MICHIGAN
ROSS SCHOOL OF BUSINESS

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
November 2014

This work cannot be used without the author's permission.
This paper can be downloaded without charge from the
Social Sciences Research Network Electronic Paper Collection:
http://ssrn.com/abstract=2518256

UNIVERSITY OF MICHIGAN

Linking Cross-Sectional and Aggregate Expected Returns*

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

November 19, 2014

Abstract

We propose a one state variable ICAPM that rationalizes the size, value, and momentum "anomalies" observed in stock returns. Our main insight is that differential covariance with news about future market discount rates drives the observed cross-sectional variation in expected returns. We find that in response to an increase in expected future market discount rates, large, growth, and recent losers company stocks outperform small, value, and recent winner stocks, respectively. Our interpretation is that such an increase in discount rates represents "bad" news for the representative investor, increasing his marginal utility of wealth. We further show that ignoring this state variable leads to drastic underestimation of the equilibrium price of "level risk" measured using bond returns. An augmented model which adds a "level" factor jointly prices both stock and bond returns.

*We thank John Cochrane, Eugene Fama, Stefano Giglio, Valentin Haddad, Lars Hansen, John Heaton, Bryan Kelly, Ralph Koijen, Diogo Palhares, Harald Uhlig 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[1] which is a linear combination of the return on the aggregate wealth portfolio and state variables which capture changes in the investment opportunity set and "outside income" (non-tradable wealth)[2]. 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.

We find that augmenting the one-factor CAPM[3] with a single additional factor, the shock to expected future excess market returns, is sufficient to rationalize the cross-section of expected returns of portfolios formed on size, book-to-market ratio, and recent performance (momentum). Since the CAPM completely fails in fitting the cross-section of portfolio sorts 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 an important difference. Instead of the typical VAR (vector auto-regression) 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. Both approaches yield consistent results but under different assumptions. Our methodology produces a very different cross-sectional patterns in factor loadings. 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.

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. In Appendix A we solve a general equilibrium model that 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.

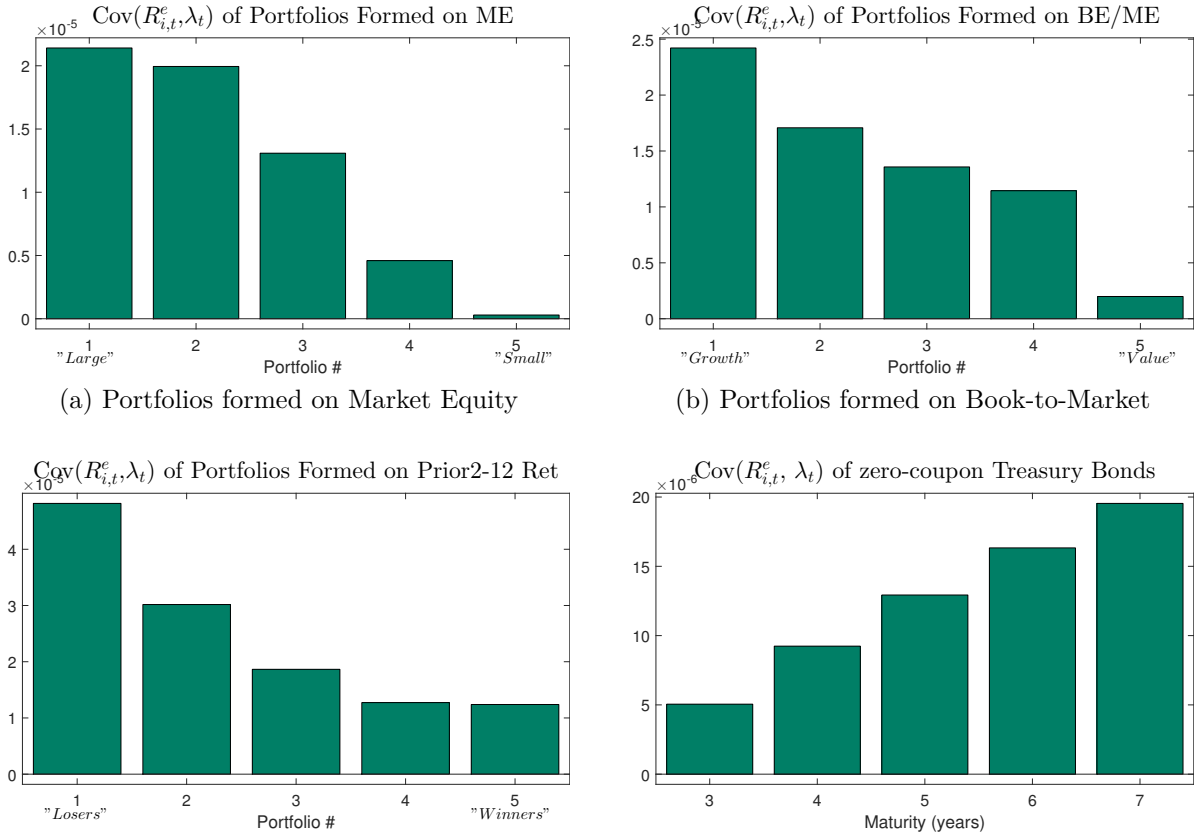
The second contribution of our paper is decomposing the average return differential between

[1] Also called a state-price density. Equivalently, no arbitrage implies the existence of a present-value function and of an equivalent martingale measure.

[2] 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)

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

Figure 1: Covariances of stock portfolios and bonds with the expected returns factor



(a) Portfolios formed on Market Equity

(b) Portfolios formed on Book-to-Market

(c) Portfolios formed on Prior 2-12m Return

(d) Zero-coupon Treasury Bonds

long and short term government bonds into a large positive return differential due to loadings on "level risk" (change in long-term yields) and a large negative return differential 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. Figure 1 shows the loadings of bonds and portfolios formed on size, book-to-market, and prior 1-year return on our expected return factor. Large stocks, growth stocks, recent losers, and long-term bonds have higher covariance with innovations in the expected excess market return than small stocks, value stocks, recent winners, and short-term bonds, respectively. 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.

Finally, we document that returns on the Fama-French factors *smb*, *hml*, and *mom* are useful forecasters of the future market excess return, significant both statistically and economically. In our sample from 1963 to 2010, the three combined greatly outperform the

dividend price ratio, $^{D}/_{P}$, in terms of R^2. This confirms the spreads in covariances seen in Figure 1 are of economically relevant magnitudes, aleviating concerns that our asset pricing exercise is simply a regression of "noise on noise."

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,

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

where $\boldsymbol{\sigma_{J,t}}$ is an $\mathbb{R}^2$-valued square-integrable utility-"volatility" process, J_t is the continuation utility for C at time t, conditional on current information, $f(C_t, J_t)$ is the flow utility, $A(J_t)$ is a variance multiplier that penalizes the variance of the utility "volatility" $\left\|\boldsymbol{\sigma_{J,t}}\right\| \equiv \boldsymbol{\sigma_{J,t}}\boldsymbol{\sigma}'_{\boldsymbol{J,t}}$, and $\boldsymbol{Z_t}$ is a vector of shocks. A pair (f, A) is called an aggregator. We use a Kreps-Porteus (Epstein-Zin-Weil) aggregator, defined as

$$f(C, J) = \frac{\delta}{\rho}\frac{C^\rho - J^\rho}{J^{\rho-1}} = \frac{\delta}{\rho}J\left[\left(\frac{C}{J}\right)^\rho - 1\right] \tag{2}$$

$$A(J) = -\frac{\alpha}{J}, \tag{3}$$

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:

$$\frac{dD}{D} = g_t dt + \boldsymbol{\sigma_D} d\boldsymbol{Z}$$

where g_t is the mean consumption growth,

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

We further extend the utility specification when the parameter α is stochastic,

$$d\alpha = \phi_\alpha(\bar{\alpha} - \alpha)\,dt + \sigma_\alpha dZ_\alpha, \tag{4}$$

This parameter can be interpreted as either time-varying risk aversion (Campbell and Cochrane (1999), Dew-Becker (2011)) or time-varying ambiguity aversion with respect to model specification (Drechsler (2013), Hansen and Sargent (2008))[4].

Finally, agent's wealth evolves according to:

$$\frac{dW}{W} = (\theta\lambda + r - C)\, dt + \theta \boldsymbol{\sigma_R} d\boldsymbol{Z}$$

where θ is the share invested in the risky asset, λ is the excess return on the risky assets, r is the risk-free rate.

2.2 Solution

Market clearing requires $C = D$, $W = P$, and $\theta = 1$. 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. *Equilibrium SDF corresponding to the problem in equations (1), (2), (3), and (4) is given by:*

$$\frac{d\Lambda}{\Lambda} = -rdt - \alpha dR_M - (\alpha - 1) \underbrace{a_\alpha}_{<0} \boldsymbol{\sigma_\alpha} d\boldsymbol{Z} - (\alpha - 1) \underbrace{a_g}_{>0} \boldsymbol{\sigma_g} d\boldsymbol{Z} \tag{5}$$

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.

Equity Risk premium is given by

$$\lambda = \lambda_0 + \alpha \times \boldsymbol{\sigma_D} \boldsymbol{\sigma_D}'$$

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

$$r = g - \lambda.$$

Proof. See Appendix A. □

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

[4]Variability in α generates time-varying prices of risk in the model. Our main results survive if we model time-varying quantity of risk (through time-varying volatility of consumption growth) instead and are therefore not exclusive to the preference specification chosen.

