Online Appendix to

Why do discount rates vary?*

Serhiy Kozak

University of Maryland, Robert H. Smith School of Business, College Park, MD

Shrihari Santosh

University of Colorado Boulder, Leeds School of Business, Boulder, CO

December 18, 2019

*Kozak (corresponding author): sekozak@umd.edu
We gratefully acknowledge helpful comments from conference and seminar participants at the European Finance Association meetings, Finance Down Under Conference, Society for Financial Studies Finance Cavalcade, American Finance Association meetings, University of Chicago, Emory University, University of Maryland, University of Michigan, University of California at Los Angeles, University of Illinois at Urbana-Champaign, and University of Wisconsin. In addition, we owe much to helpful suggestions from an anonymous referee, Nick Barberis, Mike Chernov, John Cochrane, Julien Cujean, Eugene Fama, Stefano Giglio, Valentin Haddad, Lars Hansen, John Heaton, Bryan Kelly, Ralph Koijen, Mark Loewenstein, Stefan Nagel, Diogo Palhares, Christopher Polk, Bill Schwert (the editor), Jay Shanken, Ken Singleton, and Stijn Van Nieuwerburgh. Any errors are our own.

Internet Appendix A. Comparing to VARs

We now show that Assumption 2 is weaker than what is assumed by the VAR methodology. That is, the VAR method imposes additional restrictions on the dynamic of expected returns.

Let $z_t' = \begin{bmatrix} r_{m,t} & x_t' \end{bmatrix}'$ where x_t is an N vector of state variables. On unconditionally demeaned data, the VAR representation is $z_{t+1} = Az_t + \nu_{t+1}$ and hence, $\mathrm{E}\left[R_{m,t+k}\,|\mathcal{F}_t\right] = e_1' A^k z_t$. Specializing to $k = 1$ we have $\lambda_t \equiv \mathrm{E}\left[R_{m,t+1}\,|\mathcal{F}_t\right] = e_1' A z_t$ where $e_1' = \begin{bmatrix} 1 & \mathbf{0} \end{bmatrix}'$. Notice that $\mathrm{E}\left[R_{m,t+k}\,|\mathcal{F}_t\right]$ and λ_t are both in the linear span of z_t. Let B be any $N \times (N+1)$ matrix of rank N such that

$$cov\left(\begin{bmatrix} e_1'A \\ B \end{bmatrix} z_t\right) = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \tag{A.1}$$

is block diagonal. Bz defines a basis of the orthogonal complement of λ_t in the linear span of z_t.[1] Define $\psi_t = Bz_t$. Let $D = \begin{bmatrix} e_1'A \\ B \end{bmatrix}$ and define $\tilde{z}_t = Dz_t$. Note that $Dz_t = \begin{bmatrix} \lambda_t & \psi_t' \end{bmatrix}'$. We want to find the vector θ_k such that

$$e_1' A^k z_t = \theta_k' \tilde{z} = \theta_k' D z_t \tag{A.2}$$

holds date by date. Matching coefficients on z_t and right-multiplying by D^{-1} we obtain $\theta_k' = e_1' A^k D^{-1}$. Let b_k be the first element of θ_k and c_k be the remaining elements. Then substituting in we have

$$\mathrm{E}\left[R_{m,t+k}\,|\mathcal{F}_t\right] = a_k + b_k \lambda_t + c_k' \psi_t \tag{A.3}$$

with no error term ($e_{t,t+k} = 0$). Further, b_k and c_k are fully determined by the VAR parameter A.[2] Assumption 2 imposes no restrictions on c_k and only assumes that b_k is non-negative. Except for non-negativity of b_k, Assumption 2 is strictly weaker than the VAR assumption given a set of observable variables x. In Section 3.2 we showed that $b_k = \mathrm{corr}\left(\lambda_{t+k-1},\, \lambda_t\right)$, the univariate autocorrelation of expected returns at lag k. Since this is likely to be positive at the horizons we consider, non-negativity is not restrictive.

Internet Appendix B. Optimal filtration

In this section we explore optimal construction of the proxy $\hat{\lambda}_{t:t+K}$ by studying backward linear filters using future returns. If the filter satisfies $\hat{\lambda}_{t:t+K} = \lambda_t + \xi_{t:t+K}$ and $\mathrm{E}_t\left(\xi_{t:t+K}\right) = 0$, we have $\mathrm{cov}\left(R_{i,t},\, \hat{\lambda}_{t:t+K}\right) = \mathrm{cov}\left(R_{i,t},\, \lambda_t\right)$. Our main goal is efficient estimation of $\mathrm{cov}\left(R_{i,t}, \lambda_t\right)$. Specifically, given a sample of length T, we want to find (i) a horizon K, and (ii) weights $\omega = \{\omega_k > 0\}$, which jointly minimize the mean squared error (MSE),

$$\mathrm{E}\left(\left[\widehat{\mathrm{cov}}\left(R_{i,t}, \hat{\lambda}_{t:t+K}\right) - \mathrm{cov}\left(R_{i,t}, \lambda_t\right)\right]^2\right), \tag{B.1}$$

where sample covariances are estimated as usual.

We do this in two steps: (i) for arbitrary K we derive an approximate analytic solution for ω which minimizes the MSE and (ii) we then minimize over K via simulation. To preserve tractability, we restrict our analysis to a homoskedastic setting and an AR(1) process for λ_t, $\lambda_{t+1} = \bar{\lambda} +$

[1] There are an infinite number of such bases.

[2] Alternative rotations B will generate different values for c_k without changing b_k.

$\phi\left(\lambda_t - \bar{\lambda}\right) + \nu_{t+1}$, where $\phi > 0$ is the persistence of expected returns.

B.1. Optimal weights

For fixed K and iid Gaussian data, the MSE in Eq. (B.1) is

$$\text{MSE} = \frac{1}{T-K-1}\left[\text{cov}^2\left(R_{i,t},\lambda_t\right) + \text{var}\left(R_{i,t}\right)\left[\text{var}\left(\lambda_t\right) + \text{var}\left(\xi_{t:t+K}\right)\right]\right], \tag{B.2}$$

and hence minimizing $\text{var}\left(\xi_{t:t+K}\right)$ minimizes the MSE.[3] Therefore we choose ω to minimize this error variance. Given the AR(1) assumption for λ, we can express future returns as

$$R_{m,t+k} - \bar{\lambda} = \phi^{k-1}\left(\lambda_t - \bar{\lambda}\right) + e_{t+k},$$

with $\text{E}_t\left(e_{t+k}\right) = 0$. Therefore a filter $\{\omega_k\}$ of the form

$$\hat{\lambda}_{t:t+K} = \sum_{k=1}^{K}\omega_k\left(R_{m,t+k} - \bar{\lambda}\right) \tag{B.3}$$

satisfies $\hat{\lambda}_{t:t+K} = \left(\lambda_t - \bar{\lambda}\right) + \xi_{t:t+K}$ if and only if $\sum_k \tilde{\omega}_k = 1$, where $\tilde{\omega}_k = \phi^{k-1}\omega_k$. Finding the optimal $\tilde{\omega}$ recovers the optimal ω.

