left(rx_{i,t+1},\lambda_{t+1:t+T}\right)\times\tilde{\delta}_\lambda \tag{17}$$

where $\lambda_{t+1:t+T} = E_{t+1} rx_{M,t+2:t+T+1}$ denotes expected market returns (risk premia) starting one period ahead for $T \geq 1$ periods, $\tilde{\delta}_M$ and $\tilde{\delta}_\lambda$ are two constant unconditional prices of market and expected returns risk, respectively. Appendix B.2 derives the link between these prices and the conditional ones in Equation 11 and Equation 82.

Proof. Refer to the proof of a more general Theorem 9 with an arbitrary number of factors and risk prices in Appendix B.2. □

When a moving average of future returns is used as an estimator of the expected market returns factor, we can still show consistency of covariance estimates. See Appendix B.2.1 for more details.

2.5 Empirical Relation

Equipped with the results in Theorem 4 and Theorem 5, we proceed with approximating the unconditional relation in Equation 17 to facilitate the empirical tests in the following Section 3. Appendix B.4 shows that with daily data, the unconditional pricing relation in Equation 17 can be represented as follows:

$$\forall i,t: \qquad E\left(R^e_{i,t+1}\right) = cov\left(rx_{i,t+1},rx_{M,t+1}\right)\times\tilde{\delta}_M + cov\left(rx_{i,t+1},\hat{\lambda}_{t+1:t+T}\right)\times\tilde{\delta}_{\hat{\lambda}} \tag{18}$$

where $R^e_i \equiv \frac{R_i}{R_f}$ is the level of excess returns and $\hat{\lambda}_{t+1:t+T} \equiv rx_{M,t+2:T} = \sum_{j=2}^{T} rx_{M,t+j}$. In deriving this relation, we use the fact that future realized returns can be used as an unbiased and consistent estimate of market risk premia (see Equation 87). When log returns are normally distributed, the relation is exact; otherwise it is an approximation. All random variables in Equation 18 are observed by an econometrician. We use the level of excess returns, R^e_i, on the LHS, and the covariances of log excess returns rx_i with the market rx_M and the future realized excess returns, $rx_{M,t+2:T} = \sum_{j=2}^{T} rx_{M,t+j}$ on the RHS.

3 Empirical Link Between Cross-Sectional and Aggregate Expected Returns

We estimate and test the expected return relation of Equation 18 using three sets of test assets. The first is the canonical 25 portfolios formed by a two-way sort of firms on market capitalization (ME) and book-to-market ratio (BE/ME), available at Ken French's website [6]. Lewellen et al. (2010) highlight a key issue in estimating and testing asset pricing models. When the test assets have a strong factor structure that captures much of the time-series variation as well as the cross-sectional variation in expected returns, a spurious model with many factors may still produce a remarkably good cross-sectional fit as long as the spurious factors are correlated with the "true" factors. This result is not due to sampling variation; it holds in population. A solution they propose is to add assets which increase the "dimensionality" of the test asset space.

In addition to the canonical 25 portfolios, we construct an alternative set of test assets. We include fifteen portfolios consisting of five value-weighted quintile portfolios each from independent sorts on size, book-to-market ratio, and momentum (prior 2-12)[7]. Fama and French (2008) show that sorting firms based on prior performance produces a reliable spread in average returns subsumed by neither the size effect or the book-to-market effect. Furthermore, the momentum factor, UMD (Carhart, 1997), is nearly uncorrelated with the size factor, SMB, and is negatively correlated with the book-to-market factor, HML (Fama and French, 1996). Including momentum sorted portfolios as test assets makes it decidedly more difficult for a model to fit the cross-section of expected returns. Our preferred estimation uses these fifteen portfolios; for robustness and for comparison with the literature, we perform all estimation and testing on the Fama-French 25 portfolios as well. Third, we estimate the model using 15 anomaly long-short portfolios from Novy-Marx and Velikov (2014)[8] (hereafter NMV). These capture many prominent features of the cross-section of returns[9]. The data are only available at monthly frequency. Finally, we always include the value-weight market and risk-free returns as test assets, both to show how well the model fits these assets

[6] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[7] Also available at Ken French's website.

[8] See Novy-Marx and Velikov (2014) for details on data construction. The data are available at http://rnm.simon.rochester.edu/data_lib/index.html

[9] We exclude a few high-frequency anomalies and those that did not survive in the past two decades.

and because they theoretically have the best measured factor loadings.

Following the spirit of Merton (1973) and the continuous-time model in Section 2, we use portfolio returns measured at daily frequency[10]. All returns are measured over the period 01-Aug-1963 to 31-Dec-2013. Using daily, rather than monthly, returns reduces the approximation error due to linearization of the exponential function that we rely on in deriving the estimation equation. As noted in Campbell and Vuolteenaho (2004), "July 1963 is when COMPUSTAT data become reliable and most of the evidence on the book-to-market anomaly is obtained from the post-1963 period". Furthermore, in the pre-1963 sample, the "CAPM explains the cross-section of stock returns reasonably well" (Campbell and Vuolteenaho, 2004). Since inflation estimates are not available daily[11], we use only excess log returns. These are, by construction, net of inflation (but may contain inflation risk premia).

As a proxy for the excess return on the wealth portfolio, $rx_{m,t}$, we use RmRf, the excess return on the value-weight portfolio of all common equity traded on the NYSE, AMEX, and NASDAQ. Of course the standard critique applies that there exist many assets, both traded (foreign securities) and non-traded (real-estate, human capital) that are not included in this portfolio (Roll, 1977). As discussed above, we construct $\widetilde{\lambda}_t = \sum_{i=1}^{H} rx_{m,t+i}$. For our preferred specification, we set $H = 126$ trading days, or one-half year. Though the theory of Section 2.4 implies the model should fit for all choices of H, finite length of the historic data series means that increasing H comes with a loss of precision in estimating $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$. Our results are quantitatively robust across various choices of H, from 3 months to 2 years, using daily or monthly frequency of returns (see Appendix D).

Table 1 shows the estimated covariances of asset returns with the factors. Panel A shows $Cov\left(rx_{i,t}, rx_{m,t}\right)$ with Newey-West standard errors in parentheses (252 lags; 1 year). Quintile 1 represents large firms, growth firms, and recent losers in relation to the dimensions, size, book-to-market, and momentum, respectively. Analogously, Quintile 5 represents small firms, value firms, and recent winners. The column to the right of Quintile 5 represents the Q5-Q1 spread portfolio. The FF column gives the estimates for the canonical Fama-French-Carhart (FFC) factors, SMB, HML, and UMD. The covariances match the well known pattern in market betas.

[10]We also replicated the analysis at monthly frequency and obtain very similar results (see Appendix D).
[11]Inflation is not well measured even at monthly frequency (Cecchetti, 1997).

Table 1: Covariances

	Q1	Q2	Q3	Q4	Q5	FF	Q5-Q1
			Panel A: $C_{i,M}$				
ME	9.46	9.15	8.97	8.87	7.59	-0.70	-1.87
BE/ME	9.91	8.92	8.69	8.23	8.57	-1.39	-1.34
Prior 2-12	12.05	9.98	9.35	9.24	10.69	-0.86	-1.36
			Panel B: $C_{i,\lambda}$				
ME	1.85	1.53	0.89	0.29	-0.19	-1.16	-2.04
	(2.33)	(1.44)	(0.81)	(0.25)	(-0.14)	(-1.85)	(-2.14)
BE/ME	2.07	1.50	1.06	1.10	-0.63	-1.84	-2.69
	(2.37)	(1.62)	(1.00)	(1.11)	(-0.60)	(-2.70)	(-2.80)
Prior 2-12	3.60	2.39	1.42	0.88	0.65	-2.10	-2.96
	(2.50)	(2.53)	(1.71)	(0.92)	(0.49)	(-2.01)	(-2.43)
			Panel C: $E\left[R_i^e\right]$				
ME	6.35	8.09	8.47	8.87	8.06	2.24	2.42
BE/ME	6.05	6.92	7.22	9.28	11.24	4.56	5.85
Prior 2-12	1.43	5.81	5.64	7.46	10.33	8.08	10.95

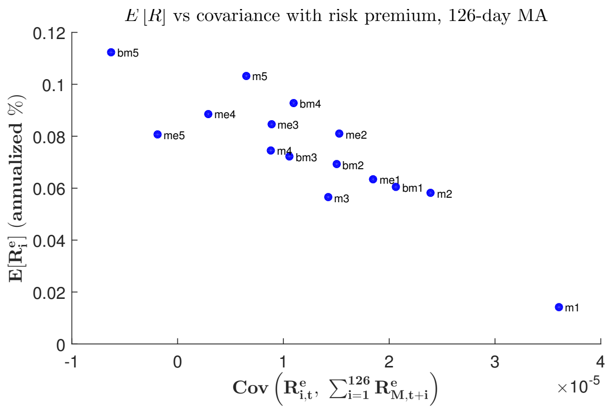
Notes: This table shows covariances and annualized mean returns estimated over 01-Aug-1963 to 31-Dec-2013. Panel A lists the covariances of portfolio returns with the market return, $C_{i,M}$. Panel B depicts the covariances of portfolio returns with the discount rate factor, $C_{i,\lambda}$. Panel C shows the the expected excess returns on each portfolio, $E\left[R_i^e\right]$. The column "FF" represent the Fama-French factors, *smb*, *hml*, and *umd*. t-statistics in parentheses are adjusted for serial correlation using Newey-West procedure with 252 lags (1 year). All covariances are scaled by 10^5.

Panel B reports $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ for the same portfolios. In all three dimensions (size, book-to-market, and momentum), $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ decreases from left to right. That is to say, when the "risk premium", λ_t, rises, small stocks are expected to fall more than large stocks, value stocks are expected to fall more than growth stocks, and recent winners are expected to fall more than recent losers. Though using realized market returns in place of expected returns produces consistent covariances estimates, they are less precisely estimated due to the extra noise present in realized returns. Still, the covariances of the spread portfolios with $\widetilde{\lambda}_t$ are statistically significantly different from zero and the covariances follow a reliable pattern, suggesting that our results are not spurious.

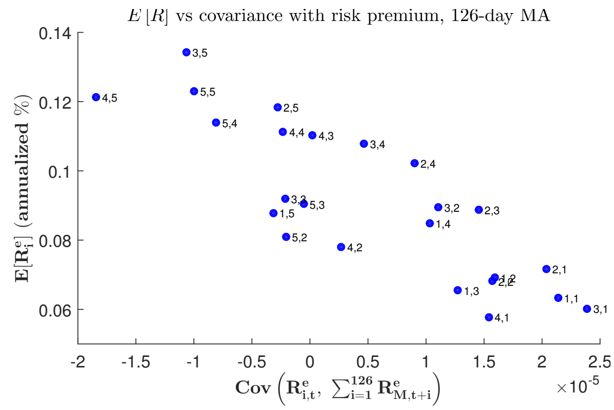
The model in Section 2 suggests these covariances could result from a "flight-to-quality" phenomenon, where the overall risk premium rises and the risk premium on "low quality" assets rises by even more. Our model is only one such motivation for time-varying expected returns. Indeed, as noted in (Shefrin, 2008, Chapter 30.3.3), the stochastic discount factor in a habit formation model (Campbell and Cochrane, 1999) is of the same general form as one based on a model of investor sentiment. Both deliver time-varying expected returns through an effectively time-varying risk-aversion parameter for the representative agent. We choose to directly model time-varying risk aversion. As we discuss below, the negative sign of δ_λ, suggests that even if these covariances are produced by sentiment driven "flight-to-quality" episodes, these are likely to be amplifications of fundamental shocks so that the risk-premium rises during aggregate "bad times". Panel C shows sample average returns which increase monotonically from left to right across quintiles, consistent with the well known size, value, and momentum phenomena. Panels B and C suggest a strong relationship between $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ and $E\left[r_{ix,t}\right]$, which can be clearly seen in Figure 1.

Figure 1 (a) plots sample values of $E\left[r_{ix,t}\right]$ vs $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ for the 15 quintile portfolios. Figure 1 (b) is the same plot for the 25 Fama-French portfolios and Figure 1 (c) shows the NMV anomalies. The graphs confirm that $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ and $E\left[r_{ix,t}\right]$ line up fairly well in the cross-section of assets, suggesting the λ_t risk factor captures the size, value, momentum effects, and NMV anomalies.

