An Information-Theoretic Approach to Estimating Risk Premia

By

Maziar M. Kazemi

B.A. Economics and Mathematics
Vassar College, 2013

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2018

Signature of Author: Signature redacted
Department of Management
May 1, 2018

Certified by: Signature redacted
Hui Chen
Associate Professor of Finance
Thesis Supervisor

Accepted by: Signature redacted
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Chair, MIT Sloan PhD Program

An Information-Theoretic Approach to Estimating Risk Premia

Maziar M. Kazemi*

Submitted to the Department of Management

on May 1, 2018 in Partial Fulfillment of the

Requirements for the Degree of Master of Science in

Management Research

Abstract

Evaluation of linear factor models in asset pricing requires estimation of two unknown quantities: the factor loadings and the factor risk premia. Using relative entropy minimization, this paper estimates factor risk premia with only no-arbitrage economic assumptions and without needing to estimate the factor loadings. The method proposed here is particularly useful when the factor model suffers from omitted variable bias, rendering classic Fama-MacBeth/GMM estimation infeasible. Asymptotics are derived and simulation exercises show that the accuracy of the method is comparable to, and frequently is higher than, leading techniques, even those designed explicitly to deal with omitted variables. Empirically, we find estimates of risk premia that are closer to those expected by financial economic theory, relative to estimates from classical estimation techniques. For example, we find that the risk premia on size, book-to-market, and momentum sorted portfolios are very close to the observed average excess returns of these portfolios. An exciting application of our methodology is to performance evaluation for active fund managers. We show that we are able to estimate a manager's "alpha" without specifying the manager's factor exposures.

*kazemi@mit.edu. This paper has greatly benefited from conversations with the following people: Hui Chen, Stefano Giglio, Daniel Greenwald, Peter Hansen, Leonid Kogan, Anna Mikusheva, Jun Pan, Adrien Verdelhan. All remaining errors are my own.

1 Introduction

Since the foundational paper of Ross (1976) developing the arbitrage pricing theory (APT), statistical linear factor models have been used by many researchers to model asset returns. Evaluation of linear factor models of this kind require estimation of two unknown quantities: the factor loadings and the factor risk premia. Traditionally, factor risk premia for non-traded assets[1] have been estimated by regressions of average excess returns on factor loadings. The coefficients in this framework correspond to the factor risk premia. It can also be shown (Section (2)) that the factor risk premia, λ, equal $-R_f\mathbb{E}[mf] = -\mathbb{E}^Q[f]$, where m is a stochastic discount factor (SDF), f is the vector of factors, R_f is the risk-free return, and $\mathbb{E}^Q[\cdot]$ refers to expectation taken under the risk-neutral measure. We will use the preceding moment condition to estimate factor risk premia.

This paper proposes using techniques from information theory to derive a semi-parametric estimate of the SDF to directly evaluate the moment condition defining factor risk premia. Importantly, our method does not require extra economic assumptions aside from those already implied by APT. That is, we only use the no arbitrage restrictions $\mathbb{E}[R_i m] = R_f$, where R_i is the return on asset i and $-R_f\mathbb{E}[fm] = \lambda$, to estimate the SDF. The change of measure from the natural to the risk-neutral probabilities is the solution of a constrained minimization problem, where each constraint corresponds to an asset/Euler equation. Thus, our estimate of the change of measure is semi-parametric: It is a function of the estimated Lagrange multipliers. The objective function is the Kullback-Leibler Information Criterion (KLIC) between the risk-neutral and natural probabilities.

Rosenberg and Engle (2002), Jackwerth and Rubinstein (1996), and Ait-Sahalia and Lo (1998) and Hutchinson, et al (1994) also uncover semiparametric/nonparametric stochastic discount factors from asset returns, but their methods require option data (the latter paper does not explicitly uncover the SDF, but it does provide a nonparametric pricing method for derivatives). The method proposed here does not require that but can easily incorporate

[1]If the factor is in the asset span then the factors expected excess return is its risk premium. See Cochrane (2005).

options into the Euler equations.

The intuition of our technique can be explained as follows. Imagine starting from some set of natural, or "true," probabilities, $\pi_1, ..., \pi_T$, over T states. Now, suppose one observes N assets for which the no arbitrage moment conditions must hold. Unless we are in a risk-neutral world, it will not be the case that $\mathbb{E}[R_i] = R_f$. The technique proposed in this paper changes each of the π_i to a π_i^Q such that $\mathbb{E}^Q[R_i] = R_f$, where $\mathbb{E}^Q[\cdot]$ means expectation under the π_i^Q. The changes in these π_i are the "smallest" necessary to still satisfy the no arbitrage constraints, and allow us to price securities as if investors are risk-neutral. That is, we make sure to incorporate no more information than that enshrined in the no arbitrage constraints. The SDF which solves this problem can be interpreted as a (locally) minimum variance positive SDF (Section (3.1)). We use the empirical likelihood as our natural probability measure, which is justified under Markov chain assumptions.

Stutzer (1995) first proposed using the change of measure implied by minimizing the KLIC subject to the Euler equations. He focuses on what he calls "information bounds," analogous to Hansen-Jagannathan (Hansen and Jagannathan, 1991) bounds. Stutzer (1996) prices options using similar methods we use here to uncover factor risk premia. Kitamura and Stutzer (2002) review information theory and econometrics, which includes the mathematical and statistical theory used here. Ghosh, et al (2016b) use the entropy based method to compare popular structural SDFs by measuring how "far" they are from ensuring the Euler equations hold. Similarly, Backus, et al (2014) use entropy to characterize the dispersion in the SDF. Ghosh, et al (2016a) use the entropy based SDF to try to explain the cross-section of expected returns. Importantly, Ghosh, et al (2016a) use the estimated SDF in single-factor model, which differs from our goal since we use the estimated SDF to evaluate the pricing implications of *any* factor model.

It is informative to consider the current method of estimating risk premia in a factor model and to determine why this method may need to be revised, as we have suggested. The standard method is to use "two-pass" or "Fama-MacBeth" (Fama and MacBeth (1973)) regressions (see Shanken (1996) for a survey).[2] These regressions exploit the time-series relationship between each asset's returns and the factors and a cross-sectional relationship between expected returns and factor risk premia. We will emphasize a key issue in factor models, omitted variable bias, which renders Fama-MacBeth infeasible, that our estimator can overcome. The omitted variables should appear in both the time-series and the cross-sectional regressions. This means that the estimates of the factor loadings and risk premia will be biased and inconsistent.

The proliferation of factor models in the literature speaks to the likelihood that many of these models are misspecified, missing relevant factors, or simply data-mined (Harvey, et al (2015)). McClean and Pontiff (2016) note that many factors found to have cross-sectional pricing ability fare much worse once the academic studies that discovered them are published. Lewellen, et al (2010) note that it is not difficult for an arbitrary set of factors to seemingly explain the cross-section of returns. Even without misspecification, Shanken (1992) notes the need for correcting the errors-in-variables problem inherent in Fama-MacBeth.

This paper is related to the literature on identifying priced factors with model misspecification. Pukthuanthong and Roll (2014) propose an empirical sorting technique to identify truly priced factors, as do Harvey and Liu (2015). Bryzgalova (2016) proposes a LASSO type penalty to avoid inclusion of spurious factors. Kleibergen and Zhan (2014) show that when macroeconomic factors are projected onto the asset span, these "mimicking portfolios" may lead to spuriously strong inference. Kleibergen (2009) shows that when the factor loadings are small or the number of test assets is large that Fama-MacBeth test statistics will be misleading. Kan and Zhang (1999a,b) show that both GMM and Fama-MacBeth tend to overreject the null that a factor is useless, and interestingly, find this problem to be exacerbated when T is large. Kan, et al (2013) develop robust to misspecification inference methods for the cross-sectional regression stage of Fama-MacBeth. These papers do not deal with omitted variable bias explicitly nor do they exploit the factor premia defining moment condition, as we do.

[2] Though GMM estimates equations jointly, it is susceptible to the same issues as Fama-MacBeth, which we will detail.

Giglio and Xiu (2017) develop a method of three-pass regression to deal with the omitted variable problem, and the aim of their paper is the most similar to ours. They propose to relate observed factors to a set of spanning "true" factors through a linear relationship. This relationship essentially adds a third regression, moving from "two-pass" (Fama-MacBeth) to "three-pass." Their methodology is thus different than ours. In particular, they exploit the "blessings of dimensionality" and show that their estimates converge as $N, T \to \infty$. Our method only needs large T to converge and shows high accuracy (as measured by root mean squared error) even when N is very small. As will be seen in the empirical section of this paper, a well-chosen set of test assets can substitute, sometimes, for a large cross-section.

