Semi-Parametric Estimation of Factor Risk Premia

Maziar M. Kazemi*

MIT Sloan

February 26, 2019

Abstract

This paper shows that factor risk premia can be consistently estimated using a semi-parametric estimate of the stochastic discount factor without requiring a correctly specified linear factor model. We use a minimum discrepancy objective function to construct a stochastic discount factor from asset returns using only the economic assumption of no arbitrage. The stochastic discount factor and factor risk-premia are estimated using only data on portfolio returns and factor realizations: The same data used when evaluating linear models. The econometrics are applications of standard extremum estimator arguments and the Delta Method, making inference simple. In simulations, the estimated risk-premia have low root mean squared errors and are comparable to classic two-pass estimates even when the model is correctly specified. Empirical estimates of popular traded factors are close to their mean excess returns. For non-traded factors, we find that intermediary leverage and consumption growth carry risk-premia, while employment growth does not. A final application shows that the estimated risk-premia can be used as an extra moment condition to discipline the creation of factor mimicking portfolios.

*kazemi@mit.edu. This paper has greatly benefited from conversations with the following people: Caio Almeida, Hui Chen, Stefano Giglio, Daniel Greenwald, Peter Hansen, Christian Julliard, Leonid Kogan, Anna Mikusheva, Jun Pan, Adrien Verdelhan. The author would like to thank seminar and conference participants at the following places: MIT Finance Lunch, SoFiE Summer School in Financial Econometrics at the University of Chicago, Yale SOM Finance Doctoral Conference, UMass Amherst Finance Brown Bag. All remaining errors are my own.
1 Introduction

Since the foundational papers of Merton (1973) and Ross (1976) developing the intertemporal CAPM (ICAPM) and arbitrage pricing theory (APT), (multiple) 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 or by looking at the average excess return of a “mimicking portfolio”. The coefficients in this framework correspond to the factor risk premia. It can also be shown that the factor risk premia, \(\lambda\), equal \(-R_f\mathbb{E}[mf]\), where \(m\) is a stochastic discount factor (SDF), \(f\) is the factor innovation, and \(R_f\) is the risk-free return. We will use the preceding moment condition to estimate factor risk premia.

The standard method to estimate linear factor models is to use “two-pass” or “-MacBeth” (Black et al. (1972), Fama and MacBeth (1973)) regressions. 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. In the presence of model-misspecification, e.g., when certain factors are mistakenly excluded, the estimates of loadings and risk-premia will be biased and inconsistent.

In this paper, we use a generalization of the Hansen-Jagannathan (Hansen and Jagannathan (1991), HJ) minimum variance SDF to estimate factor risk-premia in a way that is robust to model misspecification. In particular, we follow Almeida and Garcia (2012), Almeida and Garcia (2016), Ghosh et al. (2016), and Ghosh et al. (2018) and solve a problem similar to HJ’s but which ensures positivity of the SDF and does not focus on only the first two moments. Importantly, our method does not require extra economic assumptions aside from those already implied by HJ/APT. That is, we only use the no arbitrage restrictions (Euler equations) \(\mathbb{E}[R_i m] = R_f\), where \(R_i\) is the return on asset \(i\) and \(-R_f\mathbb{E}[fm] = \lambda\), to estimate the SDF.\(^2\) The problem is a constrained minimization problem where the con-

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

\(^2\)There is also a constraint on the mean of the SDF, but we assume the risk-free rate is observed, so this
straints correspond to the Euler equations. The Lagrange multipliers on these equations appear as the only parameters in the SDF, so our method is semi-parametric.

Our method is robust to misspecification because the moment condition defining the factor risk-premium depends only on the factor in question and a candidate SDF: The estimate is invariant to other factors we might (or might not) consider. We choose our SDF based on a number of considerations. First, the constrained problem which it solves embeds many previously used SDFs as special cases (e.g., HJ, Černý (2003), Snow (1991), and more). We use a specific form of the optimization problem which leads to an exponential tilting estimator (ET). This specific form corresponds to the Kullback-Liebler Information Criterion (KLIC) or relative entropy minimization. Second, the SDF is always positive without needing to impose non-negativity as an extra constraint. Third, the Lagrange multipliers have intuitive economic interpretations. Fourth, the optimization problem has been shown to lead to estimators with desirable econometric properties (Newey and Smith (2004), Kitamura (2006)). Fifth, subject to regularity conditions, the SDF assigns “reasonable” prices to new assets (that is, it satisfies “good-deal bounds” (Cochrane and Saa-Requejo (2000)). Finally, it requires no more information to evaluate than what would be needed to estimate the linear factor model in the first place. That is, a set of returns and set of factors provide us with all the variables we will need.\footnote{For example, one could construct an SDF from a general equilibrium model and evaluate the moment condition, but this somewhat defeats the purpose of reduced form linear models.}

In fact, the linearity of the model plays almost no role at all in the analysis. We can abstract from that functional form because our interest is in risk-premia and not factor loadings. More generally, we are not exploring how related each asset is to a source of risk: we are only after the price of risk. Any newly proposed risk-factor is typically asked two questions in a study: 1) Is there a premium for exposure to this factor? 2) Can the differences in exposure explain the differences in expected returns? These questions are often jointly answered but there is no a priori reason for this unless the researcher believes his or her asset pricing model is correct. That is, the question “Is this factor priced?” is not equivalent to constraint is easy to enforce.
"Is this factor significant in model \( a \) from the set of models \( M \)?" unless \( M \) is the universe of all possible models. By focusing on the first question, and not concerning ourselves with cross-sectional questions, we are able avoid many misspecification issues.

In simulation experiments, we assume we know the true linear factor model generating a set of returns. Even though the strength in our approach lies in its robustness to misspecification, we first consider a case where a researcher wanted to estimate factor risk-premia using traditional two-pass regressions using the correct specification. Even in this case, the performance of our ET estimator is comparable to correctly specified two-pass. In the situation where the researcher uses a factor model that is missing some priced factors, the performance of two-pass deteriorates, while ET still delivers approximately correct risk-premia. We run our simulation in situations where the number of assets is much smaller than the number of periods and where these two numbers are comparable. The results are qualitatively unchanged.

When we move to the empirical specifications, we start with traded factors. As mentioned above, the risk-premia of traded factors can in principle be estimated by simply looking at the time-series average excess return. This alternative estimation method provides us with a "sanity check" to which we can compare ET’s results. In terms of the classic Fama-French-Carhart (Fama and French (1993), Carhart (1997)) factors, ET delivers significant risk-premia that are close to the time-series averages. For the two new factors proposed in Fama and French (2015), we find that the profitability factor is not priced, but that the investment factor is. Finally, we include the market return squared as a constructed factor that might represent some form of skewness. It is not priced. In relation to two-pass, we find that ET’s estimates remain stable when we exclude certain factors or when we exclude groups of test assets (i.e., the assets used in construction of the ET SDF or used in the two-pass regressions), but that two-pass’s estimates vary quite a bit.

Since the Fama-French-Carhart factors are themselves excess returns, we can use them in the construction of the SDF. We show that the pricing implications of this ET SDF built from factors are similar to the ET SDF using all the assets. This result is encouraging, since
when we look at quarterly factors, we will want our methodology to be comparable to current practice in the literature, where researchers will typically project some macroeconomic factor onto a small set of portfolios or excess return factors, the so-called factor mimicking portfolio approach.

