Linear Approximations and Tests of Conditional Pricing Models*

Michael W. Brandt¹ and David A. Chapman²

¹Duke University and ²University of Virginia

Abstract

If a nonlinear risk premium in a conditional asset pricing model is approximated with a linear function, as is commonly done in empirical research, the fitted model is misspecified. We use a generic reduced-form model economy with moderate risk premium nonlinearity to examine the size of the resulting misspecification-induced pricing errors. Pricing errors from moderate nonlinearity can be large, and a version of a test for nonlinearity based on risk premiums rather than pricing errors has reasonable power properties after properly controlling for the size of the test. We conclude by examining the importance of moderate nonlinearity in the context of the investment-specific technology shock models of Papanikolaou (2011) and Kogan and Papanikolaou (2014).

JEL classification: C58, G12

Keywords: Conditional asset pricing model, Nonlinear risk premiums, Testing

Received May 12, 2014; accepted August 10, 2017 by Editor Alex Edmans.

1. Introduction

The empirical evidence against unconditional versions of the classic capital asset pricing model (CAPM) and the consumption CAPM (CCAPM), gathered over the past three decades, has inspired a large literature that examines conditional versions of these models; examples of this line of research include Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Lustig and van Nieuwerburgh (2005), Santos and Veronesi (2006), and Yogo (2006). In many conditional asset pricing models, loadings of asset returns on common risk factors and the associated factor risk premiums vary over time as a function of

* We would first like to thank Christian Lundblad (the referee), and Alex Edmans and Burton Hollifield (the editors) for many helpful comments. We have also received helpful input from Keith Brown, Jennifer Conrad, Joost Driessen, Jin-Chuan Duan, Wayne Ferson, Xavier Gabaix, Lorenzo Garlappi, Eric Ghysels, Larry Harris, Kevin Huang, Tim Johnson, Frank de Jong, Mark Kamstra, Raymond Kan, Ed Kane, Jon Lewellen, Andy Lo, Ludovic Phalippou, Jacob Sagi, Robert Stambaugh, and seminar participants at various universities and symposia.

© The Author(s) 2018. Published by Oxford University Press on behalf of the European Finance Association. All rights reserved. For Permissions, please email: journals.permissions@oup.com
observable state variables. We examine how inference about conditional asset pricing models is affected by misspecification of the functional form of time-varying risk premiums.

Conditional models can differ from their unconditional counterparts through time varying factor loadings and/or time-varying risk premiums. While the relative importance of these alternate channels depends on the choice of factors and on the functional form of conditioning in the model, Ferson and Harvey (1991) argue that, in a standard linear macro-factor model, time-variation in risk premiums is more important than time-variation in the factor loadings for explaining observed predictability in excess returns. Lewellen and Nagel (2006) conclude, using a methodology based on high-frequency estimation, that time-varying loadings cannot be a source of the empirical success of a conditional form of the CAPM relative to the unconditional CAPM in a cross-section of expected returns.

In a conditional asset pricing model, the form of any time-variation depends on the joint evolution of the underlying data generating process and the economic agents’ information sets. These aspects of the theoretical model are unspecified in a reduced-form theory, and they are generally unobservable; see the discussion in Hansen and Richard (1987). The econometrician must, therefore, model the time-variation in a way that is tractable. In the case of time-varying factor loadings, covariances are generally specified as functions of past returns and factor innovations, following the ARCH/GARCH modeling framework, or more simply as linear or exponentially linear functions of state-variables that are observable by the econometrician. In the case of time-varying risk premiums, the majority, if not all applications involve linear specifications in observable state-variables. This restriction seems, a priori, less credible given recent structural models that allow for technological innovation and financial sector feedback to the real economy that are inherently nonlinear. As an example, below, we consider a reduced-form model motivated by the embodied technical change model of Papanikolaou (2011). As an additional example, Roussanov (2014) uses nonparametric estimation techniques to demonstrate the empirical significance of non-linear risk premiums in a conditional consumption-based model that includes status effects.1

How sensitive are standard econometric tests of conditional asset pricing models to the assumption that time-variation in loadings and risk premiums are linear functions of observable state variables? Of course, this question can only be answered within the context of a specific assumption about the form of the true underlying risk premiums. The question is partially addressed by Ghysels (1998), who shows that when the factor loadings of a conditional pricing model exhibit structural breaks, testing the model with linear specifications can lead to statistically less reliable inference than testing an unconditional version of the model. This is true despite the fact that the data are generated by a conditional pricing model. It is reasonable to consider time varying loadings first, since variances and covariances are estimated with much greater precision than expected returns. There has, however, been no comparable analysis examining the sensitivity of statistical inferences to the linearity assumption for conditional risk premiums. Our paper fills this gap in the literature.

1 There is an older literature that considers nonlinear risk premiums in both consumption- and non-consumption-based models; see, for example, Bansal, Hsieh, and Viswanathan (1993) and Chapman (1997).
We show that, under the assumption that risk premiums can be well-approximated by a polynomial function of the underlying state variables, misspecification of the risk premiums can be reformulated as a missing-variables problem. This reformulation yields analytical expressions for the misspecification-induced mispricing as well as test statistics that can be used to detect misspecification. We then examine a reduced-form conditional asset pricing model that is an extension of the intertemporal CAPM specified in Brennan, Wang, and Xia (2004). It exhibits what we term “moderate nonlinearity” in the risk premium, as measured by the curvature of the risk premium function normalized by its first derivative function. The pricing kernel of this model depends on innovations to the market portfolio, innovations to the short rate, and innovations to the Sharpe ratio of the tangency portfolio. We view it as a representative of a large class of models that imply a time-varying capital market line.

We calibrate the model to size- and book-to-market (B/M)-sorted portfolio returns, a short-term Treasury yield, and estimates of the Sharpe ratio of the S&P 500 index, and we examine the analytical and simulated pricing errors caused by misspecifying the risk premiums as linear functions of the state variables as well as the finite-sample properties of alternative tests for nonlinear risk premiums. Our simulations show that misspecification-induced pricing errors can be large both statistically and relative to average excess returns. Given our use of the stochastic discount factor (SDF) representation, there are standard tests for the importance of omitted nonlinear terms. One test is based directly on the magnitude of the pricing errors. The other test examines the estimated risk premiums on the higher-order terms directly. We demonstrate that it is important to correct the size of the nonlinearity tests in realistic samples. In our tests, the nonlinearity test based on higher-order risk premium terms rather than on pricing errors has better finite-sample power to detect moderate nonlinearity.

As a final step in our analysis, we examine the model of embodied technological change in Papanikolaou (2011) and Kogan and Papanikolaou (2014) as an example of a structural model with nonlinearity in the risk premium function. When the model is estimated using a long sample of quarterly data on twenty-five size- and B/M-sorted portfolios from 1948 to 2016 using the standard linear approximation, there is limited evidence that the model can explain the cross-section of expected returns. However, adding the nonlinear conditional premiums terms that are a direct implication of the model results in a noticeable improvement in the cross-sectional explanatory power of the model, and it provides evidence in favor of the importance of embodied technological change in explaining measured return differences. At the same time, our results highlight the importance of using bootstrap estimates of the density function for the estimated risk premium in constructing hypothesis tests, as opposed to the asymptotic distribution.

2. Conditional Asset Pricing Models

2.1 Pricing Models in the SDF Form

Standard practice for estimating and testing conditional pricing models has evolved over time in a series of papers that include Campbell (1987), Gibbons and Ferson (1985), Harvey (1989), Shanken (1990), and Cochrane (1996), culminating in the textbook treatment of Cochrane (2005). In order to confront the standard testing methodology with a generic pricing model, we consider a class of economies with complete financial markets
and no arbitrage opportunities. The general form of the pricing kernel under the physical measure is

\[ M_{t+1} = \exp \left( r_f^t - \frac{1}{2} \mathbf{A}_t' \mathbf{A}_t + \mathbf{A}_t' \mathbf{\epsilon}_{t+1} \right), \]  

(1)

where \( r_f^t \) is the one-period risk-free return, \( \mathbf{A}_t \) is a \( K \)-vector of pricing kernel functions, and \( \mathbf{\epsilon}_{t+1} \sim \mathcal{N}(0, \mathbf{I}_K) \) is a vector of normalized common factor innovations. The time varying nature of \( \mathbf{A}_t \) makes this a conditional pricing model.

The exponential-affine structure of Equation (1) implies that the log pricing kernel is linear in the factors. We can rewrite Equation (1) as

\[ m_{t+1} \equiv \ln M_{t+1} = r_f^t - \frac{1}{2} \mathbf{A}_t' \mathbf{A}_t + \mathbf{A}_t' \mathbf{\epsilon}_{t+1}. \]  

(2)

Linearity in the normalized factor innovations has a long history in financial economics, as either a direct assumption about the pricing kernel, see Chen, Roll, and Ross (1986), as an implication of the CAPM of Sharpe, Lintner, and Mossin, or as a first-order approximation to a general consumption-based pricing model; see Breeden, Gibbons, and Litzenberger (1989) or Lettau and Ludvigson (2001). The form of Equation (1) is also common in the literature on dynamic term structure models; for example, Dai, Singleton, and Yang (2007).

The multifactor pricing kernel in Equation (2) is equivalent, in a pricing sense, to a kernel that is linear in the return to the (conditionally) mean variance efficient portfolio formed by projecting Equation (2) onto the space of marketed asset payoffs; see Hansen and Richard (1987). This (conditionally) efficient portfolio defines a unique (if the kernel is unique) maximum Sharpe ratio process, denoted \( S_t \). In the case of the kernel in Equation (1), this implies

\[ S_t = \frac{\sigma_i(M_{t+1})}{E_i(M_{t+1})} = \sqrt{\exp(\mathbf{A}_t' \mathbf{A}_t)} - 1, \]  

(3)

where the first equality follows from the definition of the Sharpe ratio and the second equality follows from the (conditional) log-normality of the pricing kernel.\(^2\)

At this level of generality, the pricing kernel innovations have no specific economic interpretation. They could represent innovations to aggregate consumption or a function of aggregate consumption, as in a habit-formation model, innovations to an aggregate market portfolio, or both. The values in the vector \( \mathbf{\epsilon}_{t+1} \) can even be innovations to the risk premium functions themselves. The normality assumption for the shocks is an important simplification, but the unconditional distribution of the pricing kernel can still be fat-tailed and/or skewed depending on the unconditional distribution of \( \mathbf{A}_t \). The pricing kernel coefficients, \( \mathbf{A}_t \), can be interpreted as risk premiums because \( \mathbf{\epsilon}_{t+1} \) are normalized to be uncorrelated and have unit variance. We will refer to \( \mathbf{A}_t \) as risk premium functions throughout the remainder of the paper.

The standard assumption used in the empirical literature is that the elements of \( \mathbf{A}_t \) are linear functions of information available to the market at time \( t \).\(^3\) This may be a reasonable assumption, but it is not easily justified for different equilibrium models, since these models

\(^2\) The proof of Equation (3) is straightforward. The first equality simply manipulates the definition of a SDF and a conditional correlation coefficient. The second equality follows immediately from the conditional log-normality of the pricing kernel in Equation (2).

\(^3\) An early example of this assumption is Shanken (1990); see also Cochrane (2005) for a textbook discussion.
are frequently silent about the precise functional form of conditional risk premiums. Indeed, Bansal, Hsieh, and Viswanathan (1993) and Chapman (1997) provide examples of both returns and consumption-based models with nonlinearities in the risk versus return trade-off. It seems reasonable that significant nonlinearities in $K_t$ would be important, but easily detected, specification errors. For example, there is a large econometric literature on detecting regime shifts and (large) structural breaks in model parameters. However, what about moderate levels of nonlinearity? Can moderate nonlinearity be detected, in practice, given noisy realized returns? Does moderate nonlinearity have an economically significant effect on measured pricing errors?

2.2 Defining “Moderate Nonlinearity”

We have, up to this point, used the term “moderate nonlinearity” without definition. Given the modifier, its definition is inherently subjective.\textsuperscript{4} Intuitively, we are interested in studying nonlinearity that is not “obviously large”—as in the case of regime shifts or structural breaks. If large nonlinearities are deemed to be important features of the economy under study, then they would clearly be modeled directly and estimated with the appropriate techniques.

In evaluating the effect of misspecification, we need to place additional restrictions on the form of the general SDF defined in Equation (1) in order to make it empirically testable. As in the existing literature, for example, Cochrane (2005) and the sources cited there, we begin by positing the existence of a set of (possibly $K$-dimensional) observable variables, $\{Z_t\}_{t=0}^{\infty}$, that adequately summarize the economically significant conditioning information in the model economy.