We run simulations for both large T and small N, and large T and large N. We find that our estimator delivers consistenly low root mean squared errors in both experiments. Strikingly, even when we model all factors as being observed, our estimation method continues to perform as well or better than three-pass and Fama-MacBeth.[3]

We develop asymptotics for our estimator of the risk premia, allowing us to test the significance of popular factors used in the literature. We present two derivations of the asymptotics. The first uses the framework of Kitamura and Stutzer (1997). The second uses a two-step procedure, first estimating a set of Lagrange multipliers and next estimating the moment condition defining factor risk premia using the delta method. The former method is the main one we use in the body of this paper. However, the two-step estimates of the risk premia are easy to compute and work well as initial guesses for one-step procedure, which is a numerical optimization problem.

We estimate the risk-premia of the factors in the popular three and five factor models of Fama and French (1993, 2015) and the four factor model in Carhart (1997), which, relative to the three factor model, includes a momentum factor. We also estimate the risk premium of the intermediary leverage factor from Adrian, et al (2014), as an example of a model with unspanned factors.

We find a statistically significant and positive risk premium for the excess return of the

[3]Shanken and Zhou (2007) undertake a detailed simulation study of beta pricing models.

market. The estimated risk premia for book-to-market, size, and momentum are all positive and significant and are close to their time series averages. These results are not true for Fama-MacBeth and three-pass regressions. Similar to Ramponi, et al (2016) we also find that the profitability and investment factors of Fama and French (2015) are not priced or are at best weakly significant. We estimate a positive and statistically significant risk premium for intermediary leverage.

An interesting application of our methodology is to fund manager performance evaluation.[4] We show that we are able to estimate the alpha of a fund manager without specifying the factor exposures of the manager. The intuition is that once we have a measure of the SDF, knowledge of the factors does not alter the pricing implications we would draw. This is a very useful fact since a key issue in performance evaluation is that an active fund manager may manage his risk in such a way so as to load on unobservable factors. Thus, traditional methods of performance evaluation incorrectly attribute to alpha this manager.

The rest of the paper is organized as follows. Section 2 reviews factor models in asset pricing and omitted variable bias and describes Fama-MacBeth and three-pass regressions. Section 3 develops the theory and estimation techniques used in this paper. Section 4 derives the asymptotic distribution of our estimator. Section 5 presents simulation results. Section 6 presents empirical results. Section 7 discusses our application to fund management. Section 8 concludes.

2 Factor Model Background

Factor models for asset returns take the form:

$$R_{t+1} = \mu + \beta' f_{t+1} + \varepsilon_{t+1} \tag{1}$$

[4]This literature is vast. See Ferson (2013) for a review.

where R is an asset's return, μ is its mean, f is a vector of (demeaned) factors, β is a constant vector of factor loadings, and ε is idiosyncratic risk.[5] The idiosyncratic risk is typically taken to be uncorrelated across factors, though some correlation is allowed under boundedness conditions. Note, importantly, that the factors do not need to be portfolios or functions of the traded assets. They may be economic factors like GDP growth, or they may be purely statistical in nature, such as the factors one would estimate via Principal Component Analysis (PCA).

Let m_t be any stochastic discount factor (SDF) (i.e., a time t measurable, positive random variable, such that $1 = \mathbb{E}[m_{t+1}R_{t+1}]$). Multiply Eq (1) by m_{t+1} and take the expectation:

$$1 = \mu\mathbb{E}[m_{t+1}] + \beta'\mathbb{E}[m_{t+1}f_{t+1}] + \mathbb{E}[m_{t+1}\varepsilon_{t+1}].$$

Noting that $\mathbb{E}[m_{t+1}] = 1/R_f$, where R_f is a risk-free return, and rearranging, we have:

$$\mu - R_f = \beta\lambda. \tag{2}$$

Here, $\lambda = -R_f\mathbb{E}[m_{t+1}f_{t+1}]$ is the factor risk premia. Importantly, the term $\mathbb{E}[m_{t+1}\varepsilon_{t+1}] = 0$. The Arbitrage Pricing Theory (APT) of Ross (1976) uses a diversification argument to show that idiosyncratic risk should receive 0 price.

The typical method of estimating λ is via a two-pass regression (Fama and MacBeth, 1973) or with the Generalized Method of Moments (GMM, Hansen, 1982). In either case, the estimation involves both equations (1) and (2). For example, in a two-pass regression, one runs time-series regression (1) n times, where n is the number of assets in one's sample. This generates a $\widehat{\beta}_i$, $i = 1, ..., n$, for each asset. Next, equation (2) is estimated cross-sectionally using these $\widehat{\beta}_i$'s as regressors and average excess returns as regressands. The estimates from this cross-sectional regression, $\widehat{\lambda}$, are the estimated factor risk premia.[6]

[5]The model can be written with modification and additional assumptions to include conditional means and time-varying loadings.

[6]Use of GMM allows one to bypass extra steps needed to correct the standard errors of the $\widehat{\lambda}$ estimate in a two-pass regression. The issue arises because the $\widehat{\beta}_i$'s are themselves estimates of true values. GMM treats both equations symmetrically, so one can simply "read off" the correct standard error from the variance-

2.1 Omitted Variables in Factor Models

The preceding discussion has taken the factors, f, as given. However, the choice of factors is not obvious. Given this uncertainty, it is likely that many proposed factor models are incorrect and have omitted variables. Aside from the usual effects of omitted variables in regression models (e.g., biased estimates), having omitted variables in the factor model, equation (1), affects equation (2) and hence the estimate for the factor risk premia.

To see this, let us assume we estimate a one factor model, while in reality there is an omitted second factor.[7]

$$R_t = \mu + \beta_1 f_{1,t} + \varepsilon_t, \quad \text{where } \varepsilon_t = \beta_2 f_{2,t} + e_t. \tag{3}$$

The Ordinary Least Squares (OLS) estimate of β_1 is:

$$\widehat{\beta}_1 = \frac{\mathrm{Cov}(f_{1,t}, R_t)}{\mathrm{Var}(f_{1,t})} = \frac{\sum_t f_{1,t}(\mu + \beta_1 f_{1,t} + \beta_2 f_{2,t} + e_t)}{\sum_t f_{1,t}^2} = \beta_1 + \beta_2 \frac{\sum_t f_{1,t} f_{2,t}}{\sum_t f_{1,t}^2} + \frac{\sum_t f_{1,t} e_t}{\sum_t f_{1,t}^2}.$$

It is the second term on the right that causes problems. In the absence of omitted variable bias, that term would not be there, and $\widehat{\beta}_1$ would be unbiased. However, when we estimate Eq.(2) using the above $\widehat{\beta}_1$ we are introducing the omitted variable bias in two ways: First, by using an inconsistent and biased estimate of β_1, and second, by having omitted variables in Eq.(2), too. That is, β_2 is missing as a regressor in the cross-sectional regression. It follows that the estimated risk premia will differ from those estimated in the correct specification.

Giglio and Xiu (2017) introduce a new method of estimating factor risk premia using three-pass regressions. They assume that the observed factors, g_t, are a linear combination of the true factors plus noise and possible measurement error:

$$g_t = \xi + \eta f_t + v_t \tag{4}$$

covariance matrix.

[7]Technically, there is a difference in the effects on our estimates if the factors in the error are priced or not. That is, it is possible to have an error term which has a factor structure, but whose factors do not carry risk premia. This distinction is not important for our discussion here.

where ξ is the measurement error, and v_t is the noise. If there are 6 true factors and only 4 observed factors, then η is 4×6. This formulation is general enough to incorporate situations where g_t is simply a subset of the true factors or a linear combination of all of them (set η equal to the concatenation of the identity matrix and a matrix of zeros of appropriate dimensions).

As the name suggests, three-pass uses the information in the three equations (1), (2), and (4) and principal components analysis to estimate the factor risk premia and overcome the issue of omitted variables bias, and, as the number of assets tends to infinity, is able to identify the factor loadings and factor risk premia.

3 Estimation of Entropy-Based Risk Neutral Probabilities

Consider the classic asset pricing equation:

$$\mathbb{E}[Rm] = 1$$

where we assume stationarity (and ergodicity) of returns and the SDF, allowing us to drop time subscripts in the unconditional expectation. Dividing through by $\mathbb{E}[m] = 1/R_f$:

$$\mathbb{E}\left[R\frac{m}{\mathbb{E}[m]}\right] = R_f.$$

Since $m/\mathbb{E}[m]$ is strictly positive and integrates to 1, we can define a probability measure Q, absolutely continuous with respect to P (the "natural" probabilities), such that:

$$\frac{m}{\mathbb{E}[m]} = \frac{dQ}{dP}.$$

We call Q the risk-neutral measure since:

$$\mathbb{E}^Q[R] = R_f.$$

That is, the risk-neutral measure allows us to price assets as if investors were risk-neutral (i.e., the SDF is a constant).