Proof. The result follows immediately from the fact that the coefficient a_α in (5) is negative, $a_\alpha < 0$ (see Appendix A). □

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

$$\begin{aligned} R^e_{i,t} &= \alpha \times cov_{t-1}\left(R^e_{i,t}, R^e_{M,t}\right) + a_g\left(\alpha - 1\right) \times cov_{t-1}\left(R^e_{i,t}, dg\right) \\ &\quad + \left(\alpha - 1\right) a_\alpha \times cov_{t-1}\left(R^e_{i,t}, d\lambda\right) \end{aligned} \tag{6}$$

where $R^e_{i,t}$ is the excess return on any test asset, $R^e_{M,t}$ is the excess return on the market portfolio, $d\lambda$ are shocks to the equity risk premium, and dg are shocks to the growth rate. The term in square brackets is negative for all plausible parameter values.

Proof. See Appendix A. □

2.3 Unconditional Pricing

We further proceed with the discrete version of the model and define $rx_{i,t+1}$ as the log return on an asset i over time period $t \to t+1$. When instanteneous excess returns $R^e_{i,t}$ are normally distributed, corresponding discrete-time excess returns $rx_{i,t+1}$ are log-normally distributed.

The model above is specified conditionally with "shocks" as factors. With two assumptions, we derive an unconditional representation in terms of "levels".

Assumption 1. *Covariances with market and expected returns factors $c_{iM} \equiv cov_t\left(rx_{i,t+1}, rx_{M,t+1}\right)$, $c_{i\lambda} \equiv cov_t\left(rx_{i,t+1}, \lambda_{t+1}\right)$, and variances $V_i = var_t\left(rx_{i,t+1}\right)$ are constant in time.*

Our goal is to capture time variation in Sharpe ratios. Empirical evidence that links conditional volatilities and conditional means is weak (Welch and Goyal, 2008). We therefore, use this assumption to focus on time variation in Sharpe ratios that is driven entirely by variation in expected returns.

Assumption 2. *Expected log market excess returns $\lambda_{t+1} = E_{t+1} rx_{M,t+2}$ follow an AR(1) process.*

This assumption, while quite restrictive, serves primarily the purpose of deriving closed-form solutions and quantifying the effects discussed below. It allows us to capture persistence in the risk premium while keeping the model very tractable. The assumption is similar in nature to those made in the literature, which typically assume that the vector of state variables follows a VAR(1) process (see, for example, Campbell (1996), Campbell and Vuolteenaho (2004)).

Assumption 3. *Consumption growth rate is constant,* $\boldsymbol{\sigma_g} = 0$.

Assumption 3 serves solely for expositional purposes in this chapter and will be relaxed in Section 4.

Theorem 4. *Given Assumption 1, Assumption 2, Assumption 3, and conditional model in Eq. 6, we obtain the following linear pricing relation:*

$$\forall i,t: \qquad E\left(rx_{i,t+1}\right) + \frac{V_i}{2} = cov\left(rx_{i,t+1}, rx_{M,t+1}\right) \times \hat{\delta}_M + cov\left(rx_{i,t+1}, \lambda_{t+1}\right) \times \hat{\delta}_\lambda \tag{7}$$

where $\hat{\delta}_M$ *and* $\hat{\delta}_\lambda$ *are two constant unconditional prices of market and expected returns risk, respectively. Section B.2 derives the link between these prices and the conditional ones in Appendix, Eq. 23.*

Proof. Refer to the proof of a more general Theorem 9 with the arbitrary number of factors and risk prices in Appendix, Section 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 result is an implication of homoskedasticity and the AR(1) structure imposed above.

2.4 Expected Returns Factor

2.4.1 Using Future Realized Returns

The unconditional model in Eq. 7 involves the expected return factor λ_{t+1} which is not directly observable by an econometrician. Empirical literature that faces this issue typically uses the residual from predictive regressions instead. Campbell (1996), Campbell and Vuolteenaho (2004) uses the VAR approach with macroeconomic and financial variables that have been shown to be good forecasters of expected returns and thus changes in investment opportunity set. This approach narrows the information set that is used to make predictions about future returns and therefore can be potentially biased. Moreover, the information that is being used is limited to a small set of forecasters that have limited ability in forecasting returns.

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. Future realized

returns as an estimate of expected returns $rx_{M,t+2} = E_{t+1}rx_{M,t+2}+\varepsilon_{t+2}$ is indeed an unbiased estimator, since for any information set $\mathcal{F}_{t+1}$ at time $t+1$,

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

True (population) conditional covariances are also equal

$$cov_t(rx_{i,t+1}, rx_{M,t+2}) = cov_t(rx_{i,t+1}, E_{t+1}rx_{M,t+2})$$

and thus an unbiased and consistent estimator of the covariance on the RHS is also an unbiased and consistent estimator of the covariance on the LHS. See Section B.4 for a more formal argument.

Our empirical approach allows us to test the model without directly estimating expected market returns. We require a restrictive AR(1) assumption, but we make no assumptions about investors' information set. Instead, our approach is based on the agent's ability to forecast returns rather than econometrician's. Unlike previous work, we do not rely on any forecasting regressions.

2.4.2 Moving Average of Future Realized Returns

One issue with using one period ahead realized returns is that the signal-to-noise ratio of such proxy is very low. If expected returns are persistent, we could cumulate sufficiently long series of future returns in order to improve the ratio. There is an obvious trade-off here: increasing the length of the cumulative sum will improve the signal-to-noise ratio at a decreasing rate. We will 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 market risk premium in Eq. 7.

Theorem 5. *When the expected returns factor is measured over long horizon, the unconditional relation in Eq. 7 still holds, with the new prices of risk that are a linear transform of the ones in Eq. 7:*

$$\forall i,t: \qquad \begin{aligned} E(rx_{i,t+1}) + \frac{V_i}{2} &= cov(rx_{i,t+1}, rx_{M,t+1}) \times \tilde{\delta}_M \\ &+ cov(rx_{i,t+1}, \lambda_{t+1:t+T}) \times \tilde{\delta}_\lambda \end{aligned} \tag{8}$$

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. Section B.2 in the Appendix derives the link between these prices and the conditional ones in Eq. 6 and Eq. 26.

Proof. Refer to the proof of a more general Theorem 10 with an arbitrary number of factors and risk prices in Appendix, Section B.2. □

In the case when the moving average of future returns is used as an estimator of the expected market returns factor, we can still show its unbiasedness. See Section B.2.1 for more details.

2.5 Price of Expected Returns Risk

We now make the following assumption that will allow us easily compare the sign on price of expected returns risk in conditional and unconditional models.

Assumption 4. *Price of the expected returns (discount rate) risk is constant, var* $(\delta_{\lambda,t}) = 0$.

This assumption states that the time-variation in expected returns is driven exclusively by the price of market risk. This implies that the expected return on a long-short portfolio of two assets that have the same loading on the market factor is not time-varying.

Assumption 4 delivers the following result:

Theorem 6. *When Assumptions 1, 2, and 4 hold, the sign of the price of expected returns risk is preserved:*

$$sign\left\{E\left[\delta_{\lambda,t}\right]\right\} = sign\left\{\tilde{\delta}_{\lambda}\right\}$$

where $\delta_{\lambda,t}$ *is the price of expected returns risk in conditional model (6) and* $\tilde{\delta}_{\lambda}$ *is the price of expected returns risk in either of conditional models in (7) or (8).*

Proof. See the proof in Appendix, Section B.3. □

2.6 Empirical Relation

Equipped with the results in Theorem 4 and Theorem 5, we proceed with approximating the unconditional relation in Eq. 8 to facilitate the empirical tests in the following Section 3. Section B.5 in the Appendix shows, that with daily data, the unconditional pricing relation in Eq. 8 can be approximated by the following relation:

$$\begin{aligned} \forall i,t: \qquad E\left(R^{e}_{i,t+1}\right) &= cov\left(rx_{i,t+1}, rx_{M,t+1}\right) \times \tilde{\delta}_{M} \\ &+ cov\left(rx_{i,t+1}, \hat{\lambda}_{t+1:t+T}\right) \times \tilde{\delta}_{\hat{\lambda}} \end{aligned} \tag{9}$$

where $R_i^e \equiv \frac{R_i}{R_f}$ 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 Eq. 27).

All the variables in Eq. 9 are observed by an econometrician. We can use the level of excess returns, R_i^e, 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.

In the Appendix, Section 3.4 we compare the empirical specification in Eq. 9 to a similar one in Campbell and Vuolteenaho (2004).

3 Empirical Link Between Cross-Sectional and Aggregate Expected Returns

We estimate and test the expected return relation of Eq. 9 using two sets of test assets. The first is the canonical 25 portfolios formed by a two-way sort of firms on market capitalization (SIZE) and book-to-market ratio (BE/ME), available at Ken French's website [5].

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.

Therefore, in addition to the canonical 25 portfolios sorted on size and book-to-market ratio, 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) [6]. 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, MOM (Mark M. 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). Therefore, 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.

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

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

Following the spirit of Merton (1973) and the continuous-time model in Section 2, we use portfolio returns measured at daily frequency[7]. All returns are measured over the period July 1963 to December 2010. Though monthly returns are available going back to 1927, we choose to use the shorter sample of daily returns. Using daily 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 because 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 (and inflation is not well measured even at monthly frequency (Cecchetti, 1997), we use only excess log returns, which, by construction, are real returns.