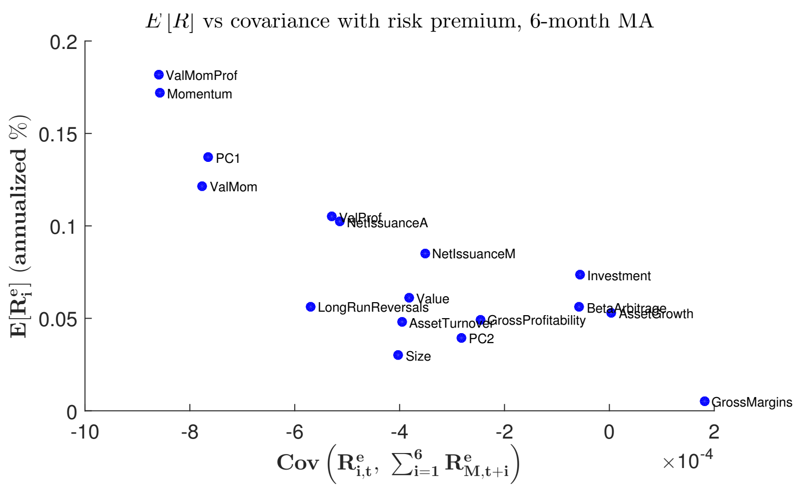
Figure 1: Univariate fit



(a) Quintile Portfolios



(b) Fama-French 25 Portfolios



(c) Anomaly Portfolios

Notes: The top left plot shows sample values of $E[r_{ix,t}]$ vs $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ for the 15 quintile portfolios: 5 size (me), 5 book-to-market (bm) and 5 momentum (m) sorted portofios. The plot in the top right panel depicts same results for the 25 Fama-French portfolios. The first number in portfolios labels refers to ME decile (1=large; 5=small); the second number corresponds to BE/ME decile (1=growth; 5=value). The bottom plot is for NMV portfolios using monthly data from August 1963 to December 2013. PC1 and PC2 are the first two principal components of NMV returns.

3.1 Estimation Results

We estimate the risk price vector $\delta = [\delta_M \ \delta_\lambda]'$ using GMM with a prespecified block-diagonal weighting matrix (Cochrane, 2001, Chapter 11.5). It is equivalent to the standard two-stage estimation procedure. $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ and $Cov\left(rx_{i,t}, rx_{M,t}\right)$ are estimated in the first stage by just-identified GMM, which yields the standard formulas for sample covariance. In the second stage, we minimize the mean-squared model pricing errors of the test assets. This is equivalent to an OLS regression of sample mean returns on the covariances estimated from the first stage. In addition to our two-factor ICAPM, we estimate the Sharpe-Lintner CAPM and well as the Fama-French model augmented with the *umd* (momentum) factor of Carhart (1997). For ease of comparison, all models are written and estimated in terms of covariances instead of regression βs. Below is a summary of the pricing equations, where δs are interpreted as risk prices (coefficients in the SDF):

$$\text{2-Factor ICAPM: } E\left[R_i^e\right] = C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda \tag{19}$$

$$\text{2-Factor ICAPM, Unrestricted: } E\left[R_i^e\right] = \alpha + C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda \tag{20}$$

$$\text{CAPM, Restricted: } E\left[R_i^e\right] = C_{i,M}\delta_M \tag{21}$$

$$\text{CAPM, Unrestricted: } E\left[R_i^e\right] = \alpha + C_{i,M}\delta_M \tag{22}$$

$$\begin{aligned}\text{4-Factor FF, Unrestricted: } E\left[R_i^e\right] &= \alpha + C_{i,M}\delta_M + C_{i,smb}\delta_{smb} + C_{i,hml}\delta_{hml} \\ &\quad + C_{i,umd}\delta_{umd}\end{aligned} \tag{23}$$

$$\text{4-Factor FF, Restricted: } E\left[R_i^e\right] = C_{i,M}\delta_M + C_{i,smb}\delta_{smb} + C_{i,hml}\delta_{hml} + C_{i,umd}\delta_{umd} \tag{24}$$

where $C_{i,X} \equiv Cov\left[rx_{i,t},\ X_t\right]$.

Estimated risk prices are given in Table 2 along with sample R^2 and mean absolute pricing errors[12]. Quantitatively, our two-factor ICAPM fits the cross-section of average returns nearly as well as the 4-factor FFC model. The estimated intercept is nearly zero, both statistically and economically. Though $Cov\left(rx_{i,t}, \widetilde{\lambda}_t\right)$ is not very well estimated for any individual test asset, the cross-sectional spread in covariances is strong enough to yield precise estimation of δ_λ. $H_0 : \delta_\lambda = 0$ is rejected for all conventional significance levels. Covariance with the expected return factor is able to capture a large portion of the cross-

[12]For all models, we don't impose the GLS restriction that the model exactly prices the factors. Imposing the restriction predictably reduces R^2 and increases $MAPE$.

Table 2: Risk Price Estimates

	α	δ_M	δ_λ	δ_{smb}	δ_{hml}	δ_{umd}	R^2	$MAPE$
2-Factor ICAPM	-	4.54	-11	-	-	-	80.9	0.959
		(4.44)	(-5.29)					
	0.812	4.17	-10.7	-	-	-	81.5	0.975
	(1.44)	(4.25)	(-5.2)					
CAPM	-	3.01	-	-	-	-	10.8	1.95
		(3.12)						
	3.12	1.7	-	-	-	-	19.2	1.95
	(4.49)	(1.82)						
4-Factor FF	-	5.54	-	3.83	12.9	8.21	87.2	0.874
		(4.66)		(1.12)	(3.84)	(5.02)		
	0.706	5.19	-	3.71	12.6	8.07	87.6	0.896
	(1.91)	(4.53)		(1.08)	(3.79)	(4.94)		

Notes: This table shows premia estimated from 01-Aug-1963 to 31-Dec-2013 for the two-factor ICAPM, the CAPM, and the augmented Fama-French model. The test assets are value-weighted quintile portfolios sorted on ME, BE/ME, and Prior2-12. α is annualized and "-" indicates that the intercept is restricted to zero. MAPE is average absolute pricing error, annualized. Moving block bootstrap t-statistics are in parentheses.

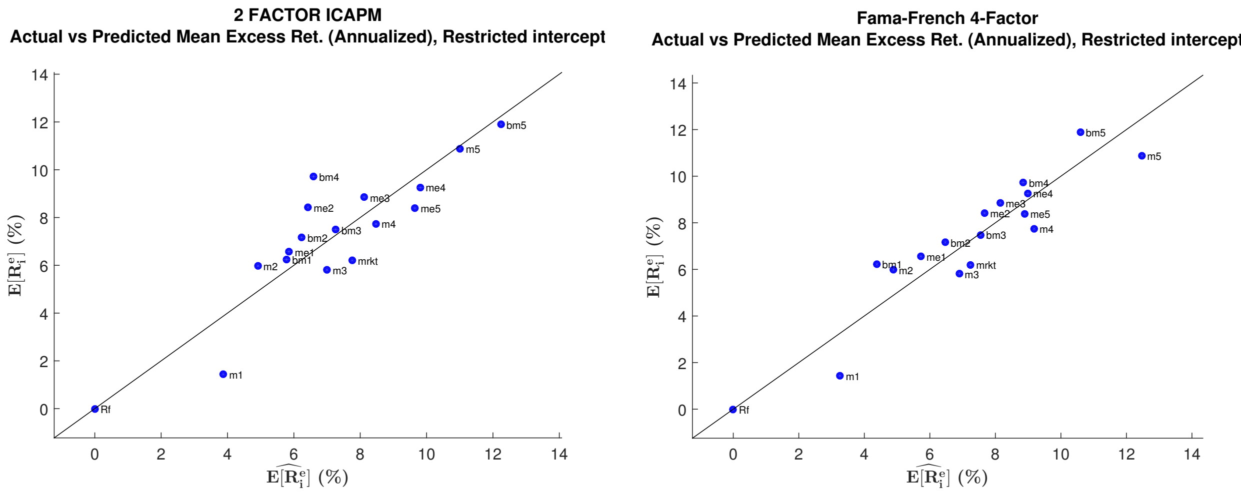
sectional variation in average returns due to the size, book-to-market, and momentum effects. Standard errors are calculated using the moving block bootstrap methodology (Horowitz, 2001) and are consistent across various choices of block size.

The cross-sectional fit of the ICAPM and 4-factor Fama-French models are given graphically in Figure 2. The graphs plot model implied mean excess returns on the horizontal and sample average returns on the vertical axis. The 45° line represents a model with perfect in-sample fit (100% R^2).

3.2 The Sign of δ_λ

Campbell and Vuolteenaho (2004) perform a similar pricing exercise but with different empirical methodology and theoretic motivation. They find a positive price of expected returns risk, though the estimate isn't statistically significant. From our estimation, the probability

Figure 2: Performance of the ICAPM and 4-Factor Fama-French models



Notes: 2-Factor ICAPM with restricted intercept on the left, and 4-factor Fama-French with restricted intercept on the right. m1-m5 correspond to momentum quintile portfolios (losers to winners). bm1-bm5 correspond to book-to-market quintiles (growth to value). me1-me5 correspond to size quintiles (large to small).

that the price of risk is positive is $P[\delta_\lambda > 0] = 0.002\%$ [13]). The alternative approaches reach very different conclusions.

We find that growth stocks and large firm stocks outperform value stocks and small firm stocks, respectively, in response to an increase in market expected returns, opposite to the pattern in Campbell and Vuolteenaho (2004). The different pattern in loadings produces a different conclusion about the compensation an investor requires for bearing the risk of time-varying expected returns. We conclude that an increase in the expected market return corresponds to a drop in the investor's utility and hence an increase in his marginal utility of wealth. This implies the investor is willing to pay *in order to eliminate* this risk from his portfolio, as predicted by our model. In contrast, Campbell and Vuolteenaho (2004) find that an investor is willing to pay *to increase his exposure* to this risk. Campbell and Vuolteenaho (2004) theoretically motivate their finding using a portfolio problem for a long-term investor who treats discount rate shocks as exogenous. It isn't surprising that an exogenous increase in expected returns (with no change in risk) is beneficial to an investor[14].

[13] p-value from a one-sided t-test of $H_0 : \lambda \geq 0$

[14] Theoretically, this depends on the coefficients of relative risk aversion and elasticity of intertemporal

We endogenize time-varying discount rates and find that a representative agent suffers when expected returns rise.

3.3 Factor Mimicking Portfolio

The two-stage OLS procedure for estimating stochastic discount factors suffers from many problems related to samples size and factor structure in the covariance of test asset returns. Lewellen et al. (2010) highlight these concerns and offer some suggestions:

1. Increase the dimensionality of the test assets relative to the dimension of the SDF.

2. Impose theoretic restrictions: "zero-beta rates should be close to the risk-free rate, the risk premium on a factor portfolio should be close to its average excess return". This is essentially using GLS instead of OLS with the factor included as a test asset.

3. Report GLS R^2 since (a) it "is completely determined by the factor's proximity to the minimum-variance boundary ... but the OLS R^2 can, in principle, be anything" and (b) "in practice, obtaining a high GLS R^2 represents a more stringent hurdle than obtaining a high OLS R^2."

4. Report confidence intervals for the cross-sectional R^2.

We already addressed issue (1) by having only one factor to "explain" three dimensions of average returns. Table 2 shows that estimates with and without restrictions on the zero-beta rate are nearly identical. Since our expected return factor is not an excess returns, we cannot directly include it as a test asset and check the restriction in (2). We can however, include a maximally correlated (mimicking) portfolio, i.e., regress the factor on the space of excess returns and use the fitted values. As shown in Cochrane (2001, Ch. 4), this yields identical OLS estimates of covariances, risk prices, pricing errors, and R^2. Because our test assets are highly correlated, in small sample the mimicking portfolio will have unrealistic extreme long/short positions. To mitigate concerns of overfitting, we instead construct $\widehat{\lambda}_t = proj\left(\lambda_t \middle| \left[\begin{array}{cccc} mrkt_t & smb_t & hml_t & umd_t \end{array}\right]\right)$. This is the linear combination of the four Fama-French-Carhart factors which has maximal correlation with our original

substitution. The result obtains for generally accepted parameter values.

Table 3: GLS Estimation

	δ_M	δ_λ	$E[r_M]$	$E[r_\lambda]$	R^2	$MAPE$
OLS	4.54	-12.7	6.87	-8.98	84.6	0.885
	(4.35)	(-4.85)				
GLS	4.49	-14	6.2	-10.1	80.8	0.96
	(4.73)	(-6.12)				

Notes: OLS estimates are from the standard two-step FM procedure. GLS restricts the model to exactly fit the market and the mimicking portfolio's in-sample average returns, ignoring pricing errors on other assets. $E[r_M]$ and $E[r_\lambda]$ are the model implied annualized expected excess returns on the market and mimicking portfolios, respectively. For GLS these are, by construction, equal to annualized sample averages.

expected return factor. Since it is an expected return we can include it as a test asset and use GLS methods.

With the factor mimicking portfolio, we can address points (2) and (3) above. For the remainder of this section, we treat the mimicking portfolio as the factor and address OLS vs GLS[15]. GLS restricts the model to exactly fit the market and the mimicking portfolio's average in-sample returns, ignoring pricing errors on other assets. Table 3 shows the estimated SDF using both methods. $E[r_M]$ and $E[r_\lambda]$ are the model implied annualized expected excess returns on the market and mimicking portfolios, respectively. For GLS these are, by construction, equal to sample averages.