To make this formulation applicable, let us move to a discrete state setting. That is,

$$R_{f,t} = \sum_{s \in S} R_{t+1}(s) \frac{\pi^Q(s)}{\pi(s)} \pi(s) \tag{5}$$

where S is the set of possible states, $\pi^Q(s)$ is the risk-neutral probability of state s, $\pi(s)$ is the natural probability of that state, and $R_{t+1}(s)$ is the return in that state. Notice, that in this discrete setting, $m/\mathbb{E}[m] \equiv dQ/dP = \pi^Q(s)/\pi(s)$. The goal here is to use no more information than equation (5) to estimate the risk-neutral probabilities.

This is a form of a linear inverse problem and can be solved as:

$$\min_{\pi^Q} I(\pi^Q || \pi), \quad \text{subject to equation (5)} \tag{6}$$

where $I(\pi^Q||\pi)$ is the Kullback-Leibler Information Criterion (KLIC):

$$I(\pi^Q, \pi) = \sum_s \pi^Q(s) \log \left(\frac{\pi^Q(s)}{\pi(s)} \right) \tag{7}$$

or more generally:

$$I(Q||P) = \int \log \left(\frac{dQ}{dP} \right) dQ.$$

This optimization problem asks what minimal adjustment to the natural probabilities is needed to ensure the Euler equation holds. The size of this adjustment is measured by the KLIC. Brown and Smith (1990) provide theoretical justification for this minimization, and in particular, show convergence of the observed outcome frequency to the solution of minimization problem (for discrete probabilities). The next subsection provides some further economic motivation for why we might be interested in minimizing the KLIC between the risk-neutral and natural probabilities.

The solution to problem (6) is known (Csiszar, 1975):

$$\frac{dQ}{dP} = \frac{e^{\gamma(R_{t+1}-R_f)}}{\mathbb{E}\left[e^{\gamma(R_{t+1}-R_f)}\right]}.$$

If there were N constraints of the form (5), then we would have:

$$\frac{dQ}{dP} = \frac{e^{\sum_{i=1}^{N}\gamma_i(R_{i,t+1}-R_f)}}{\mathbb{E}\left[e^{\sum_{i=1}^{N}\gamma_i(R_{i,t+1}-R_f)}\right]}. \tag{8}$$

Here γ is a vector of Lagrange multipliers which solves the dual problem (Ben-Tal, 1985):

$$\gamma = \arg\min \mathbb{E}\left[e^{\sum_{i=1}^{N}\gamma_i(R_{i,t+1}-R_f)}\right]. \tag{9}$$

In sample, one would solve:

$$\widehat{\gamma} = \arg\min_{\gamma} \frac{1}{T}\sum_{t=1}^{T} e^{\gamma'(R_{t+1}-R_f)}$$

And the estimated Radon-Nikodym derivative would be:

$$\left(\widehat{\frac{dQ}{dP}}\right)_t = \frac{e^{\sum_{i=1}^{N}\widehat{\gamma}_i(R_{i,t+1}-R_f)}}{\sum_{t=1}^{T}\pi_t\left[e^{\sum_{i=1}^{N}\widehat{\gamma}_i(R_{i,t+1}-R_f)}\right]}. \tag{10}$$

This change of measure is also known as exponential tilting.

3.1 Positive Variance Bound

Why do we care about the SDF which minimizes the KLIC from the natural probabilities? As we shall show, in the space of strictly positive SDFs, the KLIC minimizing one (locally) achieves the lowest variance. That is, we identify what might be called a *positive, variance bound* (Hansen and Jagannathan, 1991).

We fix the conditional mean of the SDF, which pins down the risk-free rate, $\mathbb{E}_t[m_{t+1}] = 1/R_{f,t}$, and we consider the space of positive SDFs. Recall the relation between the Radon-Nikodym derivative of the risk-neutral measure with respect to the natural measure and the SDF:[8]

$$\left(\frac{dQ}{dP}\right)_{t+1} = \frac{m_{t+1}}{\mathbb{E}_t[m_{t+1}]} = R_{f,t}m_{t+1}.$$

Consider the change of measure, $\left(\frac{dQ}{dP}\right)^*$, which minimizes the KLIC from the natural probabilities. Let $\frac{dQ}{dP}$ be any other change of measure with the same conditional mean for the SDF. Eq. (5) holds (more generally than in the discrete state setting in that equation) for both changes of measure, so they are both associated with different SDFs. We have:[9]

$$\mathbb{E}_t\left[\frac{dQ}{dP}\ln\frac{dQ}{dP}\right] \geq \mathbb{E}_t\left[\left(\frac{dQ}{dP}\right)^* \ln\left(\frac{dQ}{dP}\right)^*\right].$$

Substitute in the respective SDFs:

$$R_{f,t}\mathbb{E}_t[m_{t+1}[\ln(m_{t+1}) - \ln(R_{f,t})]] \geq R_{f,t}\mathbb{E}_t[m^*_{t+1}[\ln(m^*_{t+1}) - \ln(R_{f,t})]].$$

The risk-free rate is the same for both SDFs by assumption. The equation above simplifies to:

$$\mathbb{E}_t[m_{t+1}\ln(m_{t+1})] \geq E_t[m^*_{t+1}\ln(m^*_{t+1})].$$

Take a first-order Taylor expansion of $\ln(m_{t+1})$ around $m_{t+1} = 1$.[10]

$$\ln(m_{t+1}) \approx m_{t+1} - 1.$$

Substitute this into the previous equation and cancel the risk-free rate which reappears to

[8]Technically, we are using the density process for the change of measure, not the change of measure itself.

[9]$I(Q||P) = \int \ln \frac{dQ}{dP} dQ = \int \ln \frac{dQ}{dP}\frac{dQ}{dP} dP = \mathbb{E}\left[\frac{dQ}{dP}\ln\frac{dQ}{dP}\right]$.

[10]This is not ad hoc. Consider the equilibrium form of an SDF: $m_{t+1} = \frac{U_{t+1}}{U_t}$, where U_t is marginal utility of consumption in period t. Thus, an expansion $m_{t+1} = 1$ is the same as taking an expansion around $U_{t+1} = U_t$. This is the same log approximation used to derive a linear relationship between asset returns and consumption growth in the C-CAPM with CRRA utility.

derive:

$$\mathbb{E}_t\left[m_{t+1}^2\right] \geq \mathbb{E}_t\left[m_{t+1}^{*2}\right]. \tag{11}$$

Since the mean of the SDFs is the same, Eq. (11) implies that the variance of m_{t+1}^* is the smallest among all positive positive SDFs. Note that positivity is implicit in our use of KLIC. That is, by associating the linear pricing function with a change of measure (i.e., with a normalized strictly positive random variable), we ensure that the SDF is positive. This is *not* a property shared by the classical linear projection SDF (see Cochrane (2005) for textbook details).

Backus, et al (2014) show that the (one-period) entropy of the SDF is associated with a lower bound on mean excess returns. This is in line with the intuition we have emphasized. The lower bound for mean excess returns implied by an arbitrary SDF will be higher than that of our KLIC minimizing one. Recall that the Hansen-Jagannathan bound links standard deviations of the SDF with mean excess returns as well: Higher mean excess returns imply a larger lower bound for the standard deviation of the SDF.[11]

3.2 Don't Factor Models Already Imply an SDF?

Recall that a statistical factor model is equivalent to a SDF that is affine in the factors. One may ask how we can have an exponential tilting change of measure and an affine-in-factors SDF simultaneously. If we were in a complete markets setting, this indeed would be impossible.

The combination of (i) exponential tilting change of measure, (ii) returns are generated by a factor model, and (iii) markets are complete is mutually incompatible. This is because of positivity, which the exponential tilting change of measure possesses but which affine SDFs do not. Thus, there cannot be complete markets, since (i) and (ii) are essentially different

[11]We cannot make the exact chain of implication:

$$\sigma_t\left(m_{t+1}\right) = \sup \frac{\mathbb{E}_t\left[R_{t+1}\right] - R_f}{R_{f,t}\sigma_t\left(R_{t+1}\right)} \geq \sigma_t\left(m_{t+1}^*\right)$$

because the bound is achieved by the linear projection of the SDF onto the asset span. As mentioned above, this projection need not be positive, as the SDFs we are comparing must be.

SDFs.

One may then wish to compare the spanned components of these SDFs. That is, in incomplete markets, a SDF can be decomposed as:

$$M_t = M_{s,t} + \varepsilon_t.$$

where $M_{s,t}$ is in the asset span and ε_t is orthogonal to the asset span. However, this will not work either since both factor models and exponential tilting only give the researcher the "left-hand side." For example, the $M_{s,t}$ component will not necessarily take the special functional form we derived above for exponential tilting. Of course, all SDFs assign the same value to spanned payoffs. Thus, there is no contradiction in using exponential tilting to recover risk premia in the presence of a factor model.