Define transformed returns as $\tilde{R}_{m,t+k} = \phi^{1-k}\left(R_{m,t+k} - \bar{\lambda}\right) = \left(\lambda_t - \bar{\lambda}\right) + \tilde{e}_{t+k}$. The filter in Eq. (B.3) is given by

$$\hat{\lambda}_{t:t+K} = \sum_{k=1}^{K}\omega_k\left(R_{m,t+k} - \bar{\lambda}\right) = \sum_{k=1}^{K}\tilde{\omega}_k\tilde{R}_{m,t+k} \tag{B.4}$$

$$= \sum_{k=1}^{K}\tilde{\omega}_k\left(\lambda_t - \bar{\lambda}\right) + \sum_{k=1}^{K}\tilde{\omega}_k\tilde{e}_{t+k} \tag{B.5}$$

We now solve for the optimal filter on transformed returns. The Lagrangian for the minimization of the residual variance, $\text{var}\left(\xi_{t:t+K}\right)$, is:

$$\mathcal{L} = \frac{1}{2}\tilde{\omega}'\Sigma\tilde{\omega} - \kappa\left(\tilde{\omega}'\mathbf{1} - 1\right), \tag{B.6}$$

where $\Sigma = \text{cov}\left(\left[\begin{array}{ccc}\tilde{e}_{t+1} & \cdots & \tilde{e}_{t+T}\end{array}\right]'\right)$ is the unconditional covariance of forecast errors, $\tilde{e}_{t+k} = \phi^{1-k}e_{t+k}$. The first order conditions are

$$0 = \Sigma\tilde{\omega} - \kappa\mathbf{1} \tag{B.7}$$

$$0 = \tilde{\omega}'\mathbf{1} - 1. \tag{B.8}$$

[3] In an iid Gaussian setting, the sample covariance is unbiased and $\text{var}\left[\widehat{\text{cov}}\left(x_t, y_t\right)\right] = \frac{1}{T-1}\left(1+\rho_{x,y}^2\right)\sigma_x^2\sigma_y^2$. Substituting $R_{i,t} = x_t$ and $\hat{\lambda}_t = y_t$ we obtain the result. Persistence of λ_t and our use of overlapping observations induces positive autocorrelation in $\xi_{t:t+K}$. Eq. (B.2) thus likely understates the true MSE. In simulation, however, we find that minimizing Eq. (B.2) approximately minimizes the true MSE.

Solving we obtain the explicit solution

$$\tilde{\omega} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}, \tag{B.9}$$

which is essentially a reverse-time Kalman filter.

Uncorrelated errors In the simple case where $\tilde{e}_{t+k}$ are mutually uncorrelated, Σ is diagonal and we obtain

$$\frac{\omega_{k+1}}{\omega_k} = \phi\eta = \phi^{-1}\frac{\tilde{\omega}_{k+1}}{\tilde{\omega}_k}, \tag{B.10}$$

where $\eta = \left(\frac{\sigma_r^2 - \phi^{2k-2}\sigma_\lambda^2}{\sigma_r^2 - \phi^{2k}\sigma_\lambda^2}\right)$, $\sigma_r^2 = \operatorname{var}(R_{m,t+1})$, $\sigma_\lambda^2 = \operatorname{var}(\lambda_t)$, and we use the fact that $\operatorname{var}(e_{t+k}) = \sigma_r^2 - \phi^{2k-2}\sigma_\lambda^2$ is the unconditional variance of forecast errors at horizon k.[4] Since $\sigma_r^2 \gg \sigma_\lambda^2$, we have $\eta \approx 1$ and the optimal filter weights exhibit approximate exponential decay at rate ϕ – the persistence of expected returns. Notice that the ratios do not depend on the horizon K and hence, they hold for all K.

Correlated errors Generally, the errors will not be perfectly uncorrelated.[5] To recover the optimal filter in this case, we now prove that a population regression of λ_t on $\left\{\tilde{R}_{m,t+k}\right\}$ recovers a scaled optimal solution, $c\tilde{\omega}$.[6] In simulation with various parameter choices we can run such a "population" regression and recover $\tilde{\omega}$. For notational simplicity we assume returns are unconditionally demeaned. Let $\sigma_\lambda^2 = \operatorname{var}(\lambda_t)$ be the unconditional variance of λ. Then we have $\operatorname{cov}\left(\begin{bmatrix}\tilde{R}_{m,t+1} & \cdots & \tilde{R}_{m,t+k}\end{bmatrix}', \lambda_t\right) = \sigma_\lambda^2\mathbf{1}$. By the law of total covariance $\operatorname{cov}\left(\begin{bmatrix}\tilde{R}_{m,t+1} & \cdots & \tilde{R}_{m,t+k}\end{bmatrix}'\right) = \Sigma + \sigma_\lambda^2\mathbf{1}\mathbf{1}'$. Substituting into the standard OLS formula and using the Sherman-Morrison formula to expand the inverse of the sum, the OLS coefficients are given by

$$\tilde{\omega}^{ols} = \left(\Sigma + \sigma_\lambda^2\mathbf{1}\mathbf{1}'\right)^{-1}\sigma_\lambda^2\mathbf{1} \tag{B.11}$$

$$= \left[\Sigma^{-1} - \frac{\sigma_\lambda^2\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1}}{1 + \sigma_\lambda^2\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right]\left(\sigma_\lambda^2\mathbf{1}\right) \tag{B.12}$$

$$= \sigma_\lambda^2\left[1 - \frac{\sigma_\lambda^2\mathbf{1}'\Sigma^{-1}\mathbf{1}}{1 + \sigma_\lambda^2\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right]\left(\Sigma^{-1}\mathbf{1}\right). \tag{B.13}$$

Clearly the OLS solution, (B.13), is proportional to the optimal filter given by (B.9), and hence the ratios are equal:

$$\frac{\tilde{\omega}_{k+1}^{ols}}{\tilde{\omega}_k^{ols}} = \frac{\tilde{\omega}_{j+1}}{\tilde{\omega}_j}. \tag{B.14}$$

We now detail the simulation framework and results.

---

[4]From $R_{m,t+k} = \phi^{k-1}\lambda_t + e_{t+j}$ and $\mathrm{E}_t(e_{t+k}) = 0$ we have $\sigma_r^2 = \phi^{2k-2}\sigma_\lambda^2 + \operatorname{var}(e_{t+k})$.

[5]Consider a positive shock to λ_{t+k}. Generally this leads to a negative contemporaneous return shock; $e_{t+k} < 0$. But conditional on the increase in λ_{t+k}, future returns will be, on average, higher than expected. Hence, the time t forecast errors of future returns are likely negatively autocorrelated.

[6]These results are related to the literature on optimal forecast combination (Bates and Granger, 1969; Granger and Ramanathan, 1984).

B.2. Simulation

The starting point for our simulation is the Campbell and Shiller (1988) log-linear approximate identity with all variables unconditionally demeaned,

$$r_{m,t+1} - \mathrm{E}_t\left[r_{m,t+1}\right] = N_{cf,t+1} - N_{dr,t+1}, \tag{B.15}$$

where $r_{m,t+1} = \log\left(1 + R_{m,t+1}\right)$, $N_{dr} = \left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\sum_{s\geq 1}\rho^s r_{m,t+1+s}$, and $N_{cf} = \left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\sum_{s\geq 0}\rho^s \Delta d_{t+1+s}$. This allow us to link realized returns with shocks to expected returns, but comes at the cost of having a system which is linear in logs, not levels of returns. In order to analytically compute expectations, we restrict to a homoskedastic environment. For computational tractability, we study monthly returns.[7] For a given parameter configuration, we simulate 10,000,000 months of data from the stationary distribution of the return system and then estimate the coefficients in a regression of λ_t on $\left\{\tilde{R}_{m,t+k}\right\}$. For illustrative purposes we set the maximum future horizon to $K = 24$ months.

Given the AR(1) dynamics for λ, we can calculate discount rate news, $N_{dr,t+1}$, analytically:

$$\begin{aligned} N_{dr,t+1} &= \left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\sum_{s\geq 1}\rho^s r^m_{t+1+s} \\ &= \frac{\lambda_{t+1}}{1-\rho\phi} - \frac{\phi\lambda_t}{1-\rho\phi} = \frac{\varepsilon_{t+1}}{1-\rho\phi}. \end{aligned} \tag{B.16}$$

Calibration We perform the following steps in our calibration:

1. We set the log-linearization constant, ρ, to $0.96^{1/12}$, as usual in the literature.[8] We set $\mathrm{std}_t\left(r_{m,t+1}\right) = 4.5\%$, which is approximately 16% annualized, matching the sample data.

2. We set the unconditional standard deviation of *annual* expected returns to 3%. Since annual returns are the sum of twelve monthly returns, annual predictability maps into monthly predictability using the summation formula for an AR(1) process

$$\mathrm{var}\left(\sum_{j=1}^{J}\lambda_{t+j}\right) = \frac{\sigma_\varepsilon^2}{\left(1-\phi^2\right)}\left[J + 2\frac{\left(J-1\right)\phi^{-1} - J + \phi^{J-1}}{\left(\phi^{-1}-1\right)^2}\right]. \tag{B.17}$$

   In this setting, $J = 12$. Hence, setting the standard deviation of annual expected returns to 3% (LHS) we get a link between ϕ and $\sigma_\varepsilon^2 \equiv \mathrm{var}\left(\varepsilon\right)$.