The results are similar across methods. In particular, the model implied expected returns on the two factors (market and λ mimicking portfolio) are similar for OLS and GLS. This addresses point (2) from Lewellen et al. (2010). The GLS R^2 is mechanically lower than OLS R^2 but not substantially so. Bootstrap simulation rejects the null[16] of $R^2 = 0$ with $p \approx 1\%$. The [1%, 99%] R^2 interval under the null is $[-139\%, 81\%]$.

[15] When only a subset of test assets are used to construct the mimicking portfolio, there is no guarantee that estimated risk prices, etc will remain unchanged. Here, SMB, HML and UMD were not included in the original test assets so "anything" can happen. Still, δ and R^2 values are similar. We ignore sampling uncertainty in b_{SMB}, b_{HML}, b_{UMD} when reporting test statistics using the factor mimicking portfolio. This likely does not bias our results greatly since the Newey-West t-statistics on b_{SMB}, b_{HML}, b_{UMD} are between 2 and 3.

[16] One-sided test of $H_0: R^2 \leq 0$.

3.3.1 Testing the ICAPM

A common way to test an asset pricing model is to check whether cross-sectional pricing errors are jointly different from zero. Instead we test the ICAPM against a specific alternative, the FFC 4-factor model. If the 2-factor ICAPM is literally true we can derive the following restriction on the SDF loadings, assuming $proj\left(\lambda_t \middle\| \begin{bmatrix} mrkt_t & smb_t & hml_t & umd_t \end{bmatrix}\right)$ is the true mimicking portfolio (including other assets doesn't improve model fit[17]).

$$\begin{aligned} m_t &= \widetilde{\delta}_M \cdot mrkt_t + \delta_\lambda \lambda_t \\ &= \widetilde{\delta}_M \cdot mrkt_t + \delta_\lambda proj\left(\lambda_t \middle\| \begin{bmatrix} mrkt & smb & hml & umd \end{bmatrix}\right) + \epsilon_t && (25) \\ &\forall\ i,\ cov\left(r_i,\ \epsilon\right) = 0\ \Rightarrow \text{equivalently} \\ m_t &= \widetilde{\delta}_M \cdot mrkt_t + \delta_\lambda proj\left(\lambda_t \middle\| \begin{bmatrix} mrkt & smb & hml & umd \end{bmatrix}\right) && (26) \\ &= \widetilde{\delta}_M \cdot mrkt_t + \delta_\lambda \left[b_M \cdot mrkt_t + b_{smb} \cdot smb_t + b_{hml} \cdot hml_t + b_{umd} \cdot umd_t\right] && (27) \end{aligned}$$

The unrestricted four-factor SDF is $m_t = \delta_M \cdot mrkt_t + \delta_{smb} \cdot smb_t + \delta_{hml} \cdot hml_t + \delta_{umd} \cdot umd_t$. If the 2-factor model is strictly true, then we should have $\delta_M = \widetilde{\delta}_M + \delta_\lambda \cdot b_M$ and $\delta_i = \delta_\lambda \cdot b_i$ for $i \in \left\{ smb \quad hml \quad umd \right\}$. Table 4 shows the implied and direct coefficients on $\begin{bmatrix} mrkt & smb & hml & umd \end{bmatrix}$ in the SDF. The unrestricted coefficient on SMB is smaller than the ICAPM implied value and the coefficient on UMD is larger than its implied value. The implied and direct coefficients on HML and Market are nearly identical. This is a manifestation of the αs (pricing errors) seen in the left panel of Figure 2. SMB has a lower average return than predicted by it's covariance with λ and UMD has a positive ICAPM α. Since SDF weights are proportional to average returns when the factors are uncorrelated[18] the ICAPM αs translates directly into the difference in coefficients. The last row of Table 4 gives the p-values from Wald tests of equality of implied and unrestricted 4-factor SDF parameters[19]. There is almost no evidence in the data to reject the ICAPM in favor of the 4-factor model. This lends further support to the ICAPM, as the 4-factor model was

[17] This assumption is approximately true in the data.

[18] Market, SMB, HML, and UMD are nearly uncorrelated. The map from expected return to SDF weight depends on factor variances as well.

[19] We ignore uncertainty in the ICAPM estimates of δ_M δ_λ as well as the projection parameters b_M, b_{smb} b_{hml}, b_{hmd}. This omission biases the test in favor of rejecting the ICAPM, so it does not affect our conclusion that the data do not support rejection.

Table 4: SDF Restriction

	Restricted ZB-rate		Unrestricted ZB-rate	
	Implied	4-Factor	Implied	4-Factor
Market	5.38	5.54	4.98	5.19
SMB	4.96	3.83	4.83	3.71
HML	11.30	12.89	10.99	12.58
UMD	6.29	8.21	6.12	8.07
p-value	*0.79*		*0.76*	

Notes: Implied coefficients for market are $\widetilde{\delta}_M + \delta_\lambda \cdot b_M$ where δs are from the first row of Table 2 and b_M is from $proj\left(\lambda_t \middle| \begin{bmatrix} mrkt & smb & hml & umd \end{bmatrix}\right)$. The remaining implied coefficients are $\delta_i = \delta_\lambda \cdot b_i$ with δ_λ and b_i from the same source. p-values are from a χ^2 test of equality of implied and actual 4-factor coefficients.

specifically designed to fit these test assets.

3.4 Anomaly Portfolios

As a further test of the model, we estimate it using the cross-section of anomaly portfolios[20] defined in Novy-Marx and Velikov (2014). These represent a broad set of empirical regularities with seemingly very different fundamental drivers. However, as shown in Kozak et al. (2014), a pricing model using the first two principal components of returns produces 90% R^2 in fitting the cross-section of average returns (though much less in the time-series). This gives us hope that one (or a few) basic economic mechanisms are responsible for the variety of anomalies.

Table 5 shows parameters of the pricing models in Equations (19)-(24), estimated using monthly NMV returns[21]. We also include two principal components (labeled PC1 and PC2) as test assets. The estimated risk prices are similar to those in Table 2. Now, however, the 2-factor ICAPM significantly outperforms the 4-factor FF model in fitting the cross-section of

[20] These data are only available at monthly frequency.

[21] Because the NVM portfolios are long-short, they tend to have CAPM βs near zero. Sample β estimates are very noisy and unreliable estimates of δ_M. To address this issue, we orthogonalize each anomaly returns against the market portfolio before estimating asset pricing models. This procedure is equivalent to giving infinite weight to the market portfolio in the second stage estimation (like GLS). Without this restriction, the model R^2 slightly improves, at the cost of very poor fit for $E\left[r_M\right]$.

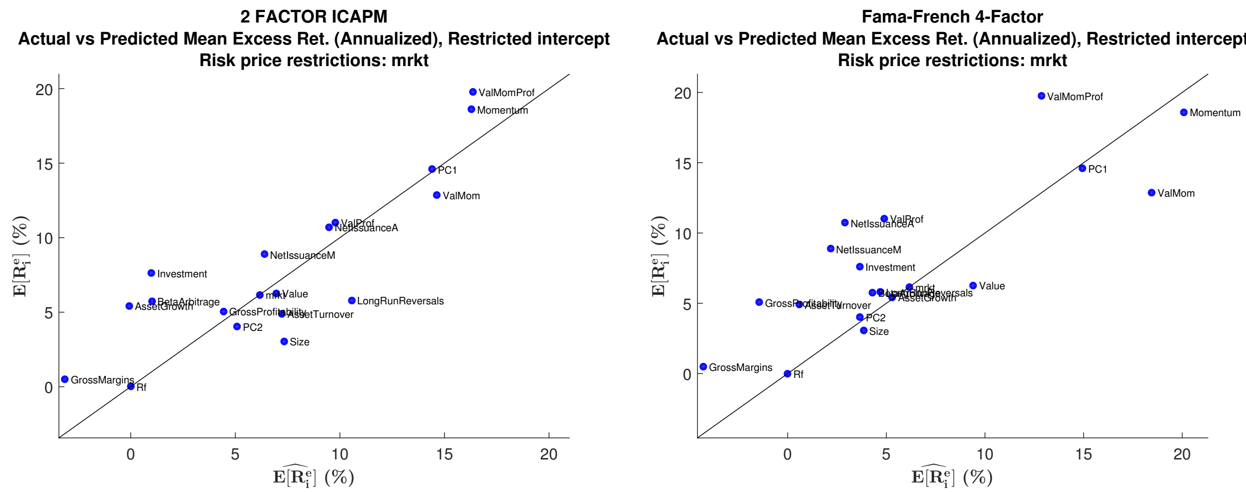
Table 5: Risk Price Estimates

	α	δ_M	δ_λ	δ_{smb}	δ_{hml}	δ_{umd}	R^2	$MAPE$
2-Factor ICAPM	-	3.86	-14.7	-	-	-	64.3	2.36
		(8.75)	(-13.5)					
	2.39	2.58	-11.5	-	-	-	72.8	2.17
	(4.9)	(5.27)	(-12.2)					
CAPM	-	2.43	-	-	-	-	-214	7.66
		(5.39)						
	7.65	-0.668	-	-	-	-	5.8	3.92
	(12.8)	(-1.21)						
4-Factor FF	-	4.42	-	2.57	9.62	6.39	37.8	3.09
		(7.53)		(2.04)	(9.65)	(10.9)		
	4.49	1.84	-	0.767	4.53	4.59	73.4	1.99
	(7.72)	(2.94)		(0.597)	(3.92)	(7.98)		

Notes: This table shows premia estimated from August 1963 to December 2013 for the two-factor ICAPM, the CAPM, and the augmented Fama-French model. The test assets are NMV anomaly portfolios. α is annualized and "-" indicates that the intercept is restricted to zero. MAPE is average absolute pricing error, annualized. Moving block bootstrap t-statistics are in parentheses.

average returns (with restricted intercept). Notably, the t-statistics on risk prices are much higher than in Table 2. This is due to the weaker factor structure in the NMV portfolios as compared to the test assets used before (quintile portfolios sorted on ME, BE/ME, and Prior2-12). Weaker cross-sectional correlation of returns results in more "effective" test assets, improving statistical power. Figure 3 shows graphically the fit of the ICAPM and 4-factor FFC models (with restricted zero-beta rate). The ICAPM is visibly superior to the 4-factor model in fitting the average NMV returns.

Figure 3: Performance of the ICAPM and 4-Factor Fama-French models



Notes: 2-Factor ICAPM with restricted intercept on the left, and 4-factor FFC with restricted intercept on the right. Assets are monthly NMV anomaly long-short portfolios from August 1963 to December 2013. PC1 and PC2 denote two largest principal components of NMV returns.

4 Bond Risks and Risk Premia

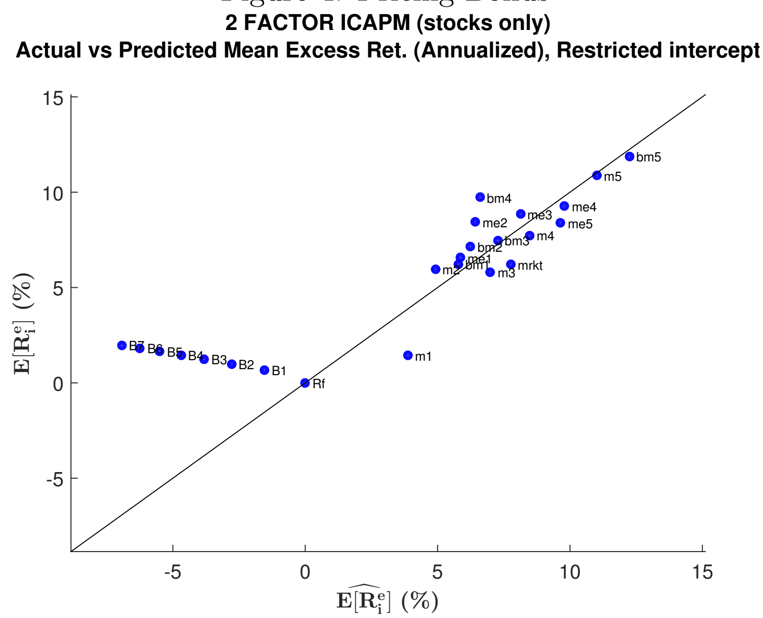
4.1 Pricing Bonds

Whereas there are numerous papers which explore risk premia separately for equities and fixed income securities, few study these assets in a unified framework[22]. We extend our analysis to include risk-free government bonds and interest rate risk. Figure 4 illustrates that the unconditional model in Equation 18 fails to price bond excess returns of maturities from 1 to 7 years. Pricing errors are big and the slope is completely wrong. Clearly, the market and expected returns factors are not sufficient to explain the risk compensation required for holding these securities.

The three factor extension in Equation 11 specifies dg or shocks to consumption growth as an additional term in the SDF. The factor helps price bonds and stocks jointly. Equation 10 allows us to substitute changes in the short term interest rate in place of dg. As an empirical proxy we use the excess return on a 1-year zero coupon Treasury bond, which is almost perfectly negatively correlated with changes in the 1-year yield. Cochrane and Piazzesi

[22] Recent work in this area includes Koijen et al. (2010).