First, consider the simple case of a single risk premium driven by a single conditioning variable; that is, $\Lambda_t = \Lambda(Z_t)$, where $\Lambda : \mathbb{R} \rightarrow \mathbb{R} \in \mathbb{C}$. We can define a curvature index, as is common in the literature on risk preferences in a utility function context, as

$$c(Z) = \frac{d^2 \Lambda}{dZ^2},$$

(4)

where (unlike in the literature on risk aversion) we do not define $\gamma(Z)$ using the negative of the ratio because there is no compelling \textit{ex ante} reason (outside of a given structural model) to consider only concave risk premium functions.

What constitutes a “moderate” range for the function $\gamma(Z)$? This is where the inherent subjectivity mentioned earlier becomes a binding issue.\textsuperscript{5} The answer is determined by the nonlinearity of the risk premium function. Clearly, $\gamma(Z) = 0$ for a $\Lambda(Z_t)$ linear. For a one-parameter exponential function, $\Lambda = \exp(aZ)$, $a \in \mathbb{R}$, $\gamma(Z) = a$. “Moderate nonlinearity” corresponds to an \textit{ex ante} restriction on the parameters of the $\Lambda(Z_t)$ such that $|\gamma(Z)| \leq c$, for some constant $c$ over some finite range of the distribution of $Z$. For example, if the risk

\textsuperscript{4} Perhaps the better part of valor would be to simply paraphrase Justice Potter Stewart in Jacobellis v. Ohio and say: “I shall not today attempt to further define … (moderate nonlinearity) … But I know it when I see it.” However, we will attempt a definition in the rest of this sub-section.

\textsuperscript{5} Even in the analogous literature on the aggregate coefficient of relative risk aversion, where there is a much clearer economic interpretation of magnitudes, there is considerable disagreement as to what constitutes a “reasonable” parameter value.
premium function is a quadratic function, then $\Lambda(Z) = \lambda_0 + \lambda_1 Z + \lambda_2 Z^2$, and if $Z$ is bounded over $[-2, 2]$, then

$$\gamma(Z) \in \left[ \frac{2\lambda_2}{\lambda_1 - 4\lambda_2}, \frac{2\lambda_2}{\lambda_1 + 4\lambda_2} \right],$$

and the moderate nonlinearity restriction requires that $\max(|\frac{2\lambda_2}{\lambda_1 - 4\lambda_2}|, |\frac{2\lambda_2}{\lambda_1 + 4\lambda_2}|) \leq c$.

If the risk premium is a function of a $K$-vector of conditioning variables, $Z$, then Equation (4) generalizes to

$$\gamma(Z) = \nabla_Z^2 L \nabla Z L,$$

where $\nabla_Z^2 L$ is the $K \times K$ Jacobian matrix of the risk premium function, with typical element $\partial^2 L/\partial Z_i \partial Z_j$ and $\nabla Z L$ is the $K$-vector gradient of the risk premium function. Equation (6) reflects the fact that curvature is now a $K$-dimensional construct. Any ex ante restriction that defines “moderate” nonlinearity must account for multiple dimensions as well. For example, we could define moderate nonlinearity as requiring that $|\max(\gamma(Z))| \leq c$, for some constant $c$ over some finite range of the distribution of $Z$.

### 2.3 Alphas Generated by Misspecification

As a simplification, we assume: (1) There is a single risk premium in the pricing kernel.\(^7\) (2) This single risk premium can be well-approximated by a set of orthonormal polynomials

$$\Lambda(Z_t) = \lambda_0 + \lambda_1 Z_t + \lambda_2 P_2(Z_t) + \lambda_3 P_3(Z_t) + \cdots,$$

where $P_i(Z_t)$ is the component of $Z_t$ that is orthogonal to $Z_t$ and $P_i(Z_t)$, for $i < j$, and normalized to have zero mean and unit variance.\(^8\) In practice, Equation (7) must be truncated at some maximum polynomial order. We let $P^j_t$ denote the vector of orthonormal polynomials from order zero to $j$ based on $Z_t$. So, $\Lambda(Z_t) = \lambda' P^j_t$, where $\lambda' \equiv (\lambda_0, \lambda_1, \ldots, \lambda_j)'$.

We further assume, consistent with standard practice, a linear structure for the natural logarithm of the asset pricing kernel in Equation (2):

$$m_{t+1} = 1 - \frac{1}{2} \Lambda_t^2 + \Lambda_t \epsilon_{t+1},$$

where we are using the single risk premium assumption, and we normalize the constant coefficient to equal 1, since we will follow the common practice of the empirical literature and estimate the model using excess returns. Again, the $K$-vector of innovations, $\epsilon_{t+1}$, drives the pricing of risk in the model.

Given that $\Lambda(Z_t)$ is itself a polynomial of order $J$, then $\Lambda_t^2$ is a polynomial of order $2J$ and there are additional higher-order terms in (the orthogonalized components of)

---

6 The extension to the case of $L$ risk premiums that depend on a $K$-vector of conditioning variables is analogous to Equation (6), only of even higher dimension.

7 The assumption of one versus multiple risk premiums is made purely for notational convenience. Our argument extends naturally to multiple risk premiums. The polynomial expansion assumption is critical for our omitted variables interpretation. However, there is a large literature in numerical analysis on the use of polynomial approximations to general functions; see Chapter 6 in Judd (1998).

8 See Chapman (1997) for another example of the use of orthonormal polynomial approximations in an asset pricing setting.
Z from the factor innovation term in Equation (8). The linearity imposed by the standard estimation of conditional pricing models of the form of Equation (2) implies the omission of the higher-order terms $P_j(Z_t)$ for $j \geq 2$ in both the $\Lambda^2_t$ and $\Lambda_t \epsilon_{t+1}$ terms of Equation (8).

Given the simple linear structure of the pricing kernel, it is possible to compute a closed-form expression for the mispricing, or alphas, generated by misspecification. These alphas are based on the true moments of the factors and excess returns. Let

$$\mathbf{F}_{t+1} = (\mathbf{F}_{1,t+1}, \mathbf{F}_{2,t+1})',$$

where $\mathbf{F}_1$ is the $J_1$-vector of linear terms in Equation (8), and $\mathbf{F}_2$ is the $J_2$-vector of higher-order terms omitted by the linear approximation, where $J = J_1 + J_2$. The covariance matrix of the factors can be partitioned accordingly. It is important to note that—by construction of the higher-order approximation—the omitted variables are orthogonal to the included factors, which significantly simplifies the analysis.

In the true model, expected excess returns are exactly linear in the betas with respect to $\mathbf{F}$, and true pricing errors are zero. If the market prices of risk are partitioned conformably, then

$$E(\mathbf{R}^\epsilon) = \begin{bmatrix} \lambda'_{1} & \lambda'_{2} \end{bmatrix} \mathbf{B},$$

where

$$\mathbf{B} = \Sigma_{\mathbf{F} \epsilon}^{-1} \Sigma_{\mathbf{FR} \epsilon}$$

is the $J \times N$ matrix of multivariate regression coefficients. In Equation (10), expected returns are expressed exactly as a function of the factors that are included in the conventional approximation, the factors that are excluded from the conventional approximation, and the covariance of these two components. Since $\Sigma_{12} = \Sigma_{21} = 0$, by construction, we can rewrite Equation (10) as

$$E(\mathbf{R}^\epsilon) = \Sigma'_{\mathbf{F} \epsilon, \mathbf{R} \epsilon} \Sigma^{-1}_{11} \lambda_1 + \Sigma'_{\mathbf{F} \epsilon, \mathbf{R} \epsilon} \Sigma^{-1}_{22} \lambda_2,$$

where $\Sigma^{-1}_{11}$ is the inverse of the partition of the covariance matrix associated with the included factors and $\Sigma^{-1}_{22}$ is the inverse of the partition of the covariance matrix associated with the excluded factors. It follows immediately that the pricing errors, alphas, induced by ignoring nonlinearity in the risk premium function are

$$\alpha = \Sigma'_{\mathbf{F} \epsilon, \mathbf{R} \epsilon} \Sigma^{-1}_{22} \lambda_2.$$
of the pricing kernel function. We will examine this issue in the specific context of the example in Section 3.

2.4 Testing for Nonlinearity

The first test for the statistical significance of nonlinearity in risk premiums is based on the estimates of alphas constructed from the unrestricted model (the specification that includes higher-order terms) and the restricted (linear) model:

\[ T(\hat{\alpha}_R - \hat{\alpha}_U)\Sigma_{\alpha}^{-1}(\hat{\alpha}_R - \hat{\alpha}_U) \sim \chi^2_n, \tag{14} \]

where \( \hat{\alpha}_U \) is the N-vector of alpha estimates under the unrestricted model, \( \hat{\alpha}_R \) is the N-vector of alphas under the restricted model, and \( \Sigma_{\alpha} \) is the estimated covariance matrix of alphas under the restricted model (see Appendix B). This statistic is analogous to the generalized method of moment (GMM)-type test statistic based on differences in the value of the objective function under the two models, and we will refer to it below as the “a-test.” If the higher-order risk premium terms are important for pricing the test assets, then their inclusion should result in a large reduction in estimated alphas, and the test statistic in Equation (14) should be large. It is important to compare the distance between the alpha vectors using a common covariance matrix in order to ensure that the test statistic is non-negative.

The second test statistic is a simple quadratic form in the estimated risk premiums:

\[ \lambda'_2 \Sigma_{\lambda_2}^{-1}\lambda_2 \sim \chi^2_{J_2}, \tag{15} \]

where \( \lambda_2 \) is the \( J_2 \)-vector of estimates of higher-order risk premiums and \( \Sigma_{\lambda_2} \) is the estimated covariance matrix of these risk premiums. Estimates of the higher-order risk premium terms that are large relative to their estimated variances and covariances are inconsistent with the null hypothesis that a linear approximation is sufficient. We will refer to this test as the “\( \lambda \)-test.”

The a- and \( \lambda \)-tests are in the class of GMM tests examined in Newey and West (1987). The a-test is in the form of the so-called “D” test, and the \( \lambda \)-test is a version of their statement of a Wald test. If the underlying problem satisfies the assumptions of Theorems 1 and 2 in Newey and West (1987), then the two tests are asymptotically equivalent; that is, their difference is \( o_p(1) \). Furthermore, in the absence of the “two-pass regression problem” (see Shanken, 1992), Proposition 4 in Newey and West (1987) states that the two tests are, in fact, identical. This proposition does not hold in our application because of the differences in the covariance matrices under errors in estimating the factor betas. Our simulation work in Section 3 and our empirical application in Section 4 demonstrate that the tests are not equal in finite sample.

The a-test and the \( \lambda \)-test follow the common practice in the empirical literature of assuming that the estimated model is correctly specified. Kan, Robotti, and Shanken (2013) (hereafter KRS) construct a test for nested models based on the cross-sectional \( R^2 \) statistic from the second-pass regression that is robust to misspecification. The (nested version of the) “KRS-test” of \( H_0 : \hat{\rho}^2_U = \hat{\rho}^2_R \) uses the fitted (cross-sectional) \( R^2 \) from each model, \( \hat{\rho}^2_i \) for \( i \in \{U, R\} \). It is defined as

\[ T(\hat{\rho}^2_U - \hat{\rho}^2_R) \sim \sum_{i=1}^{J_2} \frac{l_i}{Q_0} \hat{x}_i, \tag{16} \]

13 KRS also construct the more general test statistic for comparing non-nested models, but that is not relevant to our analysis.
where $Q_0$ is a quadratic form in the weighting matrix (e.g., ordinary least squares (OLS) or generalized least squares (GLS)) and the cross-sectional deviations in expected returns, $\eta_j$ is the $j$-th eigenvalue of the matrix product of the inverse of the appropriate partition of the model parameter covariance matrix and the covariance matrix of the additional risk premium terms in the unrestricted model, and $\tilde{x}_j$ are independent $\chi^2_1$ random variables.\(^{14}\) This distribution is dependent on the model (through the estimated weights $(\hat{\omega})$ and the weighting matrix (OLS versus GLS). However, it can be computed via simulation. As KRS note, the model ranking based on the differences in $R^2$ just re-arranges the information in the quadratic form based on the risk premiums, but there is a potential value in examining the finite-sample properties of nonlinearity tests based on misspecification-robust versions of this statistic.

3. A Calibrated Example

3.1 The Pricing Kernel

This example follows the general approach of Brennan, Wang, and Xia (2004) (hereafter, BWX) for implementing a version of the Intertemporal CAPM (ICAPM) of Merton (1973). We differ from BWX by using a discrete-time formulation of the economy and, more importantly, in the specification, identification, and calibration of the dynamics of the state variables.