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.

Table 1 below shows the estimated covariances of asset returns with the factors. Panel A shows $Cov\left(rx_{i,t}, rx_{m,t}\right)$ with GMM standard errors in parentheses. 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 last column give the estimates for the canonical Fama-French factors, SMB, HML, and MOM. The covariances match the well known pattern in market betas.

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 realized returns in place of expected returns produces consistent and unbiased covariances estimates, they will be less precisely estimated due to

[7]We also replicated the analysis at monthly frequency and obtain very similar results.

Table 1: Covariances

	Q1	Q2	Q3	Q4	Q5	FF	Q5-Q1
Panel A: $C_{i,M}$							
ME	10.25	9.62	9.26	9.20	7.82	-1.00	-2.39
BE/ME	10.66	9.63	9.21	8.99	9.31	-1.59	-1.38
Prior2-12	12.17	10.07	9.45	9.34	10.80	-0.91	-1.42
Panel B: $C_{i,\lambda}$							
ME	2.14	1.99	1.31	0.46	0.03	-1.22 (1.5)	-2.11 (1.5)
BE/ME	2.42	1.71	1.36	1.14	0.20	-1.96 (2.4)	-2.23 (2.3)
Prior2-12	4.82	3.02	1.86	1.27	1.24	-2.29 (1.5)	-3.02 (1.7)
Panel C: $E\left[R_i^e\right]$							
ME	2.94	5.11	5.52	5.51	5.23	1.63	1.40
BE/ME	2.48	3.71	4.20	5.88	7.54	4.55	4.23
Prior2-12	-2.19	3.05	3.02	5.11	7.62	7.86	7.71

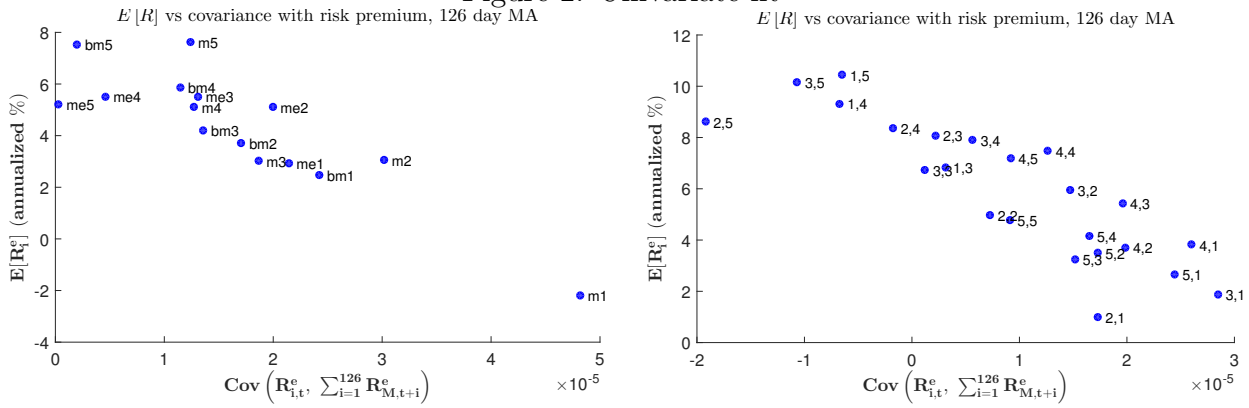
Notes: This table shows covariances and annualized mean returns estimated over 1963:07:01-2010:12:31. Panel A lists the covariances of portfolio returns with the market return, $C_{i,M}$. Panel B depicts the covariances of portfolio returns with the expected returns factor, $C_{i,\lambda}$. Panel C shows 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 *mom*. Moving-block bootstrap t-statistics in parentheses.

the extra noise present in realized returns. Still, the covariances of the spread portfolios with $\widetilde{\lambda}$ are statistically significantly different from zero and the covariances follow a reliable pattern, suggesting that we are not just "regressing noise on noise."

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. When risk aversion rises, demand for all risky assets falls, increasing their risk premia. 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 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 the sample average returns, which are monotonically increasing from left to right across quintiles, consistent with the well known size, value, and momentum premia. 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

Figure 2: Univariate fit



(a) Quintile Portfolios

(b) Fama-French 25 Portfolios

Notes: The left plot shows sample values of $E\left[r_{ix,t}\right]$ 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 right panel depicts same results for the 25 Fama-French portfolios.

be clearly seen in Figure 2. Figure 2 (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 2 (b) is the same plot for the 25 Fama-French portfolios. The graphs confirm the suspicion 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 an additional risk factor that captures the size, value, and momentum effects.

3.1 Estimation Results

We estimate risk price vector $\delta = \left[\delta_M \ \delta_\lambda\right]'$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 and 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 *mom* (momentum) factor of Mark M. Carhart (1997). For comparison, all models are written and estimated in terms of covariances instead of regression βs. Below is a summary of the pricing equations of the relevant models, where δs are interpreted as risk prices (coefficients in the SDF):

$$
\begin{aligned}
\text{2-Factor ICAPM: } E\left[R_i^e\right] &= C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda \\
\text{2-Factor ICAPM (free intercept): } E\left[R_i^e\right] &= \alpha + C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda \\
\text{CAPM, Restricted: } E\left[R_i^e\right] &= C_{i,M}\delta_M \\
\text{4-Factor FF, Unrestricted: } E\left[R_i^e\right] &= \alpha + C_{i,M}\delta_M + C_{i,smb}\delta_{smb} + C_{i,hml}\delta_{hml} + C_{i,mom}\delta_{mom} \\
\text{4-Factor FF, Restricted: } E\left[R_i^e\right] &= C_{i,M}\delta_M + C_{i,smb}\delta_{smb} + C_{i,hml}\delta_{hml} + C_{i,mom}\delta_{mom} \\
\text{where } C_{i,X} &\equiv Cov\left[rx_{i,t},\ X_t\right].
\end{aligned}
$$

Estimated risk prices are given in Table 2 along with sample R^2 and mean absolute pricing errors[8]. Quantitatively, our two-factor ICAPM fits the cross-section of average returns nearly as well as the augmented 4-factor Fama-French 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-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 (Joel L. 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 3. The graphs plot model implied average 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 that the price of risk is positive is $P\left[\delta_\lambda > 0\right] = 0.002\%$ (one-sided t-test of $H_0: \lambda \geq 0$ easily rejects the null). The alternative approaches therefore 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

[8]For the 4-factor model, 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

Model	α	δ_M	δ_λ	δ_{smb}	δ_{hml}	δ_{mom}	R^2	$MAPE$
2-Factor ICAPM	-	3.48	-10.21	-	-	-	75%	0.95%
		(3.01)	(-4.08)					
	0.26%	3.31	-9.83	-	-	-	76%	0.96%
	(0.23)	(3.02)	(-4.00)					
CAPM	-	1.53	-	-	-	-	-150%	2.86%
		(1.56)						
4-Factor FF	-	3.90	-	2.54	10.22	7.06	90%	0.61%
		(2.73)		(0.71)	(2.83)	(3.74)		
	0.13%	3.79	-	2.42	10.00	6.95	90%	0.61%
	(0.46)	(2.83)		(0.67)	(2.83)	(3.78)		

Notes: This table shows premia estimated from the 1963:07:01-2010:12:31 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

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.

3.3 Factor Mimicking Portfolio

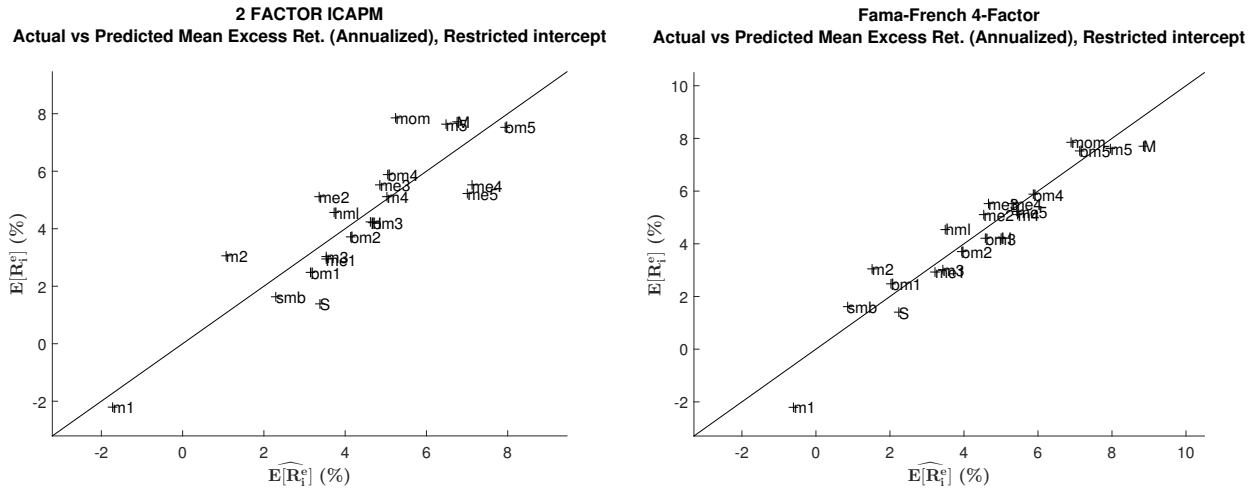
The two-stage OLS procedure for estimating stochastic discount factors suffers from many problems related to samples size and factor structure in the covariance of test asset returns. Lewellen et al. (2010) highlight these concerns and offer some suggestions:

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"

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 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). smb, hml, and mom are the canonical Fama-French-Carhart factors. S, H, and M are 5-1 quintile spread portfolios.