Figure 4: Pricing Bonds



Notes: We plot the fitted and sample mean values of expected returns in the model Equation 17. Test assets include the stock portfolios we used before as well as 7 bonds with maturities from 1 to 7 years, labeled B1-B7, respectively.

(2008) and others show that a level factor[23] is a single factor sufficient to explain average excess bond returns. Empirically, yield changes at any maturity have correlation of at least 85% with the level factor. Therefore, using convex combination of yield changes as the factor fits the cross-section of bonds returns quite well. For convenience, we term our bond return as the "level factor". The single factor bond model is:

$$E_t\left(rx_{i,t+1}\right) + \frac{1}{2}\sigma_t^2\left(rx_{i,t+1}\right) = cov_t\left(rx_{i,t+1}, rx_{B,t+1}\right) \times \delta_{B,t} \tag{28}$$

The extended model from Section 2 is:

$$\begin{aligned} E_t\left(rx_{i,t+1}\right) + \frac{1}{2}\sigma_t^2\left(rx_{i,t+1}\right) &= cov_t\left(rx_{i,t+1}, rx_{M,t+1}\right) \times \delta_{M,t} \\ &+ cov_t\left(rx_{i,t+1}, \lambda_{t+1}\right) \times \delta_{\lambda,t} \\ &+ cov_t\left(rx_{i,t+1}, rx_{B,t+1}\right) \times \delta_{B,t} \end{aligned} \tag{29}$$

[23]"Level" is approximately the average of changes in yields across all maturities.

where rx_B is the excess log return on a short maturity bond. With the same assumptions as before, the model conditions down as follows:

Theorem 6. *Given the conditional model in Equation 29, we obtain the following linear pricing relation:*

$$\forall i,t: \qquad E\left(rx_{i,t+1}\right)+\frac{V_i}{2} \;=\; cov\left(rx_{i,t+1},rx_{M,t+1}\right)\times\hat{\delta}_M + cov\left(rx_{i,t+1},\lambda_{t+1}\right)\times\hat{\delta}_\lambda + cov\left(rx_{i,t+1},rx_{B,t+1}\right)\times\hat{\delta}_B \tag{30}$$

where $\hat{\delta}_M$,$\hat{\delta}_\lambda$, and $\hat{\delta}_B$ are three constant unconditional prices of market, expected returns, and bond risk, respectively. Appendix B.2 derives the link between these prices and the conditional ones in Section 6.

Proof. This is a straightforward application of Theorem 8 in Appendix B with three factors (market, expected return, and level). □

Theorem 5 carries over to this case in a natural fashion:

Theorem 7. *When the expected returns factor is measured over long a horizon, the unconditional relation in Equation 30 still holds, with the new prices of risk that are a linear transform of the ones in Equation 30:*

$$\forall i,t: \qquad E\left(rx_{i,t+1}\right)+\frac{V_i}{2} \;=\; cov\left(rx_{i,t+1},rx_{M,t+1}\right)\times\tilde{\delta}_M + cov\left(rx_{i,t+1},\lambda_{t+1:t+T}\right)\times\tilde{\delta}_\lambda + cov\left(rx_{i,t+1},rx_{B,t+1}\right)\times\tilde{\delta}_B \tag{31}$$

where $\lambda_{t+1:t+T}=E_{t+1}rx_{M,t+2:t+T+1}$ denotes expected market returns (risk premia) starting one period ahead for $T\geq 1$ periods, $\tilde{\delta}_M$,$\tilde{\delta}_\lambda$, and $\tilde{\delta}_B$ are three constant unconditional prices of market, expected returns, and bond risk, respectively.

Proof. Refer to the proof of a more general Theorem 9 with an arbitrary number of factors and risk prices in Appendix B.2. □

Using Theorem 7 and Appendix B.4 we can show in the same way as in Section 2.5 that the following approximate relation holds:

$$\forall i,t: \qquad E\left(R^e_{i,t+1}\right) = cov\left(rx_{i,t+1}, rx_{M,t+1}\right) \times \tilde{\delta}_M + cov\left(rx_{i,t+1}, \hat{\lambda}_{t+1:t+T}\right) \times \tilde{\delta}_{\hat{\lambda}} + cov\left(rx_{i,t+1}, rx_{B,t+1}\right) \times \tilde{\delta}_B \tag{32}$$

where $R^e_i \equiv \frac{R_i}{R_f}$ is the level of excess returns and $\hat{\lambda}_{t+1:t+T} \equiv rx_{M,t+2:T} = \sum_{j=2}^{T} rx_{M,t+j}$. With these results in hand, we proceed with empirical tests of the model in Equation 32.

4.2 Data

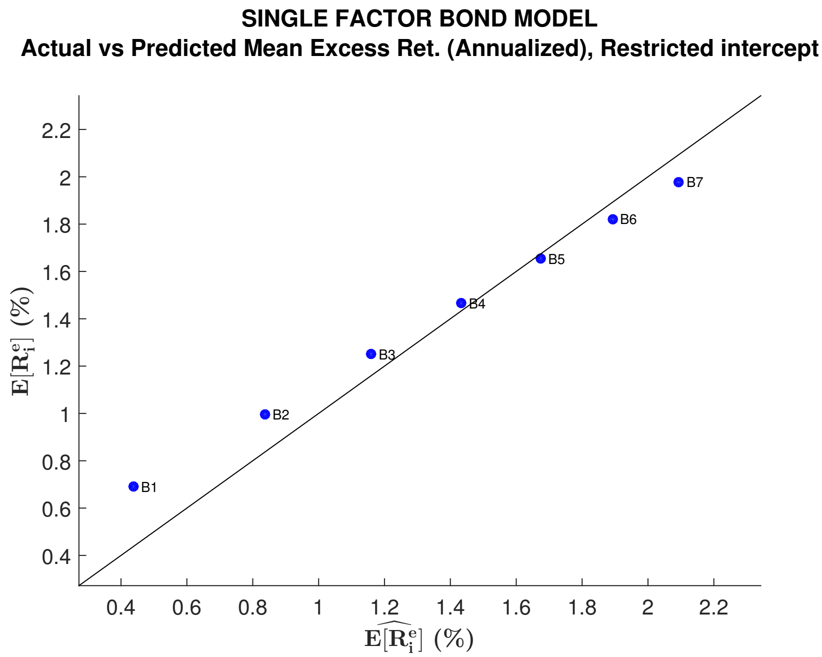
We use zero-coupon treasury yields from Gürkaynak et al. (2006) (GSW), which provides a daily constant maturity yield curve from 1961 onward. Though the data are smoothed by the use of a Svensson polynomial (extension of Nelson-Siegel), the yields are usually very close to the unsmoothed yields derived using the methodology of Fama and Bliss (1987) and "for many purposes the slight smoothing in GSW data may make no difference" (Cochrane and Piazzesi, 2008). The advantage of GSW yields is the daily observation frequency, which we have argued in Section 3 is important to our empirical strategy. Prior to 1971, the GSW yields only include maturities up to seven years. Post 1971 they includes maturities to 30 years, though there is some question of the reliability of the very long maturity yields. To match the timing of our stock data, we use maturities up to seven years, starting in 1963. To construct zero-coupon bond returns from the GSW yields, we use the daily parameter estimates available online[24]. This allows us, for example to recover the yield on a bond with 364 days to maturity. This yield is necessary for calculating the daily return on a one-year bond. Excess returns just subtract the return on a one month t-bill, the same procedure we use for stock excess returns.

4.3 Estimating the price of "level risk"

With daily excess log bond returns in hand, we estimate the model of Equation 28 using the same two-stage procedure of Section 3.1. $\tilde{\delta}_B$ is estimated to be approximately 40. The

[24]http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html

Figure 5: Univariate Bond Pricing



cross-sectional R^2 is 92% with 0.1% annualized mean absolute pricing error. Figure 5 shows graphically the good fit of the level model for bonds.

In the context of Equation 32, we argue that $\widetilde{\delta}_B$, the price of "level risk", is commonly underestimated in bond-only univariate models. It is a classic case of omitted variable bias. Equation 18 and the results of Section 3.1 suggest at least two such missing variables, $C_{i,\lambda} = 10^5 \times Cov\left[rx_{t+1}^{(i)},\ E_t\left(\sum_{i=1}^{k} rx_{M,t+i}\right)\right]$ and $C_{i,M} = 10^5 \times Cov\left[rx_{t+1}^{(i)},\ rx_{M,t+1}\right]$. Table 6 shows $C_{i,B},\ C_{i,\lambda}$ and $C_{i,M}$ across maturities. First note that $C_{i,M} \approx 0$ for all maturities. More importantly, $\forall i,\ C_{i,\lambda} \approx 12 \times C_{i,B}$. Cross-sectionally, $corr\left(C_{i,B},\ C_{i,\lambda}\right) \approx 1.0$. Since we know from Section 3.1 that $\delta_\lambda \neq 0$, the univariate level model suffers greatly from omitted variables bias. Using the estimate of $\delta_\lambda = -11$, a back-of-the-envelope calculation suggests the true $\delta_B = 40 + 12 \times 11 = 170$. In other words, the required compensation for bearing level risk is much higher than is estimated from a univariate model of bond expected returns. Treasury bonds, in addition to loading positively on level risk, also provide investors a hedge against increases in the expected return on stocks. Thus, bonds earn lower average excess returns than in the hypothetical economy where the expected market return is constant.

Table 6: Covariances

	1Y	2Y	3Y	4Y	5Y	6Y	7Y
$C_{i,B}$	0.04	0.08	0.11	0.14	0.17	0.19	0.21
$C_{i,M}$	0.02	0.02	0.03	0.04	0.06	0.08	0.10
$C_{i,\lambda}$	0.57	1.03	1.41	1.75	2.06	2.35	2.63
	(3.48)	(3.53)	(3.64)	(3.79)	(3.95)	(4.11)	(4.23)

Notes: This table shows covariances of government bonds with the bond "level" factor, $C_{i,B} = 10^5 \times Cov\left[rx_{t+1}^{(i)}, (E_{t+1} - E_t)\, rx_{t+1}^{(LT)}\right]$; the stock excess returns market factor, $C_{i,M} = 10^5 \times Cov\left[rx_{t+1}^{(i)},\ rx_{M,t+1}\right]$; and the discount rate factor, $C_{i,\lambda} = 10^5 \times Cov\left[rx_{t+1}^{(i)},\ E_t\left(\Sigma_i^k rx_{M,t+i}\right)\right]$, estimated over 01-Aug-1963 to 31-Dec-2013. t-statistics in parentheses are adjusted for serial correlation using the Newey-West procedure with 252 lags (1 year).

This intuition is formalized by estimating the 3-factor ICAPM given by Equation 32. Table 7 gives estimated risk prices (δs) from the following models:

$$\text{2-Factor ICAPM: } E\left[R_i^e\right] = C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda \quad (33)$$

$$\text{Univariate Level Risk: } E\left[R_i^e\right] = C_{i,B}\delta_B \quad (34)$$

$$\text{3-Factor ICAPM: } E\left[R_i^e\right] = C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda + C_{i,B}\delta_B \quad (35)$$

where $C_{i,X} \equiv Cov\left[rx_{i,t},\ X_t\right]$. All models are estimated with the intercept restricted to zero. The 2-factor ICAPM is estimated using only the stock portfolios from Section 3 (but both stocks and bonds are used as test assets) and hence the risk price estimates are the same as in Section 3.1. The univariate Level Risk model is estimated using only bond excess returns; bonds are also the only test assets. The 3-factor ICAPM is estimated using all assets, stock portfolios as well as bonds. Estimated values for δ_M and δ_λ are essentially unchanged in the 3-factor ICAPM (relative to the 2-factor estimates). The R^2 of the 2-factor ICAPM is so low because bonds are included as test assets (see Figure 4). Importantly, δ_B in the 3-factor ICAPM is $165 \gg 40$. This is nearly equal to the back of the envelope prediction given above.

Table 8 gives annualized percent returns by maturity in sample, as implied by the univariate Level Risk model, and as implied by the 3-factor ICAPM. Both models imply a larger term premium (spread between long and short maturity average returns) than is observed

Table 7: Risk Price Estimates

	δ_M	δ_λ	δ_B	R^2	$MAPE$
2-Factor ICAPM	4.54	-11	-	-8.4	2.5
(est. stocks only; pricing bonds & stocks)	(4.47)	(-5.33)			
Level Risk	-	-	39.8	92.1	0.105
(bonds only)			(1.69)		
3-Factor ICAPM	4.03	-9.99	165	94	0.582
(bonds and stocks)	(4.11)	(-5.56)	(5.43)		

Notes: This table shows premia estimated from the 01-Aug-1963 to 31-Dec-2013 for the 2-factor ICAPM (estimated using stock portfolios only; both bonds and stocks included as test assets), the Level Risk model (estimated using bond returns; only bonds used as test assets), and the 3-factor ICAPM (estimated using both stocks and bonds to price both). Model intercepts are restricted to zero. MAPE is average absolute pricing error, annualized. Moving block bootstrap t-statistics are in parentheses.