We move to quarterly data next, which corresponds to the frequency of the vast majority of macroeconomic and real factors proposed in the literature. Building on our previous experiments using the Fama-French-Carhart factors in the ET SDF, we construct a quarterly version to examine the risk-premia of the following factors: Leverage Factor (Adrian et al. (2014)), Consumption Growth (Breeden (1979), Lucas (1978)), Employment Growth (this is not a popular risk-factor, but it is an important macroeconomic indicator), and a “noise” factor as a final check (this factor is randomly generated to have mean and variance similar to consumption growth). We find the first two factors are priced. The Leverage Factor has an annualized risk-premium of around 2% per year. Consumption growth though statistically significant only has a risk-premium of around 0.2% per year. Thus, this seems to be in-line with the failure of the standard C-CAPM. If we wanted to explain the equity premium with consumption growth, the market “beta” on the factor would have to be around 6% (equity premium) / 0.2% ≈ 30! Employment growth and the noise factor are not priced.

As an application of the methodology, we consider imposing the the risk-premia calculated above as restrictions in the standard mimicking portfolio approach. That is, the mimicking portfolio approach is simply a linear projection, and we add one additional moment condition which says that the mean of the the mimicking portfolio excess returns should be the estimated risk-premium by ET. Note that we treat this risk-premium parametrically, since we do want the projection exercise to affect our ET estimations: we take the SDF as given, so to speak. We find that the estimated mimicking-portfolio risk-premia do get altered, though not by much.4 The risk-premia remain significant. Thus, the addition of the moment condition might be fruitful as a check on mimicking portfolio results.

Overall, our main contribution is to provide a methodology that can robustly estimate

---

4They were not drastically different from ET to begin with, in the case of these four factors at least.
factor risk-premia without imposing a specific linear factor model as our null hypothesis. This is particularly important for structural and theoretical asset pricing papers which propose a new priced component of the SDF and wish to test this empirically. Frequently, these models do not imply that, say, the Fama-French three factors should be included as controls in linear model, but they are included as controls regardless. In essence, the methodology proposed here allows the researcher to test the significance of his or her new factor in isolation.\footnote{Of course, the study of how a new factor changes the implications of existing linear models is interesting in its own right and could/should still be undertaken.}

1.1 Relation to the Literature

Rosenberg and Engle (2002), Jackwerth and Rubinstein (1996), and Aït-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 options into the Euler equations (Stutzer (1996)).

As mentioned above, the specific functional form of our objective function leads to relative entropy minimization. Thus, this paper is also related to the varied uses of entropy in the finance and economics literature. Stutzer (1995) first proposed the use of entropy to bound SDFs, and Bansal and Lehmann (1997) relate the "growth-optimal portfolio" and the lower entropy bound on the SDF. Other papers examining entropy "bounds" include Bakshi and Chabi-Yo (2012), Backus et al. (2014), Ghosh et al. (2016), and Liu (2018). None of these papers focus on estimating factor risk-premia. The paper by Ghosh et al. (2018) is most similar to this one. They look at the SDF constructed by relative entropy minimization as the single factor. They do not use this SDF to consider risk-premia of other factors.

This paper also relates to the now-massive literature on factor model econometrics and misspecification. Jagannathan and Wang (1998) showed that the estimated t-stats of the risk-premia of factors/characteristics can tend to infinity, even if the true risk-premia are zero.
in the presence of misspecification. Similarly, Kan and Zhang (1999a) and Kan and Zhang
(1999b) show related results when using GMM or two-pass with “useless” factors. There
is misleading inference even when the factors are not useless but merely weak (Kleibergen
(2009)). Hou and Kimmel (2010) argue that misspecification is in some ways even more costly
when using unspanned factors and suggest using mimicking portfolios instead of the factors
directly. Burnside (2015) shows that different normalizations of the SDF mean in linear
SDFs with excess returns can lead to identification problems. In response to these critiques,
many papers have sought to develop misspecification robust model selection and estimation
techniques (Kan and Robotti (2008), Kan et al. (2013), Bryzgalova (2015), Gospodinov
et al. (2017)). In our simulations, we will consider misspecification in the form of missing
or omitted factors, thus the most similar paper in this vein is Giglio and Xiu (2017). All
of these papers are firmly within in the linear model literature and so do not focus on the
moment condition defining risk-premia as we do. To the best of our knowledge, this is the
first application of these minimum discrepancy techniques to estimating factor risk premia.

2 Minimum Discrepancy and Motivation

This section borrows from Almeida and Garcia (2016) and Ghosh et al. (2016). We motivate
our measure of the stochastic discount factor (SDF) by showing it is the solution to a sensible
optimization problem and show that it possesses desirable properties.

2.1 Minimum Discrepancy SDF

The field of asset pricing is built around the SDF. Each proposed asset pricing model implicitly,
or explicitly, presumes a form for this object. One of the most popular diagnostics for
evaluating the admissibility of a proposed SDF is the HJ (Hansen and Jagannathan (1991))
bound. By admissible, we mean that the SDF must price a set of base assets. This bound

---

6 Gagliardini et al. (2018) provide a criterion for testing if there are omitted factors in the residuals of a linear factor model.
7 One can include positivity and other properties if he or she so wishes.
gives us a minimal variance which any SDF, with a given mean, must exceed in order to successfully price a set of test assets. Formally, the SDF is the solution to:

\[
\min_{\{m_t\}} \mathbb{E}[m_t^2] \quad \text{s.t. } \mathbb{E}[m_t R_t^e] = 0, \mathbb{E}[m_t] = a \tag{1}
\]

Here \( m_t \) is the SDF, \( R_t^e \) is a vector of excess returns, and \( a \) is the mean of the SDF, so if a risk-free asset exists, \( a = 1/R_f \). We do not include the non-negativity constraint on the SDF, but that can added at the expense of including indicator functions. The dual problem is typically easier to handle in the above optimization problem. From an empirical perspective, this problem will be semi-parametric: The estimated \( \hat{m}_t \)'s will be written as functions of Lagrange multipliers and the base assets, so the empirical implementation focuses on estimating the Lagrange multipliers.

The constraints are natural. They are, essentially, the defining properties of the SDF. But why should we be only interested in the second moment?\(^8\) Almeida and Garcia (2016) consider changing the discrepancy measure (Corcoran (1998)) from a quadratic function (which leads to minimal variance) to a general convex function, \( \phi \). In this way, one can take into account all higher moments of the SDF and not just the first two. The problem becomes:

\[
\min_{\{m_t\}} \mathbb{E}[\phi(m_t)] \quad \text{s.t. } \mathbb{E}[m_t R_t^e] = 0, \mathbb{E}[m_t] = a \tag{2}
\]

In particular, a popular choice is the Cressie-Read (Cressie and Read (1984)) (CR)\(^9\) divergence:

\[
\phi(m) = \frac{m^{\xi+1} - a^{\xi+1}}{\xi(\xi + 1)} \tag{3}
\]

The hyperparameter, \( \xi \), indexes different objective functions. For example, when \( \xi = 1 \), we get the HJ problem again.

To see how this discrepancy puts weight on all moments, consider a Taylor expansion

---

\(^8\)Recall, we fix the first moment.

\(^9\)For its use in econometrics, see Newey and Smith (2004) and the references therein.
around the mean, \( a: \)

\[
\phi(m) \approx \frac{1}{\phi'(a)} a^\xi (m - a) + \frac{a^\xi - 1}{\phi''(a)} (m - a)^2 + \frac{(\xi - 1)a^\xi - 2}{3! \phi'''(a)} (m - a)^3 + \ldots \tag{4}
\]

When \( \xi = 1 \), \( \phi^{(i)}(m) = 0 \) for all \( i \geq 3 \), where the superscript refers to the \( i \)th derivative of \( \phi(\cdot) \). So we can confirm, as claimed above, that \( \xi = 1 \) reduces to the HJ case. Importantly for our purposes, for \( \xi \leq 0 \), the weight put on every term of the expansion is non-zero. This general convex function, \( \phi(\cdot) \), allows a natural generalization of the HJ problem, and, as we will see, still retains tractability.