Let $r_t^f$ denote the yield on a one-period, default-risk-free bond from $t$ to $t+1$. The dynamics of this short rate are exogenously specified as

$$\ln r_t^f = \alpha_1 + \beta_1 \ln r_{t-1}^f + \epsilon_{1t},$$

(17)

where $\epsilon_{1t} \sim \mathcal{N}(0, \sigma^2_1)$. Equation (17) was introduced into the term structure literature in Black, Derman, and Toy (1990) and Black and Karasinski (1991). The dynamics of the log of the maximum Sharpe ratio process are given by

$$s_t = \alpha_2 + \beta_2 s_{t-1} + \beta_3 \ln r_t^f + \epsilon_{2t},$$

(18)

where $s_t \equiv \ln S_t$ and $\epsilon_{2t} \sim \mathcal{N}(0, \sigma^2_2)$. By construction, Equations (17) and (18) force the conditional correlation between $\epsilon_{1t}$ and $\epsilon_{2t}$ to equal zero, for all $t$.

Equations (17) and (18) are a linear, first-order vector autoregression in logs. Equation (18) implies that the Sharpe ratio on the conditionally mean-variance efficient portfolio can never be negative. There is no specific structural model underlying Equation (18), but there are business cycle-related movements in both short-term interest rates and estimated Sharpe ratios that are consistent with a nonzero $\beta_3$. We will estimate the parameters of Equations (17) and (18) below.

Let $r_{t+1}^M$ denote the continuously compounded return on a broad market portfolio, and let $\epsilon_{3t} \sim \mathcal{N}(0, \sigma^2_3)$ denote the innovation, at $t$, to this return. The unique pricing kernel for the economy, from $t$ to $t+1$, under the physical measure, is defined as a special case of the kernel in Equation (1):

$$M_{t+1} = \exp \left( r_t^f - \frac{1}{2} \Lambda_t^2 + \Lambda_t \omega^{-1} s_{t+1} \right),$$

(19)

\(^{14}\) The asymptotic distribution of the statistic in Equation (16) is derived in Propositions A.5 and A.6 in the Online Appendix to KRS.
where $\Lambda_i$ is a risk premium parameter, $\xi_{t+1} \equiv \delta' \epsilon_{t+1}$, $\epsilon_{t+1} = (\epsilon_{1t+1}, \epsilon_{2t+1}, \epsilon_{3t+1})'$, $\delta$ is a vector of constants, $\omega = (\delta' \Sigma \delta)^{1/2}$, and $\Sigma = I_3$ is the covariance matrix of $\epsilon_{t+1}$. The single innovation to the kernel is a weighted-average of the innovations to the market return, the risk-free rate, and the maximum Sharpe ratio.\textsuperscript{15} This is the discrete-time analog of the specification in BWX. The fact that the shock, $\zeta_t$, is normally distributed implies that $Mt$ is conditionally lognormal.

There are three economically significant restrictions imposed on the pricing kernel by Equation (19). First, we assume a single risk premium on a single composite shock. This is a simplifying assumption. The basic issue of the efficacy of the standard linear conditioning approximation carries over into a multifactor setting. Second, the market portfolio plays a specific role in the pricing kernel. This is true in a variety of common models [the CAPM, Fama and French (1993), and Epstein and Zin (1989), are examples], but it is not true of all prominent pricing models. Campbell and Cochrane (2000), Bansal and Yaron (2004), and Lustig and van Nieuwerburgh (2005) are all examples of consumption-based models in which the market portfolio (and its innovations) plays no distinct role in the pricing kernel. Third, the Sharpe ratio is a nonlinear function of the risk premium, as noted in Equation (3), which means that the pricing kernel in Equation (19) assumes that innovations to the risk premiums help determine the pricing kernel directly.

The final assumption we make in specifying the example economy is that from $t$ to $t+1$ the asset returns and pricing kernel have a joint lognormal distribution, conditional on the current realizations of the market and the factors. Under this assumption, the fundamental asset pricing equation implies a generalization of the basic moment conditions in Hansen and Singleton (1983), where marginal utility of consumption growth, in that setting, is replaced by the pricing kernel:

$$
Et(r_{it+1} + m_{t+1}) + \frac{1}{2} [\text{var}(r_{it+1}) + \text{var}(m_{t+1}) + 2\text{cov}(r_{it+1}, m_{t+1})] = 0,
$$

for $i = 1, \ldots, N$, where $r_{it+1}$ is the continuously compounded return to asset $i$. Equation (20) can be rewritten in terms of returns, functions of latent factors, and factor innovations:

$$
Et(r_{it+1}) + \frac{1}{2} \text{var}(r_{it+1}) = r_t' + \Lambda_i \text{cov}(r_{it+1}, \omega^{-1} \xi_{t+1}),
$$

for $i = 1, \ldots, N$. The covariance in Equation (21) is unconditional given the assumptions on the state variable innovations that serve as factors.

Given that the factor, $\omega^{-1} \xi_{t+1}$, is iid standard normal, Equation (21) can be rewritten in a traditional single-factor (conditional) beta pricing form:

$$
Et(r_{it+1}) + \frac{1}{2} \text{var}(r_{it+1}) - r_t' = \Lambda_i \beta_{it},
$$

where

$$
\beta_{it} = \frac{\text{cov}(r_{it+1}, \xi_{t+1})}{\omega^2}.
$$

\textsuperscript{15} In Section 3.2, we show that $\delta$ cannot be identified uniquely from the time-series regression of excess returns on the (extended) set of factors. However, the values of $\delta$ that are implicitly chosen in the regression in Equation (28) are those that best fit the set of test assets used below.
Figure 1 inverts Equation (3) to show that the risk premium is a convex function of the level of the natural logarithm of the Sharpe ratio. The first- and second-order polynomial approximations to the true risk premium function are

\[
\tilde{A}_t(1) = \psi_0 + \psi_1 s_t, \\
\tilde{A}_t(2) = \psi_0 + \psi_1 s_t + \psi_2 P_2(s_t),
\]

where \(\psi_i, i = 0, 1, 2\), are parameters of the approximation that can be fit by a simple linear regression. Figure 1 also shows the first- and second-order approximations, \(\tilde{A}_t(1)\) and \(\tilde{A}_t(2)\). It is clear that the second-order approximation is virtually exact over the entire reasonable range of \(s_t\).

\(\tilde{A}_t(2)\) is a very close approximation to the true risk premium function in this case, and we can use it to understand the meaning of moderate nonlinearity in this example. In particular, the curvature index in this case is

\[
\gamma(s_t) \approx \frac{\psi_2}{\psi_1 + \psi_2 P'_2(s_t)},
\]

where \(P'_2(s_t)\) is the first derivative of \(P_2(s_t)\). Given the estimated values of \(\psi_1\) and \(\psi_2\) and the function \(P'_2(s_t)\), for Sharpe ratios in the range \([0.1, 1.6]\), the maximum value of the curvature index is 4.5.

The simple reduced-form model described above uses the intuition of a dynamic capital market line to express a pricing kernel that can arise from a variety of underlying structural models. It generates a conditional pricing model that is nearly linear. Although there are three state variables that describe the location of the time-varying capital market line, there

![Figure 1. The risk premium on the log Sharpe ratio along with the first- and second-order approximations.](image-url)
is only one risk premium driving excess returns. The functional form of this risk premium represents the only source of nonlinearity in the pricing implications of the model. The risk premium is a convex function of the log Sharpe ratio state variable, and it can be well approximated by a second-order polynomial in the log Sharpe ratio.

We believe that this structure is ideally suited as an example of the possible pricing implications of mis-specifying moderate nonlinearity in a conditional model. On the one hand, if the model generated substantial nonlinearities, then it would be an example that could easily be dismissed as “rigged” to deliver poor performance of the standard linear approximation. On the other hand, if the model was fully linear, then the standard specification would be completely accurate. The nature and extent of the pricing misspecification introduced by the standard linear approximation approach is examined in the following subsections.

3.2 Estimation
We assume that the econometrician examining data generated by the model in Section 3.1 has access to excess returns and historical values of the market portfolio and the true state variables. The time-series regression corresponding to Equation (A.4) (in Appendix A) is

$$ r_{i,t+1}^e = a_i + b_{i,1} \xi_{t+1} + b_{i,2} s_t + b_{i,3} [\xi_{t+1} \cdot s_t] + \eta_{i,t+1}, \tag{26} $$

for $i = 1, \ldots, N$. In this case, the true factor is the innovation to the pricing kernel, $\xi_{t+1}$, the information proxy variable is the lagged value of the log Sharpe ratio, and the conditional component of the model is captured in the interaction term, $\xi_{t+1} \cdot s_t$.

Given that $\xi_{t+1} = \delta_1 \epsilon_{1,t+1} + \delta_2 \epsilon_{2,t+1} + \delta_3 \epsilon_{3,t+1}$, Equation (26) expands to

$$ R_{i,t+1}^e = a_i + \delta_1 b_{i,1} \epsilon_{1,t+1} + \delta_2 b_{i,1} \epsilon_{2,t+1} + \delta_3 b_{i,1} \epsilon_{3,t+1} + b_{i,2} s_t $$
$$ + \delta_1 b_{i,3} [\epsilon_{1,t+1} \cdot s_t] + \delta_2 b_{i,3} [\epsilon_{2,t+1} \cdot s_t] + \delta_3 b_{i,3} [\epsilon_{3,t+1} \cdot s_t] + \eta_{i,t+1}. \tag{27} $$

Since the $\delta_i$ and the original $b_i$ coefficients cannot be identified separately, it is notionally convenient to rewrite Equation (27) as

$$ r_{i,t+1}^e = a_i + d_{i,1} \epsilon_{1,t+1} + d_{i,2} \epsilon_{2,t+1} + d_{i,3} \epsilon_{3,t+1} + d_{i,4} s_t $$
$$ + d_{i,5} [\epsilon_{1,t+1} \cdot s_t] + d_{i,6} [\epsilon_{2,t+1} \cdot s_t] + d_{i,7} [\epsilon_{3,t+1} \cdot s_t] + \eta_{i,t+1}. \tag{28} $$

The second-pass cross-sectional regression is

$$ \hat{r}_i^{e} = \sum_{j=1}^7 \hat{d}_{ij} \hat{\epsilon}_{ij} + v_i, \tag{29} $$

for $i = 1, \ldots, N$, with $\hat{d}_{ij}$ denoting the fitted value of the coefficients from Equation (28) and the fitted value of $v_i$ corresponding to the measured alpha for asset $i$.

In the context of this model, the linearity assumption in the standard conditioning approach is incorrect. Given the effectiveness of the second-order linear approximation to the true risk premium, a more accurate specification is

$$ r_{i,t+1}^e = a_i + c_{i,1} \epsilon_{1,t+1} + c_{i,2} \epsilon_{2,t+1} + c_{i,3} \epsilon_{3,t+1} + c_{i,4} s_t + c_{i,5} P_2(s_t) $$
$$ + c_{i,6} [\epsilon_{1,t+1} \cdot s_t] + c_{i,7} [\epsilon_{1,t+1} \cdot P_2(s_t)] $$
$$ + c_{i,8} [\epsilon_{2,t+1} \cdot s_t] + c_{i,9} [\epsilon_{2,t+1} \cdot P_2(s_t)] $$
$$ + c_{i,10} [\epsilon_{3,t+1} \cdot s_t] + c_{i,11} [\epsilon_{3,t+1} \cdot P_2(s_t)] + \eta_{i,t+1}. \tag{30} $$
The difference between Equations (28) and (30) is the omission of the relevant factors involving \( P_2(s_t) \). There is a corresponding (expanded) cross-sectional regression that is analogous to Equation (29).

This comparison makes the source of the misspecification clear. The heart of our analysis is contained in the answers to the following questions: (i) When the true pricing errors (alphas) are zero, what are the magnitudes of the alphas introduced by misspecification of the conditional pricing relation? (ii) How do these true alphas compare with estimates of alphas that might be constructed in a sample of a realistic size drawn from this model? Finally, (iii) does a test statistic of the form in Equation (15) for omitted priced factors have any power to detect the omitted variables in a sample of realistic size drawn from this model?

3.3 How Large Are the Alphas Caused by Misspecification?
The first step in calibrating the model is to construct proxies for the state variables \( s \) and \( \ln r_f \). These state variables will be used with a set of asset returns to examine the model’s implications. We use the yield on a one-month Treasury bill, from CRSP, as a proxy for the short rate \( r_f \). The sample period is from January 1960 to December 2015 (for a total of \( T = 672 \) monthly observations). A plot of the log short rate is shown in the top panel of Figure 2. Identifying this state variable is robust to reasonable alternative short rate choices.

Identifying the log Sharpe ratio is more difficult. The model in Section 3.1 specifies the state variable as the log Sharpe ratio of the conditionally mean-variance efficient portfolio.