Table 8: Bond Expected Returns

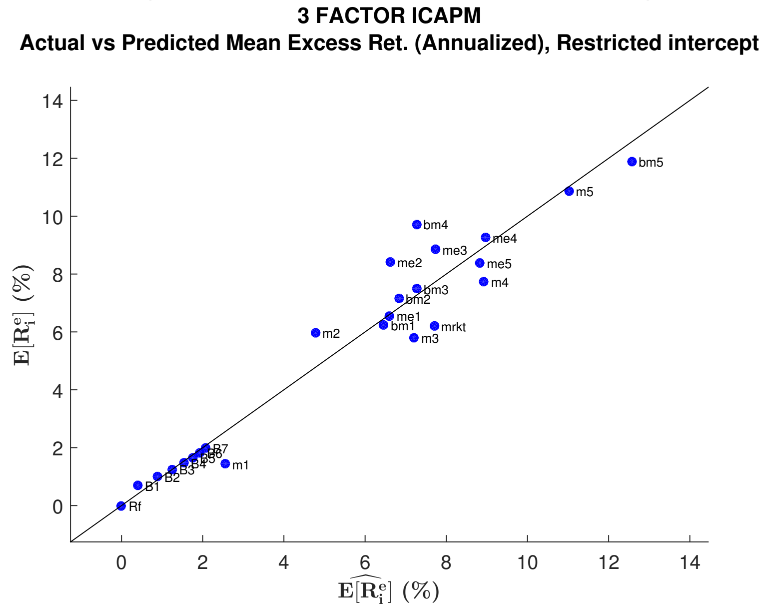
	Sample Mean	Level Risk	3-Factor ICAPM
1-year bond	0.68	0.44	0.4
2-year bond	0.97	0.83	0.88
3-year bond	1.2	1.2	1.2
4-year bond	1.4	1.4	1.5
5-year bond	1.5	1.7	1.7
6-year bond	1.6	1.9	1.9
7-year bond	1.7	2.1	2

Notes: Annualized percent returns by maturity.

in the data, with the 3-factor model performing on par with the level risk model in pricing bonds.

Figure 6 shows average returns vs 3-factor ICAPM expected returns for bonds and stock portfolios. The graphs plot model implied mean excess returns on the horizontal and sample average returns on the vertical axis. The 45° line represents a model with perfect in-sample fit (100% R^2). Stocks fit almost as well as in Figure 2 (using the 2-factor ICAPM) and bonds fit quite well. It is worth emphasizing that this result is not merely mechanical. Given two factor models, each fitting either cross-section of stocks or bonds, a combined model with all

Figure 6: Joint Stock and Bond Pricing

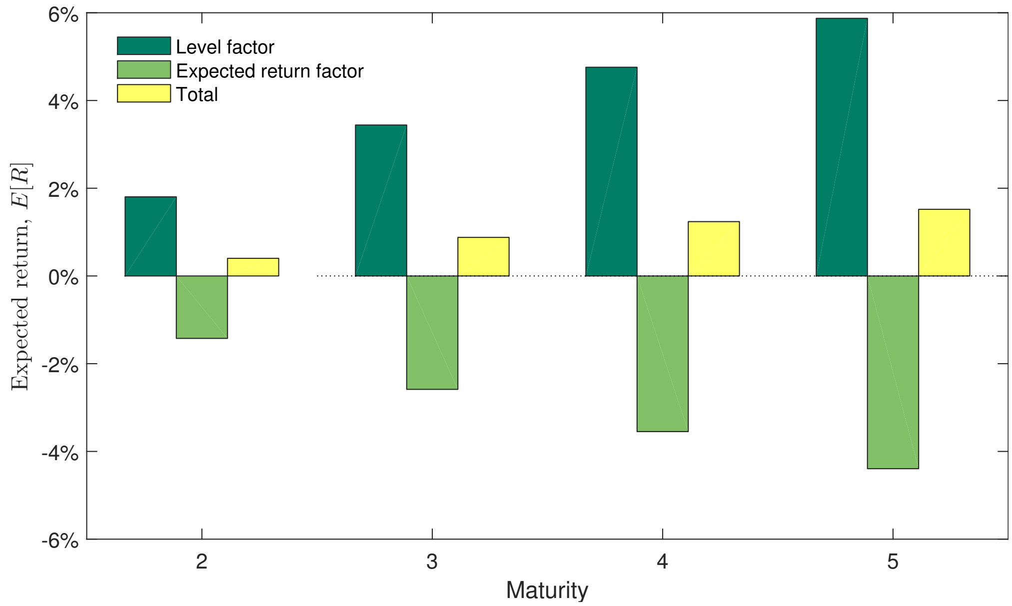


factors need not fit the joint cross-section of bonds and stocks (see Koijen et al. 2010).

Figure 7 decomposes the expected excess return on the various bonds. The premium due to market risk, $C_{i,M}$, is excluded since it is negligible for bonds. Bonds earn a large premium for loading on level risk, whereas they command a large negative premium for loading on the expected return factor. This is consistent with a "flight-to-quality" (Caballero and Krishnamurthy, 2008) interpretation where investors' appetite for risk falls and they attempt to rebalance their portfolios towards safer securities (like U.S. government debt and "good companies"). Since it is impossible for everyone to rebalance in this way at the same time, prices adjust instead of quantities. The prices of "risky" assets fall relative to the prices of "safer" assets.

Koijen et al. (2010) have a seemingly similar decomposition, albeit with a very different interpretation. Our 3-factor ICAPM as well as their model both feature a level factor and a market factor. Instead of our expected stock return factor, they use an expected bond return factor (CP from Cochrane and Piazzesi, 2005). Whereas bond returns load positively on our factor, λ, they load negatively on CP. Koijen et al. (2010) find a positive price of CP risk whereas we find a negative price of λ risk. The product of loading $\times$ risk price yields a negative number in both cases, and hence the pictures look quite similar, but with

Figure 7: Decomposition of Bond Risk Premia



different interpretation. We find that bonds hedge against increases in expected stock returns but Koijen et al. (2010) find that bonds respond negatively to increases in expected bond returns. Finally, our estimated model produces a term structure of expected returns which is somewhat steeper than in the data (Table 8). In contrast, the estimates in Koijen et al. (2010) result in a flat term structure.

5 Predicting the Future Market using Cross-Section

In our empirical methodology, we use future realized excess returns as a proxy for today's market expectation of future excess returns. We further show that this proxy is key in explaining the cross section of stock returns. This observation can be viewed from the reverse perspective. If time-varying expected returns manifest in the cross-section, the cross-section of stock returns can provide information about expected future returns. Indeed, "priced factors ... are innovations in state variables that predict future returns." (Brennan et al., 2004). It is therefore natural to ask whether cross-sectional variables can predict future returns and to what extent. Few recent papers have looked at this question. Kelly and Pruitt (2011) use the cross-section of dividend-price ratios and show that they indeed predict

Table 9: Predictability (Fama-French factors)

$R^e_{M,t+1:t+k} = a + [DP_t \; MRKT_{t-90:t} \; SMB_{t-90:t} \; HML_{t-90:t} \; UMD_{t-90:t}]\, b + \varepsilon_{M,t+1}$

	DP	$MRKT$	SMB	HML	UMD	R^2
3 months	0.018	0.049	-	-	-	0.01
	(1.3)	(0.8)				
6 months	0.036	0.0074	-	-	-	0.014
	(1.2)	(0.099)				
9 months	0.052	-0.012	-	-	-	0.019
	(1.2)	(-0.13)				
12 months	0.066	-0.047	-	-	-	0.023
	(1.2)	(-0.47)				
3 months	0.019	0.039	-0.27	-0.2	-0.19	0.069
	(1.5)	(0.47)	(-2.5)	(-2.7)	(-3.3)	
6 months	0.038	-0.024	-0.39	-0.39	-0.27	0.078
	(1.5)	(-0.26)	(-2.4)	(-2.5)	(-2.5)	
9 months	0.053	-0.077	-0.39	-0.47	-0.35	0.079
	(1.4)	(-0.67)	(-2.2)	(-2.2)	(-2.6)	
12 months	0.067	-0.1	-0.41	-0.37	-0.42	0.073
	(1.4)	(-0.75)	(-1.7)	(-1.6)	(-2.8)	

Notes: The table shows time-series predictability of the stock market risk premium using returns on SMB, HML, and UMD. Daily sample from 01-Aug-1963 to 31-Dec-2013. Newey-West t-statistics in parentheses.

future returns significantly better than the aggregate dividend-price ratio alone.

Our aim is not to construct the optimal predictor; we only want to show that predictability is indeed present and use it as a robustness check of our methodology. If future returns help to explain the cross-section, the cross-sectional returns themselves mechanically must predict future returns. We want to ensure the covariances reported in Table 1 and Figure 1 are *economically* significant. As such, we use the returns on SMB, HML, and UMD portfolios to forecast future market returns:

$$R^e_{M,t+1:t+k} = a + [DP_t \; MRKT_{t-90:t} \; SMB_{t-90:t} \; HML_{t-90:t} \; UMD_{t-90:t}]\, b + \varepsilon_{M,t+1} \qquad (36)$$

Table 10: Predictability (Anomalies)

$$R^e_{M,t+1:t+k} = a + \left[DP_t \ NMV^{PC1}_{t-90:t}\right] b + \varepsilon_{M,t+1}$$

	DP	$MRKT$	NMV_{PC1}	R^2
3 months	0.019	0.047	-	0.011
	(1.3)	(0.67)		
6 months	0.038	0.015	-	0.016
	(1.3)	(0.19)		
9 months	0.056	-0.02	-	0.022
	(1.3)	(-0.2)		
12 months	0.072	-0.035	-	0.027
	(1.3)	(-0.34)		
3 months	0.021	-0.022	-0.28	0.07
	(1.6)	(-0.34)	(-5.4)	
6 months	0.041	-0.079	-0.38	0.069
	(1.5)	(-0.93)	(-4)	
9 months	0.059	-0.12	-0.42	0.065
	(1.5)	(-1.1)	(-3.2)	
12 months	0.075	-0.14	-0.43	0.062
	(1.5)	(-1.4)	(-3.1)	

Notes: The table shows time-series predictability of the stock market risk premium using the first principal component of NMV long-short portfolios. Monthly sample from August 1963 to December 2013. Newey-West t-statistics in parentheses.

Each of the $MRKT, SMB, HML, UMD$ factors is computed using the past 90 calendar days. Results are robust to varying the lag length.

The top panel of Table 9 reports the coefficient estimates, t-statistics of estimated coefficients in Equation 36, and R^2 at various horizons, k, (3, 6, 9, and 12 months) with only the market and dividend-price ratio included as predictors. There is little evidence of return predictability at horizons up to one year, as evidence by the insignificant t-statistics and low R^2. The bottom panel shows results when including $SMB, HML,$ and UMD as additional predictors. We find that all of the coefficients for each variable at 3-9 months horizon are significant and negative and R^2 increases greatly. We conclude that covariances of FFC factors with expected future market returns are *economically* significant. Related evidence of predictability is documented by Liew and Vassalou (2000). They show that SMB and HML help forecast future rates of economic growth. We repeat the forecasting exercise using the first principal component of the NMV anomaly returns (PC1). Table 10 shows that PC1 alone has similar forecasting ability[25] to the three FFC factors combined. The statistical significance for PC1 is substantially higher, likely because $SMB, HML,$ and UMD each contain substantial idiosyncratic "noise", adding uncertainty to the estimates.

[25]As measured by R^2

6 Conclusion

In this paper we link two well-documented empirical facts: (1) time-series variation in aggregate discount rates and (2) cross-sectional dispersion in average returns. We present a model with time-varying risk aversion which generates endogenous variation in expected market returns. Investors' hedging demand with respect to discount rate shocks results in an equilibrium pricing kernel in which those shocks appear as an additional factor. In the cross-section of assets, differences in return covariance with these shocks produces differences in expected returns. Crucially, the model predicts that shocks to discount rates increase investors' marginal utility and in equilibrium have a negative price of risk.

We confront the model with return data on a large set well known asset pricing anomalies, including size, value, and momentum. Across empirical specifications, we find consistent support for a negative risk price on our expected return factor. Our theoretically motivated pricing model performs nearly as well as standard empirical factor models in fitting the cross-sectional dispersion in average returns. This is surprising since we augment the standard CAPM with a single additional factor. These empirical results and theoretical motivation contrast sharply with previous related work. Our main conclusion is that shocks to aggregate expected returns are "bad" for the representative agent, who requires higher expected returns for holding exposure to these shocks. The previous studies conclude that such shocks increase investor utility (decreasing marginal utility) and hence, higher exposure results in lower equilibrium expected returns.

Finally, we extend the model and empirical specification to jointly price stocks and bonds. We find that bonds hedge against discount rate shocks, earning negative expected returns (and downward sloping term premium). This is offset by exposure to risk-free rate shocks. Overall, the model generates a small bond risk premium and upward sloping term premium, as found in the data.