The hyperparameter \( \xi \) can be chosen by the economist. This paper will focus on the choice of \( \xi \to 0 \). This leads to the popular Exponential Tilting (ET) solution for \( m \) in the objective function. We will also briefly discuss the case \( \xi \to -1 \), which leads to Empirical Likelihood (EL) estimation.

By application of L'Hôpital's rule, one can show:

\[
\lim_{\xi \to 0} \phi(m) = m \ln m \quad \text{and} \quad \lim_{\xi \to -1} \phi(m) = \ln m
\]

The first expression is also known as the (negative of the) entropy of \( m \). The second expression leads to EL estimation which leads to non-parametric maximum likelihood (MLE). The solution to the problem equation (2) when \( \phi(m) = m \ln m \) is:

\[
m_t = \exp (\gamma' R_t^e + \alpha - 1) \tag{5}
\]

---

10\( \phi(a) = 0. \)

11Notice that once we pass the expectations operator through the approximation, the first term, corresponding to \( \phi' \) will vanish.

12There are other parameter selections that share this property, too.

13An extension not explored in this paper is a second minimization over the hyperparameter. This would involve finding an appropriate objective function, but is an interesting idea for future research.

14As mentioned in the Introduction, the objective function becomes what is known in the Information Theory literature as the Kullback-Leibler Information Criterion, or relative entropy. See the papers cited in the literature review which start directly with this objective function instead of motivating from minimum discrepancy, as we do here.
and when we solve the EL problem, we get:

$$m_t = \frac{1}{\gamma R_t + \alpha}$$  \hspace{1cm} (6)

Here $\gamma$ is the vector of Lagrange multipliers on the Euler equations, and $\alpha$ is Lagrange multiplier on the mean constraint. Previewing our estimation procedure, we know the solution for the Lagrange multipliers will satisfy the pricing constraints:

$$\mathbb{E}[m_t R_t^e] = 0 \text{ for each excess return, } R_t^e$$

and

$$\mathbb{E}[m_t] = \frac{1}{R_f}$$

These moment conditions will be used to estimate the Lagrange multipliers and, therefore, to give us a time series of estimated SDFs.\textsuperscript{15}

Because both objective functions involve the log of $m$ they will both deliver positive SDFs for any given sample of $R_t^e$. This is not true for the basic HJ problem, and imposing positivity on that problem reduces a great deal of the tractability. If we consider $m(R_t^e)$, it is clear that, being an exponential function, ET will always have $m(R_t^e) > 0$. In the case of EL, even though the SDF will positive for any given sample of excess returns which are used to estimate it, the function $m(R_t^e)$ can be negative. We do not restrict the sign of the Lagrange multipliers in our estimation procedure. There is no a priori reason to, and as we will see below, they can be related to optimal portfolio weights. Agents may very well choose to short certain portfolios, so we want to allow this. Ghosh et al. (2016) have shown that the SDF estimated by both problems are numerically quite similar and give similar prices to assets not used in their construction.\textsuperscript{16}

\textsuperscript{15}Technically, we using the dual optimization problem, instead of the original primal problem.

\textsuperscript{16}The correlation of the SDFs constructed in our paper using ET and EL are above 0.95, so we focus on ET here.
Recall that the SDF is proportional to the risk-neutral change of measure.\textsuperscript{17,18} Ghosh et al. (2016) also point out that EL has an interpretation as the non-parametric MLE of the risk-neutral density (subject to the constraints). We refer the reader to their paper and references therein for more exact derivations. Roughly, since $1/T$ is the non-parametric MLE of the natural probabilities, maximizing:

$$\frac{1}{T} \sum_{t=1}^{T} \ln m_t$$

is like finding the non-parametric MLE of $q_t$, where $q_t$ is the risk neutral density.

We will use the ET SDF in this paper because of the always-positive property\textsuperscript{19} and because of its acceptance in the finance literature.\textsuperscript{20} However, we re-emphasize that for any given sample, the solution is numerically similar to the EL problem which means ET is numerically close to MLE, a desirable property.

### 2.2 Interpreting the Lagrange Multipliers

Almeida and Garcia (2012) and Stutzer (1995) interpret the Lagrange multipliers in the solution of the minimum discrepancy problem. Almeida and Garcia (2012) consider the dual problem of the Cressie-Read discrepancy generally, while Stutzer (1995) focuses on ET. We provide a complementary interpretation related to elasticities of marginal utility.

Since, in equilibrium, the SDF is proportional to marginal utility, $MU_t$, ignoring constants of proportionality (e.g. the Lagrange multiplier on initial wealth):

$$MU_{t+1} = e^{\gamma' R_{t+1} + \alpha - 1}$$

\textsuperscript{17} $m/E[m]$ is a positive martingale with unit expectation.
\textsuperscript{18} We refer to the “real-world” probabilities as the “natural probabilities” (as opposed to the risk-neutral ones).
\textsuperscript{19} This is related the result in Almeida and Garcia (2016) that the dual problem for the Cressie-Read divergence is unconstrained in the ET case, when we impose positivity.
\textsuperscript{20} Stutzer (1995), Stutzer (1996), Kitamura and Stutzer (1997), Kitamura and Stutzer (2002), Ghosh et al. (2016), Ghosh et al. (2018).
Differentiate both sides with respect to excess return $i$:

$$\frac{\partial MU_{t+1}}{\partial R_{i,t+1}^e} = \gamma_i e^{\gamma_i R_{i,t+1}^e + \alpha - 1} = \gamma_i MU_{t+1}$$

Define $\varepsilon_i$ as the elasticity of marginal utility with respect to excess return $i$. We then see that:

$$\frac{\varepsilon_i}{R_{i,t+1}^e} = \gamma_i$$

This means that the Lagrange multiplier, $\gamma_i$, is large when the contribution of an asset, or excess return, to marginal utility is large, relative to the magnitude of the return. For example, an extremely large holding of asset $i$ in the investor’s portfolio will make consumption, and hence marginal utility, more sensitive to that asset’s returns.

This can be seen in both the constant absolute risk-aversion (CARA) case and the constant relative risk-aversion (CRRA) case. Consider a simple two-period model, where period 0 consumption is already determined. Therefore, the period 1 budget constraint is:

$$c_1 = (W_0 - c_0) \sum_{i=1}^{N} \pi_i R_{i,1}$$

where we have ignored non-asset income, $\sum_i \pi_i = 1$, and $R_{i,1}$ is the excess return of asset $i$ in period 1. For the CARA case:

$$U(c_1) = -\exp(-\alpha c_1)$$

$MU_1 = \alpha \exp(-\alpha c_1)$. We also note that $dc_1/dR_{i,1} = (W_0 - c_0) \pi_i$. Therefore,

$$\frac{\partial MU_1}{\partial R_{i,1}^e} = -\alpha MU_1 (W_0 - c_0) \pi_i$$

We can then calculate:

$$\varepsilon_i / R_{i,1}^e = -\alpha (W_0 - c_0) \pi_i = \gamma_i$$

(7)

21 The CARA case is explained in Stutzer (1995).
This example very clearly shows the connection implied above: An asset with large (in absolute value) portfolio weight will receive a large (in absolute value) Lagrange multiplier in the exponential tilting SDF.

For the CRRA case:

$$MU_1 = c_1^{-\rho}$$

and

$$\frac{\partial MU_1}{\partial R_{i,1}} = -\rho c_1^{-\rho-1} (W_0 - c_0) \pi_i$$

This leads to:

$$\frac{\varepsilon_t / R_{i,1}}{c_1 R_{i,1}} = -\rho (W_0 - c_0) \pi_i = \gamma_i$$

The numerator is the negative of the relative risk-aversion coefficient times the dollar amount invested in asset $i$. So, $\gamma_i$ is larger (in absolute value) if the dollar amount invested in the asset is large relative to its return.