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 concern of overfitting, we alternatively construct $\widehat{\lambda}_t = proj\left(\lambda_t \middle| \begin{bmatrix} mrkt_t & smb_t & hml_t & mom_t \end{bmatrix}\right)$. This is the linear combination of the four Fama-French-Carhart factors which has maximal correlation with our original expected return factor. Since it is an expected return we can include it as a test asset and use GLS methods.

If the 2-factor ICAPM is literally true we can derive the following restriction on the SDF loadings, assuming $proj\left(\lambda_t \middle| \begin{bmatrix} mrkt_t & smb_t & hml_t & mom_t \end{bmatrix}\right)$ is the true mimicking portfolio

(including other assets doesn't improve model fit[9]).

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

The unrestricted four-factor SDF is $m_t = \delta_M \cdot mrkt_t + \delta_{smb} \cdot smb_t + \delta_{hml} \cdot hml_t + \delta_{mom} \cdot mom_t$. 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 $i \in \begin{Bmatrix} smb & hml & mom \end{Bmatrix}$. Table 3 shows the implied and direct coefficients on $\begin{bmatrix} mrkt & smb & hml & mom \end{bmatrix}$ in the SDF. The unrestricted coefficient on SMB is smaller than the ICAPM implied value and the coefficient on MOM 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 3. SMB has a lower average return than predicted by it's covariance with λ and MOM has a positive ICAPM α. Since SDF weights are proportional to average returns when the factors are uncorrelated[10] the ICAPM αs translates directly into the difference SDF coefficients. Still, it is reassuring that the restriction isn't drastically violated.

Table 3: SDF Restriction

	Restricted ZB-rate		Unrestricted ZB-rate	
	Implied	4-Factor	Implied	4-Factor
Market	4.13	3.90	3.94	3.79
SMB	4.90	2.54	4.72	2.42
HML	10.67	10.22	10.28	10.00
MOM	5.97	7.06	5.75	6.95

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 $proj\left(\lambda_t \middle| \begin{bmatrix} mrkt & smb & hml & mom \end{bmatrix}\right)$. The remaining implied coefficients are $\delta_i = \delta_\lambda \cdot b_i$ with δ_λ and b_i from the same source.

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

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

[10] Market, SMB, HML, and MOM are nearly uncorrelated.

GLS[11]. 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 4 shows the estimated SDF using both. $E[r_M]$ 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.

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 substantally so. Bootstrap simulation rejects the null of $R^2 = 0$ with $p \approx 0.1\%$. Recall from Table 2 that the CAPM has a negative R^2 due to the restricted zero-beta rate and the well known issue that average returns among these assets are negatively correlated with market betas. For comparison, the 4-factor GLS R^2 is 79% (not shown).

Table 4: GLS Estimation

	δ_M	δ_λ	$E[r_M]$	$E[r_\lambda]$	R^2	$MAPE$
OLS	3.30	-10.25	3.26	-8.01	72%	0.91%
	(2.86)	(-3.81)				
GLS	3.78	-12.26	3.46	-9.58	62%	1.03%
	(2.90)	(-4.54)				

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 average in-sample 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.

3.4 Campbell-Shiller Decomposition

Assumption 2 implies that the expected returns factor can be any finite linear combination of expected future returns. In particular, it can be $\sum_{j=1}^{\infty} \rho^j r_{t+1+j}$ from the return decomposition given in Campbell and Shiller (1988) . Our empirics do not uniquely identify which one it is.

[11]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. The ICAPM R^2 drops from 73% to 69% and risk prices, δs, are similar ($\pm 10\%$). We ignore sampling uncertainty in b_{SMB}, b_{HML}, b_{MOM} 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_{MOM} are between 2 and 3.

We now compare our model to a similar one in Campbell and Vuolteenaho (2004). Campbell and Vuolteenaho (2004) starts with exploiting the Campbell and Shiller (1988) decomposition:

$$\begin{aligned} r_{t+1} - E_t r_{t+1} &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \\ &- (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &= N_{CF,t+1} - N_{DR,t+1} \end{aligned}$$

where N_{CF} denotes news to future cash flows and N_{DR} denotes news about future discount rates.

By relying on homoskedasticity assumption and linear approximations, Campbell (1993) derives an approximate discrete-time version of Merton (1973) ICAPM:

$$\begin{aligned} E_t\left(r_{i,t+1}\right) - r_{f,t+1} + \frac{\sigma_{i,t}^2}{2} &= \gamma \times cov_t\left(r_{i,t+1}, r_{M,t+1} - E_t r_{M,t+1}\right) \\ &+ (1-\gamma) \times cov_t\left(r_{i,t+1}, -N_{DR,t+1}\right) \\ &= \gamma \times cov_t\left(r_{i,t+1}, N_{CF,t+1}\right) \\ &- cov_t\left(r_{i,t+1}, N_{DR,t+1}\right) \end{aligned} \tag{10}$$

In Theorem 5 we showed that the sum of future returns over arbitrary horizon can be used as an expected returns factor. In fact, it can be easily shown that the statement holds for any weighted sum as well with market risks being scaled by a constant. In this way, the sum of discount rates in Eq. 10, $\sum_{j=1}^{\infty} \rho^j r_{t+1+j}$, is directly comparable to our expected returns factor factor λ. The formulation (10) therefore is similar to the conditional model in Eq. 6. It implies however a positive price of risk on discount rate factor (which is a scaled version of our expected returns factor with positive scaling factor).

Eq. 10 can be rewritten in terms of risk prices in Table 2:

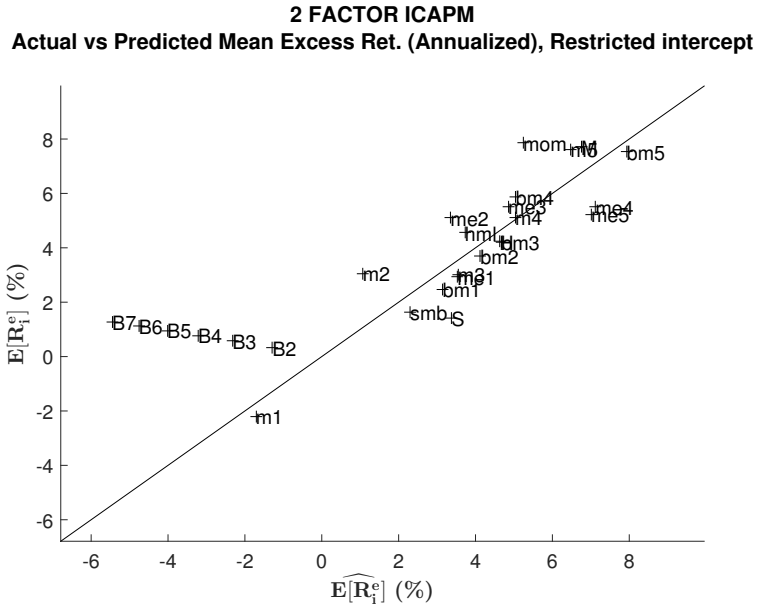
$$\begin{aligned} E_t\left(r_{i,t+1}\right) - r_{f,t+1} + \frac{\sigma_{i,t}^2}{2} &= \delta_M \times cov_t\left(r_{i,t+1}, N_{CF,t+1}\right) \\ &+ (\delta_\lambda - \delta_M) \times cov_t\left(r_{i,t+1}, N_{DR,t+1}\right) \end{aligned} \tag{11}$$

The essence of Campbell and Shiller (1988) decomposition is that unexpected stock market returns could be mechanically decomposed into a positive cash-flow news component and a negative discount-rate component, resulting in positive and negative betas, respectively. Further, Campbell and Vuolteenaho (2004) find that with Epstein-Zin (or power utility)

preferences, CAPM does not hold because the discount rate effect in Eq. 11 *is weakened* as compared to the CAPM (which holds with exponential utility, for instance). The effect is not, however, fully reversed: the price of cash-flow risk is still positive and the price of the discount-rate risk is negative. In this paper we find that the discount-rate effect is, in fact, greatly amplified when investor's risk aversion increases in bad states of the world. Table 2 shows that the loading on the discount-rate risk, δ_λ, in presence of the market returns as a second factor in the SDF is significantly negative. The price of the discount-rate risk in Eq. 11, $\delta_\lambda - \delta_M$, is therefore also strongly negative and higher in magnitude than the price of the cash-flow risk.

4 Pricing Bonds

Figure 4: Pricing Bonds



Notes: We plot the fitted and sample mean values of expected returns in the model Eq. 8. Test assets include the stock portfolios we used before as well as 5 bonds with maturities from 3 to 7 years, labeled B3-B7, respectively.

Whereas there are numerous papers which explore risk premia separately for equities and fixed income securities, few study these important assets in a unified framework[12]. We extend our analysis to include risk-free government bonds and interest rate risk.Figure 4 illustrates that the unconditional model in Eq. 8 fails to price bond excess returns of maturities from 3 to 7 years. Pricing errors are big and the slope is completely wrong. Clearly, the market

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

and expected returns factors are not sufficient to explain the risk compensation required for holding these securities.