3. We let $\phi \in \{0.8,\ 0.9,\ 0.95\}$ monthly which correspond to half-lives of 3, 6, and 12 months respectively. For each ϕ we solve for σ_ε^2 using Eq. (B.17)

4. Given a choice of ϕ, we compute $\mathrm{var}\left(N_{dr,t+1}\right) = \frac{\sigma_\varepsilon^2}{(1-\rho\phi)^2}$

5. Notice that the conditional variance of returns is given by

$$\begin{aligned} \mathrm{var}\left(r_{t+1} - \mathrm{E}\left[r_{t+1}\right]\right) &= \mathrm{var}\left(N_{cf}\right) + \mathrm{var}\left(N_{dr}\right) \\ &\quad - 2\sqrt{\mathrm{var}\left(N_{cf}\right)\cdot\mathrm{var}\left(N_{dr}\right)}\cdot\mathrm{corr}\left(N_{cf},\ N_{dr}\right). \end{aligned} \tag{B.18}$$

[7] The simulation results below require an order of magnitude more time and memory for daily analysis.
[8] See for example, Campbell and Vuolteenaho (2004) and Cochrane (2008).

Hence, the only remaining free parameter is $\text{corr}(N_{cf}, N_{dr})$, since that choice pins down $\text{var}(N_{cf})$. In single shock models like Barberis et al. (2015); Campbell and Cochrane (1999), $\text{corr}(N_{cf}, N_{dr}) = -1$. In other multiple shock papers such as Bansal and Yaron (2004); Campbell and Kyle (1993), $\text{corr}(N_{cf}, N_{dr}) = 0$. Given this range in common models, we choose values of 0 and -0.4.

6. We simulate the iid sequence N_{cf} and ε from a multivariate Gaussian distribution with zero mean and the previously computed variances. Given the values of simulated shocks , we compute the news to discount rates from Eq. (B.16) and construct the λ_t process using $\lambda_{t+1} = \bar{\lambda} + \phi\left(\lambda_t - \bar{\lambda}\right) + \varepsilon_{t+1}$, where we set $\bar{\lambda}$ equal to the sample average, $\hat{\mathrm{E}}(r_m)$.[9]

7. Next, we compute shocks to market returns based on Eq. (B.15). Having computed both one-period-ahead expected returns, λ_t, and return shocks, we can construct the full sequence of market returns $r_{m,t}$.

8. Lastly, we estimate the regression of λ_t on $\{\tilde{r}_{m,t+k}\}$ to obtain the optimal filter.

Fig. A.1 shows the approximate analytic solution, equation (B.10) as a solid blue line and OLS estimates as black points; in all cases, the approximation is quite good. For each simulation, we compute $\hat{\lambda}_{t:t+24}$ using optimal weights given by OLS estimates and using Eq. (B.10). Across simulations, the minimum time-series correlation of $\hat{\lambda}$ constructed using the alternative weights is 99.7%, further confirmation that (B.10) provides an extremely good approximation to the optimal solution. Therefore, we impose Eq. (B.10) below when searching for the optimal horizon, K.

[9] We initialize λ_0 using a random draw from the stationary distribution.

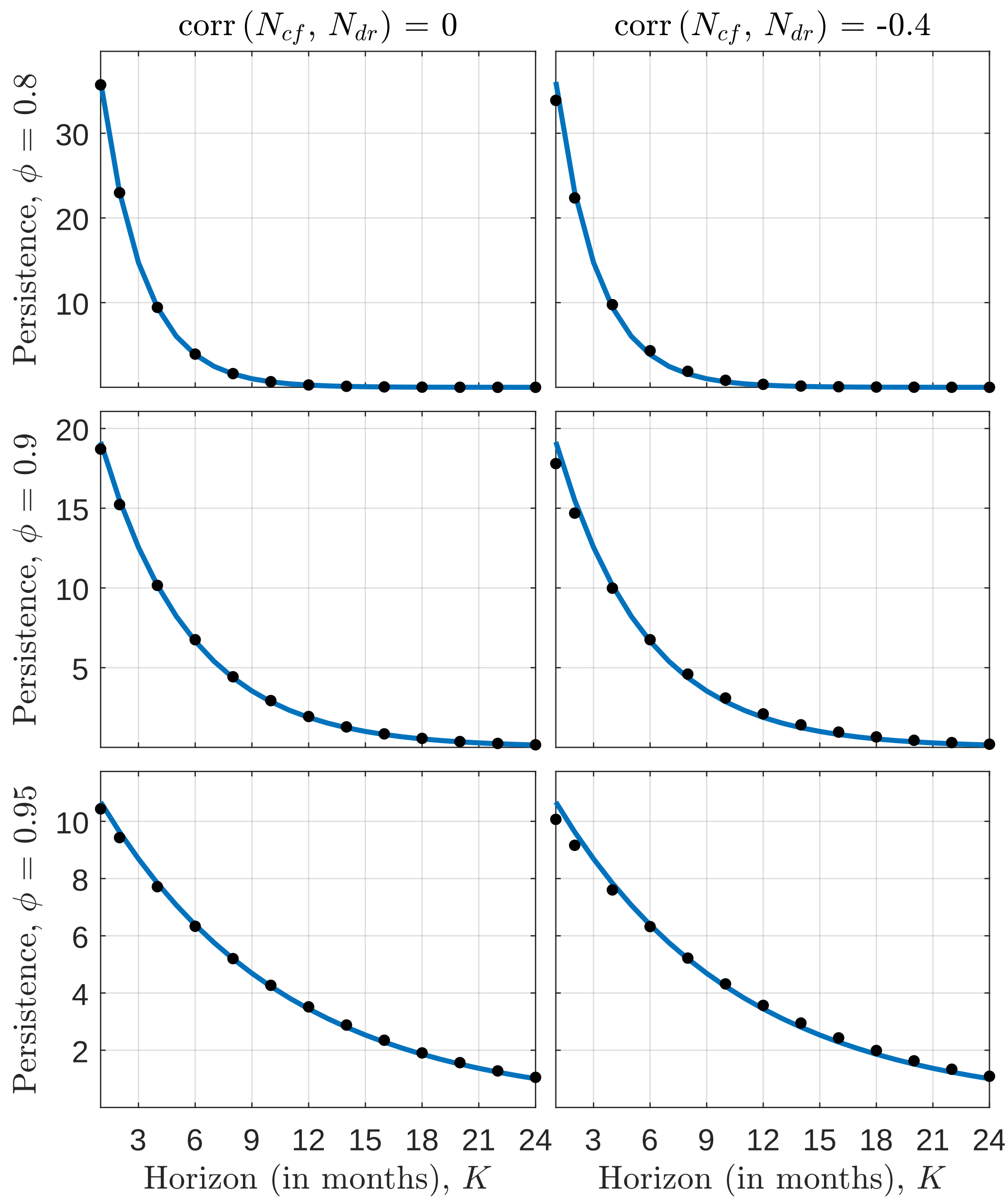

Fig. A.1. Optimal filter weights. The figure shows rescaled OLS coefficients from a regression of λ_t on $\left\{r^m_{t+k}, 1 \leq k \leq 24\right\}$ as black dots and the analytic solution, Eq. B.10, as a blue line for a variety of parameter choices. The columns correspond to corr $(N_{cf},\, N_{dr})$ equal to zero and -0.4. The rows correspond to ϕ (persistence of expected returns) equal to 0.8, 0.9 and 0.95. OLS estimates are from a simulation of 10,000,000 months.

B.2.1. Estimating covariances with finite samples

We now consider estimating $\operatorname{cov}(r_{i,t}, \lambda_t)$ using the proxy $\hat{\lambda}_{t:t+K}$, with finite data samples of length T. Based on the above analysis of optimal filters we restrict the ω_j coefficients in $\hat{\lambda}_{t:t+K}$ to satisfy (B.10). The single optimization variable now is the truncation horizon, K. For each choice of parameter values, we find the optimal K by simulation.