Recall our earlier comments about the signs of the Lagrange multipliers. Notice that, for the both the CARA and CRRA cases, assets with positive portfolio weight, $\pi_i > 0$, will receive negative Lagrange multipliers. Returning to our discussion of EL versus ET, we now see that a period of large, positive returns which are held long in optimal portfolios may cause the function, $m^{EL}(R^e_t)$ to be negative if the sample $R^e_t$ had lower returns for those assets.

### 2.3 Good Deal Bounds

The SDFs derived by our discrepancy minimization problem, including the HJ problem, can also be used as bounds on admissible SDFs. That is, if an SDF has a value of $\phi(m)$ smaller than the minimized one, it must fail to satisfy some of the Euler equations.\(^{22}\) We may also ask about bounds on asset prices. That is, after having estimated an SDF using a set of test assets in the Euler constraints, what can we say about the price of new asset.\(^{22}\) Stutzer (1995)’s early application, for example, was concerned with bounds, not that actual SDF that achieves the minimum.

---

\(^{22}\)Electronic copy available at: https://ssrn.com/abstract=3103295
Saa-Requejo (2000) pursue this question directly.

They solve:

$$\min_m \mathbb{E}[mx^n] \quad \text{s.t.} \quad \mathbb{E}[mR] = 1, \quad \mathbb{E}[m] = a, \quad m \geq 0, \quad \mathbb{E}[m^2] \leq A$$

(9)

along with the corresponding max problem. Here $x^n$ is a new asset and $A$ is some upper bound chosen by the econometrician. Solving these problems lead to “good deal bounds,” on the price of new assets. Here we ask if our minimum discrepancy SDF under ET will assign a price within these bounds to a new asset.

It is straightforward to see that the only constraint we must worry about is the second-moment bound. We might try to modify our original problem by including the second moment constraint in equation (2). It can be shown that a solution does not exist for all levels of $A$. The first-order condition to our modified problem takes the form:

$$\ln m_t - \omega m_t = \chi_t$$

where $\chi_t$ is a function of constants and excess returns at $t$ and $\omega$ is the Lagrange multiplier on the second moment constraint. The left side of this equality is maximized at $m_t = 1/\omega$. Consequently, if $\chi_t > 1/\omega$, there will exist no $m_t$ satisfying all constraints. If the econometrician raises her choice of $A$, she will eventually find a solution. Of course, if the constraint does not bind, $\omega = 0$ and there exists a solution for any positive value of $\chi_t$.

Thus, assume that $A$ is chosen sufficiently high. Then the ET SDF (and the EL SDF) satisfy all the constraints and:

$$\mathbb{E}[mx^n] \leq \mathbb{E}[m^{md}x^n] \leq \mathbb{E}[\bar{m}x^n]$$

where $\bar{m}$ and $m$ are the solutions to the upper and lower good deal bound problems, and $m^{md}$ is the ET or EL minimum discrepancy SDF. This shows that our estimated SDF is “robust” to the exclusion of some assets, under certain conditions.
3 Factor Model Background

Factor models for asset returns take the form:

\[ R_{t+1} = \alpha + \beta' f_{t+1} + \varepsilon_{t+1} \]  \hspace{1cm} (10)

where \( R \) is an asset’s return, \( \alpha \) is an intercept, \( f \) is a vector of factors, \( \beta \) is a constant vector of factor loadings, and \( \varepsilon \) is idiosyncratic risk. The idiosyncratic risk is typically taken to be uncorrelated across factors, though some correlation is allowed under boundedness conditions (Chamberlain and Rothschild (1983)). 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 consumption growth.

First, take the expectation under the natural probability measure of equation (10):

\[ \mathbb{E}[R_{t+1}] = \alpha + \beta' \mathbb{E}[f_{t+1}] \]

where \( \mathbb{E} [\varepsilon_{t+1}] = 0 \) by assumption. Second, take the risk-neutral expectation of equation (10):

\[ R_f = \alpha + \beta' \mathbb{E}^Q [f_{t+1}] \]

where the assumptions of APT allow us to claim that \( \mathbb{E}^Q [\varepsilon_{t+1}] = 0 \), and the usual Euler equation implies \( \mathbb{E}^Q [R_{t+1}] = R_f \). Subtract the two equations from one another to get:

\[ \mathbb{E}[R_{t+1}] - R_f = \beta' (\mathbb{E}[f_{t+1}] - \mathbb{E}^Q [f_{t+1}]) \equiv \beta' \lambda \]  \hspace{1cm} (11)

Here, \( \lambda = \mathbb{E}[f_{t+1}] - \mathbb{E}^Q [f_{t+1}] \) is the factor risk premia. Since our empirical and simulation exercises use demeaned factors, \( \lambda = -\mathbb{E}^Q [f_{t+1}] = -R_f \mathbb{E}[m_{t+1} f_{t+1}] \), and we will use this definition in the sequel. Importantly, the term \( \mathbb{E}[m_{t+1} \varepsilon_{t+1}] \) is 0. The APT of Ross (1976) uses a diversification argument to show that idiosyncratic risk should receive 0 price.

The typical method of estimating \( \lambda \) 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 (10) and (11). For example, in a two-pass regression, one runs time-series regression (10) \( n \) times, where \( n \) is the number of assets in one’s sample. This generates a \( \hat{\beta}_i, i = 1, ..., n \), for each asset. Next, equation (11) is estimated cross-sectionally using these \( \hat{\beta}_i \)'s as regressors and average excess returns as regressands. The estimates from this cross-sectional regression, \( \hat{\lambda} \), are the estimated factor risk premia.\(^{23}\)

### 3.1 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 (10), affects equation (11) 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.\(^{24}\)

\[
R_t = \alpha + \beta_1 f_{1,t} + \varepsilon_t, \quad \text{where } \varepsilon_t = \beta_2 f_{2,t} + e_t.
\]

(12)

The Ordinary Least Squares (OLS) estimate of \( \beta_1 \) is:

\[
\hat{\beta}_1 = \frac{\text{Cov}(f_{1,t}, R_t)}{\text{Var}(f_{1,t})} = \frac{\sum_t f_{1,t} (\alpha + \beta_1 f_{1,t} + \beta_2 f_{2,t} + e_t)}{\sum_t f_{1,t}^2} = \beta_1 + \beta_2 \frac{\sum_t f_{1,t} f_{2,t}}{\sum_t f_{1,t}^2} + \frac{\sum_t f_{1,t} e_t}{\sum_t f_{1,t}^2}.
\]

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 \( \hat{\beta}_1 \) would be unbiased. However, when we estimate

\(^{23}\)Use of GMM allows one to bypass extra steps needed to correct the standard errors of the \( \hat{\lambda} \) estimate in a two-pass regression. The issue arises because the \( \hat{\beta}_i \)'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-covariance matrix.

\(^{24}\)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.
equation (11) using the above \( \hat{\beta}_1 \) we are introducing the omitted variable bias in two ways: First, by using an inconsistent and biased estimate of \( \beta_1 \), and second, by having omitted variables in equation (11), too. That is, \( \beta_2 \) 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.

Theoretically speaking, the issue can explained using the notation from above. If the omitted factors carry risk premia, then the condition \( \mathbb{E}[m_{t+1}\varepsilon_{t+1}] = 0 \) fails. There is some part of \( \varepsilon \) that receives non-zero price. Note that the natural expectation part, \( \mathbb{E}[\varepsilon_{t+1}] = 0 \), is trivial to ensure.

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:

\[
g_t = \xi + \eta f_t + v_t
\]  

(13)

where \( \xi \) is the measurement error, and \( v_t \) is the noise. For example, if there are 6 true factors and only 4 observed factors, then \( \eta \) is \( 4 \times 6 \). 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 \( \eta \) 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 (10), (11), and (13) 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.