We use Eq. 6 to motivate a third factor – the growth rate of the economy. The factor helps price bonds and stocks jointly. To proxy for the changes in the growth rate of the economy, we use the return on a long-maturity government bond. The yield on a long-term bond is related to the "level" factor in bonds, which has been shown to to be the only priced factor in the cross-section of nominal bonds (Cochrane and Piazzesi, 2008).

The extended model from Section 2 is:

$$\begin{aligned} E_t\left(rx_{i,t+1}\right)+\frac{1}{2}\sigma_t^2\left(rx_{i,t+1}\right) &= cov_t\left(rx_{i,t+1},rx_{M,t+1}\right)\times\delta_{M,t} \\ &+ cov_t\left(rx_{i,t+1},\lambda_{t+1}\right)\times\delta_{\lambda,t} \\ &+ cov_t\left(rx_{i,t+1},rx_{B,t+1}\right)\times\delta_{B,t} \end{aligned} \tag{12}$$

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

Theorem 7. *Given Assumption 1 and Assumption 2 and conditional model in Eq. 12, we obtain the following linear pricing relation:*

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

$$+ cov\left(rx_{i,t+1},\lambda_{t+1}\right)\times\hat{\delta}_\lambda$$

$$+ cov\left(rx_{i,t+1},rx_{B,t+1}\right)\times\hat{\delta}_B \tag{14}$$

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

Proof. This is a straightforward application of Theorem 9 in Operationalizing the Model, Section B.2 with two factors (market and bond) and zero price of bond risk. □

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

Theorem 8. *When the expected returns factor is measured over long a horizon, the unconditional relation in Eq. 13 still holds, with the new prices of risk that are a linear transform of the ones in Eq. 13:*

$$\forall i,t: \qquad E\left(rx_{i,t+1}\right)+\frac{V_i}{2} = cov\left(rx_{i,t+1},rx_{M,t+1}\right)\times\tilde{\delta}_M \tag{15}$$
$$+ \; cov\left(rx_{i,t+1},\lambda_{t+1:t+T}\right)\times\tilde{\delta}_\lambda$$
$$+ \; cov\left(rx_{i,t+1},rx_{B,t+1}\right)\times\tilde{\delta}_B \tag{16}$$

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. Section B.2 in the Operationalizing the Model derives the link between these prices and the conditional ones in Eq. 12.

Proof. Refer to the proof of a more general Theorem 10 with an arbitrary number of factors and risk prices in Operationalizing the Model, Section B.2. □

Using Theorem 8 and Section B.5 in the Operationalizing the Model we can show in the same way as in Section 2.6 that the following approximate relation holds:

$$\forall i,t: \qquad E\left(R^e_{i,t+1}\right) = cov\left(rx_{i,t+1},rx_{M,t+1}\right)\times\tilde{\delta}_M \tag{17}$$
$$+ \; cov\left(rx_{i,t+1},\hat{\lambda}_{t+1:t+T}\right)\times\tilde{\delta}_{\hat{\lambda}}$$
$$+ \; cov\left(rx_{i,t+1},rx_{B,t+1}\right)\times\tilde{\delta}_B$$

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 Eq. 17.

5 Bond Risks and Risk Premia

5.1 "Level" factor

Most popular term structure models feature factors typically labelled "level", "slope", and "curvature" Ang and Piazzesi (2003), Diebold and Li (2006), Diebold et al. (2005), Litterman and Scheinkman (1991). In these models, the level factor induces parallel shifts in the log-yield curve. In the estimated affine term structure model of Cochrane and Piazzesi (2008), unconditional expected bond excess returns are described by $E\left(rx^{(i)}_{t+1}\right)=\delta_B\times cov\left(rx^{(i)}_{t+1},\left(E_{t+1}-E_t\right)level_{t+1}\right)$[13]. The single factor driving expected returns is the shock to the "level" of yields.

[13] rx^i_{t+1} is the log excess return on an i year bond for period $t+1$

By construction, the yield on a hypothetical bond of infinite maturity is proportionate to the level factor. Approximating the long-end of the yield curve to be flat we can write

$$\begin{aligned} ln\left(P_{t+1}^{T-1}/P_t^T\right) &= \frac{T-1}{T}\left(y_t^T - y_{t+1}^{T-1}\right) \\ &\approx \frac{T-1}{T}\left(y_t^{LT} - y_{t+1}^{LT}\right) \end{aligned}$$

where P_t^T is the price of a T year bond at time t, and y_t^{LT} is the log yield-to-maturity of a "long-term" bond at time t. The left hand side is the log return on the long-term bond and the right hand side is a multiple of the change in the level factor (change in the long-term yield). This shows that the return on a long-term bond is essentially perfectly negatively correlated with the level factor.

Written in terms of shocks we have,

$$(E_{t+1} - E_t)\left[ln\left(P_{t+1}^{T-1}/P_t^T\right)\right] \approx (E_{t+1} - E_t)\left[\frac{T-1}{T}\left(y_t^{LT} - y_{t+1}^{LT}\right)\right]$$

where we have equated the unexpected return on the long-term bond with the shock to the level factor. The model of Cochrane and Piazzesi (2008) can be recast as

$$E\left(rx_{t+1}^{(i)}\right) = \widetilde{\delta}_B \times cov\left(rx_{t+1}^{(i)}, (E_{t+1} - E_t)\, r_{t+1}^{(LT)}\right).$$

Since the model is written in real terms but we only observe nominal bond returns, we substitute[14] $rx_{t+1}^{(LT)}$, the excess return on the long-term bond, in place of $r_{t+1}^{(LT)}$, yielding

$$\begin{aligned} E\left(rx_{t+1}^{(n)}\right) &= \widetilde{\delta}_B \times cov\left(rx_{t+1}^{(i)}, (E_{t+1} - E_t)\, rx_{t+1}^{(LT)}\right) \\ &= \widetilde{\delta}_B C_{n,B} \end{aligned} \tag{18}$$

5.2 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, 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

[14] This assumes the Fisher identity holds for short-term bonds.

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 daily zero-coupon bond returns from the GSW yields, we must interpolate between the given maturities. We use linear interpolation though this is technically inconsistent with the functional form used by GSW. For daily (or even monthly) returns, this introduces negligible error since the GSW function is smooth and the weight given to the nearest whole year maturity is essentially 1 (actually 251/252). Finally, we must choose how to deal with timing. We use the convention of 252 trading days per year and treat each trading day as being 1 "day" after the previous trading day. This means that we have essentially eliminated weekends and trading holidays from the calendar. This introduces measurement error in the returns that is reduced as the measurement horizon increases (from daily to monthly to quarterly, etc). Excess returns just subtract the log return on a one month t-bill, the same procedure we use for stock excess returns.

5.3 Estimating the price of "level risk"

With daily excess log bond returns in hand, we estimate the model of using the same two-stage procedure of Section 3.1. $\widetilde{\delta}_B$ is estimated to be 3.6. The cross-sectional R^2 is 95% with an annualized mean absolute pricing error of 0.09%. Figure 5 shows graphically the good fit of the model.

In the context of Eq. 8, we argue that $\widetilde{\delta}_B$, the price of "level risk", is commonly underestimated when using Eq. 18 and bond excess returns. It is a classic case of "omitted variable bias"[15]. Eq. 8 and the results of Section 3.1 suggest at least two such missing variables, $C_{i,\lambda} = 10^6 \times Cov\left[rx_{t+1}^{(i)},\ E_t\left(\sum_{i=1}^{k} rx_{M,t+i}\right)\right]$ and $C_{i,M} = 10^6 \times Cov\left[rx_{t+1}^{(i)},\ rx_{M,t+1}\right]$. Table 5 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, $\forall i,\ C_{i,\lambda} \approx 1.4 C_{i,B}$. Cross-sectionally, $corr\left(C_{i,B},\ C_{i,\lambda}\right) \approx 1$. Since we know Section 3.1 that $\delta_\lambda \neq 0$, the univariate level model suffers greatly from omitted variables bias. Using the estimate of $\delta_\lambda = 10.3$, a back-of-the-envelope calculation suggests the true $\delta_B = 3 + 1.4 \times 10.3 = 17.4$. 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 returns than in the hypothetical economy where the expected market return is constant.