![Figure 2](image.png)

**Figure 2.** The time series of the pricing kernel state variables. The log short rate in panel (a) and the log Sharpe ratio in panel (b).
that prices all traded assets. This portfolio is unobservable, and it is the focus of most modern asset pricing research. Furthermore, even if this asset was observable, its moments would need to be estimated in order to construct the Sharpe ratio. Our approach to calibrating the dynamics of the log Sharpe ratio is to treat the market portfolio as an observable proxy for the conditionally mean-variance efficient portfolio and then to estimate the dynamics of the log Sharpe ratio from historical data on the market portfolio. While this proxy is clearly imperfect, it is not unreasonable to assume that the log Sharpe ratio dynamics of the unobservable conditionally mean-variance efficient portfolio is similar to that of the observable market portfolio, especially for our calibration purposes.

More specifically, we identify the log Sharpe ratio state variable using the filtering approach in Brandt and Kang (2004). They assume that the continuously compounded excess return on the CRSP value-weighted index has both a time-varying conditional mean and a time-varying conditional volatility. The values of these moments are recovered from the realized data using an approximation to the true likelihood based on the Kalman filter. Smoothed estimates of the conditional moments are then constructed based on the full sample of return data. Our estimate of the log Sharpe ratio is constructed from these smoothed estimates. This filtering procedure also recovers a (full sample) estimate of the shock to the market portfolio, \( \epsilon_3 \), that can be used in calibrating the pricing kernel.\(^{16}\) A plot of the log Sharpe ratio, from 1960 to 2015, is shown in the bottom panel of Figure 2.

Panel A of Table I presents some simple full-sample summary statistics for the empirical estimates of the two state variables. The continuously compounded return on the Treasury bill averaged 4.4% per year over this period with a standard deviation of 2.9% per year. It is slightly right-skewed and fat-tailed relative to the normal distribution, and it is highly persistent.\(^{17}\) The log of the Sharpe ratio averages about –0.58 with a standard deviation of 0.47. The log Sharpe ratio is slightly right-skewed with tails that are substantially thinner than a normal distribution.

Given the time series of the state variables, we fit Equations (17) and (18) to obtain estimates of the parameters that describe these dynamics. Since this system is linear in logs, all of the parameters are estimated by OLS with robust standard errors used for inference. Panel B of Table I presents these results. The log Sharpe ratio is significantly positively related to its own lagged value and not significantly related to the contemporaneous level of the log short rate. The estimates of the short rate parameters are consistent with the large literature on fitting univariate time-series models to short-term interest rate data.

The fitted residuals from Equations (17) and (18) do not satisfy the model assumption of being iid draws from a bivariate standard normal density. This is not surprising, given the evidence in Figure 2 and the extensive literature on the volatility dynamics of both stock returns and short-term interest rates. In order to calibrate our model with a simple error structure, we filter the residuals from the regressions in Table I using the dynamic

---

16 By construction, these return shocks are normalized to have a standard deviation of 1.

17 The \( t \)-test reported in Panel B of Table IV for the null of \( \beta_1 = 1 \) is only suggestive. There is a large literature on the failings of standard asymptotic theory in this estimation setting. An augmented Dickey–Fuller test for a unit root in the short rate (log Sharpe ratio), using an optimal automatic lag length selection algorithm, has a fitted value of –2.135 (–6.493), with a \( p \)-value of 0.234 (0.00).
conditional correlation multivariate GARCH model of Engle (2002). The transformed shocks correspond more closely to a constant identity matrix covariance matrix for $(\epsilon_1, \epsilon_2)'$. They are combined with $\epsilon_3$ [which is already normalized and orthogonalized with respect to $(\epsilon_1, \epsilon_2)'$].

We calibrate asset returns to the twenty-five portfolios formed by Fama and French (1993) on the basis of market capitalization and B/M ratio. The monthly portfolio returns are value-weighted and continuously compounded in excess of the continuously compounded yield on a one-month Treasury bill. The sample period is from, again, January 1960 to December 2015. The first four sample moments of excess returns are reported in Table II. Holding market capitalization constant, the average excess returns increase from low B/M to high B/M portfolios. In our sample period, there is not a consistent size effect

### Table I. State variables

**Notes:** The sample period is from January 1960 to December 2015 ($T = 672$ months). $\ln(\text{Short Rate})$ is the continuously compounded yield on the one-month Treasury bill, quoted annually and not in percent. The $\ln(\text{Sharpe Ratio})$ is the natural logarithm of the Sharpe ratio extracted from the S&P 500 using the filtering method in Brandt and Kang (2004). Panel A contains simple summary statistics of the full sample. Skewness is the sample estimate of the third central moment standardized by the cube of the sample standard deviation. Kurtosis is the sample estimate of the fourth central moment standardized by the sample standard deviation raised to the fourth power. It is not excess kurtosis relative to the normal distribution. AR(1) is the sample estimate of the first-order autocorrelation coefficient. In Panel B, the parameters are estimated using ordinary least squares. The parameter estimates are for Equations (17) and (18) in the text. Robust standard error estimates, based on the Newey–West estimator are in parentheses. The brackets are $t$-tests for $\beta_i = 0$, $i = 1, 2$, $\beta_1 = 1$, for $i = 1, 2$, and $\beta_3 = 0$. The estimates in this table are constructed using the programs ReturnMoments.m and rp_est.m.

#### Panel A: Summary statistics

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(\text{Short Rate})$</td>
<td>0.044</td>
<td>0.029</td>
<td>0.585</td>
<td>3.702</td>
<td>0.982</td>
</tr>
<tr>
<td>$\ln(\text{Sharpe Ratio})$</td>
<td>-0.576</td>
<td>0.470</td>
<td>0.564</td>
<td>0.572</td>
<td>0.555</td>
</tr>
</tbody>
</table>

#### Panel B: Regression results

<table>
<thead>
<tr>
<th>Series</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\gamma$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\sigma_3$</th>
<th>$\sigma_4$</th>
<th>$\sigma_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(\text{Short Rate})$</td>
<td>0.983</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>$\ln(\text{Sharpe Ratio})$</td>
<td>-0.245</td>
<td>-0.245</td>
<td>-0.245</td>
<td>-0.245</td>
<td>-0.245</td>
<td>-0.245</td>
<td>-0.245</td>
<td>-0.245</td>
</tr>
</tbody>
</table>

18 This model was fit to the data using Matlab code distributed by the Economics department of the University of California at San Diego.
for a given B/M rank. Excess return volatility declines with size, and it also declines with the B/M ratio but only up to the fourth B/M quintile. The returns to all series are negatively skewed and exhibit excess kurtosis (relative to the normal distribution.). The estimates in this table are constructed using the program ReturnMoments.m.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 BM1</td>
<td>-0.055</td>
<td>8.011</td>
<td>-0.538</td>
<td>5.684</td>
</tr>
<tr>
<td>S1 BM2</td>
<td>0.496</td>
<td>6.914</td>
<td>-0.450</td>
<td>6.203</td>
</tr>
<tr>
<td>S1 BM3</td>
<td>0.562</td>
<td>6.023</td>
<td>-0.615</td>
<td>6.089</td>
</tr>
<tr>
<td>S1 BM4</td>
<td>0.803</td>
<td>5.712</td>
<td>-0.555</td>
<td>6.676</td>
</tr>
<tr>
<td>S1 BM5</td>
<td>0.892</td>
<td>6.036</td>
<td>-0.669</td>
<td>6.742</td>
</tr>
<tr>
<td>S2 BM1</td>
<td>0.164</td>
<td>7.280</td>
<td>-0.720</td>
<td>5.453</td>
</tr>
<tr>
<td>S2 BM2</td>
<td>0.531</td>
<td>6.064</td>
<td>-0.865</td>
<td>6.843</td>
</tr>
<tr>
<td>S2 BM3</td>
<td>0.679</td>
<td>5.479</td>
<td>-0.484</td>
<td>6.803</td>
</tr>
<tr>
<td>S2 BM4</td>
<td>0.762</td>
<td>5.263</td>
<td>-0.777</td>
<td>6.387</td>
</tr>
<tr>
<td>S2 BM5</td>
<td>0.785</td>
<td>6.111</td>
<td>-0.809</td>
<td>6.541</td>
</tr>
<tr>
<td>S3 BM1</td>
<td>0.241</td>
<td>6.707</td>
<td>-0.717</td>
<td>5.320</td>
</tr>
<tr>
<td>S3 BM2</td>
<td>0.607</td>
<td>5.523</td>
<td>0.910</td>
<td>7.021</td>
</tr>
<tr>
<td>S3 BM3</td>
<td>0.576</td>
<td>5.053</td>
<td>-0.769</td>
<td>5.825</td>
</tr>
<tr>
<td>S3 BM4</td>
<td>0.711</td>
<td>4.924</td>
<td>-0.615</td>
<td>5.813</td>
</tr>
<tr>
<td>S3 BM5</td>
<td>0.849</td>
<td>5.651</td>
<td>-0.768</td>
<td>6.621</td>
</tr>
<tr>
<td>S4 BM1</td>
<td>0.379</td>
<td>5.964</td>
<td>-0.582</td>
<td>5.340</td>
</tr>
<tr>
<td>S4 BM2</td>
<td>0.431</td>
<td>5.191</td>
<td>-0.940</td>
<td>7.345</td>
</tr>
<tr>
<td>S4 BM3</td>
<td>0.539</td>
<td>5.026</td>
<td>-0.885</td>
<td>7.309</td>
</tr>
<tr>
<td>S4 BM4</td>
<td>0.725</td>
<td>4.815</td>
<td>-0.529</td>
<td>5.451</td>
</tr>
<tr>
<td>S4 BM5</td>
<td>0.646</td>
<td>5.702</td>
<td>-0.694</td>
<td>5.796</td>
</tr>
<tr>
<td>S5 BM1</td>
<td>0.340</td>
<td>4.688</td>
<td>-0.485</td>
<td>4.884</td>
</tr>
<tr>
<td>S5 BM2</td>
<td>0.400</td>
<td>4.462</td>
<td>-0.627</td>
<td>5.502</td>
</tr>
<tr>
<td>S5 BM3</td>
<td>0.433</td>
<td>4.306</td>
<td>-0.537</td>
<td>5.697</td>
</tr>
<tr>
<td>S5 BM4</td>
<td>0.372</td>
<td>4.690</td>
<td>-0.931</td>
<td>8.014</td>
</tr>
<tr>
<td>S5 BM5</td>
<td>0.529</td>
<td>5.339</td>
<td>-0.453</td>
<td>4.471</td>
</tr>
</tbody>
</table>

The calibrated values of the alphas generated by the model in Section 3.1 are constructed by computing the required inputs: (i) the factor covariance matrices, \( \Sigma_{11}, \Sigma_{22}, \) and \( \Sigma_{12} \); (ii) the covariance matrices of the factors with the excess returns, \( \Sigma_{F1R^e} \) and \( \Sigma_{F2R^e} \); and (iii) the market risk premium parameters, \( \lambda_1 \) and \( \lambda_2 \). In this setting,

\[
F_1_{(7\times 1)} = (\epsilon_{1,t+1}, \epsilon_{2,t+1}, \epsilon_{3,t+1}, s_t, \epsilon_{1,t+1} \cdot s_t, \epsilon_{2,t+1} \cdot s_t, \epsilon_{3,t+1} \cdot s_t)'
\]

\[
F_2_{(4\times 1)} = (P_2(s_t), \epsilon_{1,t+1} \cdot P_2(s_t), \epsilon_{2,t+1} \cdot P_2(s_t), \epsilon_{3,t+1} \cdot P_2(s_t))'
\]

are the included and omitted factors, respectively.

The proxies for the factors are constructed from the fitted (and normalized) shocks to the measured state variables, the shock to the market portfolio, and the lagged levels of the
Given time-series estimates for the factors, it is straightforward to construct point estimates of the necessary parameters. The purpose of our estimation is to calibrate the model and not to propose the model as a thorough description of the cross-section of expected returns. The parameters are estimated using Fama–MacBeth with a single time-series regression to recover the factor loadings and monthly cross-sectional regressions to estimate the risk premiums and cross-sectional pricing errors. Since the monthly risk premiums and pricing errors are not iid, we estimate standard errors using a robust covariance matrix (Newey–West) with an asymptotically optimal bandwidth selection.

Table III contains the estimates of the risk premiums from the model that include second-order approximations in the conditioning information. The innovations to the short rate and the Sharpe ratio have positive risk premiums. The other point estimates are predominantly negative. The Shanken-adjusted standard errors are typically three times as large.

Following the recommendation in Ferson, Sarkissian, and Simin (2003), we stochastically detrend the (de-meaned) conditioning variable, \( s_t \), using a 12-month moving average filter.
large as their uncorrected (but robust) counterparts. There is some evidence consistent with significant risk premiums using only robust standard errors, but this disappears after accounting for measurement error in the factor loadings.

The estimates of the alphas (pricing errors) from the full model are shown in Table IV. The model prices nineteen (of the twenty-five) portfolios with an average alpha of 10% or less of the corresponding average excess return. The smallest size and B/M portfolio has a point estimate of (absolute) alpha is nearly 50% of the level of the average excess returns. Six of the twenty-five estimated alphas are large relative to the robust standard errors, but none of the alphas is statistically significant after applying the Shanken correction.