3.3 The Natural Probabilities

Having considered the question "Why do we care about minimizing KLIC?" we now ask "How can we operationalize the theory?" A natural follow-up question to ask is "What are the natural probabilities?" That is, what is P? Stutzer (1996) advocates using the empirical likelihood. This changes the more abstract notion of state s to period t in a sample of, say, T, returns. So, for example, equation (5) becomes:

$$R_{f,t} = \sum_{t=1}^{T} R_{t+1}(t) \frac{\pi^Q(t)}{1/T} \frac{1}{T}.$$

There are two main reasons to use the empirical likelihood as the natural probabilities. First, if the underlying data generating process is an ergodic Markov chain, the empirical likelihood is a consistent estimator of the invariant distribution of the chain. It is also the fastest converging non-parametric, consistent estimator.

Second, using the empirical likelihood reduces the KLIC distance[12] to Shannon's Entropy,

[12]KLIC is not really a distance measure in the formal sense of the word since it is not symmetric, that is, $I(Q||P) \neq I(P||Q)$ necessarily.

and thus the optimization problem satisfies the maximum entropy principle. It formalizes the notion of starting from complete randomness and using only the information in equation (5) to determine the new probability measure. This can be seen by writing out the KLIC objective function:

$$I\left(\pi, T^{-1}\mathbf{1}_{T\times 1}\right) = \sum_{t=1}^{T} \pi_t \ln(T\pi_t) = \sum_{t=1}^{T} \pi_t \ln \pi_t + \ln(T) \sum_{t=1}^{T} \pi_t.$$

By definition of a probability measure:

$$\sum_{t=1}^{T} \pi_t = 1$$

so:

$$I\left(\pi, T^{-1}\mathbf{1}_{T\times 1}\right) = \sum_{t=1}^{T} \pi_t \ln \pi_t + \ln(T).$$

The final term on the right does not depend on π_t, and so minimizing the KLIC in this case is equivalent to minimizing $\sum_t \pi_t \ln \pi_t$, which is equivalent to maximizing $-\sum_t \pi_t \ln \pi_t$. This expression is the definition of the entropy of π. Notice that Eq.(10) can then be written as:

$$\widehat{\pi}_t^Q = \frac{\pi_t e^{\widehat{\gamma}' R_{t+1}}}{\sum_{t=1}^{T} \pi_t e^{\widehat{\gamma}' R_{t+1}}} = \frac{e^{\widehat{\gamma}' R_{t+1}}}{\sum_{t=1}^{T} e^{\widehat{\gamma}' R_{t+1}}}$$

since $\pi_t = T^{-1}$ cancels out between the numerator and denominator.

The use of KLIC as the objective function has an axiomatic foundation in Hobson (1971). That paper shows that if a function satisfies certain properties, then it must be proportional to KLIC. These properties are (i) the function should be equal to zero if and only if the two arguments of $I(\cdot||\cdot)$ are the same; (ii) a mere relabeling or rearrangement of the states should not change the value of the function; (iii) if a probability measure π assigns positive probability to m states, and probability measure π^* assigns positive probability to $n \leq m$ states, then $I(\pi^*||\pi)$ should be increasing in m (and decreasing in n); (iv) there is a complicated "composition rule" which is beyond the scope of this paper.

4 Asymptotics

This section presents the asymptotics of our estimator of the factor risk premia using an application of the results of Kitamura and Stutzer (1997). The Appendix contains a proof showing the asymptotics of the estimator when estimation is done in two steps: First, the Lagrange multipliers, γ, are estimated, and second, the moments condition $\lambda = -\mathbb{E}^Q[f_t]$ is estimated using the delta method.

Kitamura and Stutzer (1997) write the Euler equations as:

$$\mathbb{E}^Q[F(\beta, \gamma)] = 0$$

where each element of the column vector F, F_i, is an Euler equation. The new parameter here, β, is a vector of parameters we might be interested in estimating. Considering the Euler equations in the previous sections (that is, the excess returns on assets over the risk-free rate), and assuming the risk-free rate is known, we have the $\beta = \emptyset$, since $\mathbb{E}^Q[R_i - R_f] = 0$, does not rely on any parameters beyond the Lagrange multipliers.

However, note that our moment condition of interest, $\lambda = -\mathbb{E}^Q[f_t]$, can also be included in their definition of $F(\beta, \gamma)$. Thus, we can "stack" the factor risk premia equations with excess return constraints and call this new vector $F(\beta, \gamma)$, where $\beta = \lambda$. Kitamura and Stutzer show that the estimator solves the following problem:

$$(\widehat{\lambda}_T, \widehat{\gamma}_T) = \arg\max_{\beta} \min_{\gamma} \left[\frac{1}{T} \sum_{t=1}^{T} e^{\gamma' \widehat{F}(t,\lambda)} \right] \tag{12}$$

where

$$\widehat{F}(t, \lambda) \equiv \sum_{k=-K}^{K} \frac{1}{2K+1} F(x_{t-k}, \lambda)$$

and $x_t = (R_t^e, f_t)$, $K \to \infty$ as $T \to \infty$, and $K^2/T \to 0$. Here K is a smoothing parameter. Under regularity conditions (see their paper), it can be shown that

$$\sqrt{T} \left(\widehat{\lambda}_T - \lambda^* \right) \xrightarrow{d} N(0, V)$$

where λ^* the true risk premia and $V = \left(D'S^{-1}D\right)^{-1}$. Here:

$$D = \mathbb{E}\left[\frac{\partial F(x, \lambda^*)}{\partial \lambda'}\right] = \begin{bmatrix} \mathbf{0}_{N \times Nf} \\ I_{Nf \times Nf} \end{bmatrix}$$

where $I_{Nf \times Nf}$ is the $Nf \times Nf$ identity matrix, and Nf is the number of factors. Lastly,

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left[F\left(x_t, \lambda^*\right) F\left(x_{t-j}, \lambda^*\right)'\right]$$

which can be estimated by the method of Newey and West (1987).

In order to identify the correct mini-max solutions to Eq.(12), it is important to start with a "close" guess. By examining graphs of the objective function, we have observed that the function is non-convex on regions of the parameter space. A choice of initial guess which worked in every simulation run is the one proposed in the Appendix. This estimate is also computationally fast to estimate.

5 Simulation Results

This section compares the performance of two-pass, three-pass, and the exponential tilting methods in a set of simulation studies. The risk-free rate is set to the average monthly risk-free rate from 1963-2017. The number of simulations in each experiment is 1,000. The true number of factors generating returns is five. We calibrate these factors to the five Fama-French factors (Fama and French (2015)). The true betas are generated from a multivariate normal random variable with mean and covariance matrix calibrated to the estimated loadings of the 25 book-to-market sorted portfolios from Kenneth French's website on the five factors.[13]

The first set of results assumes all factors are known, so omitted variable bias is not a problem. The factors are demeaned for simplicity. If they were not, we would have an

[13] We ran a simulation where the loadings came from a Beta(0.5,0.5) distribution as well as a simulation where there is 0 true cross-sectional dispersion in betas. The results are qualitatively unchanged.

additional term in the data generating process for returns. This process is:

$$R_{i,t} = R_f + \sum_{j=1}^{5} \beta_{i,j}(f_{j,t} + \gamma_j) + u_{i,t} \tag{13}$$

where $u_{i,t}$ is iid noise. To confirm that Eq.(13) leads to a "beta-pricing" model as desired, pass the expectations operator through to get:

$$\overline{R}_i - R_f = \sum_{j=1}^{5} \beta_{i,j}(\mathbb{E}[f_{j,t}] + \gamma_j) + \mathbb{E}[u_{i,t}] = \sum_{j=1}^{5} \beta_{i,j}\gamma_j$$

since both $f_{j,t}$ and $u_{i,t}$ are mean 0.

In the second set of simulations, we assume the observed factors are:

$$g_t = \eta f_t$$

where

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

That is, we assume only the first three factors are observed, even though the returns were generated by all five.

Table (1) shows root mean squared errors (RMSEs) for two different pairs of periods, T, and assets, N, when all five factors are observed. The first panel shows results for $T = 1000$ and $N = 25$. Fama-MacBeth actually outperforms three-pass in RMSE for all factors. Exponential tilting and Fama-MacBeth are comparable, with the former having smaller RMSEs for two of the five factors. The first panel of Table (2) shows the bias for the same simulations. Interestingly, almost all factors are estimated with a slightly negative bias. More importantly, the bias is small for all estimators.

The second panels of Tables (1) and (2) show RMSEs and biases when $T = 600$ and $N = 200$. The main thing to note is that all estimation methods now have smaller RMSEs.

Three-pass has the largest improvement between panels. Exponential tilting's performance improved enough so that the RMSE of that procedure is now less than or equal to the RMSE of Fama-MacBeth for three of the five factors.

Table 1: RMSE, All Factors Observed

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	T=1000, N=25				
Three-Pass	0.00375	0.00151	0.00185	0.00287	0.00234
Fama-MacBeth	0.00208	0.00051	0.00060	0.00100	0.00152
Exponential Tilting	0.00039	0.00069	0.00137	0.00164	0.00219
	T=600, N=200				
Three-Pass	0.00194	0.00040	0.00043	0.00031	0.00164
Fama-MacBeth	0.00140	0.00026	0.00034	0.00040	0.00088
Exponential Tilting	0.00017	0.00033	0.00049	0.00045	0.00102

This table shows the root mean squared error for each factor in 1000 simulations. The top panel corresponds to the case when $T = 1000$ and $N = 25$, while the bottom panel corresponds to the case when $T = 600$ and $N = 200$. Each row corresponds to estimation technique and each column corresponds to a factor.