[15] See Wooldridge (2002, Chapter 4.4)

Figure 5:

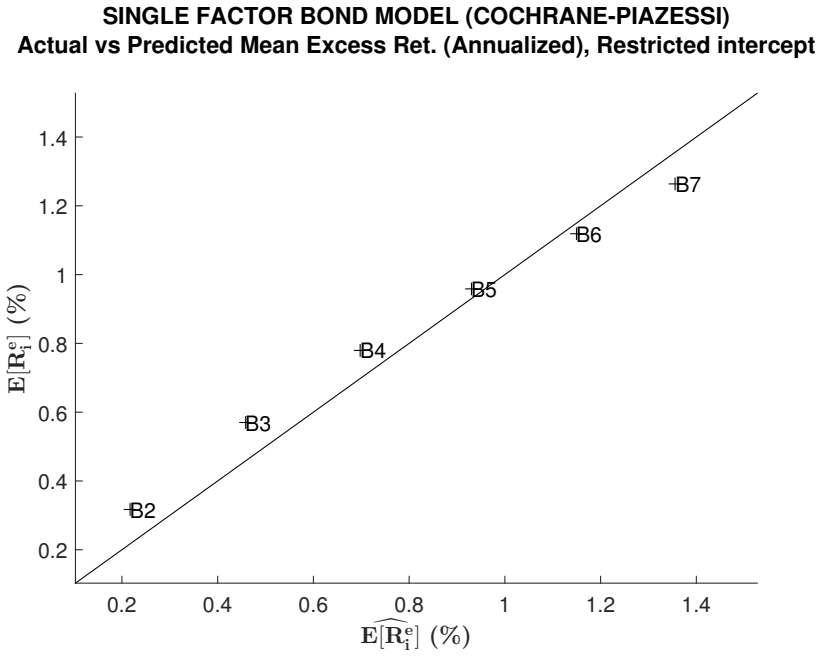


Table 5: Covariances

	2Y	3Y	4Y	5Y	6Y	7Y
$C_{i,B}$	0.29	0.60	0.92	1.22	1.51	1.78
$C_{i,\lambda}$	0.51	0.92	1.29	1.63	1.95	2.26
$C_{i,M}$	0.01	0.03	0.07	0.12	0.17	0.24

$$\begin{aligned} C_{i,B} &= 10^5 \times Cov\left[rx_{t+1}^{(i)}, (E_{t+1} - E_t)\, rx_{t+1}^{(LT)}\right] \\ C_{i,\lambda} &= 10^5 \times Cov\left[rx_{t+1}^{(i)},\ E_t\left(\Sigma_i^k rx_{M,t+i}\right)\right] \\ C_{i,M} &= 10^5 \times Cov\left[rx_{t+1}^{(i)},\ rx_{M,t+1}\right] \end{aligned}$$

This intuition is formalized by estimating the 3-factor ICAPM given by Eq. 15. Table 6 gives estimated risk prices (δs) from the following models:

$$\begin{aligned} \text{2-Factor ICAPM: } E\left[R_i^e\right] &= C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda \\ \text{Univariate Level Risk: } E\left[R_i^e\right] &= C_{i,B}\delta_B \\ \text{3-Factor ICAPM: } E\left[R_i^e\right] &= C_{i,M}\delta_M + C_{i,\lambda}\delta_\lambda + C_{i,B}\delta_B \\ \text{where } C_{i,X} &\equiv Cov\left[rx_{i,t},\ X_t\right] \end{aligned}$$

Table 6: Risk Price Estimates

Model	δ_M	δ_λ	δ_B	R^2	$MAPE$
2-Factor ICAPM	3.48	-10.21	-	2%	1.75%
	(3.01)	(-4.03)			
Level Risk	-	-	3.01	94%	0.07%
			(1.21)		
3-Factor ICAPM	3.10	-9.60	15.12	88%	0.73%
	(2.84)	(-3.97)	(3.47)		

Notes: This table shows premia estimated from the 1963:07:01-2010:12:31 for the 2 ICAPM (estimated using stock portfolios), the Level Risk model (estimated using bond returns), and the 3-factor ICAPM (estimated using both stocks and bonds). Model intercepts are restricted to zero. MAPE is average absolute pricing error, annualized. Moving block bootstrap t-statistics are in parentheses

Table 7: Bond Expected Returns

Maturity	Sample Mean	Level Risk	3-Factor ICAPM
2	0.32	0.22	-0.12
3	0.57	0.46	0.08
4	0.78	0.70	0.42
5	0.96	0.93	0.79
6	1.12	1.15	1.15
7	1.26	1.36	1.49

Notes: Annualized percent returns by maturity.

. All models are estimated with the intercept restricted to zero. The 2-factor ICAPM is estimated using only the stock portfolios from Section 3 and hence the risk price estimates are the same as in Section 3.1. The univariate Level Risk model is estimated using only the bond excess returns. 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 though they are excluded from the estimation of risk prices (see Figure 4). Importantly, δ_B in the 3-factor ICAPM is $\approx 16 \gg 3$. This is similar to the back of the envelope prediction given above.

Table 7 gives annualized percent returns by maturity in sample, implied by the univariate Level Risk model, and implied by the 3-factor ICAPM. Both models imply as larger term premium (spread between long and short maturity average returns) than is observed in the data, with the 3-factor model performing somewhat worse, but still quite well.

Figure 6 shows average returns vs 3-factor ICAPM expected returns for bonds and stock portfolios. The graphs plot model implied average 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 as well as in Figure 3 (using the 2-factor ICAPM) and bonds fit quite well.

Figure 6:

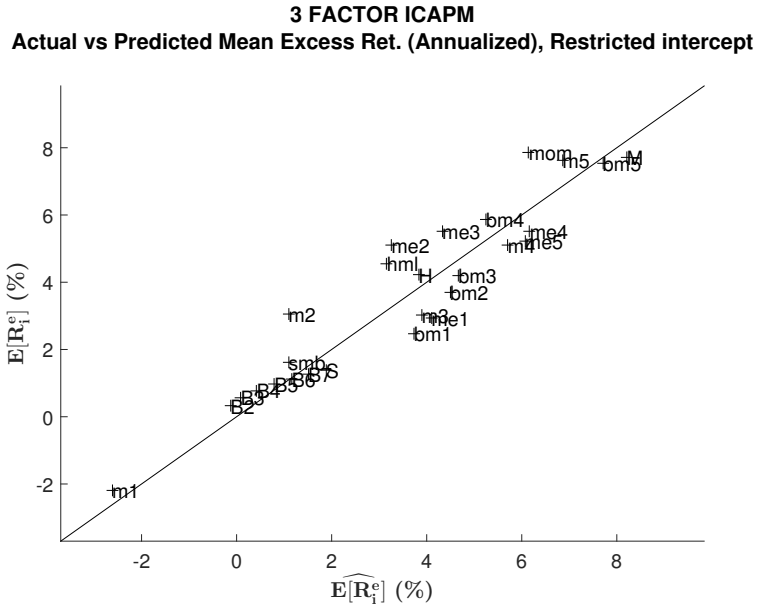
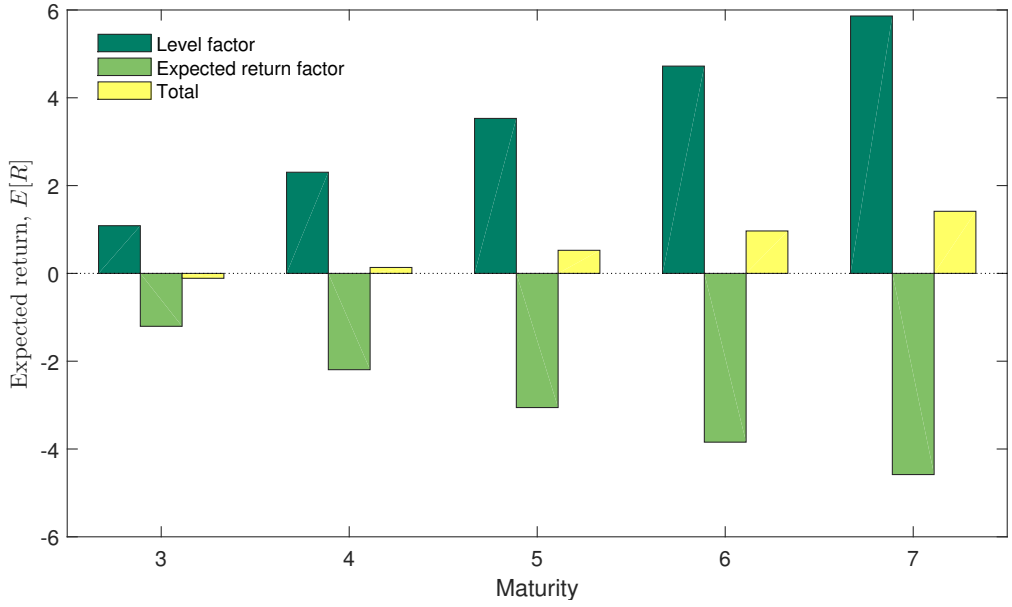


Figure 7 decomposes the expected excess return on the various bonds. The premium due to market risk, $C_{i,M}$, is excluded since it is negligible for bonds. Bonds earn a large premium for loading on level risk, whereas they command a large negative premium for loading on the expected return factor. 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 opposite

interpretation. We find that bonds hedge against increases in expected stock returns but Koijen et. al. 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. In contrast, the estimates in Koijen et al. (2010) results in a flat term structure (no term premium).

Figure 7: Decomposition of Bond Risk Premia



6 Predicting the Future Market using Cross-Section

In our empirical methodology, we used the future realized excess returns as a proxy for today's market expectation of future excess returns. We further showed that this proxy was key in explaining the cross section of stock returns. The motivation came from Merton (1973) ICAPM, who documented that when investment opportunity set, or expected returns, in particular, are time varying, investors hedge the risk with regard to the source of variation. If test assets have differential loadings on these hedging factors, the corresponding investors' hedging demands get reflected in the cross-section of asset returns.