The optimal K depends not only on the properties of aggregate returns, but on the particular test asset return, $r_{i,t}$, and the dynamics of risk prices. Since the 2-factor model (Eq. (C.3)) produces similar risk price estimates as our main 3-factor model, we use that model to set expected returns in the simulation thereby reducing the number of simulation parameters. So that $r_{i,t}$ has properties resembling our test assets (long-short anomaly portfolios), we set it equal to the λ-mimicking portfolio plus noise. In this case we can analytically derive the true $\text{cov}\,(r_{i,t}, \lambda_t)$. Finally, we set the dynamics of risk prices to match the sample Sharpe ratio of the λ-mimicking portfolio.

For convenience, we recast the conditional pricing equation as

$$\text{E}_t\left[r_{i,t+1}\right] = \lambda_t \beta_{i,m} + \zeta_t \beta_{i,\nu}, \tag{B.19}$$

where the β are multivariate regression coefficients from a conditional regression of $r_{i,t+1}$ on $r_{m,t+1}$ and λ_{t+1}. We set ζ_t so that the Sharpe ratio of the λ-mimicking portfolio matches the Sharpe ratio we find in the data, approximately -0.25 monthly.

Using the law of total covariance and the fact that $r_{i,t}$ is the λ-mimicking portfolio (plus "unpriced" noise), we have the true covariance is $\text{cov}\,(r_{\varepsilon,t}, \lambda_t) = \left(1 + \frac{\phi \frac{\text{E}[\zeta_t]}{\text{E}[\lambda_t]}}{1-\phi^2}\right)\sigma_\varepsilon^2$. Given parameters, we compute the MSE of $\hat{\text{cov}}\left(r_{i,t}, \hat{\lambda}_{t:t+K}\right)$ across 1,000,000 simulations of 600 months (50 years) each. Fig. A.2 shows the MSE for various estimation methods Equal weight indicates that all scaled weights, $\tilde{\omega}_k$, are equal and optimal indicates that weights are determined by (B.10).

Fig. A.2 shows the log-MSE for various estimation methods based on monthly simulation, using the parameters $\phi = 0.9$ and $\text{corr}\,(N_{cf},\, N_{dr}) = 0$.[10] *Equal weight* indicates that all scaled weights, $\tilde{\omega}_k$, are equal, and *optimal* indicates that weights are determined by (B.10). *Non-overlapping* indicates that estimation uses $\lfloor T/K \rfloor$ "unique" observations, whereas overlapping uses $T - K$ observations. The plots are normalized by subtracting the log-MSE for $K = 1$, which is, by construction, the same across methods. All methods are nearly unbiased (not shown). It is well known that using overlapping data introduces a moving-average structure to the variable of interest, even if the underlying process is iid. Hence, $T - K$ overstates the number of "independent" observations. Our simulated MSE naturally accounts for this.

The first important observation from Fig. A.2 is that with non-overlapping data, the MSE increases regardless of the weighting method. Intuitively, this makes sense. Since the persistence, $\phi < 1$, further out observations have lower and lower signal-to-noise ratio. With uncorrelated errors, doubling the horizon K reduces the noise in $\hat{\lambda}_{t:t+K}$ by less than half. At the same time the effective sample size is halved, doubling the variance of the estimate.

For overlapping observations, the minimum MSE is indicated with a ⧫. The optimal K for both weighting schemes substantially reduces the MSE relative to using $K = 1$ month. Further, optimal weighting non-trivially reduces the estimator's MSE relative to equal weighting. Across a range of parameter choices, the improvement is around 10% (not shown). Most importantly, the objective function is remarkably flat for optimal weighting, but more sensitive for equal weighting. This is because optimal weighting places a majority of the total weight on small k (short-horizon), regardless of the maximum horizon K. Hence, optimal weighting reduces the minimum MSE and additionally is robust across sub-optimal choices of K.

[10] Results are similar for other parameter choices.

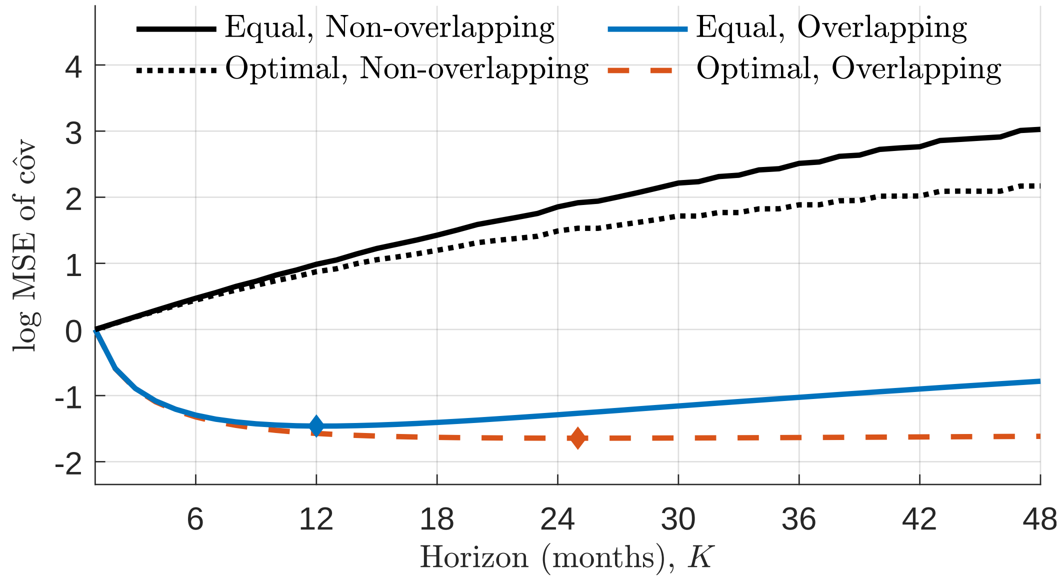

Fig. A.2. Small sample MSE. The figure shows the log mean squared error of various estimators of covariance with λ. All values are normalized relative to the MSE using $K = 1$ month.

B.3. Finite sample properties of estimated risk prices

Given the above assumptions, the standard two-stage (Fama-Macbeth) estimator generates consistent estimates of average risk prices. However, in small sample there may be a substantial errors-in-variables bias in the estimate of $\mathrm{E}\left(\delta_t^\lambda\right)$. Since $\widehat{\mathrm{cov}}\left(R_{i,t}, \hat{\lambda}_{t:t+K}\right)$ contains substantial sampling error, the estimate of $\mathrm{E}\left(\delta_t^\lambda\right)$ likely biased toward zero. We quantify this severity of this concern via simulation to to recover the small sample distribution of estimated risk prices. For comparison, we also estimate risk prices using the VAR methodology.

To simulate the finite sample distribution of estimated risk prices, we start with the above framework. We set $\phi = 0.96$ monthly which implies annual persistence equal 0.61, matching the autocorrelation of adjusted dp in Lettau and Van Nieuwerburgh (2007). To generate a cross section of test assets, we let $\beta_{i,m}$ vary from 0.5 to 1.5 in steps of 0.25 and we independently let $\beta_{i,\nu}$ vary symmetrically around zero to produce a spread in expected returns equal to twice the market premium to generate 25 portfolios which have substantial variation in both dimensions, but most of the cross-sectional R^2 coming from our factor (as in the data). Finally, we add iid noise to each test asset return with variance chosen so that time-series of a regressions of asset returns on the market produced a (cross-sectional) average R^2 of 85%, again similar to what we find in the data.

As a baseline scenario, we provide the econometrician with the state variable $z_t = \frac{1}{\theta}\left(\lambda_t - \overline{\lambda}\right)$, an arbitrarily rescaled version of demeaned expected returns. Hence, she must estimate θ from a regression of $r_{m,t+1}$ on z_t in order to recover expected returns.[11] Given the AR(1) dynamics for λ_t and availability of z_t, both our estimator and the VAR-based estimator of risk prices are consistent and asymptotically normal (with the same variance). Their small sample distributions are, however, quite different.