Table 2: Bias, All Factors Observed

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	T=1000, N=25				
Three-Pass	-0.00349	-0.00145	-0.00162	-0.00286	-0.00230
Fama-MacBeth	-0.00060	-0.00007	-0.00006	0.00002	-0.00045
Exponential Tilting	0.00029	-0.00060	-0.00130	-0.00161	-0.00217
	T=600, N=200				
Three-Pass	-0.00180	-0.00035	-0.00030	-0.00011	-0.00163
Fama-MacBeth	-0.00110	-0.00013	-0.00005	0.00003	-0.00069
Exponential Tilting	0.00003	-0.00018	-0.00028	-0.00014	-0.00090

This table shows the average bias in estimation for each factor in 1000 simulations. The top panel corresponds to the case when $T = 1000$ and $N = 25$, while the bottom panel corresponds to the case when $T = 600$ and $N = 200$. Each row corresponds to estimation technique and each column corresponds to a factor.

Now we consider the case where only the first three factors are observed. The top panel of table (3) shows RMSEs for $T = 1000$ and $N = 25$. First, note that the RMSEs for Fama-MacBeth, as expected, have increased. The RMSEs for three-pass are virtually unchanged, and the RMSEs for exponential tilting are literally unchanged. This latter findings is expected. A factor's estimated risk premium using exponential tilting is unaffected by the other factors being priced. The other factors do matter for the covariance matrix of estimated risk premia, as can be seen in the previous section. The top panel of table (4) shows that the biases are still slightly negative but very small.

The bottom panel of table (3) shows RMSEs when $T = 600$ and $N = 200$. Once again, three-pass experiences the biggest improvement, whereas Fama-MacBeth actually delivers slightly higher RMSEs for two of the factors compared to the top panel. Exponential tilting, as expected, delivers the same small RMSEs from the bottom panel of Table (1).

It is instructive to the note that the first factor (the "market factor") has the highest covariance with omitted factors, and hence Fama-MacBeth has highest RMSE for this estimate. The third factor has low covariance with the omitted factors, and so Fama-MacBeth delivers a RMSE comparable to those of three-pass and exponential tilting. The main takeaway is that exponential tilting has good performance under a variety of scenarios: both with and without observing all factors and under situations where $N \ll T$.

Table 3: RMSEs, Two Factors Unobserved

	Factor 1	Factor 2	Factor 3
	T=1000, N=25		
Three-Pass	0.00375	0.00151	0.00185
Fama-MacBeth	0.00446	0.00103	0.00060
Exponential Tilting	0.00039	0.00069	0.00137
	T=600, N=200		
Three-Pass	0.00194	0.00040	0.00043
Fama-MacBeth	0.00489	0.00105	0.00056
Exponential Tilting	0.00017	0.00033	0.00049

This table shows the root mean squared error for each factor in 1000 simulations. The top panel corresponds to the case when $T = 1000$ and $N = 25$, while the bottom panel corresponds to the case when $T = 600$ and $N = 200$. Each row corresponds to estimation technique and each column corresponds to a factor.

Table 4: Bias, Two Factors Unobserved

	Factor 1	Factor 2	Factor 3
	T=1000, N=25		
Three-Pass	-0.00349	-0.00145	-0.00162
Fama-MacBeth	-0.00425	-0.00093	-0.00015
Exponential Tilting	0.00029	-0.00060	-0.00130
	T=600, N=200		
Three-Pass	-0.00180	-0.00035	-0.00030
Fama-MacBeth	-0.00484	-0.00103	0.00045
Exponential Tilting	0.00003	-0.00018	-0.00028

This table shows the average bias in estimation for each factor in 1000 simulations. The top panel corresponds to the case when $T = 1000$ and $N = 25$, while the bottom panel corresponds to the case when $T = 600$ and $N = 200$. Each row corresponds to estimation technique and each column corresponds to a factor.

6 Empirical Results

6.1 Main Results

For our main empirical tests, we estimate the risk premia of the factor in the popular three, four, and five factor models of Fama and French (1993), Carhart (1997), and Fama and French (2015), respectively. We also estimate the risk premium for the intermediary leverage factor of Adrian, et al (2014), which is a non-traded factor.

The three factor model includes $RMRF$, the excess return of the market over the risk-free rate, SMB, the excess return of small capitalization stocks over large capitalization ones, and HML, the excess return of high book-to-market ("value") firms over low book-to-market ("growth") ones.[14] The four factor model adds the excess return of a momentum portfolio: previous "winners" over "losers." The five factor model adds two new factors to the three factor model: profitability and investment.

It is important to realize that the factors, aside from intermediary leverage, are themselves portfolios of stocks. Thus, their risk premia should be their expected returns. That is, the

[14]See Kenneth French's website for a detailed description.

risk premia of the value factor, HML_t, is $\mathbb{E}[HML_t]$. We can compare our estimates of the risk premia to the realized average return, $\overline{HML_t}$. This last quantity need not be exactly equal to the true expectation, but only serves as a benchmark. For non-traded factors, we have no such benchmark.

The data depend on the model being tested (due to data availability). For the three and four factor models, the data are monthly from 1927:01-2017:06. For the five factor model, the data are monthly from 1963:07-2017:06. For the intermediary leverage model, the data are quarterly from 1968:Q1-2009:Q4. Our test assets consist of 80 portfolios: 25 portfolios sorted on value and market cap, 25 portfolios sorted on value and momentum, and 30 industry portfolios. All return data, including excess return factor data, are from Kenneth French's website. The intermediary leverage factor data are from Tyler Muir's website.

Table (5) shows estimated risk premia and the associated t-statistics for each factor using Fama-MacBeth, three-pass, and exponential tilting. For the factors which are excess returns, the factor's average return is also listed. The average returns and risk premia have been multiplied by 100 to make them interpretable as percentages. The standard errors of the sample means have been adjusted for possible auto and cross-correlation using the method of Newey and West (1987).

Exponential tilting finds a positive and significant risk premium for the market excess return in every factor model. The risk premium estimates barely change across models. Three-pass never assigns a significant risk premium to the market, and Fama-MacBeth sometimes assigns a positive risk premium and sometimes a negative one. Exponential tilting consistently delivers positive and significant risk premia for SMB and HML with risk premia close to their observed average excess returns. Both three-pass and Fama-MacBeth estimate a risk premium close to the observed average excess return for SMB, but this is not the case for HML, where signs and significances change from model to model. Momentum is assigned a positive and significant risk premium by all three methods, but exponential tilting estimates a risk premium closest to observed empirical average excess return. Fama-MacBeth finds risk premia closest to the empirical excess returns for the profitability and investment factors, but these factors are at best weakly significant across estimation techniques. Lastly, all three methods find a strongly significant and positive risk premium for intermediary leverage, though the estimate from Fama-MacBeth is almost eight times larger than those of three-pass and exponential tilting.

Taken together, these results show that exponential tilting estimator delivery empirically and theoretically plausible estimates of factor risk premia.

The appendix contains a GMM application of exponential tilting. That is, we note that the first-order condition for the Lagrange multipliers and $\lambda = -\mathbb{E}\left[f_t(dQ/dP)\right]$ are moment conditions that can be included along with the usual moment conditions used to estimate factor models via GMM:

$$\mathbb{E}\left[\varepsilon_{i,t}\right] = \mathbb{E}\left[\varepsilon_{i,t} f_{j,t}\right] = 0, \quad i = 1, ..., N, \ j = 1, ..., N_f \tag{14}$$

where N_f is the number of factors. The results are similar. When we include only the moment conditions implied by exponential tilting, we estimate risk premia close the empirical average excess returns. However, once we include the moment conditions in (14), the estimates of the risk premia change and are further from the excess returns. This is because λ now needs to approximately satisfy (14) as well, which is not a valid moment condition if the model is misspecified.

6.2 The Information in Excess Return Factors

Excess return factors, such as those in the four factor model, are supposed to explain the returns of portfolios using other portfolios (the factors). We may try to formalize this idea by essentially reversing the process we have used in the previous subsection.