This observation can be viewed from the opposite perspective. If expected returns manifest in the cross-section, the cross-section of stock returns can itself provide information about future expected returns. Indeed, "Priced factors ... are innovations in state variables that predict future returns." (Brennan et al., 2004). It is therefore natural to ask a question whether cross-sectional variables can predict future returns and to what extent. Few recent papers have looked at this question. Kelly and Pruitt (2011) uses the cross-section

Table 8: Predictability

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

Horizon (k)	DP	$MRKT$	SMB	HML	MOM	R^2
3 months	1.43	0.79				1.23
6 months	1.39	0.26				1.86
9 months	1.39	0.08				2.51
12 months	1.41	-0.30				3.05
3 months	1.58	0.67	-3.04	-2.72	-3.23	8.48
6 months	1.58	0.08	-2.57	-2.17	-2.13	8.21
9 months	1.56	-0.37	-2.29	-1.98	-2.25	8.50
12 months	1.56	-0.46	-1.68	-1.45	-2.43	8.18

Notes: t-statistics and R^2 from predictive regressions at various horizons

of dividend-price ratios to show that they indeed predict the future returns well beyond the aggregate dividend-price ratio variable. They develop a statistical procedure that uses information from the cross-section to predict aggregate returns.

Our paper does not aim to construct the optimal predictor; we merely want to show that predictability is indeed present and use it as a robustness check of our methodology. In particular, we want to make sure the covariances reported in Table 1 and Figure 2 are *economically* significant. As such, we use the returns on SMB, HML, and MOM portfolios to forecast future market returns. If future returns help to explain the cross-section, the cross-sectional returns themselves mechanically should predict future returns. An important question is whether this predictability is economically significant and thus can be credibly exploited in the cross-sectional tests. We test this with the following regression:

$$R^e_{M,t+1:t+k} = a + [DP_t \; MRKT_{t-90:t} \; SMB_{t-90:t} \; HML_{t-90:t} \; MOM_{t-90:t}]\, b + \varepsilon_{M,t+1} \quad (19)$$

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

The top panel of Table 8 reports the t-statistics of estimated coefficients in Eq. 19 and R^2 at various horizons (3, 6, 9, and 12 months) with only the market and dividend-price ratio included as predictors. There is little evidence of return predictibility 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 MOM as additional predictors. We find that all of the coefficients for each variable at 3-9 months horizon are highly significant and negative. This means that each of SMB, HML, and MOM predict future market return *independently*. Also, when either of them tank, the market is predicted to go up. Recall

that SMB, HML, and MOM typically fall in times of high marginal utility, when the risk premium is high. The results in Table 8 fully support this interpretation. The table shows that the R^2 improves substantially by including the SMB, HML, and MOM predictors. We therefore conclude that SMB, HML, and MOM are *economically* significant predictors of future market returns. Similar evidence of predictability was documented by Liew and Vassalou (2000). They show that SMB and HML help forecast future rates of economic growth.

References

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregressino of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*, 745–787.

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. (1993). Intertemporal Asset Pricing without Consumption Data. *The American Economic Review*, 487–512.

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

Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies 1*(3), 195–228.

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.

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

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

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

Cochrane, J. 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.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Diebold, F. X., M. Piazzesi, and G. D. Rudebusch (2005, May). Modeling Bond Yields in Finance and Macroeconomics. *AEA Papers and Proceedings 95(2)*, 415–420.

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.

Joel L. Horowitz (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.

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.

Litterman, R. and J. Scheinkman (1991). Common factors affecting bond returns. *Journal of Fixed Income 1*, 51–61.

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

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

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.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. The MIT press.

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 Eq. 2 and Eq. 3. Although such a representation requires computing an additional variance term in Eq. 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, $\frac{1}{2}\frac{\alpha}{J}\left\|\boldsymbol{\sigma_{J,t}}\right\|^2$ 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,

$$J_t = \mathbb{E}_t\left(\int_t^\infty \left[f\left(C_\tau, J_\tau\right) + \frac{1}{2}A\left(J_\tau\right) \parallel J_{\boldsymbol{X}}\left(\boldsymbol{X_\tau}, \tau\right)\sigma_{\boldsymbol{X}}\left(\boldsymbol{X_\tau}, W_\tau, \tau\right) \parallel^2\right] d\tau\right), \tag{20}$$

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

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

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

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

First-order conditions

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

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:

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

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

$$F(\alpha, g) = exp(a_0 + a_\alpha \alpha + a_g g).$$

Plug everything in:

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

Equalize coefficients near α, g, and const:

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

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:

$$\begin{aligned}
\lambda &= \alpha\boldsymbol{\sigma_D}\boldsymbol{\sigma_D}' - \left(a_\alpha\boldsymbol{\sigma_D}\boldsymbol{\sigma_\alpha}' + a_g\boldsymbol{\sigma_D}\boldsymbol{\sigma_g}'\right) \\
&= \text{const} + \alpha \times \boldsymbol{\sigma_D}\boldsymbol{\sigma_D}'
\end{aligned}$$

Risk-free rate:

$$\begin{aligned}
r = \frac{1}{dt}\mathbb{E}\left[\frac{dC}{C}\right] - \lambda &= g - \lambda \\
&= \text{const} + g - \alpha \times \boldsymbol{\sigma_D}\boldsymbol{\sigma_D}'
\end{aligned}$$

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

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

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

where operator $\mathcal{L}(\cdot)$ 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:

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

Given the expression for λ and r above, this can be rewritten as:

$$\begin{aligned} R^e_{i,t} = \alpha \times \text{cov}_{t-1}\left(R^e_{i,t}, R^e_{M,t}\right) &+ \left(\alpha - 1\right)\left[a_\alpha\left(\boldsymbol{\sigma_D}\boldsymbol{\sigma_D'}\right)^{-1} + a_g\right] \times \text{cov}_{t-1}\left(R^e_{i,t}, d\lambda\right) \\ &+ a_g\left(\alpha - 1\right) \times \text{cov}_{t-1}\left(R^e_{i,t}, dr\right) \end{aligned}$$

Note that since λ and r are negatively correlated, the sign of λ risk may change due to inclusion of factor r instead of g.

B Operationalizing the Model

B.1 SDF

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

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

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

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

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

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

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

B.2 Unconditional Pricing Relation

Let the conditional model be

$$\begin{aligned} E_t\left(rx_{i,t+1}\right)+\frac{1}{2}var_t\left(rx_{i,t+1}\right) &= cov_t\left(rx_{i,t+1},f_{t+1}\right)\times\delta_{f,t} \\ &+ cov_t\left(rx_{i,t+1},\lambda_{t+1}\right)\times\delta_{\lambda,t} \end{aligned} \tag{21}$$

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 5. *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

$$\begin{aligned} E_t\left(rx_{i,t+1}\right)+\frac{V_i}{2} &= c_{if}^{'}\times\delta_{f,t}+c_{i\lambda}^{'}\times\delta_{\lambda,t} \\ &\equiv C_i\times D_t \end{aligned} \tag{22}$$

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

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

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

Theorem 9. *Given the assumptions 1 and 2 and the conditional model (21), we obtain the following linear pricing relation:*

$$\begin{aligned} E\left(rx_{i,t+1}\right)+\frac{V_i}{2} &= cov\left(rx_{i,t+1},f_{t+1}\right)\times\hat{\delta}_f \\ &+ cov\left(rx_{i,t+1},\lambda_{t+1}\right)\times\hat{\delta}_\lambda \\ &\equiv cov\left(rx_{i,t+1},\left[\begin{array}{c}f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times\widetilde{D} \\ where\ \widetilde{D} &= \Theta^{-1}E\left[D_t\right] \\ \Theta &= \mathbb{I}_{2k}+\Phi_{f\lambda}\times\left[\begin{array}{cc}\mathbb{I}_k & \Lambda^{'}\end{array}\right] \\ \Phi_{f\lambda} &= \left[\begin{array}{c}\Gamma_f\Sigma_f+\Gamma_{f\lambda}\Sigma_{f\lambda}^{'}\\ \Gamma_\lambda\Sigma_{f\lambda}^{'}+\Gamma_{f\lambda}\Sigma_f\end{array}\right] \end{aligned} \tag{23}$$

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 **??** to get

$$
\begin{aligned}
E\left(rx_{i,t+1}\right)+\frac{V_i}{2} &= cov\left(rx_{i,t+1},f_{t+1}\right)\times E\left[\delta_{f,t}\right]+cov\left(rx_{i,t+1},\lambda_{t+1}\right)\times E\left[\delta_{\lambda,t}\right] \\
&- cov\left(E_trx_{i,t+1},E_tf_{t+1}\right)\times E\left[\delta_{f,t}\right]+cov\left(E_trx_{i,t+1},E_t\lambda_{t+1}\right)\times E\left[\delta_{\lambda,t}\right] \quad (24)
\end{aligned}
$$

Using Assumption 5, we can write $E_t\lambda_{t+1}=\Lambda_0+\Lambda\lambda_t$. Then

$$
\begin{aligned}
E\left(rx_{i,t+1}\right)+\frac{V_i}{2}=C_i\times D &= cov\left(rx_{i,t+1},\left[\begin{array}{c} f_{t+1} \\ \lambda_{t+1}\end{array}\right]\right)\times D \\
&- cov\left(E_trx_{i,t+1},E_tf_{t+1}\right)\times\left[\begin{array}{cc}\mathbb{I}_k & \Lambda'\end{array}\right]\times D \quad (25)
\end{aligned}
$$