Taken as a whole, the empirical evidence is consistent with our model in which increases in investor risk aversion, or "flight-to-quality", generate time-varying aggregate discount rates. In the model and data, cross-sectional heterogeneity in sensitivity to these shocks generates the observed spread in average returns. Like much of the literature, we examine the equilibrium conditions of an asset pricing model. It remains uncertain as to precisely what generates the different covariances of asset returns with discount rate shocks.

References

Brennan, M., A. Wang, and Y. Xia (2004). Estimation and test of a simple model of intertemporal capital asset pricing. *The Journal of Finance 59*(4), 1743–1776.

Caballero, R. and A. Krishnamurthy (2008). Collective risk management in a flight to quality episode. *The Journal of Finance 63*(5), 2195–2230.

Campbell, J. (1996). Understanding Risk and Return. *Journal of Political Economy 104*, 298–345.

Campbell, J. and T. Vuolteenaho (2004). Bad Beta, Good Beta. *American Economic Review 94*(5), 1249–1275.

Campbell, J. Y. and J. H. Cochrane (1999). By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy 107*(2), pp. 205–251.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance 52*, 57–82.

Cecchetti, S. (1997). Measuring short-run inflation for central bankers. *Review* (May), 143–155.

Cochrane, J. H. (2001). *Asset Pricing, 2001.* Princeton University Press.

Cochrane, J. H. and M. Piazzesi (2005). Bond Risk Premia. *American Economic Review*, 138–160.

Cochrane, J. H. and M. Piazzesi (2008). Decomposing the yield curve. *Graduate School of Business, University of Chicago, Working Paper*.

Dew Becker, I. (2011). A model of time-varying risk premia with habits and production.

Drechsler, I. (2013). Uncertainty, Time-Varying Fear, and Asset Prices. *The Journal of Finance 68*(5), 1843–1889.

Duffie, D. and L. Epstein (1992a). Asset pricing with stochastic differential utility. *Review of Financial Studies 5*(3), 411–436.

Duffie, D. and L. Epstein (1992b). Stochastic differential utility. *Econometrica: Journal of the Econometric Society*, 353–394.

Fama, E. and R. Bliss (1987). The information in long-maturity forward rates. *The American Economic Review*, 680–692.

Fama, E. and K. French (2008). Dissecting anomalies. *The Journal of Finance 63*(4), 1653–1678.

Fama, E. F. and K. R. French (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance 51*, 55–84.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2006, June). The U.S. Treasury Yield Curve: 1961 to the Present.

Hansen, L. P. and T. J. Sargent (2008). *Robustness*. Princeton university press.

Horowitz, J. L. (2001). *Handbook of Econometrics - Chapter 52: The Bootstrap*, Volume 5. Elsevier B.V.

Kelly, B. and S. Pruitt (2011). Market Expectations in the Cross Section of Present Values.

Koijen, R., H. Lustig, and S. Van Nieuwerburgh (2010). The cross-section and time-series of stock and bond returns. Technical report, National Bureau of Economic Research.

Kozak, S. (2013). *Endogenous risk-taking and risk premia dynamics*. The University of Chicago.

Kozak, S., S. Nagel, and S. Santosh (2014). Interpreting factor models.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*, 175–194.

Liew, J. and M. Vassalou (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics 57*(2), 221–245.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867–887.

Novy-Marx, R. and M. Velikov (2014). A taxonomy of anomalies and their trading costs. Technical report, National Bureau of Economic Research.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of financial economics 4*(2), 129–176.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of finance 19*(3), 425–442.

Shefrin, H. (2008). *A behavioral approach to asset pricing*. Academic Press.

Skiadas, C. (2009). *Asset pricing theory*. Princeton Univ Pr.

Treynor, J. (1961). Toward a theory of market value of risky assets.

Appendix

A The Model

Utility function Duffie and Epstein (1992a,b) show finding an ordinally equivalent aggregator $\left(\bar{f}, \bar{A}\right)$ is possible such that $\bar{A} = 0$, a *normalized* aggregator. Most papers that employ Epstein-Zin-Weil preferences use this type of aggregator.

We take a different approach and use an *unnormalized* aggregator as defined explicitly in Equation 2 and Equation 3. Although such a representation requires computing an additional variance term in Equation 1, it allows us to separate the effect of elasticity of intertemporal substitution (EIS) and risk aversion in the stochastic differential utility. In particular, the first term, $f(C, J)$, depends only on EIS, whereas the second term, $\frac{1}{2}\frac{\alpha}{J}\|\boldsymbol{\sigma_{J,t}}\|^2$ depends only on risk aversion and is linear in it (it might depend on the EIS indirectly through the $\boldsymbol{\sigma_{J,t}}$ term, however).

HJB equation Assume complete markets. A representative investor in this economy maximizes his utility over consumption,

$$J_t = \mathbb{E}_t\left(\int_t^\infty \left[f\left(C_\tau, J_\tau\right) + \frac{1}{2}A\left(J_\tau\right) \| \, J_{\boldsymbol{X}}\left(\boldsymbol{X_\tau}, \tau\right)\sigma_{\boldsymbol{X}}\left(\boldsymbol{X_\tau}, W_\tau, \tau\right) \|^2\right] d\tau\right), \tag{37}$$

where $\boldsymbol{X} = (\alpha,\, g,\, W)$ is a vector of aggregate state variables.

$$U\left(C_\tau\right) = f\left(C_\tau, J_\tau\right) - \frac{1}{2}\frac{\alpha}{J}\left\|J_{\boldsymbol{X}}\boldsymbol{\sigma_X}\right\|^2 \tag{38}$$

The Hamilton-Jacobi-Bellman (HJB) equation for the planner's problem is given by

$$\begin{aligned} 0 \;=\; \sup_{C,\theta} \;\; \Big\{ & U\left(C,\, J\right) + J_W\mathbb{E}\left(dW\right)/dt + J_\alpha\mathbb{E}\left(d\alpha\right)/dt + J_g\mathbb{E}\left(dg\right)/dt + \\ & + \frac{1}{2}J_{WW}\mathbb{E}\left(dW^2\right)/dt + \frac{1}{2}J_{\alpha\alpha}\mathbb{E}\left(d\alpha^2\right)/dt + \frac{1}{2}J_{gg}\mathbb{E}\left(dg^2\right)/dt \\ & + J_{W\alpha}\mathbb{E}\left(dW d\alpha\right)/dt + J_{Wg}\mathbb{E}\left(dW dg\right)/dt + J_{\alpha g}\mathbb{E}\left(d\alpha dg\right)/dt\Big\}. \end{aligned} \tag{39}$$

First-order conditions

$$[C]: \quad f_C \;=\; J_W \tag{40}$$

$$\begin{aligned} [\theta]: \quad 0 \;=\; & \frac{\alpha}{J}J_W^2 W^2\theta\boldsymbol{\sigma_R}\boldsymbol{\sigma_R'} + J_W W\lambda + J_{W\alpha}W\boldsymbol{\sigma_R}\boldsymbol{\sigma_\alpha'} \\ & + J_{Wg}W\boldsymbol{\sigma_R}\boldsymbol{\sigma_g'} + J_{WW}W\theta^2\boldsymbol{\sigma_R}\boldsymbol{\sigma_R'} \end{aligned} \tag{41}$$

Value function guess Find a solution of the form $J(W,\, \alpha,\, g) = W \times F\left(\alpha,\, g\right)$. Partial derivatives are: $J_W = F$, $J_{WW} = 0$, $J_\alpha = J\frac{F_\alpha}{F}$, $J_g = J\frac{F_g}{F}$, $J_{W\alpha} = F_\alpha$, $J_{Wg} = F_g$, $J_{\alpha\alpha} = J\frac{F_{\alpha\alpha}}{F}$,

$J_{gg} = J\frac{F_{gg}}{F}$, $J_{\alpha g} = J\frac{F_{\alpha g}}{F}$.

The first order-condition with respect to consumption implies constant consumption-to-wealth ratio:

$$f_C = \delta \frac{J}{C} = F \quad \Longrightarrow \quad \frac{C}{W} = \delta \tag{42}$$

and hence $\boldsymbol{\sigma_R} = \boldsymbol{\sigma_D}$. Market clearing requires $C = D$, $W = P$, and $\theta = 1$.

Further assume EIS=1, then $f(C, J) = \delta(lnC - lnJ)J = J\delta[ln\delta - lnF]$. Guess the solution of the form

$$F(\alpha, g) = exp(a_0 + a_\alpha \alpha + a_g g). \tag{43}$$

Plug everything in:

$$\begin{aligned} \delta[ln\delta - (a_0 + a_\alpha\alpha + a_g g)] - \frac{1}{2}\alpha \left\|\boldsymbol{\sigma_D} + a_\alpha\boldsymbol{\sigma_\alpha} + a_g\boldsymbol{\sigma_g}\right\|^2 + (g - \delta) \\ +a_\alpha\phi_\alpha(\bar{\alpha} - \alpha) + a_g\phi_g(\bar{g} - g) + a_\alpha\theta\boldsymbol{\sigma_R}\boldsymbol{\sigma}'_{\boldsymbol{\alpha}} + a_g\theta\boldsymbol{\sigma_R}\boldsymbol{\sigma}'_{\boldsymbol{g}} \\ +\frac{1}{2}a_1^2\boldsymbol{\sigma_\alpha}\boldsymbol{\sigma}'_{\boldsymbol{\alpha}} + \frac{1}{2}a_2^2\boldsymbol{\sigma_g}\boldsymbol{\sigma}'_{\boldsymbol{g}} + a_1a_2\boldsymbol{\sigma_\alpha}\boldsymbol{\sigma}'_{\boldsymbol{g}} \quad &= \quad 0 \end{aligned} \tag{44}$$

Equalize coefficients near α, g, and const:

$$\begin{aligned} \text{const:} \quad 0 \quad &= \quad \delta ln\delta - \delta a_0 - \delta + a_\alpha\phi_\alpha\bar{\alpha} + a_g\phi_g\bar{g} + a_\alpha\boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{\alpha}} + a_g\boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{g}} \\ &\quad +\frac{1}{2}a_1^2\boldsymbol{\sigma_\alpha}\boldsymbol{\sigma}'_{\boldsymbol{\alpha}} + \frac{1}{2}a_2^2\boldsymbol{\sigma_g}\boldsymbol{\sigma}'_{\boldsymbol{g}} + a_1a_2\boldsymbol{\sigma_\alpha}\boldsymbol{\sigma}'_{\boldsymbol{g}} \end{aligned} \tag{45}$$

$$\alpha: \quad 0 \quad = \quad -a_\alpha(\delta + \phi_\alpha) - \frac{1}{2}\left\|\boldsymbol{\sigma_D} + a_\alpha\boldsymbol{\sigma_\alpha} + a_g\boldsymbol{\sigma_g}\right\|^2 \tag{46}$$

$$g: \quad 0 \quad = \quad 1 - a_g(\delta + \phi_g) \tag{47}$$

The second equation immediately implies that $a_\alpha < 0$ and the third equations implies $a_g = \frac{1}{\delta+\phi_g} > 0$.

Asset prices Excess return:

$$\lambda \quad = \quad \alpha\boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{D}} - \left(a_\alpha\boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{\alpha}} + a_g\boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{g}}\right) \tag{48}$$

$$\quad = \quad \lambda_0 + \alpha \times \boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{D}} \tag{49}$$

Risk-free rate:

$$r = ER - \lambda = \delta + \frac{1}{dt}\mathbb{E}\left[\frac{dC}{C}\right] - \lambda \quad = \quad \delta + g - \lambda \tag{50}$$

$$\quad = \quad \lambda_0 + g - \alpha \times \boldsymbol{\sigma_D}\boldsymbol{\sigma}'_{\boldsymbol{D}} \tag{51}$$

SDF The SDF is given by (Duffie and Epstein, 1992b):

$$\frac{d\Lambda}{\Lambda} \equiv \bar{f}_V\left(C, \bar{J}\right)dt + \frac{d\bar{f}_C\left(C, \bar{J}\right)}{\bar{f}_C\left(C, \bar{J}\right)}. \tag{52}$$

$$\begin{aligned}\frac{d\Lambda}{\Lambda} &= -rdt + \mathcal{L}\left(dln f_C\left(C, J\right) - \alpha dln J\right)d\boldsymbol{Z} &(53)\\ &= -rdt - \alpha dR_M - \left(\alpha - 1\right)\underbrace{a_\alpha}_{<0}\boldsymbol{\sigma_\alpha} d\boldsymbol{Z} - \left(\alpha - 1\right)\underbrace{a_g}_{>0}\boldsymbol{\sigma_g} d\boldsymbol{Z} &(54)\end{aligned}$$

where operator $\mathcal{L}\left(\cdot\right)$ denotes a vector of loading on shocks and $dR_M \equiv \boldsymbol{\sigma_R} d\boldsymbol{Z}$.