Since these factors are excess returns, we may use them as the test assets in constructing our change of measure. That is, our Euler equations now take the form:

$$\mathbb{E}^Q\left[f_t\right] = 0$$

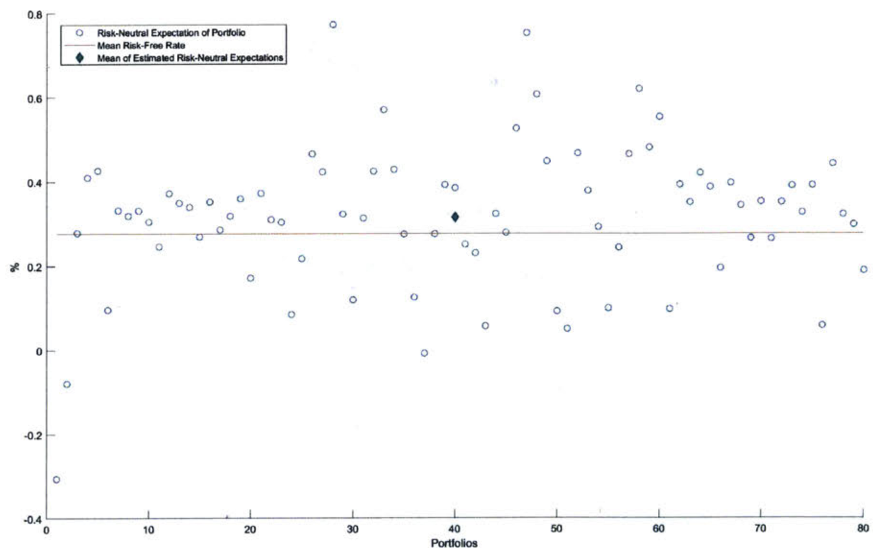
where f_t is either $RMRF, SMB, HML$, or MOM. Now, instead of estimating factor risk premia, we are interested in evaluating the risk-neutral expectation of an arbitrary portfolio, R_t. We know that:

$$\mathbb{E}^{Q}\left[R_t\right] = R_f$$

for any given portfolio. We take the risk-neutral expectation of the 80 portfolios from the previous subsection. By doing this we are asking if the information in the returns of the four excess return factors is enough to accurately estimate the risk-neutral change of measure, where accuracy is judged by how closely this change of measure brings the expectation of a portfolio's return to the risk-free rate.

Figure 1 displays the results. Each bubble represents $\mathbb{E}^{Q}[R_t]$ for one of the 80 portfolios. The red line is the average risk-free rate over the sample (monthly from 1927:01-2017:06). The filled in diamond is the average over all 80 risk-neutral expectations, $(1/80)\sum_{i=1}^{80}\mathbb{E}^{Q}\left[R_{i,t}\right]$. While there is some dispersion, the most of the bubbles are clustered around the red line. The diamond is at 0.32% per month while the red line is at 0.28% per month. These results suggest that in future applications of this estimation methodology, researchers may not need a large-cross section of assets to estimate the change of measure. This is in line with the simulation results in previous section, where we showed that even when using only 25 assets, exponential tilting had small estimation errors on average.

Figure 1: Estimated Risk-Free Rates



Estimated risk-free rates using four excess return factors to estimate risk-neutral change of measure.

Table 5: Estimated Risk Premia of Factors from the Literature

				Fama-MacBeth		Three-Pass		Exponential Tilting	
	Factor	Ave. Ret	T-Stat	RP	T-Stat	RP	T-Stat	RP	T-Stat
Fama and French (1993)	RMRF	0.65310	3.73110	-0.67646	-2.92017	-0.09483	-0.36244	0.96467	7.32051
	SMB	0.21605	2.09680	0.24331	2.33859	0.25663	2.64404	0.24583	4.74175
	HML	0.38891	3.31280	0.13163	1.12948	-0.10345	-1.13599	0.34154	4.86921
Carhart (1997)	RMRF	0.65310	3.73110	0.01429	0.06342	-0.09483	-0.36244	0.96453	7.31942
	SMB	0.21605	2.09680	0.18777	1.80572	0.25663	2.64404	0.24583	4.74177
	HML	0.38891	3.31280	0.32690	2.84410	-0.10345	-1.13599	0.34155	4.86930
	MOM	0.65627	4.77420	0.75224	5.00811	0.26497	1.94579	0.61114	5.50304
Fama and French (2015)	RMRF	0.51886	2.84100	-0.35990	-1.16438	0.10488	0.45991	0.93805	6.77514
	SMB	0.25591	1.97810	0.28618	2.34681	0.24729	1.82596	0.24778	3.20319
	HML	0.34750	2.62750	0.03115	0.25528	0.12520	0.97995	0.34579	4.29604
	RMW	0.24725	2.40290	0.16176	1.09431	0.09362	1.29856	0.00475	0.06116
	CMA	0.29360	3.21810	0.33250	1.64405	0.09183	1.09767	0.13032	1.86241
Adrian, et al (2014)	Leverage	NA		8.37409	2.89948	1.14305	2.14876	1.45719	2.34564

This table shows estimated risk premia and t-statistics for popular factor models from the literature as estimated by using Fama-MacBeth regressions (columns 4-5), three-pass regressions (columns 6-7) and exponential tilting (columns 8-9). The left column identifies the paper whose factor model is being tested. The next column lists the identity of each factor. The third column lists the average return of the factor if the factor is an excess return portfolio. Standard errors for the mean excess returns were calculated using the method of Newey and West (1987). All other standard errors are calculated using the method implied by the column heading. All risk premia and average returns have been multiplied by 100 to make the numbers interpretable as percentages.

7 Application to Fund Manager Performance Evaluation

In this section, we detail a key application of our estimation methodology. Performance evaluation of asset managers is a long-researched question and coming up with a robust way to assess a manager's performance is important for investors deciding how to allocate their money. A manager is said to generate alpha if his or her portfolio has excess returns above and beyond those implied the portfolio's exposure to the priced factors. That is, manager i's alpha is defined to be:

$$\alpha_i = \mathbb{E}[R_i] - R_f - \beta_i' \lambda.$$

Using a sample of fund returns, we could calculate:

$$\widehat{\alpha}_i = \overline{R}_i - R_f - \widehat{\beta}_i' \widehat{\lambda}$$

where $\widehat{\lambda}$ are the estimated factor risk premia which are estimated from a set of test assets and $\widehat{\beta}_i$ are estimated by a time-series regression of fund i's returns on the factors. It follows that the econometrician needs to take a stand on what the appropriate factor model for the manager is.

We see here another instance in which omitted variables (missing factors) can cause problems in our estimation goals. By an appropriate choice of securities, an active fund manager can influence how his portfolio loads on certain factors. For example, as an extreme case, the manager may have a significant loading (beta) on only one risk factor. If this factor is unobserved by the econometrician, the risk premium on that factor will be measured as alpha. More formally, consider a situation where there are two true factors generating returns, but the fund manager's portfolio only loads on the second factor:

$$R_{i,t} = R_f + \beta_{i,1} \left(f_{1,t} + \lambda_1 \right) + \beta_{i,2} \left(f_{2,t} + \lambda_2 \right) + u_{i,t}.$$

Now consider the "alpha" of the manager if we believed there was only one factor, f_1:

$$\widetilde{\alpha}_i = \mathbb{E}\left[R_i\right] - R_f - \beta_{i,1}\lambda_1 = \mathbb{E}\left[R_i\right] - R_f$$

since we have assumed the manager's returns only load on the second factor. But using our factor model above, we see:

$$\mathbb{E}[R_i] = R_f + \beta_{i,1}\lambda_1 + \beta_{i,2}\lambda_2$$

since the f_i and error term are mean zero. Thus, in reality,

$$\widetilde{\alpha}_i = \beta_{i,2}\lambda_2$$

Using our exponential tilting measure, we circumvent the omitted factor problem by estimating the risk-neutral change of measure directly from asset returns. Recall (see, e.g. Cochrane (2005)) that a factor model for returns implies a SDF which is affine in the factors. Thus, if the econometrician already knows the SDF, knowledge of the factors does not change the price he or she would assign to a security or portfolio. To see how this affects our measure of a fund's alpha, take the risk-neutral expectation of the example two factor model we used above, but explicitly include an α term (it may be 0, positive, or negative):

$$\mathbb{E}^Q[R_i] = R_f + \alpha + \beta_{i,1}\left(\mathbb{E}^Q[f_{1,t}] + \lambda_1\right) + \beta_{i,2}\left(\mathbb{E}^Q[f_{2,t}] + \lambda_2\right).$$

But, recall that $-E^Q[f] = \lambda$ for any factor priced f. Thus,

$$\mathbb{E}^Q[R_i] = R_f + \alpha.$$

Therefore, once we are able to estimate the risk-neutral expectation of a portfolio's return, we will be able to estimate the portfolio's alpha. This is the exact same problem we tackled in previous sections. That is, instead of estimating the (negative) risk-neutral expectation of a factor and calling this the factor's risk premium, we are estimating the risk-neutral expectation of a manager's excess returns and calling this his or her alpha.