In the simulations to follow, we will compare the estimates from Fama-MacBeth (FM), Giglio and Xiu (GX), and ET methods.

\[25\] Even if the factors do not carry risk premia, \( \text{Cov}(f, R) \neq \beta_1 \), still, so the issue remains, though the interpretation is different.
3.2 Discussion: Factor Models and Incomplete Markets

By examining the moment condition $-R_f E[mf]$ (and its sample counterpart), we see why our point estimate of the risk premium will be unaffected by the exclusion of other factors in the model. The only random variables appearing in these moments are the factor we wish to price and the change of measure/SDF. The latter, recall, is a function of a set of excess returns. Neither the factor whose risk premium we are estimating nor any other factor appears in this object. Even if the factors we are interested in are in excess return form (e.g. Fama-French factors), we should not include them in the SDF. The SDF is constrained to correctly price all assets in its formation, so the problem becomes trivial if we include the factors in the SDF.

Similarly, this moment condition is unaffected by the inclusion of other factors in the model. This is important because the previous misspecification literature (see references in the introduction) has shown that the inclusion of a weak factor can make estimation of loadings and premia of even strong, observed factors difficult. For example, Giglio and Xiu (2017) provide arguments and a modified method from their main one to handle weak latent factors (i.e., small eigenvalue principal components).

Consider a set $L$ of linear asset pricing models, all of which contain our random variable of interest, $f_t$. We are interested in whether $f_t$ carries a risk premium. For the present paper and discussion, we are not concerned about explaining the cross-section of returns or finding the “best” model. Given the results on factor model misspecification cited before, it is possible that every model in $L$ is misspecified, and, hence, our tests would lead to inconsistent estimates of $f_t$’s risk premium. In fact, depending on the form of misspecification, we might assign risk premia of different magnitudes, signs, and significances across models. All of this can happen regardless of whether $f_t$ is priced or not. Thus, we see the issue: “$f_t$ carries a risk premium” is not equivalent to “$f_t$ is a factor in linear model $l \in L$” yet in the case of non-traded factors, this has been the traditional way of testing the first clause. The first statement implies the second, but not vice versa, if the set of models $L$ does not include the universe of feasible models.
The previous comments apply to the point estimates of the risk premia. We will see in the simulation exercises that follow that these point estimates indeed do not change when we exclude factors. However, standard errors depend on the factors included, since the variance-covariance matrix of the extremum estimator depends on all the moment conditions used.

Finally, one should consider this method as an incomplete markets method. It is well known (see Back (2017)) that a linear factor model implies an SDF that is affine in the factors, which does not coincide with our SDF. A linear SDF in complete markets is difficult to reconcile with no arbitrage, as Dybvig and Ingersoll (1982) show, unless one is prepared to make assumptions about the range of values the factors can take or to impose strict assumptions on utility. Therefore, it is natural to consider an incomplete market setting, where our ET SDF is an approximation to one of infinitely many possible SDFs that agree on the asset span.

4 Estimation

We focus on the dual of problem (2). When there are $N$ excess returns and a risk-free return, the dual problem has $N + 1$ parameters to estimate: the Lagrange multipliers on the Euler equations and the Lagrange multiplier on the mean constraint. Conversely, the primal problem, for a time series of length $T$, has $T$ objects to estimate: the realization of $m_t$ for all $t = 1, 2, ..., T$.

In the special case of ET, the dual problem is:

$$\max_{\gamma, \alpha} \alpha \frac{R_f}{\hat{R}_t} - \mathbb{E} \left[ e^{\gamma' \hat{R}_t + \alpha - 1} \right]$$

(14)

This is a well-behaved concave program and so extremum estimation techniques can be fruitfully applied.\textsuperscript{26} Let $\gamma_1 = (\gamma', \alpha)'$ and “hatted” quantities denote estimates. Then:

$$\sqrt{T} (\hat{\gamma}_1 - \gamma_1) \xrightarrow{d} N(0, \Omega)$$

(15)

\textsuperscript{26}See Almeida and Garcia (2012) for more proofs.
where $\Omega$ is the asymptotic covariance matrix. Consistent estimators of it are simple to construct.

Let $m_t = \exp (\gamma'R_t^e + \alpha - 1)$. Then one can verify that the moment conditions needed to identify the parameters $\gamma$ and $\alpha$ are simply:

$$
\mathbb{E} [m_t R_i^e] = 0, \ i = 1, ..., N; \ \mathbb{E} [m_t] = \frac{1}{R_f}
$$

Using these moments, one follows the textbook arguments in, e.g. Hayashi (2000), to construct the variance-covariance matrix $\Omega$.\(^{27}\)

The estimate of the risk premium of a factor is:

$$
\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} m_t(\gamma_1) \left( -\frac{f_t}{R_f} \right)
$$

(16)

There are multiple ways to proceed. One could estimate equation (16) jointly with the Euler equations and mean constraint, since this equation is also a moment condition. Alternatively, one could apply the Delta Method, viewing $\hat{\lambda}$ as a function of $\gamma_1$. However, it is also worth noting that not all the Lagrange multipliers will be significant in the construction of the SDF.\(^{28}\) To reduce noise, it is fruitful to first eliminate constraints that do not bind (i.e., do not include excess returns that have insignificant Lagrange multipliers).

Once the subset of “important” assets has been identified, we use the Delta Method option to estimate the risk premia of the factors. Ultimately, this choice was made for economic rather than statistical reasons. Estimating the SDF separately from the risk premia ensures the same SDF is being used for each pricing test. We do not want risk premia moments to influence the moments identifying the Lagrange multipliers. In the appendix we display results where 1) Everything is estimated jointly without pruning the insignificant excess returns, and 2) The pruning is undertaken but the second step is still joint. The takeaway is that point estimates are stable while standard errors do change. The qualitative inference

\(^{27}\)We apply the method of Newey and West (1987) to estimate the sample counterpart of this matrix.

\(^{28}\)As we show in the empirical section, many are not significant at the 90% cutoff.
is stable.

## 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 goal here is to use a situation where the risk-premia of the factors and the number of factors generating returns is known to the econometrician.

We calibrate the true factors in the simulation to the four Fama-French-Carhart ([Fama and French](1993), [Carhart](1997)) model. These factors are the excess return of the market over the risk-free rate ($RMRF_t$), the excess return of small capitalization stocks over large capitalization stocks ($SMB_t$), the excess return of high book-to-market stocks over low book-to-market stocks ($HML_t$), and the excess return of stocks with highest returns in the past year over those with the lowest returns in the past year ($UMD_t$). The frequency is monthly, and the time period is 1927M1-2017M6.

In order to simulate the model, we must assign true betas in the factor model. We first download 80 portfolios from Ken French’s website: 25 sorted on size and book-to-market, 25 sorted on momentum, and 30 industries. We calculate the betas on these portfolios. From this matrix of betas, we calculate the mean beta on each factor as well as the four-by-four covariance matrix of betas across factors. Within each simulation, for each simulated asset, we generate that asset’s betas from a multivariate Normal distribution with mean and covariance matrix as calculated above.

Similarly, we calculate the time-series mean of each factor and the four-by-four covariance matrix of factor returns. Within each simulation we generate a time-series of length $T$, where $T$ will take different values as described below, with the factors being generated from a multivariate Normal distribution with mean and contemporaneous covariance matrix as described above.

---

29See Kenneth French’s data library for more detailed descriptions.
The number of assets generated in each simulation will be either 200 or 25. To investigate the role of omitted factors, we compare the estimation results of our three methods (FM, GX, ET) when all factors are observed and when only the first two factors (\(RMRF_t\) and \(SMB_t\)) are observed. The number of simulations in each experiment is 1000 and the value of \(T\) will be either 600 or 1000. Since the simulated factors are based on traded factors, we set the true risk-premia equal to the time-series averages of the factors.