When risk aversion $\alpha \geq 1$, the price of α risk is negative and the price of g risk is positive.

The SDF implies a three-factor model:

$$R^e_{i,t} = \alpha \times \text{cov}_{t-1}\left(R^e_{i,t}, R^e_{M,t}\right) + a_\alpha\left(\alpha - 1\right) \times \text{cov}_{t-1}\left(R^e_{i,t}, d\alpha\right) + a_g\left(\alpha - 1\right) \times \text{cov}_{t-1}\left(R^e_{i,t}, dg\right) \tag{55}$$

B Operationalizing the Model

B.1 SDF

We assume that asset returns are log normally distributed and specify an SDF of the form

$$-m_{t+1} = r^f_t + \frac{1}{2}\Lambda^{'}_t \Sigma \Lambda_t + \Lambda^{'}_t \varepsilon_{t+1} \tag{56}$$

where m_t is a log of an SDF, r^f_t is the log nominal risk free rate, ε_{t+1} is a $N \times 1$ vector of shocks, and Λ_t is the $N \times 1$ vector of market prices of risk associated with these shocks at time t. Errors ε_{t+1} are assumed to be i.i.d. and standard normally distributed.

No-arbitrage implies

$$\begin{aligned}1 &= E_t\left[M_{t+1}R_{t+1}\right] = E_t\left[e^{m_{t+1}+r_{t+1}}\right] &(57)\\ 0 &= E_t\left[m_{t+1}\right] + E_t\left[r_{t+1}\right] + \frac{1}{2}\sigma^2_t\left(m_{t+1}\right) &(58)\\ &+ \frac{1}{2}\sigma^2_t\left(r_{t+1}\right) + cov_t\left(r_{t+1}, m_{t+1}\right) &(59)\end{aligned}$$

where $r = log\left(R\right)$. Since $E_t\left[m_{t+1}\right] + \frac{1}{2}\sigma_t^2\left(m_{t+1}\right) = -r_t^f$, we get

$$E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma_t\left(rx_{t+1}\right) = cov_t\left(rx_{t+1}, \varepsilon_{t+1}\right)\Lambda_t \tag{60}$$

where $rx = r - r^f$ is excess return on an asset.

B.2 Unconditional Pricing Relation

Let the conditional model be

$$\begin{aligned} E_t\left(rx_{i,t+1}\right) + \frac{1}{2}var_t\left(rx_{i,t+1}\right) &= cov_t\left(rx_{i,t+1}, f_{t+1}\right) \times \delta_{f,t} \\ &+ cov_t\left(rx_{i,t+1}, \lambda_{t+1}\right) \times \delta_{\lambda,t} \end{aligned} \tag{61}$$

where $f_{t+1} = \left[f_{t+1}^{(1)}, f_{t+1}^{(2)}, ...f_{t+1}^{(k)}\right]'$ denotes k factors that are log excess returns, $\lambda_{t+1} = E_{t+1}f_{t+2}$ denotes expected log returns (risk premia) on those returns starting one period ahead; $\delta_{f,t}$ and $\delta_{\lambda,t}$ are of size $k \times 1$ and denote corresponding factor risk prices.

Assumption 2. *All covariances* $c_{if}^{'} \equiv cov_t\left(rx_{i,t+1}, f_{t+1}\right)$*,* $c_{i\lambda}^{'} \equiv cov_t\left(rx_{i,t+1}, \lambda_{t+1}\right)$*, and variance* $V_i = var_t\left(rx_{i,t+1}\right)$ *are constant in time.*

With this notation in hand, we can rewrite

$$\begin{aligned} E_t\left(rx_{i,t+1}\right) + \frac{V_i}{2} &= c_{if}^{'} \times \delta_{f,t} + c_{i\lambda}^{'} \times \delta_{\lambda,t} \\ &\equiv C_i \times D_t \end{aligned} \tag{62}$$

where $C_i = \left[\begin{array}{cc} c_{if}^{'} & c_{i\lambda}^{'} \end{array}\right]$ and $D_t = \left[\begin{array}{c} \delta_{f,t} \\ \delta_{\lambda,t} \end{array}\right]$.

Assumption 3. *Risk premia* $\lambda_{t+1} = E_{t+1}f_{t+2}$ *follow a VAR(1) process:*

$$\lambda_{t+1} = \Lambda_0 + \Lambda\lambda_t + \Omega_{\lambda,t+1} \tag{63}$$

where $\Omega_\lambda$are mean-zero errors uncorrelated with λ_t.

Theorem 8. *Given the assumptions 1 and 2 and the conditional model (61), we obtain the following* *linear pricing relation:*

$$
\begin{aligned}
E\left(rx_{i,t+1}\right)+\frac{V_i}{2} &= cov\left(rx_{i,t+1},f_{t+1}\right)\times\hat{\delta}_f && (64)\\
&+ cov\left(rx_{i,t+1},\lambda_{t+1}\right)\times\hat{\delta}_\lambda\\
&\equiv cov\left(rx_{i,t+1},\left[\begin{array}{c} f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times\widetilde{D} && (65)\\
\textit{where } \widetilde{D} &= \Theta^{-1}E\left[D_t\right]\\
\Theta &= \mathbb{I}_{2k}+\Phi_{f\lambda}\times\left[\begin{array}{cc}\mathbb{I}_k & \Lambda'\end{array}\right] && (66)\\
\Phi_{f\lambda} &= \left[\begin{array}{c}\Gamma_f\Sigma_f+\Gamma_{f\lambda}\Sigma'_{f\lambda}\\ \Gamma_\lambda\Sigma'_{f\lambda}+\Gamma_{f\lambda}\Sigma_f\end{array}\right] && (67)
\end{aligned}
$$

where $\mathbb{I}_{2k}$ *denotes an identity matrix of size* $2k\times 2k$, $\Sigma_f = var\left[f_{t+1}\right]$, $\Sigma_{f\lambda}=\left[\begin{array}{cccc} c_{f_1\lambda} & c_{f_2\lambda} & ... & c_{f_k\lambda}\end{array}\right]'$, $\Gamma_{f\lambda}=cov\left[\delta_f,\delta_\lambda\right]$, $\Gamma_f=var\left[\delta_f\right]$, $\Gamma_\lambda=var\left[\delta_\lambda\right]$.

Proof. Take expectations and apply the law of total covariance to the RHS of Equation 61 to get

$$
\begin{aligned}
E\left(rx_{i,t+1}\right)+\frac{V_i}{2} &= cov\left(rx_{i,t+1},f_{t+1}\right)\times E\left[\delta_{f,t}\right]+cov\left(rx_{i,t+1},\lambda_{t+1}\right)\times E\left[\delta_{\lambda,t}\right]\\
&- cov\left(E_t rx_{i,t+1},E_t f_{t+1}\right)\times E\left[\delta_{f,t}\right]-cov\left(E_t rx_{i,t+1},E_t\lambda_{t+1}\right)\times E\left[\delta_{\lambda,t}\right] \quad (68)
\end{aligned}
$$

Using Assumption 2, we can write $E_t\lambda_{t+1}=\Lambda_0+\Lambda\lambda_t$. Then

$$
\begin{aligned}
E\left(rx_{i,t+1}\right)+\frac{V_i}{2}=C_i\times D &= cov\left(rx_{i,t+1},\left[\begin{array}{c} f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times D\\
&- cov\left(E_t rx_{i,t+1},E_t f_{t+1}\right)\times\left[\begin{array}{cc}\mathbb{I}_k & \Lambda'\end{array}\right]\times D && (69)
\end{aligned}
$$

where $D=E\left[\begin{array}{c}\delta_{f,t}\\ \delta_{\lambda,t}\end{array}\right]$.

We now use Equation 62 to substitute the expressions for two covariates in $cov\left(E_t rx_{i,t+1},E_t f_{t+1}\right)$ term

$$
cov\left(c'_{if}\times\delta_{f,t}+c'_{i\lambda}\times\delta_{\lambda,t}-\frac{V_i}{2},\Sigma_f\times\delta_{f,t}+\Sigma_{f\lambda}\times\delta_{\lambda,t}-\frac{1}{2}V_f\right) \quad (70)
$$

$$
=c'_{if}var\left[\delta_f\right]\Sigma_f+c'_{i\lambda}var\left[\delta_\lambda\right]\Sigma'_{f\lambda}+c'_{if}cov\left[\delta_f,\delta_\lambda\right]\Sigma'_{f\lambda}+c'_{i\lambda}cov\left[\delta_f,\delta_\lambda\right]\Sigma_f \quad (71)
$$

$$
=C_i\times\left[\begin{array}{c}\Gamma_f\Sigma_f+\Gamma_{f\lambda}\Sigma'_{f\lambda}\\ \Gamma_\lambda\Sigma'_{f\lambda}+\Gamma_{f\lambda}\Sigma_f\end{array}\right] \quad (72)
$$

$$
\equiv C_i\times\Phi_{f\lambda} \quad (73)
$$

where $V_f=\left[V_{f_1},V_{f_2},...,V_{f_k}\right]'$.

Plugging this back in Equation 69 and collecting terms on the LHS and RHS gives

$$C_i \times \left[\mathbb{I}_{2k} + \Phi_{f\lambda} \times \begin{bmatrix} \mathbb{I}_k & \Lambda' \end{bmatrix}\right] \times D \tag{74}$$

$$= cov\left(rx_{i,t+1}, \begin{bmatrix} f_{t+1} \\ \lambda_{t+1} \end{bmatrix}\right) \times D \tag{75}$$

Denote

$$\Theta = \mathbb{I}_{2k} + \Phi_{f\lambda} \times \begin{bmatrix} \mathbb{I}_k & \Lambda' \end{bmatrix} \tag{76}$$

and assume that Θ is invertible. Then

$$C_i = cov\left(rx_{i,t+1}, \begin{bmatrix} f_{t+1} \\ \lambda_{t+1} \end{bmatrix}\right) \times \Theta^{-1} \tag{77}$$

and hence

$$E\left(rx_{i,t+1}\right) + \frac{V_i}{2} = C_i \times D \tag{78}$$

$$= cov\left(rx_{i,t+1}, \begin{bmatrix} f_{t+1} \\ \lambda_{t+1} \end{bmatrix}\right) \times \left\{\Theta^{-1} D\right\} \tag{79}$$

$$= cov\left(rx_{i,t+1}, \begin{bmatrix} f_{t+1} \\ \lambda_{t+1} \end{bmatrix}\right) \times \tilde{D} \tag{80}$$

□

Theorem 9. *When factor risk premia is measured over long horizon*

$$\begin{aligned} E\left(rx_{i,t+1}\right) + \frac{V_i}{2} &= cov\left(rx_{i,t+1}, f_{t+1}\right) \times \tilde{\delta}_f \\ &+ cov\left(rx_{i,t+1}, \lambda_{t+1:t+T}\right) \times \tilde{\delta}_\lambda \end{aligned} \tag{81}$$

the unconditional relation in Equation 65 holds with Θ given by

$$\Theta = \mathbb{I}_{2k} + \Phi_{f\lambda} \times \begin{bmatrix} \mathbb{I}_k & \mathbb{L} \end{bmatrix} \tag{82}$$

where $\lambda_{t+1:t+T} = E_{t+1} f_{t+2:t+T+1}$ denotes expected returns (risk premia) on factors f starting one period ahead for $T \geq 1$ periods, $\mathbb{L} = \Lambda \times \left(\mathbb{I}_k - \Lambda\right)^{-1} \left(\mathbb{I}_k - \Lambda\right)^T$.

Proof. Using Assumption 3[26],

[26] AR(1) assumption is sufficient condition but by no means necessary. Given sufficient persistence of risk premia, moving average of future realized returns is a good non-parametric proxy for today's risk premium. Therefore, $E_t \lambda_{t+1}$ will be approximately proportional to $E_t f_{t+1}$ and the the proof of Theorem 1 carries over without the AR(1) assumption. The assumption was made primarily for expositional reasons and in order to quantify the factor of proportionality.

$$
\begin{aligned}
E_t \lambda_{t+1:t+T} &= \sum_{i=1}^{T} E_t \lambda_{t+i} && (83)\\
&= const + \sum_{i=1}^{T} \Lambda^i \lambda_t && (84)\\
&= const + \Lambda \times (\mathbb{I}_k - \Lambda)^{-1} (\mathbb{I}_k - \Lambda)^{\top} && (85)
\end{aligned}
$$

Plugging this into Equation 68, the proof of Theorem 8 carries over with Λ replaced by $\mathbb{L} = \Lambda \times (\mathbb{I}_k - \Lambda)^{-1} (\mathbb{I}_k - \Lambda)^{\top}$. □

B.2.1 Consistency of the moving average estimator

For any information set $\mathcal{F}_t$,

$$
rx_{M,t+2:T} = \sum_{j=2}^{T} rx_{M,t+j} = E\left[\sum_{j=2}^{T} rx_{M,t+j} \,|\mathcal{F}_t\right] + \sum_{j=2}^{T} \varepsilon_{M,t+j} \tag{86}
$$

and $E\left[\sum_{j=2}^{T} \varepsilon_{M,t+j} \,|\mathcal{F}_t\right] = 0$.