For the VAR method, we first estimate θ from the regression $r_{m,t+1} = a + \theta z_t + \varepsilon_{m,t+1}$. Then

[11] To focus on sampling uncertainty in $\mathrm{E}\left(\delta_t^\lambda\right)$ caused by estimation of θ, we endow the econometrician with knowledge of ϕ, the persistence of z.

Table A.1. Small sample properties

Simulated population moments for risk prices estimators using the methodology in Section 3.3 and the traditional VAR method. In the baseline scenario, the single state variable driving expected returns is observable. In the misspecified scenario, the observed variable is contaminated with iid noise having $\frac{1}{8}$ the variance of the true state variable.

	BIAS	SD	SKEW	KURT
Baseline:				
$\hat{\lambda}_{t:t+K}$	-0.33	0.56	1.51	4.29
VAR	0.13	1.90	-0.93	18.86
Misspecified:				
$\hat{\lambda}_{t:t+K}$	-0.35	0.58	1.52	4.03
VAR	0.32	3.12	-0.14	17.32

for each asset we compute $C_{i,\lambda} = \text{cov}\left(r_{i,t}, \hat{\theta}z_t - \phi\hat{\theta}z_{t-1}\right)$ and $C_{i,m} = \text{cov}\left(r_{i,t}, r_{m,t} - \hat{\theta}z_{t-1}\right)$. In the second stage, we estimate risk prices from a cross-sectional regression of sample mean returns on first-stage covariances. For our method, we construct $\hat{\lambda}_{t:t+K} = \sum_{k=1}^{K} \omega_k \left(R_{m,t+k}\right)$ using the optimal K, where the ω_k decay at rate ϕ. Then for each asset we compute $\text{cov}\left(r_{i,t}, \hat{\lambda}_{t:t+K}\right)$ and $C_{i,m} = \text{cov}\left(r_{i,t}, r_{m,t} - \hat{\theta}z_{t-1}\right)$. We then forecast $r_{i,t}$ and $\hat{\lambda}_{t:t+K}$ using z_{t-1} and compute the sample covariance of these forecasts. This is covE. Then we set $C_{i,\lambda} = \text{cov}\left(r_{i,t}, \hat{\lambda}_{t:t+K}\right) - \text{covE}$.[12] Finally, we estimate risk prices from a CS regression of sample mean returns on first-stage covariances. In this setup, our estimator is not only consistent for the sign of δ_λ, but the magnitude as well.

We explore the robustness of the two methods by considering the case where the econometrician observes only a noisy version of the true state variable, $\tilde{z}_t = z_t - \eta_t$. The simulations below use iid noise with variance equal to 1/8 the variance of z.[13] Note that noise in the observed state variable is equivalent to the case of an unobserved additional state variable.[14] In this case, both methods will deliver inconsistent estimates since the required assumptions are not satisfied. However, their small sample bias and error may not be equally affected by the measurement error in $\tilde{z}_t$.

Since the sample moments of estimated risk prices may not exist, we assess bias and dispersion by 1% trimmed mean and trimmed standard deviation, respectively. We measure bias as $\text{E}\left(\hat{\delta}_\lambda\right)/\delta_\lambda - 1$ and dispersion as $\text{std}\left(\hat{\delta}_\lambda\right)/\delta_\lambda$ so that both measures are scaled by the true value of δ_λ. Skewness and excess kurtosis are measured as usual on the trimmed sample.

The first two rows of Table A.1 present statistics for the baseline scenario where z is measured perfectly. Our estimator does indeed suffer from moderate downward bias (towards zero) whereas the VAR method exhibits only a slight upward bias. However, the standard deviation of the VAR estimator is greater than three times that of our estimator. Further, the VAR estimator exhibits substantial kurtosis. Both methods exhibit little skewness. Importantly, estimated risk prices of discount rate shocks from two methods are nearly uncorrelated across simulation iterations (not shown), suggesting that the two methods are complementary in small samples.

[12] The small sample properties of estimated risk prices are essentially unchanged if we set covE = 0, showing that our method can be quite robust to misspecification of the state vector.

[13] This results in approximately 90% correlation between z and $\tilde{z}$. We have also simulated the case where η has the same persistence as z (with orthogonal innovations) and obtain quantitatively similar results.

[14] The dynamics of $r_{m,t}$, $\tilde{z}_t$, and η_t admit a first-order VAR representation.

The next two rows show statistics for the misspecified scenario, where the econometrician is missing a state variable. The bias and dispersion of our estimator are essentially unaffected by the noise in z. The VAR estimator, now exhibits upward bias (away from zero) of the same magnitude as the downward bias in our estimator. Moreover, its standard deviation has increased to more than five times that of our estimator. Finally, bias and dispersion using our methodology are essentially unchanged if we set $\text{cov}(\text{E}) = 0$ (not shown), suggesting that our estimates are relatively insensitive to the choice of state variables.

Internet Appendix C. Robustness

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

C.1. Estimating covariance of expectations

In our main estimation, we set $\text{cov}(\text{E}) = \text{cov}\left(\text{E}_{t-1}[R_{i,t}], \text{E}_{t-1}\left[\hat{\lambda}_{t:t+K}\right]\right)$ to zero, which implies that expected conditional covariances equals unconditional covariances. Of course, this assumption may not be true and hence our risk price estimates may be (even asymptotically) biased. To address this potential concern, we rely on Assumption 3 which states that given the common (linear) predictors of $R_{i,t}$ and $\hat{\lambda}_{t:t+K}$, we can recover $\text{cov}(\text{E})$ as $\text{cov}\left(\text{E}[R_{i,t}| z_{t-1}], \text{E}\left[\hat{\lambda}_{t:t+K}\middle| z_{t-1}\right]\right)$.

We choose z as the VAR state variables from Bansal et al. (2014) which are the realized market variance, log price-dividend ratio, log dividend growth rate, term spread, default spread, and long-term real interest rate. Notice that Assumption 3 requires only that z is a sufficient subset of the predictors of $\hat{\lambda}_{t:t+K}$, whereas Bansal et al. (2014) assume z follows a VAR(1) and fully summarizes forecasts of $R_{m,t+k}$ at all horizons. To construct a daily time-series of z, we carry observed values forward for variables which are only available at lower than daily frequency.

Using z, we estimate the time-series of $\text{E}\left[\hat{\lambda}_{t:t+K}\middle| z_{t-1}\right]$ and $\text{E}[R_{i,t}| z_{t-1}]$ for each test asset then compute, $\text{cov}(\text{E})$, the unconditional covariance of these forecasts.[15] Given the unconditional covariance estimates from Section 4 we estimate $C_{i,\lambda}$ as

$$C_{i,\lambda} \equiv \text{E}\left(C_t^{i,\lambda}\right) = \text{cov}\left(R_{i,t}, \hat{\lambda}_{t:t+K}\right) - \text{cov}(\text{E}). \tag{C.1}$$

Finally we estimate the 3-factor model Eq. (11) using this new estimate of $C_{i,\lambda}$. Table A.2 shows results for the three sets of test assets. Estimated risk prices and model fit are quite similar to those in Sections 4.3 and 4.4, suggesting that our negative estimate of $\hat{\delta}_\lambda$ is not an artifact of ignoring $\text{cov}(\text{E})$ when estimating $C_{i,\lambda}$.

C.2. Future market return horizon and holding period

Table A.3 shows estimated δ_λ and cross-sectional R^2 using alternative horizons, K, ranging from six months to three years and alternative persistence, $\phi = \frac{\omega_{j+1}}{\omega_j}$, ranging from 0.8 to 0.97

[15] For assets that have positive conditional covariance with the market, $\text{cov}(\text{E})$ is likely biased upward in finite samples. For example, with $K = 1$ and using the market itself as a test asset, $\frac{\text{cov}(\text{E})}{\text{var}(R_m)}$ is the R^2 from a predictive regression of market returns using z. In-sample "overfitting" implies these statistics are upward biased.

Table A.2. Risk price estimates: covariance of expectations

This table shows risk prices estimates for the 3-factor model, Eq. 11, with cov (E) estimated as described in, Section C.1, using three sets of test assets. α is annualized (in %) and "-" indicates that the value is restricted to zero. MAPE is average absolute pricing error, annualized. The sample is daily from August 1963 to December 2015.