Table IV. Fitted alphas

Notes: The data are from January 1961 to December 2015. The model is estimated using Fama–MacBeth with a second-order approximation to the risk premium function; that is, using Equations (28) and (29) in the text. They are reported in percent per month. The standard error column labeled [1] uses the Newey–West estimate of the long-run covariance matrix with automatic bandwidth selection. The standard error column labeled [2] adds the Shanken correction term for variability introduced in the first-pass regression to the Newey–West estimator. “Rel. $a_i$” for a given portfolio is defined as the portfolio’s alpha divided by the average excess return to the portfolio. The estimates in this table are constructed using the program rp_est_FSS_FM.m.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$a$</th>
<th>[1]</th>
<th>[2]</th>
<th>Rel. $a_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 BM1</td>
<td>-0.167</td>
<td>0.059</td>
<td>0.206</td>
<td>-0.508</td>
</tr>
<tr>
<td>S1 BM2</td>
<td>0.144</td>
<td>0.063</td>
<td>0.218</td>
<td>0.163</td>
</tr>
<tr>
<td>S1 BM3</td>
<td>-0.007</td>
<td>0.037</td>
<td>0.128</td>
<td>-0.007</td>
</tr>
<tr>
<td>S1 BM4</td>
<td>0.106</td>
<td>0.062</td>
<td>0.213</td>
<td>0.089</td>
</tr>
<tr>
<td>S1 BM5</td>
<td>0.095</td>
<td>0.051</td>
<td>0.178</td>
<td>0.074</td>
</tr>
<tr>
<td>S2 BM1</td>
<td>-0.038</td>
<td>0.037</td>
<td>0.127</td>
<td>-0.069</td>
</tr>
<tr>
<td>S2 BM2</td>
<td>0.053</td>
<td>0.037</td>
<td>0.129</td>
<td>0.058</td>
</tr>
<tr>
<td>S2 BM3</td>
<td>0.074</td>
<td>0.055</td>
<td>0.190</td>
<td>0.070</td>
</tr>
<tr>
<td>S2 BM4</td>
<td>-0.028</td>
<td>0.034</td>
<td>0.117</td>
<td>-0.025</td>
</tr>
<tr>
<td>S2 BM5</td>
<td>0.024</td>
<td>0.023</td>
<td>0.081</td>
<td>0.021</td>
</tr>
<tr>
<td>S3 BM1</td>
<td>-0.048</td>
<td>0.047</td>
<td>0.164</td>
<td>-0.077</td>
</tr>
<tr>
<td>S3 BM2</td>
<td>-0.065</td>
<td>0.051</td>
<td>0.176</td>
<td>-0.066</td>
</tr>
<tr>
<td>S3 BM3</td>
<td>0.036</td>
<td>0.058</td>
<td>0.200</td>
<td>0.037</td>
</tr>
<tr>
<td>S3 BM4</td>
<td>0.096</td>
<td>0.061</td>
<td>0.210</td>
<td>0.087</td>
</tr>
<tr>
<td>S3 BM5</td>
<td>0.053</td>
<td>0.034</td>
<td>0.119</td>
<td>0.043</td>
</tr>
<tr>
<td>S4 BM1</td>
<td>-0.039</td>
<td>0.054</td>
<td>0.186</td>
<td>-0.051</td>
</tr>
<tr>
<td>S4 BM2</td>
<td>-0.001</td>
<td>0.040</td>
<td>0.140</td>
<td>-0.002</td>
</tr>
<tr>
<td>S4 BM3</td>
<td>0.088</td>
<td>0.054</td>
<td>0.185</td>
<td>0.095</td>
</tr>
<tr>
<td>S4 BM4</td>
<td>0.094</td>
<td>0.054</td>
<td>0.185</td>
<td>0.085</td>
</tr>
<tr>
<td>S4 BM5</td>
<td>-0.273</td>
<td>0.117</td>
<td>0.404</td>
<td>-0.265</td>
</tr>
<tr>
<td>S5 BM1</td>
<td>-0.065</td>
<td>0.054</td>
<td>0.185</td>
<td>-0.090</td>
</tr>
<tr>
<td>S5 BM2</td>
<td>0.019</td>
<td>0.044</td>
<td>0.152</td>
<td>0.024</td>
</tr>
<tr>
<td>S5 BM3</td>
<td>0.088</td>
<td>0.039</td>
<td>0.136</td>
<td>0.108</td>
</tr>
<tr>
<td>S5 BM4</td>
<td>-0.083</td>
<td>0.040</td>
<td>0.137</td>
<td>-0.109</td>
</tr>
<tr>
<td>S5 BM5</td>
<td>-0.157</td>
<td>0.063</td>
<td>0.219</td>
<td>-0.172</td>
</tr>
</tbody>
</table>
In Table V, we present estimates of the differences in alphas from the nonlinear and linear models. These differences form the numerator of the quadratic form in the alpha-based nonlinearity test. The final column shows the difference in alphas as a percentage of average absolute excess returns. These differences can be large, both absolutely and relative to average absolute excess returns. The pricing errors for the linear model tend to be larger in the smaller three size quintiles (seen as negative $\Delta \alpha$ differences), but the main point is that adding the second-order terms to the pricing kernel has an impact on the measured pricing errors.

### 3.4 A Monte-Carlo Study

Do tests of the form of Equation (14) or (15), described in Section 2, have reasonable size and power against a specific local alternative? If not, then the only symptom of nonlinearity are the measured pricing errors, which generally reflect a wide range of possible model misspecification, not only nonlinearity of the risk premiums.

**Table V. Change in alpha between the nonlinear and linear models**

*Notes:* The data are monthly from January 1961 to December 2015. Alpha changes are quoted in percent per month, and they are constructed by subtracting the alphas from the linear model from the alphas generated by the nonlinear approximation. “Rel. $\Delta \alpha$” are defined as the difference divided by the absolute value of the associated average excess returns. The estimates in this table are constructed using the programs `rp_est_FSS_FM.m` and `rp_est_fss_lin.m`.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$\Delta \alpha$</th>
<th>Rel. $\Delta \alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 BM1</td>
<td>0.077</td>
<td>0.235</td>
</tr>
<tr>
<td>S1 BM2</td>
<td>−0.013</td>
<td>−0.015</td>
</tr>
<tr>
<td>S1 BM3</td>
<td>0.018</td>
<td>0.019</td>
</tr>
<tr>
<td>S1 BM4</td>
<td>−0.043</td>
<td>−0.036</td>
</tr>
<tr>
<td>S1 BM5</td>
<td>−0.047</td>
<td>−0.037</td>
</tr>
<tr>
<td>S2 BM1</td>
<td>−0.026</td>
<td>−0.048</td>
</tr>
<tr>
<td>S2 BM2</td>
<td>0.099</td>
<td>0.108</td>
</tr>
<tr>
<td>S2 BM3</td>
<td>−0.018</td>
<td>−0.017</td>
</tr>
<tr>
<td>S2 BM4</td>
<td>−0.036</td>
<td>−0.032</td>
</tr>
<tr>
<td>S2 BM5</td>
<td>0.031</td>
<td>0.027</td>
</tr>
<tr>
<td>S3 BM1</td>
<td>−0.074</td>
<td>−0.118</td>
</tr>
<tr>
<td>S3 BM2</td>
<td>−0.088</td>
<td>−0.089</td>
</tr>
<tr>
<td>S3 BM3</td>
<td>−0.008</td>
<td>−0.008</td>
</tr>
<tr>
<td>S3 BM4</td>
<td>−0.037</td>
<td>−0.034</td>
</tr>
<tr>
<td>S3 BM5</td>
<td>0.025</td>
<td>0.020</td>
</tr>
<tr>
<td>S4 BM1</td>
<td>−0.046</td>
<td>−0.061</td>
</tr>
<tr>
<td>S4 BM2</td>
<td>0.079</td>
<td>0.097</td>
</tr>
<tr>
<td>S4 BM3</td>
<td>0.083</td>
<td>0.090</td>
</tr>
<tr>
<td>S4 BM4</td>
<td>0.015</td>
<td>0.014</td>
</tr>
<tr>
<td>S4 BM5</td>
<td>−0.013</td>
<td>−0.012</td>
</tr>
<tr>
<td>S5 BM1</td>
<td>0.039</td>
<td>0.054</td>
</tr>
<tr>
<td>S5 BM2</td>
<td>−0.017</td>
<td>−0.022</td>
</tr>
<tr>
<td>S5 BM3</td>
<td>−0.001</td>
<td>−0.002</td>
</tr>
<tr>
<td>S5 BM4</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>S5 BM5</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>
The general steps in the structure of our Monte-Carlo experiment are:

1. Generate 5,000 independent simulated sample paths of an economy in which returns are generated according to the model economy calibrated to the point estimates of size and B/M-sorted returns. Excess returns on the different portfolios are constructed by:
   a. Simulating values for the model state variables, 
      \( (st, \ln rf_t)' \), using the estimated parameter values in Table IV and bivariate standard normal shocks.
   b. Along each simulated sample path, realized excess returns are generated as the sum of expected returns and unexpected returns. Time-varying expected returns are generated according to the model, the calibrated parameter values, and the realized values of the state variables generated in Step #1a; that is, using Equations (22) and (24):
      \[
      E_t(r_{i,t+1}) + \frac{1}{2} \operatorname{var}_t(r_{i,t+1}) - r_t^{*} = [\psi_0 + \psi_1 st + \psi_2 P_2(s_t)]\beta_t,
      \]
      for each \( i \) in \( N \), where \( \beta_t \) is defined in Equation (23). Unexpected returns are generated at each date in the simulation by drawing from a multivariate normal distribution with mean zero and a diagonal covariance matrix with the nonzero elements of the matrix set equal to the estimates of the diagonal elements of the covariance matrix of the shocks from Equation (26).
   c. Along each simulated path of the state variables, we generate returns by a separate linear version of the model economy in order to assess the actual size of the nonlinearity tests.

2. Each simulation is of length 675 months, and the state variables and realized excess returns are collected after a “burn-in” period of 6,075 months designed to reduce the influence of initial conditions.\(^2^0\)

3. For each simulated path:
   a. We estimate a standard linearized version of the conditional model, as in Equations (26) and (29). The model is estimated using a standard Fama–MacBeth algorithm. The alphas (pricing errors) for each asset and the estimated risk premiums in each simulation are stored for later examination.
   b. We also estimate a version of an approximate model that includes second-order terms in the conditioning variables. We construct the test statistics in Equations (14) and (15)—both with and without a correction for estimation of the betas in the first-pass time-series regression—for the null hypothesis that the higher-order terms are all jointly equal to zero. We store these results for later examination (of the power of the tests against the specific alternative).

4. Repeat Steps (1)–(3) using more nonlinearity in the pricing kernel by slightly adjusting the higher-order parameter values to generate a slightly larger spread in the simulated test asset returns.