We quickly evaluate the merits of this proposed procedure in a simulation exercise. We use the same setting as in the section on simulation above. Now, we also add a manager whose returns only load on the fifth factor. For simplicity, we take his beta on this factor to be one. As in the section on simulation, we assume only the first three factors are observed. We regress the manager's returns on the three observed factors to get an estimate of $\widehat{\beta}_i = \left[\widehat{\beta}_{i,1}, \widehat{\beta}_{i,2}, \widehat{\beta}_{i,3}\right]'$. We then estimate the manager's alpha as $\widehat{\alpha} = \overline{R}_i - R_f - \widehat{\beta}_i' \widehat{\lambda}$ where the factor risk premia are estimated by classic Fama-MacBeth. We also use exponential tilting to estimate $\widetilde{\alpha} = \widehat{\mathbb{E}}^Q [R_i] - R_f$. Table (6) shows the RMSE and mean alpha over 500 simulations using the two methods. The manager in this exercise has no alpha, however notice that the mean alpha estimated using Fama-MacBeth is 0.41% a month (the simulation is calibrated to monthly data). That translates to an alpha over 4% a year. Exponential tilting estimates an average alpha of less than 1/4 of that. Also, the RMSE is two times smaller for exponential tilting.

Table 6: Simulation Results for Estimates of Manager's Alpha

Method	RMSE	Mean Alpha
$\overline{R} - R_f - \widehat{\beta\lambda}$	0.0043	0.41
$\widehat{\mathbb{E}}^Q[R] - R_f$	0.002	0.0978

This table shows the RMSE and mean estimated alpha over 500 simulations. The top row shows the estimated alpha using Fama-MacBeth. The second row shows the estimated alpha using exponential tilting. The mean alpha has been multiplied by 100.

8 Conclusion

This paper has developed and evaluated a new method for estimating factor risk premia. The theory built on results from information theory, casting the estimation problem into a simple set of moment conditions defining the factor risk premia. These moment conditions are valid whether or not the model is missing certain factors. Building on the work of Kitamura and Stutzer (1997), we derived asymptotic standard errors for the estimates of the risk premia. Simulation exercises showed that the accuracy of our estimation method ("exponential tilting") was comparable, and frequently higher, than that of Fama-MacBeth. Finally, empirical results showed that exponential tilting delivers estimated risk premia that are in line with those predicted by financial economic theory. For example, the risk premia of excess return factors are very closed to their observed average excess returns.

It goes without saying that there are many more factor models which could be subjected to the techniques developed in this paper. For example, it would be important to know how a comprehensive list of "priced" and "unpriced" factors, as determined by exponential tilting, compares to similar lists, as determined by the recent literature on model misspecification. Similarly, the application to performance evaluation can be taken to the data. For example, mutual funds have consistently been found to under-perform certain benchmarks. It would be interesting to see if our methodology further confirms this result.

References

[1] Adrian, T., Etula, E., and Muir, T. Financial intermediaries and the cross-section of asset returns. *The Journal of Finance 69*, 6 (2014), 2557–2596.

[2] Aït-Sahalia, Y., and Lo, A. W. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance 53*, 2 (1998), 499–547.

[3] Backus, D., Chernov, M., and Zin, S. Sources of entropy in representative agent models. *The Journal of Finance 69*, 1 (2014), 51–99.

[4] Brown, D. E., and Smith, R. L. A correspondence principle for relative entropy minimization. *Naval Research Logistics (NRL) 37*, 2 (1990), 191–202.

[5] Bryzgalova, S. Spurious factors in linear asset pricing models. LSE manuscript, 2015.

[6] Carhart, M. M. On persistence in mutual fund performance. *The Journal of finance 52*, 1 (1997), 57–82.

[7] Cochrane, J. H. *Asset Pricing:(Revised Edition)*. Princeton university press, 2005.

[8] Csiszár, I. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability* (1975), 146–158.

[9] Fama, E. F., and French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*, 1 (1993), 3–56.

[10] Fama, E. F., and French, K. R. A five-factor asset pricing model. *Journal of Financial Economics 116*, 1 (2015), 1 – 22.

[11] Fama, E. F., and MacBeth, J. D. Risk, return, and equilibrium: Empirical tests. *Journal of political economy 81*, 3 (1973), 607–636.

[12] Ferson, W. E., et al. Investment performance: A review and synthesis. *Handbook of Economics and Finance 2* (2012), 969–1010.

[13] Ghosh, A., Julliard, C., and Taylor, A. P. An information-theoretic asset pricing model. Working paper, 2016a.

[14] Ghosh, A., Julliard, C., and Taylor, A. P. What is the consumption-capm missing? an information-theoretic framework for the analysis of asset pricing models. *The Review of Financial Studies 30*, 2 (2016b), 442–504.

[15] Giglio, S., and Xiu, D. Inference on risk premia in the presence of omitted factors. Tech. rep., National Bureau of Economic Research, 2017.

[16] Hansen, L. P. Large sample properties of generalized method of moments estimators. *Econometrica* (1982), 1029–1054.

[17] Hansen, L. P., and Jagannathan, R. Implications of security market data for models of dynamic economies. *Journal of political economy 99*, 2 (1991), 225–262.

[18] Harvey, C. R., and Liu, Y. Lucky factors. Working paper, 2016.

[19] Harvey, C. R., Liu, Y., and Zhu, H. and the cross-section of expected returns. *The Review of Financial Studies 29*, 1 (2016), 5–68.

[20] Hobson, A. *Concepts in statistical mechanics.* CRC Press, 1971.

[21] Hutchinson, J. M., Lo, A. W., and Poggio, T. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance 49*, 3 (1994), 851–889.

[22] Jackwerth, J. C., and Rubinstein, M. Recovering probability distributions from option prices. *The Journal of Finance 51*, 5 (1996), 1611–1631.

[23] Kan, R., Robotti, C., and Shanken, J. Pricing model performance and the two-pass cross-sectional regression methodology. *The Journal of Finance 68*, 6 (2013), 2617–2649.

[24] Kan, R., and Zhang, C. Gmm tests of stochastic discount factor models with useless factors. *Journal of Financial Economics 54*, 1 (1999a), 103–127.

[25] Kan, R., and Zhang, C. Two-pass tests of asset pricing models with useless factors. *the Journal of Finance 54*, 1 (1999b), 203–235.

[26] Kitamura, Y., and Stutzer, M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica* (1997), 861–874.

[27] Kitamura, Y., and Stutzer, M. Connections between entropic and linear projections in asset pricing estimation. *Journal of Econometrics 107*, 1 (2002), 159–174.

[28] Kleibergen, F. Tests of risk premia in linear factor models. *Journal of econometrics 149*, 2 (2009), 149–173.

[29] Kleibergen, F., and Zhan, Z. Mimicking portfolios of macroeconomic factors. Tech. rep., Technical report, Brown University Working Paper, 2014.

[30] Lewellen, J., Nagel, S., and Shanken, J. A skeptical appraisal of asset pricing tests. *Journal of Financial economics 96*, 2 (2010), 175–194.

[31] McLean, R. D., and Pontiff, J. Does academic research destroy stock return predictability? *The Journal of Finance 71*, 1 (2016), 5–32.

[32] Newey, W. K., and West, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 3 (1987), 703–708.

[33] Pukthuanthong, K., and Roll, R. A protocol for factor identification. Working Paper, 2014.

[34] Raponi, V., Robotti, C., and Zaffaroni, P. Ex-post risk premia and tests of multi-beta models in large cross-sections. Tech. rep., Technical report, Imperial College London, 2016.

[35] Rosenberg, J. V., and Engle, R. F. Empirical pricing kernels. *Journal of Financial Economics 64*, 3 (2002), 341–372.

[36] Ross, S. A. The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 3 (1976), 341–360.

[37] Shanken, J. On the estimation of beta-pricing models. *The review of financial studies 5*, 1 (1992), 1–33.

[38] Shanken, J. Statistical methods in tests of portfolio efficiency: A synthesis. *Handbook of statistics 14* (1996), 693–711.

[39] Shanken, J., and Zhou, G. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics 84*, 1 (2007), 40–86.

[40] Singleton, K. J. *Empirical dynamic asset pricing: model specification and econometric assessment.* Princeton University Press, 2009.

[41] Stutzer, M. A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics 68*, 2 (1995), 367–397.

[42] Stutzer, M. A simple nonparametric approach to derivative security valuation. *The Journal of Finance 51*, 5 (1996), 1633–1652.

A Alternate Proof of Asymptotics

This section derives asymptotics for the risk premia when the estimation procedure is done sequentially. That is, first we estimated the Lagrange multipliers using the excess return Euler constraints. Then we estimate the factor risk premia as a function of these Lagrange multipliers using the delta method. The standard errors in this section may differ from those derived in the body of the paper due to the two-step nature of this estimation procedure. If the first-order approximation inherent in the delta method leaves out non-negligible variation in estimates of the Lagrange multipliers, then the standard errors will be understated.[15] The advantage of the method presented here is that it is computationally faster than the one in the body. In fact, the point estimates for the risk premia from the method in the Appendix and very close to the ones estimated by the one-step procedure. Thus, we use the two-step estimates as our "guesses" when initializing the one-step procedure.