	α	δ^m	δ^f	δ^λ	R^2	$MAPE$
Quintiles	-	1.47	1.54	-9.07	0.943	0.516
	0.0737	1.45	1.52	-9.05	0.943	0.515
FF25	-	1.42	1.41	-8.87	0.667	1.28
	2.36	0.876	0.591	-7.71	0.695	1.35
Anomalies	-	1.52	1.88	-15.5	0.317	4.99
	0.051	1.5	1.87	-15.5	0.317	4.99

monthly when constructing $\widehat{\lambda} = \sum_{k=1}^{K} \omega_k R_{m,t+k}$.[16] All estimates restrict the zero-beta rate. Panel A uses daily returns, as in our primary estimation. The estimated δ_λ and cross-sectional R^2 are quite stable across horizon. The t-statistics on δ_λ are similar across horizon and choice of ϕ. Panel B shows the results of repeating the exercise using monthly returns. In an iid serially uncorrelated model, the SDF coefficients should be approximately identical across holding period. The point estimates and patterns are similar to the daily results, confirming that our negative estimate of δ_λ is not driven by artifacts of daily data.

[16]The values of ϕ we choose correspond to approximately 3m, 6m, 12m, and 24m half-life for discount-rate shocks in an AR(1) setting.

Table A.3. Alternative construction of $\hat{\lambda}$

Estimated risk price of discount rate factor and cross-sectional R^2 for alternative choices of persistence and horizon in computing, $\hat{\lambda}$, the weighted sum of future market returns. Test assets are value-weighted quintile portfolios sorted on ME, BE/ME, Prior2-12, and duration, as well as 1-7 year zero-coupon Treasury bonds. Panels A and B use daily and monthly excess returns, respectively. Stationary block bootstrap t-statistics are in parentheses.

		6m	12m	18m	24m	30m	36m
		Panel A: Daily					
0.80:	δ_λ	-5.6	-5.5	-5.4	-5.7	-5.5	-5.6
	t-stat	(-4.2)	(-4.0)	(-4.2)	(-4.2)	(-4.3)	(-4.4)
	R^2 (%)	96	96	96	96	96	95
0.90:	δ_λ	-7.3	-8.1	-8.2	-8.1	-8.1	-8.4
	t-stat	(-4.2)	(-4.4)	(-4.5)	(-4.3)	(-4.0)	(-4.3)
	R^2 (%)	96	97	96	96	96	95
0.95:	δ_λ	-8.6	-11	-12	-11	-12	-13
	t-stat	(-4.1)	(-4.2)	(-4.3)	(-4.4)	(-3.7)	(-3.9)
	R^2 (%)	97	97	96	95	94	94
0.97:	δ_λ	-9.2	-12	-15	-14	-16	-17
	t-stat	(-4.1)	(-4.2)	(-4.3)	(-3.9)	(-3.7)	(-4.1)
	R^2 (%)	97	97	95	95	93	92
		Panel B: Monthly					
0.80:	δ_λ	-7.7	-7.8	-7.5	-7.8	-7.4	-7.8
	t-stat	(-5.1)	(-5.3)	(-5.1)	(-5.4)	(-5.4)	(-5.3)
	R^2 (%)	85	85	85	85	83	83
0.90:	δ_λ	-8.6	-10	-10	-10	-9.7	-10
	t-stat	(-4.9)	(-5.0)	(-4.9)	(-5.3)	(-4.7)	(-5.0)
	R^2 (%)	85	85	84	84	83	82
0.95:	δ_λ	-9.4	-13	-15	-14	-14	-15
	t-stat	(-5.0)	(-5.3)	(-5.4)	(-5.3)	(-5.2)	(-5.1)
	R^2 (%)	85	84	84	84	82	82
0.97:	δ_λ	-9.8	-14	-18	-17	-18	-19
	t-stat	(-5.1)	(-5.4)	(-5.3)	(-5.1)	(-4.9)	(-4.4)
	R^2 (%)	85	84	84	84	82	81

C.3. Real interest rates

Our primary estimation uses changes in the nominal 3-month Treasury bill yield as a proxy for changes in the short rate. To address concerns that our factor is contaminated by changes in expected inflation, we compute expected inflation using the "adaptive expectations" method of Cieslak and Povala (2015) with a horizon of 100 months. Assuming the Fisher equation holds, we

Table A.4. Risk price estimates: real interest rate

This table shows risk prices estimates for the 3-factor model, Eq. 11, using three sets of test assets. Changes in real interest rates are computed using the Fisher equation and inflation expectations constructed according to the method in Cieslak and Povala (2015). α is annualized (in %) and "-" indicates that the value is restricted to zero. MAPE is average absolute pricing error, annualized. Stationary block bootstrap t-statistics are in parentheses. The sample is from August 1963 to December 2015.

	α	δ^m	δ^f	δ^λ	R^2	$MAPE$
Quintiles	-	2.4	1.49	-8.1	97.1	0.402
		(2.6)	(4.1)	(-4.1)		
	0.166	2.35	1.43	-8.07	97.1	0.403
	(0.5)	(2.6)	(4.0)	(-4.1)		
FF25	-	1.98	1.9	-11.7	79.8	1.05
		(2.2)	(3.9)	(-4.0)		
	1.55	1.55	1.34	-10.9	81	1.07
	(2.7)	(1.8)	(2.7)	(-3.8)		
Anomalies	-	2.6	2.17	-15.6	58	3.67
		(1.8)	(4.3)	(-6.8)		
	-0.602	2.79	2.38	-15.7	58.5	3.76
	(-1.4)	(1.9)	(4.2)	(-6.8)		

have that changes in the real rate can be computed as

$$\Delta r_{f,t} = \Delta i_t - \Delta \mathrm{E}_t\left(\pi\right), \tag{C.2}$$

where r_f is the real short-rate, i is the nominal short-rate, and π is inflation. Since CPI is only available monthly, we compute daily changes in expected inflation by first linearly interpolating expected inflation within a month, then computing daily differences. We re-estimate the pricing model (11) using changes in real rates constructed this way. The results, shown in Table A.4, are nearly identical to the previous estimates.

C.4. Two-Factor model

We present risk prices estimates from a two-factor model,

$$\mathrm{E}\left[R_{i,t+1}\right] = \delta_m C_{i,m} + \delta_\lambda C_{i,\lambda} \tag{C.3}$$

and use only stock portfolios (and the risk-free asset) as test assets. The results, shown in Table A.5, are nearly identical to the previous estimates, showing that identification of δ_λ does not depend on inclusion of the risk-free rate factor in the pricing model or bond returns in the test assets.

Table A.5. Risk price estimates: 2-factor model

This table shows risk prices estimates for the 2-factor model, equation Eq. C.3 using three sets of test assets. α is annualized (in %) and "-" indicates that the value is restricted to zero. MAPE is average absolute pricing error, annualized. Stationary block bootstrap t-statistics are in parentheses. The sample is from August 1963 to December 2015.

	α	δ^m	δ^λ	R^2	$MAPE$
Quintiles	-	2.93	-7.98	91	0.51
		(3.4)	(-4.2)		
	0.144	2.87	-7.95	91	0.516
	(0.3)	(3.4)	(-4.2)		
FF25	-	2.65	-10.7	61.5	1.2
		(2.8)	(-3.9)		
	1.75	1.94	-10.2	63.1	1.18
	(2.1)	(2.2)	(-3.8)		
Anomalies	-	3.42	-15.7	60.7	4.16
		(2.6)	(-6.7)		
	-0.619	3.68	-15.7	61.3	4.24
	(-1.3)	(2.7)	(-6.7)		

C.5. Covariance of factor exposures and risk prices

Our primary estimation strategy requires Assumption 1 that risk prices are orthogonal to test asset covariances with the factors, which allows us to condition down the pricing relation. This assumption may not hold, potentially biasing our estimates. To address this issue, we use the method of mimicking portfolios, which does not require Assumption 1. For the factors $\hat{\lambda}_{t:t+K}$ and $\Delta R_{f,t}$, we estimate factor mimicking portfolio weights from an OLS regression

$$f_t = a + \beta' R_t + \varepsilon_t,$$

where f_t is the factor realization and R_t is the vector of returns on the fifteen quintile portfolios and the seven zero-coupon Treasury bonds.[17] The factor mimicking portfolio return is then $\tilde{f}_t = \beta' R_t$. Since the market portfolio is traded, it is its own mimicking portfolio. We collect the three mimicking portfolio returns in the vector $\tilde{F}_t$. Substituting the mimicking portfolios into the SDF, Eq. (1), we have

$$M_{t+1} \approx R_{f,t}^{-1} - \boldsymbol{b}_t' \left(\tilde{F}_{t+1} - \mathrm{E}_t \left[\tilde{F}_{t+1} \right] \right), \tag{C.4}$$

where $\boldsymbol{b}_t' = \left[\begin{array}{ccc} \delta_t^m & \delta_t^\lambda & \delta_t^f \end{array} \right]$.