---

\(^{20}\) The unconditional means of the state variables are used as the initial conditions for each simulation. In order to mitigate problems caused by the near unit root properties of the estimated short rate, we discard sample paths whose first value after the burn-in period is more than three standard deviations from the unconditional mean of the short rate. This adjustment is important because the nonlinearity becomes more important in the tails of the distribution, and it is therefore designed to make our results conservative.
The first column of Table VI shows the average excess returns of the test portfolios in the baseline parameterization. These are the cross-simulation averages of the time-series average excess returns along each simulated sample path. The standard deviation of excess returns reported in the second column is the cross-simulation average of the time-series standard deviation computed along each simulated path. The simulated excess returns of the characteristics-sorted portfolios in Table VI are generally reasonable. For the most part, they are comparable in magnitude to the actual returns reported in Table II (albeit somewhat larger), and there is a value premium. However, as with the actual data over this sample

Table VI. Excess returns in the simulated example economies

Notes: This table reports average excess returns, average standard deviation of excess returns, and alphas—both absolute and relative—from 5,000 simulations of length $T = 675$ months of the model economy in Section 3.1 calibrated to size- and B/M-sorted portfolio returns. Each entry corresponds to the cross-simulation average of the time-series average of each component of excess returns or alphas. Alphas are constructed from the application of the linear approximation to the two parameterizations of the nonlinear model. The baseline parameterization uses the coefficient estimates from Table I to estimate the nonlinear factors. The more nonlinear parameterization provides a slight increase in the nonlinear risk premium parameters. OLS is used in fitting the second-pass cross-sectional regression. True alphas are set to zero in the simulations. These estimates are constructed in the programs Sim01_fss_FM.m and Sim02_fss_FM.m.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Baseline Mean</th>
<th>Baseline Standard deviation</th>
<th>More nonlinear Mean</th>
<th>More nonlinear Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 BM1</td>
<td>0.484</td>
<td>8.298</td>
<td>-0.461</td>
<td>8.301</td>
</tr>
<tr>
<td>S1 BM2</td>
<td>0.741</td>
<td>6.824</td>
<td>0.230</td>
<td>6.826</td>
</tr>
<tr>
<td>S1 BM3</td>
<td>0.950</td>
<td>6.034</td>
<td>0.587</td>
<td>6.036</td>
</tr>
<tr>
<td>S1 BM4</td>
<td>1.175</td>
<td>6.195</td>
<td>1.127</td>
<td>5.719</td>
</tr>
<tr>
<td>S1 BM5</td>
<td>1.175</td>
<td>7.171</td>
<td>1.247</td>
<td>6.196</td>
</tr>
<tr>
<td>S2 BM1</td>
<td>0.590</td>
<td>7.171</td>
<td>-0.207</td>
<td>7.172</td>
</tr>
<tr>
<td>S2 BM2</td>
<td>0.861</td>
<td>5.858</td>
<td>0.503</td>
<td>5.860</td>
</tr>
<tr>
<td>S2 BM3</td>
<td>0.988</td>
<td>5.333</td>
<td>0.833</td>
<td>5.334</td>
</tr>
<tr>
<td>S2 BM4</td>
<td>1.171</td>
<td>5.004</td>
<td>1.279</td>
<td>5.005</td>
</tr>
<tr>
<td>S2 BM5</td>
<td>1.140</td>
<td>5.826</td>
<td>1.837</td>
<td>5.828</td>
</tr>
<tr>
<td>S3 BM1</td>
<td>0.678</td>
<td>6.530</td>
<td>0.119</td>
<td>6.529</td>
</tr>
<tr>
<td>S3 BM2</td>
<td>1.057</td>
<td>5.533</td>
<td>1.009</td>
<td>5.334</td>
</tr>
<tr>
<td>S3 BM3</td>
<td>0.927</td>
<td>4.860</td>
<td>0.795</td>
<td>4.861</td>
</tr>
<tr>
<td>S3 BM4</td>
<td>1.001</td>
<td>4.969</td>
<td>1.220</td>
<td>4.970</td>
</tr>
<tr>
<td>S3 BM5</td>
<td>1.177</td>
<td>5.360</td>
<td>1.770</td>
<td>5.363</td>
</tr>
<tr>
<td>S4 BM1</td>
<td>0.808</td>
<td>5.889</td>
<td>0.285</td>
<td>5.889</td>
</tr>
<tr>
<td>S4 BM2</td>
<td>0.819</td>
<td>5.166</td>
<td>0.525</td>
<td>5.168</td>
</tr>
<tr>
<td>S4 BM3</td>
<td>0.838</td>
<td>4.984</td>
<td>0.584</td>
<td>4.985</td>
</tr>
<tr>
<td>S4 BM4</td>
<td>1.011</td>
<td>4.681</td>
<td>1.081</td>
<td>4.682</td>
</tr>
<tr>
<td>S4 BM5</td>
<td>1.295</td>
<td>5.549</td>
<td>1.777</td>
<td>5.550</td>
</tr>
<tr>
<td>S5 BM1</td>
<td>0.797</td>
<td>4.684</td>
<td>0.222</td>
<td>4.686</td>
</tr>
<tr>
<td>S5 BM2</td>
<td>0.770</td>
<td>4.176</td>
<td>0.625</td>
<td>4.175</td>
</tr>
<tr>
<td>S5 BM3</td>
<td>0.729</td>
<td>3.957</td>
<td>0.623</td>
<td>3.958</td>
</tr>
<tr>
<td>S5 BM4</td>
<td>0.837</td>
<td>4.868</td>
<td>1.200</td>
<td>4.867</td>
</tr>
<tr>
<td>S5 BM5</td>
<td>1.068</td>
<td>5.387</td>
<td>1.442</td>
<td>5.389</td>
</tr>
</tbody>
</table>
period, there is little or no evidence of a consistent size effect. The second pair of columns shows the same simulated moments for the more nonlinear simulation.

Table VII shows the alphas and relative alphas from both parameterizations of the model. By construction, the true alphas are zero in the correctly specified model. Again, these are the cross-sectional averages of the time-series average from each of the 5,000 simulations. In the baseline simulation, the average alphas are largest (in absolute value) for the lowest B/M portfolio in the first two size quintiles only. The largest (in absolute value) alphas are around 0.25% per month (3.0% per year). Relative to average expected returns, the largest alphas are between (roughly) 20% and 50% of average excess returns. The cross-sectional standard deviation of average alphas across the different portfolios ranges between 0.09% and 0.12% per month. Since these are the cross-simulation moments, the mean and standard deviations indicate that a substantial number of the 5,000 simulations generated large alphas. Relative alphas are more variable than alphas, indicating that there are many sample paths where estimated alphas are large relative to average excess returns. For the lowest B/M assets, the cross-sectional standard deviations of the distributions of relative pricing errors are huge.

Is it possible to detect the nonlinear components of risk premiums using the nonlinearity test statistics in Equations (14) and (15), respectively? Before we consider the power of the test against the null of moderate nonlinearity in the risk premiums, it is important to establish the actual size of the different test statistics. In order to do this, we calibrate a fully linear version of the model in Section 3.1, one that conforms to the null of zero coefficients in higher-order risk premium terms. We then simulate 5,000 sample paths of length $T_{sim} = 675$ months and compute the different versions of the test statistic along each sample path. The resulting simulated distributions are then stored for later examination.

Table VIII contains estimates of the actual size of the nonlinearity tests from these simulations. Panel A examines the risk premium test ($k$-test), estimated using risk premium covariances based on both simple time-series estimates, with and without a Shanken correction, and methods that are robust to serial correlation and heteroskedasticity in the measured risk premiums, again, with and without a Shanken correction. Panel B (Panel C) contains the same calculations for the $a$-test (KRS test).

As Panel A of Table VIII shows, using either a simple or a robust estimate of the covariance matrix, the uncorrected tests both have a smaller right tail than the corresponding asymptotic distribution resulting in too few rejections of the true null at either the 5% or 1% nominal levels. At the 10% nominal level, the test statistic based on the simple covariance matrix estimate has an actual size of 1.5%, which declines to 0.1% at the 5% nominal level. So, using the nominal size would reject the true null too often. The $k$-test with robust variance–covariance (VCV) but without the Shanken correction performs slightly better than the version with the simple VCV. The actual sizes of both of the Shanken-corrected risk premium tests appear to be quite distorted relative to the asymptotic

---

21 There are some simulated assets with low levels of average returns; for example, small size and low B/M (see Table VI). In these cases, the denominator of the relative alpha may be close to zero in some simulated return series, and this has the effect of “blowing up” even a relative small estimated alpha. As a result, we do not emphasize the relative alpha results for the six assets with the lowest average excess return in Table VII.

22 The robust covariance matrix estimates use the Newey–West estimator with the asymptotically optimal lag length choice.
The apparent overcorrection in the estimated covariance matrix may follow from the simple iid nature of volatility in the example economy.

In the $t$-tests examined in Panel B of Table VIII, the results are quite different. All of the $t$-tests—whether estimated with or without a robust VCV and with or without a Shanken correction—severely over-reject the true null at all conventional significance levels. Panel C

Table VII. Simulation results for $\alpha$ in the example economies.

Notes: This table reports average excess returns, average standard deviation of excess returns, and alphas—both absolute and relative—from 5,000 simulations of length $T = 675$ months of the model economy in Section 3.1 calibrated to size- and B/M-sorted portfolio returns. Each entry corresponds to the cross-simulation average of the time-series average of each component of excess returns or alphas. The baseline parameterization uses the coefficient estimates from Table I to estimate the nonlinear factors. The more nonlinear parameterization provides a slight increase in the nonlinear risk premium parameters. OLS is used in fitting the second-pass cross-sectional regression. True alphas are zero. Misspecification alphas (based on the true moments of the data) are shown in the first column of Table IV. These estimates are constructed in the programs Sim01_fss_FM.m and Sim02_fss_FM.m.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Baseline Mean</th>
<th>Baseline Standard deviation</th>
<th>Relative Mean</th>
<th>Relative Standard deviation</th>
<th>More nonlinear Mean</th>
<th>More nonlinear Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 BM1</td>
<td>-0.254</td>
<td>0.117</td>
<td>-0.544</td>
<td>0.280</td>
<td>-0.713</td>
<td>0.441</td>
</tr>
<tr>
<td>S1 BM2</td>
<td>-0.128</td>
<td>0.108</td>
<td>-0.177</td>
<td>0.151</td>
<td>-0.393</td>
<td>0.352</td>
</tr>
<tr>
<td>S1 BM3</td>
<td>0.006</td>
<td>0.102</td>
<td>0.005</td>
<td>0.108</td>
<td>-0.184</td>
<td>0.325</td>
</tr>
<tr>
<td>S1 BM4</td>
<td>0.095</td>
<td>0.101</td>
<td>0.087</td>
<td>0.093</td>
<td>0.181</td>
<td>0.308</td>
</tr>
<tr>
<td>S1 BM5</td>
<td>0.145</td>
<td>0.107</td>
<td>0.123</td>
<td>0.090</td>
<td>0.252</td>
<td>0.326</td>
</tr>
<tr>
<td>S2 BM1</td>
<td>-0.208</td>
<td>0.112</td>
<td>-0.361</td>
<td>0.206</td>
<td>-0.631</td>
<td>0.424</td>
</tr>
<tr>
<td>S2 BM2</td>
<td>-0.045</td>
<td>0.102</td>
<td>-0.055</td>
<td>0.120</td>
<td>-0.237</td>
<td>0.323</td>
</tr>
<tr>
<td>S2 BM3</td>
<td>0.042</td>
<td>0.097</td>
<td>0.042</td>
<td>0.098</td>
<td>-0.001</td>
<td>0.299</td>
</tr>
<tr>
<td>S2 BM4</td>
<td>0.179</td>
<td>0.096</td>
<td>0.153</td>
<td>0.081</td>
<td>0.325</td>
<td>0.306</td>
</tr>
<tr>
<td>S2 BM5</td>
<td>0.130</td>
<td>0.105</td>
<td>0.113</td>
<td>0.091</td>
<td>0.665</td>
<td>0.442</td>
</tr>
<tr>
<td>S3 BM1</td>
<td>-0.156</td>
<td>0.108</td>
<td>-0.234</td>
<td>0.167</td>
<td>-0.460</td>
<td>0.374</td>
</tr>
<tr>
<td>S3 BM2</td>
<td>0.098</td>
<td>0.101</td>
<td>0.091</td>
<td>0.095</td>
<td>0.122</td>
<td>0.296</td>
</tr>
<tr>
<td>S3 BM3</td>
<td>0.008</td>
<td>0.093</td>
<td>0.007</td>
<td>0.101</td>
<td>-0.019</td>
<td>0.283</td>
</tr>
<tr>
<td>S3 BM4</td>
<td>0.058</td>
<td>0.095</td>
<td>0.057</td>
<td>0.094</td>
<td>0.291</td>
<td>0.317</td>
</tr>
<tr>
<td>S3 BM5</td>
<td>0.176</td>
<td>0.100</td>
<td>0.149</td>
<td>0.083</td>
<td>0.667</td>
<td>0.418</td>
</tr>
<tr>
<td>S4 BM1</td>
<td>-0.074</td>
<td>0.102</td>
<td>-0.094</td>
<td>0.129</td>
<td>-0.366</td>
<td>0.352</td>
</tr>
<tr>
<td>S4 BM2</td>
<td>-0.067</td>
<td>0.099</td>
<td>-0.084</td>
<td>0.123</td>
<td>-0.208</td>
<td>0.298</td>
</tr>
<tr>
<td>S4 BM3</td>
<td>-0.063</td>
<td>0.095</td>
<td>-0.077</td>
<td>0.114</td>
<td>-0.182</td>
<td>0.287</td>
</tr>
<tr>
<td>S4 BM4</td>
<td>0.065</td>
<td>0.092</td>
<td>0.064</td>
<td>0.090</td>
<td>0.192</td>
<td>0.286</td>
</tr>
<tr>
<td>S4 BM5</td>
<td>0.247</td>
<td>0.105</td>
<td>0.190</td>
<td>0.080</td>
<td>0.640</td>
<td>0.391</td>
</tr>
<tr>
<td>S5 BM1</td>
<td>-0.074</td>
<td>0.096</td>
<td>-0.095</td>
<td>0.122</td>
<td>-0.431</td>
<td>0.360</td>
</tr>
<tr>
<td>S5 BM2</td>
<td>-0.106</td>
<td>0.089</td>
<td>-0.139</td>
<td>0.118</td>
<td>-0.135</td>
<td>0.268</td>
</tr>
<tr>
<td>S5 BM3</td>
<td>-0.134</td>
<td>0.088</td>
<td>-0.185</td>
<td>0.124</td>
<td>-0.126</td>
<td>0.273</td>
</tr>
<tr>
<td>S5 BM4</td>
<td>-0.053</td>
<td>0.094</td>
<td>-0.065</td>
<td>0.113</td>
<td>0.295</td>
<td>0.356</td>
</tr>
<tr>
<td>S5 BM5</td>
<td>0.114</td>
<td>0.098</td>
<td>0.106</td>
<td>0.091</td>
<td>0.455</td>
<td>0.361</td>
</tr>
</tbody>
</table>

distribution. The apparent overcorrection in the estimated covariance matrix may follow from the simple iid nature of volatility in the example economy.