Recall that the Lagrange multipliers in the entropy minimization problem can be solved as

$$\widehat{\gamma} = \arg\min_{\gamma} \frac{1}{T}\sum_{t=1}^{T} e^{\gamma' R_t^e}$$

where γ is an $N \times 1$ vector of multipliers, R_t^e is an $N \times 1$ vector of excess returns, and N is the number of returns/Euler equations. This problem is convex, so the minimization is

[15] I thank Anna Mikusheva for this insight.

well-defined. The first-order condition is:

$$\frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} e^{\widehat{\gamma}' R_t^e} = 0.$$

We may follow standard extremum estimator arguments.[16] Take a mean value expansion around the population solution, γ_0, of the first order condition:

$$0 = \frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} e^{\gamma_0' R_t^e} + \left[\frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} R_t^e e^{\overline{\gamma}' R_t^e}\right] (\widehat{\gamma} - \gamma_0) \tag{15}$$

where $\overline{\gamma}$ is a convex combination of γ_0 and $\widehat{\gamma}$.

Assumption A.1

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} e^{\gamma_0' R_t^e} \xrightarrow{d} N(0, \Omega)$$

where

$$\Omega = \sum_{j=-\infty}^{\infty} \mathbb{E}\left[\left(R_t^{e\prime} e^{\gamma_0' R_t^e}\right)\left(R_{t-j}^{e\prime} e^{\gamma_0' R_{t-j}^e}\right)'\right].$$

Also,

$$\frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} R_t^e e^{\overline{\gamma}' R_t^e} \xrightarrow{p} \mathbb{E}\left[R_t^{e\prime} R_t^e e^{\gamma_0' R_t^e}\right] \equiv B.$$

Rearranging Eq. (15) and multiplying through by $\sqrt{T}$, we have:

$$\sqrt{T}(\widehat{\gamma} - \gamma_0) = -\left[\frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} R_t^e e^{\overline{\gamma}' R_t^e}\right]^{-1}\left[\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} R_t^{e\prime} e^{\gamma_0' R_t^e}\right].$$

Taking $T \to \infty$, we have:[17]

$$\sqrt{T}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} N\left(0, B^{-1}\Omega B^{-1\prime}\right). \tag{16}$$

[16]See, e.g., Singleton (2009) for details.
[17]We use the method of Newey and West (1987) to estimate Ω.

Next, recall that our estimator for the factor risk premia is:

$$\widehat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}(-f_t)\left(\widehat{\frac{dQ}{dP}}\right)_t$$

where

$$\left(\widehat{\frac{dQ}{dP}}\right)_t = \frac{e^{\widehat{\gamma}' R_t^e}}{\frac{1}{T}\sum_{t=1}^{T} e^{\widehat{\gamma}' R_t^e}}$$

and f_t is a $K \times 1$ vector of factors. This estimator is a function of $\widehat{\gamma}$, and so we apply the Delta Method. Define:

$$\frac{1}{T}\sum_{t=1}^{T}(-f_t)\left(\frac{dQ}{dP}\right)_t = \frac{\frac{1}{T}\sum_{t=1}^{T}(-f_t)e^{\gamma' R_t^e}}{\frac{1}{T}\sum_{t=1}^{T} e^{\gamma' R_t^e}} \equiv d(\gamma)$$

where the change of measure is evaluated at the γ.

Notice that

$$\begin{aligned} d'(\gamma_0) = &\left[\frac{1}{T}\sum_{t=1}^{T}(-f_t)R_t^{e'} e^{\gamma_0' R_t^e}\right]\left[\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_0' R_t^e}\right]^{-1} \\ &- \left[\frac{1}{T}\sum_{t=1}^{T}(-f_t)e^{\gamma_0' R_t^e}\right]\left[\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_0' R_t^e}\right]^{-2}\left[\frac{1}{T}\sum_{t=1}^{T} R_t^{e'} e^{\gamma_0' R_t^e}\right]. \end{aligned}$$

The term $\frac{1}{T}\sum_{t=1}^{T} R_t^{e'} e^{\gamma_0' R_t^e}$ after the minus sign goes to 0 as $T \to \infty$, since this converges to the population first order condition for γ_0. We use the following assumption:

Assumption A.2

$$\left[\frac{1}{T}\sum_{t=1}^{T}(-f_t)R_t^{e'} e^{\gamma_0' R_t^e}\right]\left[\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_0' R_t^e}\right]^{-1} \xrightarrow{p} D$$

where

$$D \equiv \mathbb{E}\left[(-f_t)R_t^{e'} e^{\gamma_0' R_t^e}\right] / \mathbb{E}\left[e^{\gamma_0' R_t^e}\right]$$

Thus an application of the Delta Method to $d(\widehat{\gamma}) \equiv \widehat{\lambda}$, yields:

$$\widehat{\lambda} = \lambda + d'(\overline{\gamma})(\widehat{\gamma} - \gamma_0). \tag{17}$$

Rearranging Eq. (17), multiplying by $\sqrt{T}$, and taking $T \to \infty$ we find:

$$\sqrt{T}\left(\widehat{\lambda} - \lambda\right) \xrightarrow{d} N\left(0, DB^{-1}\Omega\left(B^{-1}\right)' D'\right).$$

B GMM Estimation of Exponential Tilting

The methodology used in this paper used two sets of moment conditions to estimate two sets of parameters: the Lagrange multipliers, γ, and the prices of risk, λ. Though we did not use the Lagrange multipliers directly in any tests, it is necessary to estimate them since the estimate of the change of measure is a function of γ.

The moment conditions were the first-order conditions for the Lagrange multipliers:

$$\mathbb{E}\left[\sum_{t=1}^{T} R_t^e e^{\gamma' R_t^e}\right] = 0$$

and risk-premia moments:

$$\lambda = -\mathbb{E}\left[\sum_{t=1}^{T} f_t \frac{dQ}{dP}_t\right].$$

We can use these moment conditions in a GMM framework that includes the typical moment conditions used to estimate the factor loadings and risk premia in factor models. To remind ourselves of what these are, recall that the factor model can be written as:

$$R_{i,t+1} = R_f + \left(\sum_{j=1}^{N_f} \left(f_{j,t+1} + \lambda_j\right)\right) \beta_i' + \varepsilon_{i,t+1}$$

for each asset i. We impose:

$$\mathbb{E}\left[\varepsilon_{i,t}\right] = \mathbb{E}\left[f_{j,t+1}\varepsilon_{i,t+1}\right] = 0, \quad \forall i, j.$$

We will use the same $N = 80$ portfolios we used in the main empirical section in the body of the paper. As factors, we choose the three factors of Fama and French (1993). In total, we have N moment conditions from the first-order conditions for the Lagrange multipliers, N_f = number of factors (3 Fama-French factors in this example) moment conditions from the definition of factor risk premia, and $N + N_f N$ moments from the original GMM formulation of the factor model.

As a first check, we estimate the factor risk premia using only the moments implied by exponential tilting. That is, we have $N + N_f$ moments corresponding to the N Lagrange multipliers and the N_f factor risk premia.

We use two-step GMM where the initial weight matrix is the identity and the second weight matrix is the inverse of the moment covariance matrix estimated in the first step.

Panel A of Table 7 displays the estimated risk premia when the only moment conditions used are the first-order conditions for the Lagrange multipliers and moment conditions defining factor risk premia. The first two columns include the average excess returns of the three traded factors and t-statistics where the standard errors are corrected for correlation using the Newey-West procedure with 10 lags. We first notice that the estimated risk premia are close to the average excess returns. Panel B shows the estimated risk premia when we also include the standard GMM moment conditions for factor models. While the risk premia remain significant, we see that they are further from average excess returns in the first column. Note that λ, the factor risk premia, appear in all the moment conditions except for the Lagrange multipliers' first-order conditions. Thus, requiring them to satisfy the standard GMM moment conditions changes the estimates.

Table 7: GMM and Exponential Tilting

	Panel A: Exponential Tilting Moments Only			
Factor	Mean Excess Return	T-Stat	Estimated Risk Premium	T-Stat
RMRF	0.65	3.7311	0.7	42.7766
SMB	0.22	2.0968	0.24	16.2249
HML	0.39	3.3127	0.34	16.9321
	Panel B: Full Set of Moments			
Factor	Mean Excess Return	T-Stat	Estimated Risk Premium	T-Stat
RMRF	0.65	3.7311	0.56	71.3301
SMB	0.22	2.0968	0.38	32.3745
HML	0.39	3.3127	0.45	36.0392

Panel A shows the mean excess return and t-statistic on each of the three traded factors. Standard errors are corrected for correlations using the Newey-West procedure with 10 lags. The risk premia are estimated using two-step GMM where the moment conditions are the first-order conditions for the Lagrange multipliers and the moment condition defining the factor risk premia. Panel B estimates risk premia using the same moment conditions as in Panel A and also includes the moment conditions $\mathbb{E}[\varepsilon_{i,t}] = \mathbb{E}[f_{j,t}\varepsilon_{i,t}] = 0 \;\forall\, i, j$.