Since the factors are traded, $\boldsymbol{b}_t$ must equal $R_{f,t}^{-1} \Sigma_t^{-1} \mu_t$, where $\Sigma_t = \mathrm{cov}_t \left(\tilde{F}_{t+1} \right)$ and $\mu_t = \mathrm{E}_t \left[\tilde{F}_{t+1} \right]$. In what follows, we assume $R_{f,t}^{-1} \approx 1$ at a daily horizon so that $\boldsymbol{b}_t \approx \Sigma_t^{-1} \mu_t$. Suppose Σ_t is observable by an econometrician. We decompose $\tilde{F}_{t+1}$ as $\tilde{F}_{t+1} = \mu_t + \nu_{t+1}$ with $\mathrm{E}_t (\nu_{t+1}) = 0$. Then $\Sigma_t^{-1} \tilde{F}_{t+1} = \boldsymbol{b}_t + \Sigma_t^{-1} \nu_{t+1}$ with $\mathrm{E}_t \left(\Sigma_t^{-1} \nu_{t+1} \right) = 0$. Taking unconditional expectations we have

[17] We obtain identical results if we include the market portfolio return.

Table A.6. Risk price estimates: mimicking portfolios

This table shows risk prices estimates for the 3-factor model, Eq. C.4 using the mimicking portfolio method. The first row shows estimates which use the sample covariance matrix estimated over the past 63 trading days (1 quarter) returns as the conditional covariance matrix. The Risk Metrics method uses exponentially decaying weights (at a rate of 0.95 per day) as opposed to the equal weights in the sample covariance matrix. We impose the null hypothesis $\delta = 0$ when computing t-statistics, which do not account for uncertainty in the mimicking portfolio weights. The sample is from August 1963 to December 2015.

	δ^m	δ^f	δ^λ
Sample Covariance	4.31	1.33	-8.70
	(2.7)	(2.8)	(-4.3)
Risk Metrics	3.16	1.04	-9.69
	(1.8)	(1.9)	(-4.3)

$\mathrm{E}\left[\Sigma_t^{-1}\tilde{F}_{t+1}\right] = \mathrm{E}\left[\boldsymbol{b}_t\right]$. Therefore, if Σ_t is observable, the sample average $\hat{\mathrm{E}}\left[\Sigma_t^{-1}\tilde{F}_{t+1}\right]$ is an unbiased (and consistent) estimator of the unconditional average risk prices, $\mathrm{E}\left[\boldsymbol{b}_t\right]$.

In a continuous-time diffusion setting, Σ_t is observable. However, discrete time, microstructure noise and possible price jumps break this result. Hence, we must estimate Σ_t.[18] We do so in two ways. First, we use the sample covariance matrix computed using the past 63 trading days (one quarter). Second, we use the Risk Metrics exponentially weighted moving average method. The sample covariance matrix of a vector x can be written as

$$\frac{1}{N-1}\sum_{n=0}^{N-1}\left(x_{t-n}-\bar{x}\right)\left(x_{t-n}-\bar{x}\right)',$$

whereas the Risk Metrics estimator is

$$\sum_{n=0}^{N-1} w_n\left(x_{t-n}-\bar{x}\right)\left(x_{t-n}-\bar{x}\right)',$$

where $w_0 = 0.05$ and $w_n = 0.95w_{n-1}$. For large N, the weights sum to approximately unity. In practice, we use 252 trading days (1 year) to avoid losing too many observation at the beginning of the sample. Table A.6 shows estimated risk prices, $\hat{\mathrm{E}}\left[\Sigma_t^{-1}\tilde{F}_{t+1}\right]$, and t-statistics using the two methods of constructing Σ_t.[19] The two sets of estimates are very close and also quite similar to those in Table 2. This result confirms our negative estimate of $\mathrm{E}\left(\delta_t^\lambda\right)$ without relying on Assumption 1.

Internet Appendix D. Stylized models

We present simplified models of stochastic risk aversion and stochastic volatility to illustrate the key mechanisms.

[18] If the error in our estimate of Σ_t^{-1} is mean zero and uncorrelated with μ_t, the resulting estimator is still unbiased and consistent.

[19] The t-statistics do not account for uncertainty in the mimicking portfolio weights.

D.1. Time-varying risk aversion model

Consider the dynamic problem of a representative investor with Epstein-Zin preferences and unitary elasticity of intertemporal substitution ($\psi = 1$) and time varying coefficient of risk aversion (γ_t). Denote his wealth at time t as W_t, the vector of additional state variables at t as $X_t = \{\alpha_t\}$, and let $\alpha_t \equiv 1 - \gamma_t$, $\rho \equiv 1 - \frac{1}{\psi}$. The investor's value function $J(W_t, X_t)$ is given by

$$\begin{aligned} J(W_t, X_t) &= \max_{\{C,\Theta_t\}} \lim_{\rho \to 0} \left\{ (1-\delta) C_t^\rho + \delta \left(\mathrm{E}_t \left[J(W_{t+1}, X_{t+1})^{\alpha_t} \right] \right)^{\frac{\rho}{\alpha_t}} \right\}^{\frac{1}{\rho}} \\ &= \max_{\{C,\Theta_t\}} \left\{ C_t^{1-\delta} \left(\mathrm{E}_t \left[J(W_{t+1}, X_{t+1})^{\alpha_t} \right] \right)^{\frac{\delta}{\alpha_t}} \right\}, \end{aligned} \tag{D.1}$$

where C_t is her consumption at time t and θ_t is a vector of weights allocated to each asset in his portfolio. Let R_{t+1} denote a vector of returns on n assets available to an investor and let $r_t \equiv \ln(R_{t+1})$. His budget constraint is given by

$$W_{t+1} = (W_t - C_t)\, \theta_t' R_{t+1}. \tag{D.2}$$

We assume log consumption growth $\Delta c_{t+1} \equiv \ln\left(\frac{C_{t+1}}{C_t}\right)$ is iid, the risk aversion parameter follows an AR(1) processes, and their shocks are uncorrelated:

$$\Delta c_{t+1} = \mu_c + \sigma_c \varepsilon_{t+1}^c \tag{D.3}$$

$$\gamma_{t+1} = \mu_\gamma + \phi_\gamma \gamma_t + \sigma_\gamma \varepsilon_{t+1}^\gamma. \tag{D.4}$$

Solution Market clearing requires $C_t = D_t$ and $W_t = P_t$. Since preferences are homogeneous of degree one in wealth, we define

$$J(W_t, X_t) = \phi(X_t)\, W_t \equiv \phi_t W_t. \tag{D.5}$$

Taking the log of Eq. (D.1) and evaluating first-order conditions for consumption yields:

$$C_t = (1-\delta)\, W_t. \tag{D.6}$$

A unitary elasticity of substitution therefore implies that consumption is proportional to wealth, i.e. that agents possess a form of (rational) myopia in consumption and savings decisions. Optimal portfolio choice is fully dynamic, however, unless risk aversion is also unity (Giovannini and Weil, 1989).