Returns for asset \(i\) in each simulation are generated from:

\[
R_{i,t} = R_f + \sum_{k=1}^{4} \beta_{i,k} \left( \tilde{f}_{k,t} + \lambda_k \right) + u_{i,t}
\]  

(17)

where \(R_f\) is the risk-free rate, calculated as the mean risk-free rate over the sample period of the factors, \(\beta_{i,k}\) is asset \(i\)’s beta on factor \(k\), generated as described above, \(\tilde{f}_{k,t}\) is the demeaned return of factor \(k\), \(\lambda_k\) is factor \(k\)’s risk-premium, and \(u_{i,t} \sim N(0, 1/33)\) is a noise term.\(^{30}\)

Table (1), Panel A shows root mean squared errors (RMSEs) for two different pairs of periods, \(T\), and assets, \(N\), when all four factors are observed. The first four columns show results for \(T = 600\) and \(N = 200\). The final four columns show results for \(T = 1000\) and \(N = 25\). When both \(T\) and \(N\) are large and all four factors are observed, none of the methods performs significantly better or worse than the others. In fact, ET has higher RMSE in every case, but the differences are negligible. When \(N\) is small and \(T\) is large, the RMSEs are larger across methods, as expected. However, the deterioration in performance is smaller for ET. We see now that it outperforms three-pass for all factors and Fama-MacBeth for one factor. The main takeaway is, however, when all factors are observed, traditional Fama-MacBeth works well. Panel A of Table (2) shows the average bias for the same experiments as Table (1). The results are similar in that no method seems to be exceptional and Fama-MacBeth works fine.

Now consider Panel B of Table (1), where we have displayed results for the case where

\(^{30}\)We choose 1/33 as the variance since this make the volatility of the simulated returns and actual returns similar.
the final two factors, \( HML_t \) and \( UMD_t \), are omitted by the econometrician. As in the panel above, the first two columns refer to a large \( N \) and large \( T \) case, and the final two columns refer to the small \( N \) and large \( T \) case. First note that both three-pass and ET deliver identical RMSEs to the case where all factors are observed. This is to be expected, as the methods are designed to be robust to this type of misspecification. Note that Fama-MacBeth has larger RMSEs than in the fully-observed case. In the first two columns, Fama-MacBeth as larger RMSEs than both methods, and three-pass slightly outperforms ET. When \( N \) is small, we see ET outperforms the other two methods. All the methods have larger RMSEs when \( N \) is smaller, but ETs’ increases relatively less.

Panel B of Table (2) show average biases in the case where only the first two factors are observed. Once again, the results for three-pass and ET are identical to the case where all factors are observed. The bias of Fama-MacBeth has increased across all four columns, occasionally reaching around 1% per year.

Overall, these simulation exercises show that ET performs well in models with missing factors, and even if the factor model is correctly specified, ET performs comparably to Fama-MacBeth. Importantly, and differently from three-pass, ET’s performance is still quite good when the number of test assets is small.
Table 1: RMSE (Annualized Percentage)

Panel A: All Factors Observed

|                  | N=200, T=600 | N=25,T=1000 |
|------------------|--------------|-------------|
|                  | RMRF        | SMB         | HML         | UMB         | RMRF     | SMB     | HML     | UMB     |
| Three-Pass       | 0.277       | 0.284       | 0.65        | 0.873       | 1.338    | 1.488   | 3.691   | 8.374   |
| Fama-MacBeth     | 0.28        | 0.285       | 0.514       | 0.474       | 0.763    | 0.692   | 1.15    | 1.163   |
| ET               | 0.284       | 0.365       | 0.645       | 0.773       | 0.588    | 0.711   | 1.967   | 3.742   |

Panel B: Two Factors Unobserved

|                  | N=200, T=600 | N=25,T=1000 |
|------------------|--------------|-------------|
|                  | RMRF        | SMB         |             |             | RMRF     | SMB     |         |         |
| Three-Pass       | 0.277       | 0.284       |             |             | 1.338    | 1.488   |         |         |
| Fama-MacBeth     | 0.558       | 1.066       |             |             | 1.518    | 1.787   |         |         |
| ET               | 0.284       | 0.365       |             |             | 0.588    | 0.711   |         |         |

This table shows the annualized, in percentage, root mean squared error (RMSE) of the factor risk premia as calculated by the three methods: three-pass, Fama-MacBeth, and ET. Panel A displays results for the case where all four factors are observed. The first four columns of this panel display results for simulations with $N = 200$ and $T = 600$, and the final four columns display results for when $N = 25$ and $T = 1000$. Panel B displays results for when only the first two factors are observed. The first two columns of this panel display results for simulations with $N = 200$ and $T = 600$, and the final two columns display results for when $N = 25$ and $T = 1000$. 

Electronic copy available at: https://ssrn.com/abstract=3103295
Table 2: Average Bias (Annualized Percentage)

Panel A: All Factors Observed

|                  | N=200, T=600 | N=25, T=1000 |
|------------------|--------------|--------------|
|                  | RMRF | SMB | HML | UMB | RMRF | SMB | HML | UMB |
| Three-Pass       | -0.023 | 0.08 | -0.448 | -0.763 | -0.412 | 0.88 | -3.492 | -7.809 |
| Fama-MacBeth     | -0.014 | -0.011 | -0.07 | -0.173 | 0.016 | -0.004 | -0.065 | -0.03 |
| ET               | -0.008 | 0.05 | -0.312 | -0.555 | -0.17 | 0.36 | -1.809 | -3.599 |

Panel B: Two Factors Unobserved

|                  | N=200, T=600 | N=25, T=1000 |
|------------------|--------------|--------------|
|                  | RMRF | SMB | RMRF | SMB |
| Three-Pass       | -0.023 | 0.08 | -0.412 | 0.88 |
| Fama-MacBeth     | -0.284 | 0.925 | -0.463 | 0.939 |
| ET               | -0.008 | 0.05 | -0.17 | 0.36 |

This table shows the annualized, in percentage, average bias of the estimates factor risk premia as calculated by the three methods: three-pass, Fama-MacBeth, and ET. Panel A displays results for the case where all four factors are observed. The first four columns of this panel display results for simulations with \( N = 200 \) and \( T = 600 \), and the final four columns display results for when \( N = 25 \) and \( T = 1000 \). Panel B displays results for when only the first two factors are observed. The first two columns of this panel display results for simulations with \( N = 200 \) and \( T = 600 \), and the final two columns display results for when \( N = 25 \) and \( T = 1000 \).

6 Empirical Results

6.1 Monthly Factors

In this first subsection, we examine the estimated risk-premia of seven monthly factors. Six of these factors are traded (i.e., portfolios) factors, and one is recently proposed non-linear transformation of a traded factor. We compare the results using ET to Fama-MacBeth.\(^{31}\)

For the traded factors, we estimate the factor risk premia for the four factors used in the previous section: \( RMRF, SMB, HML, \) and \( UMD \), as well as the two new factors from Fama and French (2015) corresponding to operating profitability (\( RMW_t \)) and investment (\( CMA_t \)). The non-traded factor is the square of the market return, which would deliver the co-skewness of the market with the SDF as a price of risk, if it was priced. Our selection of

\(^{31}\)The results for three-pass are similar to ET, and are excluded for space.
traded factors is because 1) They have been shown to adequately explain the cross-section of expected returns, though not perfectly, of course; and 2) their risk premia should be approximately equal to the mean excess returns, which we can estimate by historical time-series average. We choose the square of the market to demonstrate the ease of estimating more complicated risk-premia. In this case, the mean of the factor return does not provide an estimate of the true risk premium.

We use monthly data from 1967M7-2017M6. We use 135 sorted portfolios from Kenneth French’s website as test assets. For ET, these will be used to construct the semi-parametric SDF. For Fama-MacBeth, these will be used in the two-pass procedure. The portfolios sorts are: 10 size, 10 book-to-market, 10 dividend-price, 25 size by book-to-market, 25 size by momentum, 10 momentum, 25 size by short-term reversal, 10 short-term reversal, 10 industries.