Similarly, true unconditional covariances are equal,

$$
cov\left(rx_{i,t+1}, rx_{M,t+2:T}\right) = cov\left(rx_{i,t+1}, E\left[\sum_{j=2}^{T} rx_{M,t+j} \,|\mathcal{F}_t\right]\right) + \underbrace{cov\left(rx_{i,t+1}, \sum_{j=2}^{T} \varepsilon_{M,t+j}\right)}_{=0} \tag{87}
$$

and thus a consistent estimator of the covariance on the RHS is also a consistent estimator of the covariance on the LHS.

B.3 Covariances

Conditional Covariances are Equal:

$$
\begin{aligned}
cov\left(R_{i,t}, E\left[R_{m,t+1} \,|\mathcal{I}_t\right] |\mathcal{I}_{t-1}\right) &= && (88)\\
E\left\{R_{i,t} \cdot E\left[R_{m,t+1} \,|\mathcal{I}_t\right] |\mathcal{I}_{t-1}\right\} - E\left\{R_{i,t} \,|\mathcal{I}_{t-1}\right\} \cdot E\left\{E\left[R_{m,t+1} \,|\mathcal{I}_t\right] |\mathcal{I}_{t-1}\right\} &= && (89)\\
E\left\{E\left[R_{i,t} \cdot R_{m,t+1} \,|\mathcal{I}_t\right] |\mathcal{I}_{t-1}\right\} - E\left\{R_{i,t} \,|\mathcal{I}_{t-1}\right\} \cdot E\left\{E\left[R_{m,t+1} \,|\mathcal{I}_t\right] |\mathcal{I}_{t-1}\right\} &= && (90)\\
E\left\{R_{i,t} \cdot R_{m,t+1} \,|\mathcal{I}_{t-1}\right\} - E\left\{R_{i,t} \,|\mathcal{I}_{t-1}\right\} \cdot E\left\{E\left[R_{m,t+1} \,|\mathcal{I}_t\right] |\mathcal{I}_{t-1}\right\} &= && (91)\\
E\left\{R_{i,t} \cdot R_{m,t+1} \,|\mathcal{I}_{t-1}\right\} - E\left\{R_{i,t} \,|\mathcal{I}_{t-1}\right\} \cdot E\left\{R_{m,t+1} \,|\mathcal{I}_{t-1}\right\} &= && (92)\\
cov\left(R_{i,t}, R_{m,t+1} \,|\mathcal{I}_{t-1}\right) & && (93)
\end{aligned}
$$

Unconditional Covariances are Equal

$$\begin{aligned}
cov\left(R_{i,t}, E\left[R_{m,t+1}\left|\mathcal{I}_t\right.\right]\right) &= \quad (94)\\
E\left\{cov\left(R_{i,t}, E\left[R_{m,t+1}\left|\mathcal{I}_t\right.\right]\left|\mathcal{I}_{t-1}\right.\right)\right\} - E\left\{R_{i,t}\left|\mathcal{I}_{t-1}\right.\right\}\cdot E\left\{E\left[R_{m,t+1}\left|\mathcal{I}_t\right.\right]\left|\mathcal{I}_{t-1}\right.\right\} &= \quad (95)\\
E\left\{cov\left(R_{i,t}, R_{m,t+1}\left|\mathcal{I}_{t-1}\right.\right)\right\} - E\left\{R_{i,t}\left|\mathcal{I}_{t-1}\right.\right\}\cdot E\left\{E\left[R_{m,t+1}\left|\mathcal{I}_t\right.\right]\left|\mathcal{I}_{t-1}\right.\right\} &= \quad (96)\\
E\left\{cov\left(R_{i,t}, R_{m,t+1}\left|\mathcal{I}_{t-1}\right.\right)\right\} - E\left\{R_{i,t}\left|\mathcal{I}_{t-1}\right.\right\}\cdot E\left\{R_{m,t+1}\left|\mathcal{I}_{t-1}\right.\right\} &= \quad (97)\\
cov\left(R_{i,t}, R_{m,t+1}\right) & \quad (98)
\end{aligned}$$

B.4 Empirical Relation

Assume that the log excess returns are normally distributed, with constant variance and constant risk free rate,

$$rx_t \equiv log\left(R_t^e\right) \backsim N\left(\mu_t, \sigma^2\right) \tag{99}$$

where $R^e \equiv \frac{R}{R_f}$. Then conditional expected excess returns are given by

$$\begin{aligned}
E_{t-1}\left(R_t^e\right) &= exp\left(\mu_t + \frac{\sigma^2}{2}\right) && (100)\\
&\simeq \mu_t + \frac{\sigma^2}{2} && (101)
\end{aligned}$$

Relative errors of this approximation when using *daily* returns are negligible (less than 0.03% for typical test assets used). Taking the unconditional expectations, this conditions down to

$$ER^e = \mu + \frac{\sigma^2}{2} \tag{102}$$

So when estimating a pricing equation, we use log returns to estimate covariances and simple returns to estimate the LHS of Equation 65, $E\left(rx_{i,t+1}\right) + \frac{V_i}{2}$.

C Bootstrap

We construct standard errors for risk prices using the moving block bootstrap procedure as follows. There are N test assets, k factors, and T periodic observations. All moments are sample moments taken as expectations across T. The general model is $r_t = C'\lambda + \varepsilon_t$. C is an $N \times k$ matrix of univariate covariances, $Cov\left(r_t,\, f_t\right)$, where f_t are the k factors. Notice the model is homoskedastic. λ is the vector of risk prices, and ε_t is the vector of pricing errors. The null hypothesis is that $\lambda = 0$ and $E\left[\varepsilon_t\right] = 0$. The alternative is $\lambda \neq 0$.

Bootstrap procedure:

1. Estimate $\widehat{C}$ and $\widehat{\lambda}$ via usual two-stage regression

2. Construct $\widetilde{r}_t = r_t - E[r_t]$

   (a) $\widetilde{r}_t$ is satisfies the null hypothesis of risk-neutrality and maintains all other properties of the true DGP which are shared with the null. In particular, $Cov(\widetilde{r}_t,\, f_t) = Cov(r_t,\, f_t)$

3. Let L be the bootstrap window width. Let $X = \begin{bmatrix} \widetilde{r}_1' & f_1' \\ \vdots & \vdots \\ \widetilde{r}_T' & f_T' \\ \widetilde{r}_1' & f_1' \\ \vdots & \vdots \\ \widetilde{r}_L' & f_L' \end{bmatrix}$. To generate bootstrap sample i, randomly draw j from $U[1,\, T]$. Let $s_j = X(i : i+L,\, :)$ in Matlab's indexing convention. Append s_j to X_i, which is initialized as $[\emptyset]$. Repeat until X_i is of length T. Unless T/L is an integer, the process yields a bootstrap sample of incorrect length. Build X_i to be at least length T then trim.

4. Estimate the two-stage regression on sample X_i and save the estimate $\widehat{\lambda}_i$

5. Repeat B times (we use 100,000 replications). The estimated $\widehat{\lambda}_i$ should be approximately mean zero, and $Std\left(\widehat{\lambda}_i\right) \approx SE\left(\widehat{\lambda}\right)$

6. Perform usual asymptotic tests

D Robustness

We present additional results showing the sensitivity of our results to changes in specification (or lack thereof).

D.1 Monthly Estimation

Table 11 presents risk price estimates using monthly returns on our primary test assets. The ICAPM estimated parameters and model fit are very similar to the daily results in Table 2. The 4-factor FF model fit has improved to nearly perfect, but the estimated risk prices (δ_{smb}, δ_{hml}, δ_{umd}) are half of the corresponding values in Table 2. In an i.i.d serially uncorrelated model, the SDF coefficients should be identical no matter what the frequency of observation[27]. This result suggests the 4-factor model is overfit, and hence, the estimates are not consistent across frequency.

[27] Subject to log-linearization error

Table 11: Risk Price Estimates

	α	δ_M	δ_λ	δ_{smb}	δ_{hml}	δ_{umd}	R^2	$MAPE$
2-Factor ICAPM	-	3.28	-9.39	-	-	-	78.3	0.995
		(8.39)	(-6.64)					
	1.39	2.75	-9.41	-	-	-	79.9	1.01
	(3.76)	(6.62)	(-6.66)					
CAPM	-	2.76	-	-	-	-	30.9	1.63
		(7.16)						
	1.31	2.25	-	-	-	-	32.3	1.72
	(3.48)	(5.71)						
4-Factor FF	-	4	-	1.66	6.67	4.46	94.9	0.469
		(7.52)		(1.34)	(8.36)	(10)		
	0.0855	3.96	-	1.68	6.65	4.45	94.9	0.475
	(1.1)	(7.52)		(1.35)	(8.27)	(9.97)		

Notes: This table shows risk prices estimated using monthly returns from August 1963 to December 2013 for the two-factor ICAPM, the CAPM, and the augmented Fama-French model. The test assets are value-weighted quintile portfolios sorted on ME, BE/ME, and Prior2-12. α is annualized and "-" indicates that the intercept is restricted to zero. MAPE is average absolute pricing error, annualized. Moving block bootstrap t-statistics are in parentheses

D.2 Future Market Return Horizon

Table 12 shows estimated δ_λ and cross-sectional R^2 using alternative horizons, T, to define $\widehat{\lambda} = \sum_{j=2}^{T} rx_{M,t+j}$ ranging from three months to two years (using daily return data). All estimates restrict the zero-beta rate. δ_λ declines almost monotonically with T, which is expected since $cov\left(rx_{i,t+1}, \hat{\lambda}_{t+1:t+T}\right)$ increases with T and hence δ_λ must decline. Cross-sectional R^2 are fairly stable across horizon, with a peak at one year.

Table 12: Alternative Horizons (Daily Returns)

	3m	6m	9m	12m	15m	18m	21m	24m
δ_λ	-12	-11	-9.6	-10	-8.4	-9	-5.8	-5.3
	(-4.3)	(-5.3)	(-5.1)	(-5.8)	(-5.3)	(-5.5)	(-4.8)	(-4.9)
R^2	63	81	83	88	79	71	71	75

Notes: Estimated risk price of discount rate factor and cross-sectional R^2 for alternative choices of moving average horizon. Data are daily returns with value-weighted quintile portfolios sorted on ME, BE/ME, and Prior2-12 as test assets. Moving block bootstrap t-statistics are in parentheses

Table 13 shows the results of repeating the exercise using monthly returns. The point estimates and patterns are similar, confirming that our results aren't driven by the choice of horizon for future market returns.

Table 13: Alternative Horizons (Monthly Returns)

	3m	6m	9m	12m	15m	18m	21m	24m
δ_λ	-14	-9.4	-7.2	-7.3	-7.2	-7.3	-5.3	-5.4
	(-7.6)	(-6.5)	(-6.1)	(-6.8)	(-6.9)	(-7.3)	(-6.9)	(-7.1)
R^2	79	78	72	75	73	70	73	75

Notes: Estimated risk price of discount rate factor and cross-sectional R^2 for alternative choices of moving average horizon. Data are monthly returns with value-weighted quintile portfolios sorted on ME, BE/ME, and Prior2-12 as test assets. Moving block bootstrap t-statistics are in parentheses

D.3 Fama-French 25

Our main results are presented using value-weighted quintile portfolios sorted on ME, BE/ME, and Prior2-12. Table 14 gives estimates using daily returns on the Fama-French 25 portfolios sorted on ME and BE/ME. As before, the 4-factor model has better fit than the ICAPM but at the expense of less stable estimates across horizons and test assets.

Table 14: Risk Price Estimates

	α	δ_M	δ_λ	δ_{smb}	δ_{hml}	δ_{umd}	R^2	$MAPE$
2-Factor ICAPM	-	4.65	-9.3	-	-	-	66	1.39
		(4.1)	(-3.87)					
	2.22	3.6	-8.7	-	-	-	68.7	1.36
	(2.31)	(3.5)	(-3.84)					
CAPM	-	4	-	-	-	-	-7.85	2.47
		(3.74)						
	5.46	1.53	-	-	-	-	10.5	2.26
	(3.59)	(1.44)						
4-Factor FF	-	7.31	-	4.23	17.7	16.4	79.3	1.02
		(4.51)		(1.36)	(4.12)	(2.1)		
	0.931	6.78	-	4.1	17.2	15.8	79.7	1.03
	(2.25)	(4.4)		(1.33)	(4.05)	(2.05)		

Notes: This table shows premia estimated using monthly returns from 01-Aug-1963 to 31-Dec-2013 for the two-factor ICAPM, the CAPM, and the augmented Fama-French model. The test assets are the 25 portfolios sorted on ME and BE/ME. α is annualized and "-" indicates that the intercept is restricted to zero. MAPE is average absolute pricing error, annualized. Moving block bootstrap t-statistics are in parentheses