Substituting Eq. (D.5) into Eq. (D.1) we have:

$$\phi_t = (1-\delta)^{1-\delta}\, \delta^\delta \left(\mathrm{E}_t \left[\phi_{t+1}^{\alpha_t} (R_{m,t+1})^{\alpha_t} \right] \right)^{\frac{\delta}{\alpha_t}} \tag{D.7}$$

which implies

$$\mathrm{E}_t \left[B_t \phi_{t+1}^{\alpha_t} (R_{m,t+1})^{\alpha_t} \right] = 1, \tag{D.8}$$

where $B_t = \left(\frac{(1-\delta)^{1-\delta} \delta^\delta}{\phi_t} \right)^{\frac{\alpha_t}{\delta}}$.

Taking the log of Eq. (D.7) and guessing that $\ln \phi_t = a_0 + a_1 \gamma_t$ we have:

$$\begin{aligned} a_0 + a_1 \gamma_t &= (1-\delta)\ln(1-\delta) + \delta\left(a_0 + a_1\left[\mu_\gamma + \phi_\gamma \gamma_t\right]\right) \\ &\quad + \delta \mu_c + \frac{1}{2}\delta(1-\gamma_t)\left(a_1^2 \sigma_\gamma^2 + \sigma_c^2\right), \end{aligned} \tag{D.9}$$

where we use the fact that $\mathrm{E}_t(r_{m,t+1}) = \mu_c - \ln \delta$ (follows from Eq. (D.2)). It then follows that

$$a_1 = -\frac{1}{2}\delta \frac{a_1^2 \sigma_\gamma^2 + \sigma_c^2}{1 - \delta \phi_\gamma} < 0. \tag{D.10}$$

The portfolio problem of the investor is given by

$$\max_{\theta_t} \mathrm{E}_t\left[\phi_{t+1}^{\alpha_t}\left(\theta_t' R_{t+1}\right)^{\alpha_t}\right] \tag{D.11}$$

subject to $\theta' 1_n = 1$. The relevant first-order conditions are:

$$\mathrm{E}_t\left[\phi_{t+1}^{\alpha_t}\left(R_{m,t+1}\right)^{\alpha_t - 1}\left(R_{i,t+1} - R_{1,t+1}\right)\right] = 0, \tag{D.12}$$

where $R_{1,t+1}$ denotes the return on any asset , for instance, the risk-free rate. Combining this equation with Eq. (D.8), we obtain the Euler equation for any asset:

$$\mathrm{E}_t\left[B_t \phi_{t+1}^{\alpha_t}\left(R_{m,t+1}\right)^{\alpha_t - 1} R_{i,t+1}\right] = 1. \tag{D.13}$$

Assuming returns are jointly log-normal and computing conditional expectations, it follows that the risk premium on any asset is given by:

$$\mathrm{E}_t R_{i,t+1}^e \approx \gamma_t \mathrm{cov}_t\left(R_{i,t+1}^e, R_{m,t+1}^e\right) + a_1(\gamma_t - 1)\,\mathrm{cov}_t\left(R_{i,t+1}^e, \gamma_{t+1}\right), \tag{D.14}$$

where $\mathrm{E}_t\left(R_{i,t+1}^e\right)$ is the risk premium of asset i (in excess of the riskless rate). Specializing this to the market return we obtain the expression for the market risk premium:

$$\mathrm{E}_t\left(R_{m,t+1}^e\right) \approx \gamma_t \sigma_m^2. \tag{D.15}$$

Substituting this into Eq. (D.14) we have that

$$\mathrm{E}_t R_{i,t+1}^e \approx \gamma_t \mathrm{cov}_t\left(R_{i,t+1}^e, R_{m,t+1}^e\right) + \frac{a_1}{\sigma_m^2}(\gamma_t - 1)\,\mathrm{cov}_t\left(R_{i,t+1}^e, \lambda_{t+1}\right), \tag{D.16}$$

meaning the price of discount-rate risk, δ_t^λ, is negative if and only if the coefficient of relative risk aversion γ_t is above 1.

D.2. Stochastic volatility model

We consider the same setup as in the previous section except with constant risk aversion and stochastic volatility as in Bansal and Yaron (2004):

$$\sigma_{c,t+1}^2 = \mu_\sigma + \phi_\sigma \sigma_{c,t}^2 + \sigma_\sigma \varepsilon_{t+1}^\sigma, \tag{D.17}$$

with uncorrelated shocks to consumption and volatility. Taking logs of the Bellman equation and guessing that $\ln \phi_t = a_0 + a_1 \sigma_{c,t}^2$ we have:

$$\begin{aligned} a_0 + a_1 \sigma_{c,t}^2 &= (1-\delta)\ln(1-\delta) + \delta\left(a_0 + a_1\left[\mu_\sigma + \phi_\sigma \sigma_{c,t}^2\right]\right) \\ &\quad + \delta\mu_c + \frac{1}{2}\delta(1-\gamma)\left(a_1^2\sigma_\sigma^2 + \sigma_{c,t}^2\right), \end{aligned} \tag{D.18}$$

where we use the fact that $\mathrm{E}_t(r_{m,t+1}) = \mu_c - \ln\delta$ (follows from Eq. (D.2)). It follows that

$$a_1 = -\frac{1}{2}\frac{\delta(\gamma-1)}{1-\delta\phi_\sigma} < 0. \tag{D.19}$$

As before, it follows that the risk premium on any asset is given by:

$$\mathrm{E}_t R_{i,t+1}^e \approx \gamma \mathrm{cov}_t\left(R_{i,t+1}^e, R_{m,t+1}^e\right) + a_1(\gamma-1)\mathrm{cov}_t\left(R_{i,t+1}^e, \sigma_{c,t+1}^2\right). \tag{D.20}$$

Specializing this to the market return we obtain the expression for the market risk premium

$$\mathrm{E}_t\left(R_{m,t+1}^e\right) \approx \gamma\sigma_{c,t}^2. \tag{D.21}$$

Substituting this into Eq. (D.20) we have that

$$\mathrm{E}_t R_{i,t+1}^e \approx \gamma \mathrm{cov}_t\left(R_{i,t+1}^e, R_{m,t+1}^e\right) + a_1\frac{(\gamma-1)}{\gamma}\mathrm{cov}_t\left(R_{i,t+1}^e, \lambda_{t+1}\right). \tag{D.22}$$

Again, the price of discount-rate risk, δ_t^λ, is negative if and only if the coefficient of relative risk aversion γ is above 1. Bansal, Kiku, Shaliastovich, and Yaron (2014) and Campbell, Giglio, Polk, and Turley (2018) consider a heteroskedastic version of Campbell (1996). Their models predict that future volatility news, N_V, is proportional to news to discount rates, N_{DR}. When this restriction is imposed, it leads to the negative price of discount rate risk in a model that combines the two terms.

References

Bansal, R., Kiku, D., Shaliastovich, I., Yaron, A., 2014. Volatility, the macroeconomy, and asset prices. Journal of Finance 69, 2471–2511.

Bansal, R., Yaron, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Barberis, N., Greenwood, R., Jin, L., Shleifer, A., 2015. X-CAPM: an extrapolative capital asset pricing model. Journal of Financial Economics 115, 1–24.

Bates, J. M., Granger, C. W., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

Campbell, J., Shiller, R., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J. Y., 1996. Understanding risk and return. Journal of Political Economy 104, 298–345.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Campbell, J. Y., Giglio, S., Polk, C., Turley, R., 2018. An intertemporal CAPM with stochastic volatility. Journal of Financial Economics 128, 207–233.

Campbell, J. Y., Kyle, A. S., 1993. Smart money, noise trading, and stock price behaviour. Review of Economic Studies 60, 1–34.

Campbell, J. Y., Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review 94, 1249–1275.

Cieslak, A., Povala, P., 2015. Expected returns in Treasury bonds. Review of Financial Studies 24, 2859–2901.

Cochrane, J. H., 2008. The dog that did not bark: a defense of return predictability. Review of Financial Studies 21, 1533–1575.

Giovannini, A., Weil, P., 1989. Risk aversion and intertemporal substitution in the capital asset pricing model. Working paper 2824, National Bureau of Economic Research, Cambridge, MA .

Granger, C. W., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3, 197–204.

Lettau, M., Van Nieuwerburgh, S., 2007. Reconciling the return predictability evidence. Review of Financial Studies 21, 1607–1652.