As mentioned in the estimation section, we first estimate the Lagrange multipliers on the 135 excess return portfolios. We then find those with with (absolute value of) t-statistics greater than $\Phi^{-1}(0.9)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function (cdf).\(^{32}\) This leaves 84 excess return portfolios.\(^{33}\) For consistency, we also estimate the FM risk-premia with the same portfolios.

It is interesting to look at which portfolios have extreme gammas, that is, which gammas are very large in absolute value. Seven of the top 10 largest gammas are on portfolios with some sort of “momentum” aspect (e.g., size and momentum, size and short-term reversal, short-term reversal on its own). Five of the bottom 10 smallest gammas are on portfolios with some sort of momentum aspect. Recall that the signs of the gammas are the opposite of what the approximate portfolio holdings are (see Section 2). Thus, it seems investors take long-short positions in momentum based portfolios.

Table (3) displays the five largest and smallest gammas and their associated portfolio names. For intuition, we also provide the implied portfolio weight for a CARA investor with

\(^{32}\)Results are robust to lower/higher cutoffs.

\(^{33}\)Note that the Lagrange multiplier corresponding to the mean constraint is always included, so that there are 85 parameters in the SDF.
risk-aversion equal to 3 and $100 invested (see equation (7)). The number of the portfolio refers to its location in the sort. For example, 6 in a single sort, means the 6th decile portfolio. In double sorts on size, if we think of a 5x5 matrix, with size quintiles counting rows and the second variable’s quintiles counting columns, then the numbering is “column-wise.” For example, the number 12 means 3rd quintile for size and 2nd quintile for the second variable.

The size and short-term reversal sort appears in both the set of largest and smallest Lagrange multipliers. For the smallest gamma portfolios, the 23rd and 24th portfolio appear. These are in final quintile of size, which means these are the largest stocks. Similarly, these stocks are also in the final quintile of the short-term reversal sort, so these are stocks that performed well recently. Curiously, the 5th portfolio in the size and short-term reversal sort is in the set of the largest gammas. This is the 1st quintile of size, which means these are small stocks. However, this is also the last quintile for short-term reversal, meaning these stocks performed well recently. Consequently, we can say by rule-of-thumb that the size effect seems to dominate. We leave drawing more definite conclusions for future research.
Table 3: Largest and Smallest Lagrange Multipliers

Panel A: Five Largest Gamma Portfolios

| Name                           | Number | Gamma       | Implied Portfolio Weight |
|--------------------------------|--------|-------------|--------------------------|
| Short-Term Reversal            | 6      | 71.38788    | -0.237959587             |
| Size                           | 4      | 54.1692     | -0.180564006             |
| Size and Book-to-Market        | 22     | 52.99484    | -0.176649473             |
| Short-Term Reversal            | 5      | 48.00274    | -0.160009135             |
| Size and Short-Term Reversal   | 5      | 46.74275    | -0.155809173             |

Panel B: Five Smallest Gamma Portfolios

| Name                           | Number | Gamma       | Implied Portfolio Weight |
|--------------------------------|--------|-------------|--------------------------|
| Size and Short-Term Reversal   | 23     | -72.299     | 0.240996526              |
| Size and Momentum              | 24     | -66.366     | 0.221220084              |
| Book-to-Market                 | 3      | -46.5175    | 0.155058291              |
| Size and Short-Term Reversal   | 24     | -45.5634    | 0.151878059              |
| Book-to-Market                 | 4      | -41.673     | 0.13891012               |

This table shows the portfolios that are weighted most heavily in the construction of the SDF. Each panel prints the name of the portfolio as defined in Kenneth French’s website. The number of the portfolio refers to its location in the sort. For example, 6 in a single sort, means the 6th decile portfolio. In double sorts on size, if we think of a 5x5 matrix, with size quintiles counting rows and the second variable’s quintiles counting columns, then the numbering is “column-wise.” For example, the number 12 means 3rd quintile for size and 2nd quintile for the second variable. The third columns displays the estimated gamma (Lagrange multiplier). The final column displays the implied portfolio weight for a CARA investor with risk-aversion 3 and invested amount $100.
Table 4: Risk-Premia of Monthly Factors

### Risk-Premia: All Seven Factors

| Factor      | $\widehat{E}[f]$ | S.E. | T-Stat | FM RP | S.E. | T-Stat | ET RP | S.E. | T-Stat |
|-------------|------------------|------|--------|-------|------|--------|-------|------|--------|
| Market      | 6.089            | 2.282| 2.668  | 2.572 | 3.493| 0.736  | 6.448 | 4.066| 1.586  |
| Size        | 2.772            | 1.61 | 1.722  | 2.854 | 1.525| 1.871  | 3.214 | 2.201| 1.46   |
| Value       | 4.455            | 1.633| 2.729  | 3.593 | 1.411| 2.546  | 5.098 | 2.726| 1.87   |
| Momentum    | 8.246            | 2.096| 3.934  | 8.39  | 2.092| 4.01   | 8.495 | 5.157| 1.647  |
| Profitability| 3.009           | 1.292| 2.33   | 6.314 | 1.802| 3.504  | 0.524 | 1.828| 0.287  |
| Investment  | 3.846            | 1.121| 3.429  | -3.148| 1.758| -1.791 | 2.494 | 1.62 | 1.539  |
| Market$^2$  |                 |      |        |       | 2.776| 0.648  | 4.282 | -0.218| 0.353 | -0.616 |

### Risk-Premia: Fama-French Three Factors Only

| Factor      | $\widehat{E}[f]$ | S.E. | T-Stat | FM RP | S.E. | T-Stat | ET RP | S.E. | T-Stat |
|-------------|------------------|------|--------|-------|------|--------|-------|------|--------|
| Market      | 6.221            | 2.265| 2.746  | 1.947 | 3.386| 0.575  | 6.581 | 3.994| 1.648  |
| Size        | 2.756            | 1.597| 1.726  | 2.138 | 1.529| 1.399  | 3.2   | 2.178| 1.469  |
| Value       | 4.617            | 1.624| 2.844  | 2.167 | 1.45 | 1.494  | 5.191 | 2.743| 1.893  |
| Market$^2$  |                 | -2.716| 0.905 | -3    | -0.199| 0.342  | -0.582 |      |        |

This table shows the risk-premia of selected monthly factors as estimated by Fama-MacBeth and ET. Panel A shows results when we include all seven factors in the estimation. Panel B uses only the Fama-French Three Factors (and the market squared). For traded factors, we display the mean excess return, standard error, and t-stat. Fama-MacBeth standard errors and t-stats are constructed as described in Fama and MacBeth (1973). For ET we use the methods described in the Estimation section of this paper. There are 618 observations in Panel A and 624 in Panel B. All risk-premia and standard errors are in percentage, annualized.

Table 4 displays results for estimating the factor risk-premia of our monthly factors using both FM and ET. Panel A displays results for the five Fama-French factors, the momentum factor, and the market factor squared. For the traded factors, we also display the mean excess return along with its standard error and t-statistic. The FM risk-premia and standard errors are calculated as described in Fama and MacBeth (1973). The ET risk-premia and standard errors are calculated as described in the Estimation section of this paper. Panel B displays results for when we only include the original Fama and French (1993) three factors and the market squared. In terms of FM, this means the other factors are not included in the time-series first-stage, and hence their betas are not included in the cross-sectional second stage. For ET, this means that when we perform the Delta Method, the function $\lambda(\gamma_1) = -\widehat{E} [F_t m_t(\gamma_1)]$ does not include the omitted factors in $F_t$.\(^{34}\)

\(^{34}\)As mentioned, this affects standard errors through the Jacobian of the lambda function.