In the $t$-tests examined in Panel B of Table VIII, the results are quite different. All of the $t$-tests—whether estimated with or without a robust VCV and with or without a Shanken correction—severely over-reject the true null at all conventional significance levels. Panel C
of Table VIII shows that the GLS version of the KRS has finite-sample size that is quite consistent with its asymptotic counterpart.

Given our findings in Table VIII, it makes sense to examine the power of only the versions of the tests that have reasonable size properties—which leads us to eliminate the $\lambda$-tests from further analysis. In addition, we do not rely on the critical values for the tests based on the asymptotic $\chi^2$-distribution. Instead, we use the simulated-based critical values tabulated in the construction of Table VIII. We present these results in Table IX. Panel A presents the results from the baseline nonlinear model parameterization, while Panel B shows the power of the tests to distinguish the linear from the more nonlinear parameterization.

The simplest version of the $\lambda$-test, with no attempt at correcting for non-iid homoskedastic errors and no Shanken correction—and the GLS version of the KRS test have comparable power in detecting nonlinear risk premiums in the baseline parameterization, but this power is low—essentially not different from the (empirical) sizes of the tests. The $\lambda$-test

<table>
<thead>
<tr>
<th>Panel A: Risk premium test ($\lambda$-test)</th>
<th>Nominal size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>Simple VCV w/o Shanken correction</td>
<td>0.015</td>
</tr>
<tr>
<td>Robust VCV w/o Shanken correction</td>
<td>0.027</td>
</tr>
<tr>
<td>Simple VCV w/Shanken correction</td>
<td>0</td>
</tr>
<tr>
<td>Robust VCV w/Shanken correction</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Alpha test ($\alpha$-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple VCV w/o Shanken correction</td>
</tr>
<tr>
<td>Robust VCV w/o Shanken correction</td>
</tr>
<tr>
<td>Simple VCV w/Shanken correction</td>
</tr>
<tr>
<td>Robust VCV w/Shanken correction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: KRS-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
</tr>
<tr>
<td>GLS</td>
</tr>
</tbody>
</table>
based on the robust VCV performs slightly worse. In the more nonlinear parameterization of the model, the power of both $\lambda$-tests increases substantially to over 80% at an empirical size of 10%. The GLS version of the KRS test, on the other hand, actually declines in their frequency of identifying the true alternative of nonlinear risk premiums.

In summary, the Monte-Carlo results suggest that it is difficult to detect modest nonlinearity, but the $\lambda$-tests appear to provide reasonable power against more pronounced nonlinearity after correcting the finite-sample sizes of these statistics.

4. Application: Asset Pricing in a Model of Technological Innovation

This example is based on the general equilibrium production economy in Papanikolaou (2011) and its partial equilibrium counterpart in Kogan and Papanikolaou (2014). The model has implications for the role of technological innovation in determining different moments of asset returns, including the value premium. Among other things, it implies that
the pricing kernel depends on innovations to both disembodied (denoted $x_i$) and embodied productivity shocks (denoted $\xi$) attached to specific vintages of capital. The model is formulated in continuous-time and the SDF evolves as

$$\frac{d\pi_t}{\pi_t} = -r_{f,t} dt - \gamma_x(\omega_t) dB_t^x - \gamma_\xi(\omega_t) dB_t^\xi,$$  

(32)

where $r_{f,t}$ is the instantaneous risk-free rate, $dB_t^x$ is the innovation to disembodied (labor-augmenting) productivity shock, and $dB_t^\xi$ is the innovation to the embodied technology shock.

Given Equation (32), the instantaneous excess return to any asset $i$ is

$$E_t \left[ \frac{dS_{i,t} + D_{i,t} dt}{S_{i,t}} - r_{f,t} dt \right] = \gamma_x(\omega_t) \text{cov}_t \left( \frac{dS_{i,t}}{S_{i,t}}, dB_t^x \right) + \gamma_\xi(\omega_t) \text{cov}_t \left( \frac{dS_{i,t}}{S_{i,t}}, dB_t^\xi \right),$$

(33)

where $S_{i,t}$ is the ex-dividend price of asset $i$ and $D_{i,t}$ is the instantaneous dividend payout rate for asset $i$. This has the standard interpretation of decomposing the excess return into prices of risk [$\gamma_x(\omega_t)$ and $\gamma_\xi(\omega_t)$] and measures of risk, defined as the (conditional) covariances of instantaneous ex-dividend returns with the innovations to the two shocks.\(^{24}\) The state variable determining (in part) the market prices of risk is $\omega_t$, a linear function of the two technology shocks and the log-level of physical capital in the economy. In the general equilibrium model of Kogan, Papanikolaou, and Stoffman (2017), $\omega_t$ “summarizes all relevant information for the model’s non-linear dynamics.”\(^{25}\) The interesting point for our purposes is the conditional—and possibly nonlinear—nature of the model-implied risk premiums.

We will actually estimate the simpler, but qualitatively similar, model from Papanikolaou (2011) where the pricing equation analogous to Equation (33) is

$$E_t \left[ \frac{dS_{i,t} + D_{i,t} dt}{S_{i,t}} - r_{f,t} dt \right] = \gamma_x \text{cov}_t \left( \frac{dS_{i,t}}{S_{i,t}}, dB_t^x \right) + \gamma_\xi(\omega_t) \text{cov}_t \left( \frac{dS_{i,t}}{S_{i,t}}, dB_t^\xi \right),$$

(34)

where $\gamma_x$ is the constant price of market-wide risk affecting all firms and $\gamma_\xi(\omega_t)$ is the non-linear price of embodied (investment-specific) technology shocks. In this case, the state variable, $\omega_t$, is correlated with the economy’s log investment-to-consumption ratio; so, following Kogan and Papanikolaou (2014), we use this as an empirical proxy in our conditional estimation.\(^{26}\) The embodied productivity shock is proxied, as in Papanikolaou (2011), by using a return spread on investment versus consumption sector portfolios.

\(^{24}\) In the partial equilibrium model of Kogan and Papanikolaou (2014), Equation (32) is simplified by assuming that the exogenous SDF has constant measures of risk $\gamma_x$ and $\gamma_\xi$. Kogan, Papanikolaou, and Stoffman (2017) embed technological innovation in a general equilibrium, incomplete markets model. In this setting, of course, the SDF is not unique, but Kogan, Papanikolaou, and Stoffman (2017) use a projection of individual agents’ SDFs onto the information filtration to select a specific SDF. The pricing kernel in this setting cannot be expressed in closed-form (see their Proposition 1 and the related section in their Appendix), but it is nonlinear and conditional on both the stationary component of the technology shocks and on moments of the wealth distribution of households in the economy.


\(^{26}\) Kogan and Papanikolaou (2014) use the log ratio of non-residential private investment to consumption of nondurables plus services.
Essentially, this portfolio serves as the projection of the non-marketed factor onto the space of marketed returns.\textsuperscript{27}

The “beta” version of the extended security market line in Equation (34) is

\[ E_t[R_{i,t+1}] = \lambda_{0,t} + \lambda_1 \beta_{i,t}^c + \lambda_2, \]

where \( R_{i,t+1} \) is the excess return to one of \( N \)-test assets measured from \( t \) to \( t+1 \), \( \lambda_{0,t} \) is a (possibly) time-varying intercept (reflecting possible measurement error in the risk-free rate proxy), \( \lambda_1 \) is a scalar (constant) risk premium on market-wide risk, \( \lambda_2,t \) is the conditional risk premium on embodied technological change, and \( \beta_{i,t}^c \) for \( j \in \{ x, \xi \} \) is the conditional beta of asset \( i \) with respect to a source of risk.

If we apply the standard approximations from Section 2.3 and Appendix 5 to estimate the conditional model, then we get the following specification:

\[ \lambda_{0,t} = a_0 + a_1 \Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \]

\[ \lambda_1 = b_0, \]

\[ \lambda_2,t = c_0 + c_1 \Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \]

where \( \Delta \ln \left( \frac{C_t}{C_{t+1}} \right) \), and by the law of iterated expectations, the expanded unconditional pricing model is

\[ E[R_{i,t+1}] = \lambda_0 + \lambda_1 \beta_i. \]

\( \beta_i \) are the betas of asset \( i \) with respect to the expanded set of factors formed from interacting the constant and the consumption growth factors in Equation (35) with the conditioning variable \( \Delta \ln \left( \frac{C_t}{C_{t+1}} \right) \). These factors can be denoted as

\[ F_{TC}^{\text{TC}} (5 \times 1) = (1, \Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \Delta \ln \left( \frac{C_t}{C_{t+1}} \right))^t, \]

where the change in the aggregate consumption growth, \( \Delta \ln \left( \frac{C_t}{C_{t+1}} \right) \), is the proxy for the market-wide risk, \( \Delta \ln \left( \frac{C_t}{C_{t+1}} \right) \), is a proxy for technological innovation based on the spread in the equity returns between investment and consumption sector firms.\textsuperscript{29} In constructing Equation (38), we impose the constraint that the market-wide risk premium is constant.

There is no simple connection between the true conditional risk premium, \( \lambda_2,t \), and the (relevant) coefficients \( \lambda \) in Equation (37). The computation of \( \lambda_2,t \) from \( \lambda \) requires knowledge of the conditional covariance of the expanded factors.\textsuperscript{30} In order to illustrate the impact of potential nonlinearities in risk premiums, we consider two additional sets of factors:

Quadratic:

\[ F_{TC}^{\text{TC}} (4 \times 1) = (P_2(\Delta \ln \left( \frac{C_t}{C_{t+1}} \right), P_2(\Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \Delta \ln \left( \frac{C_t}{C_{t+1}} \right))^t, \]

Cubic:

\[ F_{TC}^{\text{TC}} (4 \times 1) = (P_2(\Delta \ln \left( \frac{C_t}{C_{t+1}} \right), P_2(\Delta \ln \left( \frac{C_t}{C_{t+1}} \right), P_3(\Delta \ln \left( \frac{C_t}{C_{t+1}} \right), P_3(\Delta \ln \left( \frac{C_t}{C_{t+1}} \right), \Delta \ln \left( \frac{C_t}{C_{t+1}} \right))^t, \]

\textsuperscript{27} See Appendix C for a description of the data construction.

\textsuperscript{28} Equation (34) also suggests that factor loadings are time-varying. We would want to account for this in a general estimation of the technological innovations model. However, in the interest of focusing on the importance of nonlinear conditional risk premiums, we have assumed constant betas in our conditional model.

\textsuperscript{29} See Appendix C.

\textsuperscript{30} See, for example, the discussion in Lettau and Ludvigson (2001).
where \( P_2(\Delta ic_t) \) is the component of \( \Delta ic_t^2 \) that is orthogonal to \( \Delta ic_t \), normalized to have zero mean and the same unconditional volatility as \( \Delta ic_t \) and \( P_3(\Delta ic_t) \) is defined analogously for \( \Delta ic_t^3 \).

The results of estimating the technological change model, using both \( F^TC \) and an extended model using \( (F^TC, F^{3TC})' \) and \( (F^TC, F^{TC})' \), are shown in Table X. The data consist of \( T = 275 \) quarters of data, from the second quarter of 1948 to the fourth quarter of 2016.\(^{31}\) The test assets consist of the twenty-five size- and B/M-sorted portfolios that have become the standard in empirical work.

In addition to the point estimates of the risk premium coefficients, we report two sets of “\( t \)-statistics” formed from the ratio of the estimated coefficient and the coefficient’s standard error based on a Fama–MacBeth procedure using a covariance matrix estimator that is robust to deviations from the iid assumption. “\( t \)-stat1” is the ratio formed from using a robust covariance matrix estimator alone, while “\( t \)-stat2” adds the additional terms due to Shanken’s correction for measurement error in the first-pass beta estimates.