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

We now use Eq. 22 to substitute the expressions for two covariates in $cov\left(E_trx_{i,t+1},E_tf_{t+1}\right)$ term

$$
\begin{aligned}
&cov\left(c_{if}'\times\delta_{f,t}+c_{i\lambda}'\times\delta_{\lambda,t}-\frac{V_i}{2},\Sigma_f\times\delta_{f,t}+\Sigma_{f\lambda}\times\delta_{\lambda,t}-\frac{1}{2}V_f\right) \\
&=c_{if}'var\left[\delta_f\right]\Sigma_f+c_{i\lambda}'var\left[\delta_\lambda\right]\Sigma_{f\lambda}'+c_{if}'cov\left[\delta_f,\delta_\lambda\right]\Sigma_{f\lambda}'+c_{i\lambda}'cov\left[\delta_f,\delta_\lambda\right]\Sigma_f \\
&=C_i\times\left[\begin{array}{c}\Gamma_f\Sigma_f+\Gamma_{f\lambda}\Sigma_{f\lambda}' \\ \Gamma_\lambda\Sigma_{f\lambda}'+\Gamma_{f\lambda}\Sigma_f\end{array}\right] \\
&\equiv C_i\times\Phi_{f\lambda}
\end{aligned}
$$

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

Plugging this back in Eq. 25 and collecting terms on the LHS and RHS gives

$$
\begin{aligned}
&C_i\times\left[\mathbb{I}_{2k}+\Phi_{f\lambda}\times\left[\begin{array}{cc}\mathbb{I}_k & \Lambda'\end{array}\right]\right]\times D \\
&=cov\left(rx_{i,t+1},\left[\begin{array}{c}f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times D
\end{aligned}
$$

Denote

$$
\Theta=\mathbb{I}_{2k}+\Phi_{f\lambda}\times\left[\begin{array}{cc}\mathbb{I}_k & \Lambda'\end{array}\right]
$$

and assume that Θ is invertible. Then

$$
C_i=cov\left(rx_{i,t+1},\left[\begin{array}{c}f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times\Theta^{-1}
$$

and hence

$$
\begin{aligned}
E\left(rx_{i,t+1}\right)+\frac{V_i}{2} &= C_i\times D \\
&= cov\left(rx_{i,t+1},\left[\begin{array}{c}f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times\left\{\Theta^{-1}D\right\} \\
&= cov\left(rx_{i,t+1},\left[\begin{array}{c}f_{t+1}\\ \lambda_{t+1}\end{array}\right]\right)\times\tilde{D}
\end{aligned}
$$

□

Theorem 10. *When factor risk premia is measured over long horizon*

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

the unconditional relation in Eq. 23 holds with Θ *given by*

$$\Theta=\mathbb{I}_{2k}+\Phi_{f\lambda}\times\left[\begin{array}{cc}\mathbb{I}_k & \mathbb{L}\end{array}\right] \tag{26}$$

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 6[16],

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

Plugging this into Eq. 24, the proof of Theorem 9 carries over with Λ replaced by $\mathbb{L}=\Lambda\times\left(\mathbb{I}_k-\Lambda\right)^{-1}\left(\mathbb{I}_k-\Lambda\right)^{\top}$. □

B.2.1 Unbiasedness of the moving average estimator

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

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

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

Similarly, true unconditional covariances are equal,

$$cov\left(rx_{i,t+1},rx_{M,t+2:T}\right)=cov\left(rx_{i,t+1},E\left[\sum_{j=2}^{T}rx_{M,t+j}\left|\mathcal{F}_t\right.\right]\right)+\underbrace{cov\left(rx_{i,t+1},\sum_{j=2}^{T}\varepsilon_{M,t+j}\right)}_{=0} \tag{27}$$

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

[16] 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_tf_{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.

B.3 Price of Expected Returns Risk

Theorem 11. *When Assumptions 1, 2, and 4 hold, the sign of the price of expected returns risk is preserved:*

$$sign\left\{E\left[\delta_{\lambda,t}\right]\right\} = sign\left\{\hat{\delta}_{\lambda}\right\}$$

where $\delta_{\lambda,t}$ is the price of expected returns risk in conditional model in **??** *and $\hat{\delta}_{\lambda}$ is the price of expected returns risk in either of conditional models in Eq. 7 and Eq. 8.*

Proof. Apply Theorem 9 to this special case as follows:

$$\Phi_{f\lambda} = \begin{bmatrix} \Gamma_f \Sigma_f + \Gamma_{f\lambda}\Sigma_{f\lambda}' \\ \Gamma_{\lambda}\Sigma_{f\lambda}' + \Gamma_{f\lambda}\Sigma_f \end{bmatrix} = \begin{bmatrix} var\left[rx_M\right] var\left[\delta_M\right] \\ 0 \end{bmatrix}$$

Therefore,

$$\begin{aligned} \Theta &= \mathbb{I}_{2k} + \Phi_{f\lambda} \times \begin{bmatrix} \mathbb{I}_k & \Lambda' \end{bmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{bmatrix} var\left[rx_M\right] var\left[\delta_M\right] \\ 0 \end{bmatrix} \times \begin{bmatrix} 1 & \Lambda \end{bmatrix} \\ &= \begin{bmatrix} 1 + var\left[rx_M\right] var\left[\delta_M\right] & \Lambda var\left[rx_M\right] var\left[\delta_M\right] \\ 0 & 1 \end{bmatrix} \\ \Theta^{-1} &= \begin{bmatrix} \left(1 + var\left[rx_M\right] var\left[\delta_M\right]\right)^{-1} & -\frac{\Lambda var[rx_M] var[\delta_M]}{1+var[rx_M] var[\delta_M]} \\ 0 & 1 \end{bmatrix} \end{aligned}$$

where Λ is an AR(1) coefficient on the lagged price of market risk.

The derivation implies that the transformation $\tilde{D} = \Theta^{-1} D$ preserves the sign and the magnitude of the price of expected returns risk. □

B.4 Covariances

Conditional Covariances are Equal:

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

Unconditional Covariances are Equal

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

B.5 Empirical Relation

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

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

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

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

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

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

So when estimating a pricing equation, we use log returns to estimate covariances and simple returns to estimate the LHS of Eq. 23, $E\left(rx_{i,t+1}\right) + \frac{V_i}{2}$.

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 $\widetilde{r}_t = r_t - E\left[r_t\right]$
   (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, $Cov\left(\widetilde{r}_t,\, f_t\right) = Cov\left(r_t,\, f_t\right)$

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

D.1 Alternative measure of expected returns

We show that our results are robust to using volatility of stock returns instead of future realized returns as a factor.

Table 9: Risk Price Estimates

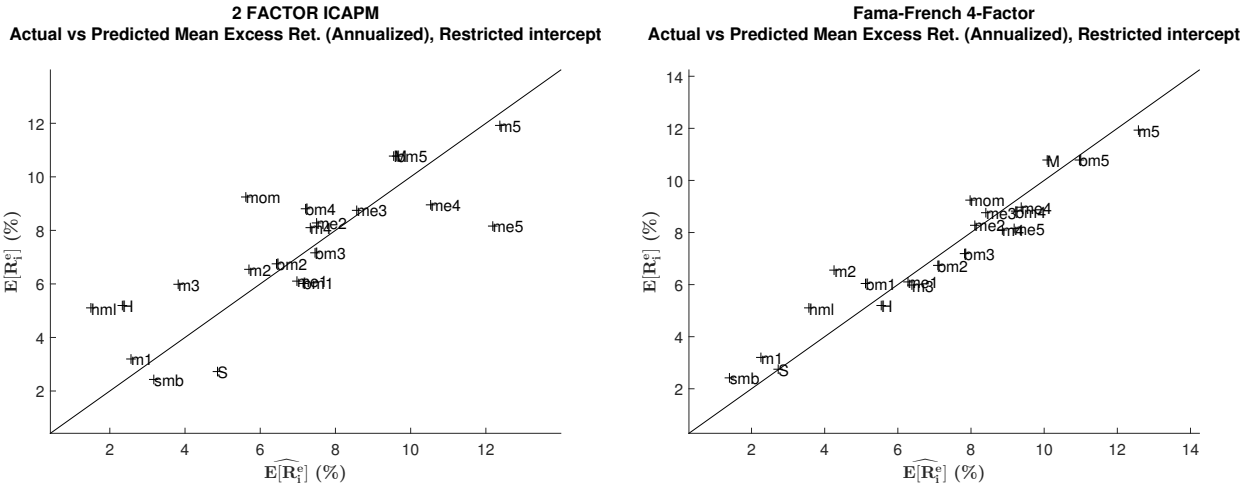
Model	α	δ_M	δ_λ	δ_{smb}	δ_{hml}	δ_{mom}	R^2	$MAPE$
2-Factor ICAPM	-	3.37	-197.45	-	-	-	45%	1.39%
		(3.30)	(-3.56)					
	2.57%	2.23	-137.07	-	-	-	63%	1.18%
	(2.36)	(1.98)	(-2.30)					
CAPM	-	2.84	-	-	-	-	-238%	3.36%
		(2.99)						
4-Factor FF	-	5.65	-	3.86	11.88	8.43	88%	0.65%
		(4.32)		(1.09)	(3.49)	(4.66)		
	1.80%	4.31	-	2.24	8.88	6.95	94%	0.48%
	(4.45)	(3.15)		(0.62)	(2.57)	(3.68)		

Notes: This table shows premia estimated from the 1963:07:01-2010:12:31 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

We use rolling one month daily volatility as our factor λ. Table 9 shows estimated risk prices and Figure 8 depicts the overall fit of the model. Even though the fit deteriorates, it is still significant and potent in explaining the cross-section of portfolio sorts.

We also estimated volatility using GARCH(1,1) model and perform the same analysis. The results are essentially identical.

Figure 8: 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). smb, hml, and mom are the canonical Fama-French-Carhart factors. S, H, and M are 5-1 quintile spread portfolios.