The results in the three columns labeled “Linear” in Panel A of Table X show that: (i) the point estimate of the constant risk premium on the disembodied technical change, \( x_t \), is negative but not statistically significant; (ii) the linear risk premium, \( \Lambda_{2,t} \) in Equation (36), has a point estimate for the intercept term that is negative, but the point estimate of \( c_1 \) (of 0.02) indicates that the risk premium is increasing in the level of the embodied technical change, \( \zeta_t \); (iii) the statistical significance of these linear risk premium coefficients (using the asymptotic distribution) is dependent on whether or not the Shanken correction is used in estimating standard errors. When the correction is made, both coefficients appear to be highly significant (again, according to the asymptotic distribution).

The next three columns in Panel A of Table X, labeled “Quadratic,” add squared terms to the estimate of \( \Lambda_{2,t} \), the conditional time-varying risk premium.\(^{32}\) There seems to be some support, in the coefficient estimates, for the importance of the quadratic terms in describing \( \Lambda_{2,t} \). In particular, the point estimates of the quadratic coefficients are both negative and large relative to the magnitude of the linear term coefficients.\(^{33}\) The statistical evidence appears to be stronger for the quadratic terms in that the \( t \)-statistics on the coefficients are significant (or borderline significant) whether or not the Shanken correction is used. Finally, the point estimates (and \( t \)-statistics) in the final three (Cubic) columns show no evidence of the importance of third-order terms in \( \Lambda_{2,t} \).

In addition to the coefficient estimates in Panel A, we also compute the \( \lambda \)-tests with the robust covariance estimate for—separately—quadratic and cubic terms in the risk premium function both with and without the Shanken correction in Panel B of Table X.\(^{34}\) These results largely confirm the findings from the individual coefficient estimates, but they highlight the importance of using empirical, as opposed to asymptotic, \( p \)-values in assessing

\(^{31}\) Given the size of our cross-section of test assets, we need as long a time series of returns and factors as possible; see the simulation results in Ferson and Forster (1994).

\(^{32}\) Notice that the coefficients on the linear terms do change slightly with the addition of the quadratic regressors. This happens because, while the quadratic terms have been orthogonalized with respect to the lower order terms of \( \Delta ic_t \), they have not been orthogonalized with respect to the other regressors.

\(^{33}\) Since the quadratic term has been rescaled to have the same size as the linear terms, this finding does not simply reflect the relative sizes of linear and quadratic terms.

\(^{34}\) We use these statistics because of their size and power properties from Tables VIII and IX.
These statistics. If we examine the tests for the quadratic risk premium terms, we see that both statistics are highly significant using the asymptotic $\chi^2_2$ distribution. However, when we construct an estimate of the empirical distributions of these tests using a block bootstrap algorithm that accommodates dependence in the Fama–MacBeth residuals, the results change significantly and neither quadratic risk premium test statistic is significant at conventional levels. The $k$-tests for the cubic terms are weaker using both the asymptotic and empirical significance levels, consistent with the findings in Part A of Table X.

As a final look at the model performance, we plot the realized versus the predicted (or expected) returns in both the linear and nonlinear versions of the model. These results are
There does appear to be a noticeably smaller spread of pricing errors in moving from the linear to the quadratic model. For example, in the linear specification there are four (of the twenty-five) test assets whose average alpha is greater than 6% per year, while there is only one average alpha above this threshold in the quadratic risk premium model. However, there is no obvious improvement in moving from the quadratic to the cubic risk premium.

5. Conclusions

If conditional risk premiums are nonlinear, what happens when a standard linear approximation is used to estimate the model? This question is important because theoretical models rarely provide guidance to an econometrician on how to handle the unobservable dynamics of conditioning information. We have shown, under a general approximation to a nonlinear model, that using a linear approximation introduces an omitted variables bias associated with the higher-order terms in the nonlinear approximation. We derived the forms of the pricing error bias, and we examined alternative forms of specification tests for nonlinearity. While the tests that we examine are asymptotically equivalent, their finite-sample properties may be very different.

We conduct a Monte-Carlo study to quantify the possible importance of moderate nonlinearity on estimation results based on the assumption of linear conditioning. Our findings suggest that, despite the general noise in estimating expected returns, misspecifying the form of the conditional risk premium can contribute significantly to measured pricing errors. The Monte-Carlo results also provide guidance as to the form of the nonlinearity.

This observation is actually made, not from Figure 3 directly, but from the tabular version of the same data constructed in the program TCmodelest.m.
test that is most likely to be useful in practice: a direct test of nonlinear risk premiums using an explicit correction for errors-in-variables in estimating the factor loadings in the first-pass time-series regression. We believe that these results are representative of an interesting class of practical problems, but they do not demonstrate the shape of the power function against arbitrary alternative hypotheses to linearity. Rather, they suggest the importance of testing for nonlinearity in specific settings using the finite-sample properties of alternative tests.

Finally, we examine the implications of misspecifying the risk premium in an example of a recent structural asset pricing. Our empirical analysis of the embodied technological change model of Papanikolaou (2011) demonstrates the importance of allowing for nonlinearity in evaluating the model’s performance.

Appendix A: The “Standard” Testing Approach

The standard testing approach consists of four steps:

**Step 1:** Start with the fundamental asset pricing equation:

\[ E(M_{t+1}R_{t+1}^e | \mathcal{F}_t) = 0, \]  

(A1)

where \( R_{t+1}^e \) is an \( N \)-vector of marketed asset returns from \( t \) to \( t+1 \) in excess of a risk-free return proxy, \( 0 \) is an \( N \)-vector of ones, and \( \mathcal{F}_t \) denotes the information used by the market in determining prices at time \( t \).

**Step 2:** Use a linear structure for the asset pricing kernel:

\[ M_{t+1} = 1 + f_{t+1} b_t, \]  

(A2)

as either an exact representation of the model being tested (including possibly a log transformation) or as an approximation to the true model. The use of excess returns results in normalizing the conditional mean of the pricing kernel to 1. \( M_{t+1} \) is a conditional pricing model if the vector of pricing kernel coefficients, \( b_t \), is time-varying with the information in \( \mathcal{F}_t \).

The essence of the linear conditioning approximation is contained in the following step:

**Step 3:** Reduce the information set, \( \mathcal{F}_t \), to a \( K \)-vector of observable variables, \( Z_t \), and write the pricing kernel coefficient, \( b_t \), as

\[ b_t = B Z_t, \]  

(A3)

where \( B \) is a \( K \times L \) matrix of constant coefficients.\(^{36}\)

It is important to note that the linearity assumption in Equation (A.3) is distinct from the linear form of the pricing kernel. There is often no obvious reason to impose linearity in the true model state variables—much less linearity in the proxies \( Z \)—on the pricing kernel coefficients. The use of the linearity assumption is a direct result of the econometrician’s inability to specify a complete structure for the dynamics of investors’ information sets.

The implications of the fundamental asset pricing Equation (A.1) have now been reduced to a set of linear moments that can be estimated using either the GMM directly or the

---

36 If the first element of \( Z \) is 1, then there is an intercept in the pricing kernel. As Cochrane (2005) notes, if \( Z_t \) contains all measurable transformations of all elements of the random vector whose minimal \( \sigma \)-algebra defines \( \mathcal{F}_t \), then this is not an assumption but rather a mathematical fact. The assumption is that a specific finite set of \( Z \) is an adequate approximation to this larger set.
two-pass estimation (beta pricing) approach that evolved for linear pricing models prior to the explicit introduction of GMM estimators. As Cochrane (2001) and Jagannathan and Wang (2002) proved, these two approaches are mathematically identical as long as the GMM moment conditions are chosen appropriately.

**Step 4:** Construct the (standard) test statistics used to evaluate linear pricing models. The first step in this estimation scheme is a multivariate time-series regression of excess returns on the model factors:

$$R_{t+1}^e = a_0 + BF_{t+1} + \eta_{t+1}, \quad (A4)$$

for \(t = 1, \ldots, T\), where \(F_{t+1} \equiv (f_{t+1} \otimes Z_t)\), \(B\) is the \(N \times KL\) matrix of regression coefficients (betas), and \(\eta_{t+1}\) is an \(N\)-vector of residual returns. The second step in the estimation is a cross-sectional regression consistent with the following unconditional moments for excess returns:

$$E(R_{t+1}^e) = B\lambda + \alpha, \quad (A5)$$

where \(\lambda\) is an \(LK\)-vector of constant risk premiums and \(\alpha\) is a vector of pricing errors. The test statistic for evaluating the joint significance of all pricing errors is

$$T\hat{\lambda}'\hat{\Sigma}_\lambda^{-1}\hat{\lambda} \sim \chi^2_{KL}, \quad (A6)$$

where \(\hat{\Sigma}_\lambda\) is the covariance matrix for the risk premiums.

Individual pricing errors and individual risk premiums can be tested using the square root of the elements on the main diagonal of \(\hat{\Sigma}_\lambda\) or \(\hat{\Sigma},\) respectively. These models are also commonly evaluated heuristically by examining the size (and structure) of individual pricing errors and by computing average cross-sectional \(R^2\) statistics.

**Appendix B: Covariance Matrices**

The covariance matrix for \(\alpha\) is defined as:

$$\Sigma_\alpha = T^{-1}[\hat{\Sigma}_\eta - \hat{B}(\hat{B}'\hat{\Sigma}_\eta^{-1}\hat{B})^{-1}\hat{B}'](1 + \hat{\lambda}'\hat{\Sigma}_F^{-1}\hat{\lambda}), \quad (A8)$$

where \(\hat{\Sigma}_\eta\) is the sample estimate of the covariance matrix for \(\eta,\) \(\hat{\Sigma}_F\) is the sample estimate of the covariance matrix of the factors \(F.\) The term in parentheses is a correction factor due to Shanken (1992) that accounts for the fact that the cross-sectional regressors are generated from the first-pass time-series estimation.

If OLS is used in the cross-sectional regression, then there are two commonly used versions of the covariance matrix of the risk premiums. The first estimate is

$$\hat{\Sigma}_\lambda = T^{-1}[(\hat{B}'\hat{B})^{-1}\hat{B}\hat{\Sigma}_\eta\hat{B}'(\hat{B}'\hat{B})^{-1} + \hat{\Sigma}_F], \quad (A9)$$

where \(\hat{\Sigma}_\eta\) is the estimated covariance matrix of the time-series shocks and \(\hat{\Sigma}_F\) is the estimated factor covariance matrix. The second estimate uses the Shanken correction

$$\hat{\Sigma}_\lambda = T^{-1}[(\hat{B}'\hat{B})^{-1}\hat{B}'\hat{\Sigma}_\eta\hat{B}(\hat{B}'\hat{B})^{-1}(1 + \hat{\lambda}'\hat{\Sigma}_F^{-1}\hat{\lambda}) + \hat{\Sigma}_F]. \quad (A10)$$
There are analogs to Equations \((A.9)\) and \((A.10)\) when GLS is used in the cross-sectional regression. They are

\[
\widehat{\Sigma}_k = T^{-1}[(\mathbf{B}^\prime \Sigma^{-1}_\eta \mathbf{B})^{-1} + \widehat{\Sigma}_f],
\]

(A11)

and

\[
\widehat{\Sigma}_k = T^{-1}[(\mathbf{B}^\prime \Sigma^{-1}_\eta \mathbf{B})^{-1}(1 + \hat{\lambda}^\prime \Sigma^{-1}_f \hat{\lambda}) + \widehat{\Sigma}_f].
\]

(A12)

### Appendix C: The Investment Shock Proxy

Papanikolaou (2011) [Equations (29) and (3) and the discussion in Section V] and Kogan and Papanikolaou (2014) [Equation (26) and Section 3.1] establish the correspondence in the model between the embodied investment-specific technology shock and the return spread between investment and consumption sector portfolios. Essentially, this portfolio serves as the projection of the non-marketed factor onto the space of marketed returns.

In order to estimate a conditional model using a conventional cross-section of test assets, we need a longer time series of the spread returns than those used in Papanikolaou (2011) and Kogan and Papanikolaou (2014).\(^{37}\) We use the ten industry return series available from Ken French’s website (at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The investment sector return is constructed as the equally weighted average of the value-weighted returns to the manufacturing, energy, hi-tech, and telecom industries. The consumption sector returns are constructed as the equally weighted average of the value-weighted returns to the consumer durables, consumer nondurables, and shops industries.\(^{38}\)

A comparison of the constructed return differences, at the annual frequency, between our spread returns and the spread returns used in Papanikolaou (2011) and Kogan and Papanikolaou (2014) over the period from 1964 to 2008 (matching the earlier papers) is shows that the two series move together with an unconditional correlation coefficient of 0.59. Monthly continuously compounded returns are converted to quarterly by multiplying one plus the monthly returns within the quarter.

### References


---

37 See Ferson and Forster (1994).

38 See French’s website for the mapping between the industry definitions and SIC codes.
Brennan, M. J., Wang, A. W., and Xia, Y. (2004): Estimation and test of a simple model of inter-
Kan, R., Robotti, C., and Shanken, J. (2013): Pricing model performance and the two-pass cross-