29
Looking at Panel A, we see that ET estimates risk-premia that are quite close the time-series averages for the first four factors. All four of them are significant at the 90% level as well. While FM does well with the size, value, and momentum factors, perhaps surprisingly, it is quite far off on the market factor. Neither method estimates a risk premium close to the time-series average for profitability, but note that ET does not find this to be a priced (significant) factor anyway, so it is not surprising that the point estimate is nearer zero. FM assigns a significant risk-premium. The investment factor receives a significant and negative risk-premium from FM, while ET assigns a positive and significant risk-premium, though the point estimate is further from the time-series average than for other factors. Finally, ET does not find any significant risk-premium for the market squared, whereas FM does.

As expected in Panel B, the point estimates for ET are practically unchanged.\(^{35}\) Also, though we know the standard errors will change, they do not change by very much, and all significances (or lack thereof) are maintained. The FM estimates, on the other hand, have changed by quite a bit (more than a percentage point in some cases). The risk-premium for the market is still far from the time-series average and the market squared is still significant, but with a flipped sign.

In anticipation of the next subsection, we also consider a different form of omission. Table (5) shows estimated risk-premia for the same factors as above, except now we exclude the 25 size and book-to-market portfolios in our set of base assets. Thus, the total number of Euler equations is 110 now. The key takeaway is that estimated risk-premia and significances are almost unchanged from the original sample results. Curiously, the value factor seems to be priced “better” than before, but note that the 95% confidence intervals do overlap, so one should not think the performance has necessarily improved. The profitability factor now has a negative estimated risk-premium, but it is still insignificant, so this is immaterial.

The reason this experiment motivates the next subsection is that when we look at quarterly factors (all of which are non-traded), we will take the path of most empirical researchers and use fewer test assets. In fact, the more common procedure for quarterly, non-traded fac-

\(^{35}\)As noted in the table notes, the sample sizes are slightly different due to data availability. If they were the same, the point estimates would be unchanged.
tors is to form mimicking portfolios by projecting the factors onto a (small) set of assets or traded factors.\textsuperscript{36}

Table 5: Estimated Risk-Premia Excluding 25 Size and Book-to-Market Portfolios

| Factor     | $\hat{E}[f_t]$ | S.E. | T-Stat | ET RP | S.E. | T-Stat |
|------------|----------------|------|--------|-------|------|--------|
| Market     | 6.089          | 2.282| 2.668  | 6.285 | 3.697| 1.7    |
| Size       | 2.772          | 1.61 | 1.722  | 2.76  | 2.051| 1.345  |
| Value      | 4.455          | 1.633| 2.729  | 4.317 | 2.492| 1.732  |
| Momentum   | 8.246          | 2.096| 3.934  | 8.195 | 5.475| 1.497  |
| Profitability | 3.009    | 1.292| 2.33   | -0.091| 1.455| -0.062 |
| Investment | 3.846          | 1.121| 3.429  | 2.905 | 1.391| 2.088  |
| Market$^2$ | 0              | 0    | 0      | -0.17 | 0.236| -0.721 |

This table shows the risk-premia of our selected monthly factors when the set of test assets does not include the 25 size and book-to-market sorted portfolios. That is, in the construction of the SDF, those portfolios are not included as constraints.

6.2 Reducing the Cross-Section

When we move to quarterly data, we will drastically reduce $N$, the number of test assets used in estimation of the SDF. We will use three portfolios as test assets: the Fama-French three portfolios, $RMRF$, $SMB$, and $HML$.\textsuperscript{37} This choice is due to two reasons. First, we want to showcase that we can drastically reduce the cross-section, and still achieve good results. That is not to say a researcher should simply throw away data. Second, these portfolios are easily available at quarterly frequency. In this section however, we will stick to monthly frequency, but will construct the SDF from those three portfolios.

Note that these portfolios were previously factors, but since they are excess returns, they can be used in the Euler equations. Importantly, it is vacuous to now try to estimate the risk-premia: We have constructed the SDF to price them perfectly, so to speak. In general,

\textsuperscript{36}See discussions in Balduzzi and Robotti (2008), Breeden et al. (1989), Huberman et al. (1987), and Kleibergen and Zhan (2018).

\textsuperscript{37}We do not perform our two-step procedure, since we are not worried about having many insignificant Lagrange multipliers. Interestingly, the size factor is not significant. Thus, a three parameter SDF (market, value, and mean constraint) suffices Regardless, results are robust to including or excluding size.
the technique used here works for any factor whose price is known. If the econometrician knows $E[m_t f_t]$ then he or she may include it in the Euler equations.

We also construct the SDF as in the previous section: Using our full 135 portfolios. In order to compare the pricing of the two SDFs, we take advantage of the fact that not all portfolios are used in the construction of the first SDF. That is, recall that we only retain a subset of assets with significant Lagrange multipliers. The remaining assets (50 of them) are then what we will seek to “price.” In general, we know that:

$$E[m_R] = 1$$

for any return. Call our original SDF $m$ and our new, reduced-$N$ SDF $m_N$. Ideally, we would have:

$$E[m_R] = E[m_N R] = 1$$

or:

$$E[R(m - m_N)] = 0$$

We form the sample version of the above moment conditions. Let $g_t(R_t, m_t, m_{N,t})$ be the $50 \times 1$ of sample means of $R_{i,t} (m_t - m_{N,t})$ for $i = 1, ..., 50$. Form the Wald statistic:

$$T g_t(R_t, m_t, m_{N,t})' \hat{S}^{-1} g_t(R_t, m_t, m_{N,t})$$

where $\hat{S}$ is the sample estimate of the long-run covariance matrix of $\sqrt{T} g_t$ and $T$ is the time-series length.\(^{38}\) Note that these moments were not used in the construction of the SDFs, so there is no need for a degrees of freedom adjustment. For simplicity, we ignore the sampling variation in the Lagrange multipliers and treat the SDFs as given. Then, under the null, the above Wald statistic is $\chi^2(50)$.

The 95% critical value for this distribution is 67.5. The value of our Wald statistic is 4.3517. We easily fail to reject the null that moments are zero. Let $c^*$ be the value of our

\(^{38}\)We use the Newey-West method to calculate $\hat{S}$.
Wald statistic and \( \chi \sim \chi^2(50) \). Then, \( \Pr(\chi \geq c^*) = 0.6986 \).

Note that it is possible that both SDFs assign some systematic bias to a particular set of stocks, but this bias must be the same for both SDFs. We are testing if the pricing is the same, not if the pricing is exactly correct for both. For example, consider the exact same procedure but replace the moments with \( \mathbb{E}[Rm - 1] \), that is, a test that our original SDF prices all excluded assets correctly. With an abuse of notation, we have \( \Pr(\chi \geq c^*) = 0.9999 \). For the reduced-\( N \) SDF, \( \Pr(\chi \geq c^*) = 0.4898 \). Naturally, the larger \( N \) SDF has smaller Wald statistic, but neither Wald statistic leads to a rejection of the null.

6.3 Non-Traded Factors

In this subsection, we will examine three non-traded factors. The first is consumption growth, which forms the backbone of consumption based asset pricing models (Breeden (1979), Lucas (1978)). The second is the leverage factor of Adrian et al. (2014). We choose this factor because Kleibergen and Zhan (2018) use it as their test case to show how mimicking portfolios can lead to spurious significance in risk-premia for a non-traded factor when that factor has low covariance with the base assets. The final factor is employment growth, which is less connected with asset pricing theory, but is an important macroeconomic aggregate. We also test a noise factor: A randomly generated sequence of numbers, which should carry no risk price.

We construct our semi-parametric SDF using the quarterly Fama-French factors from Ken French’s website. For the risk-free rate, we compound the monthly risk-free rate used in the previous subsection. Consumption is real non-durable per capita consumption, seasonally adjusted from FRED. The leverage factor comes from Tyler Muir’s website. Employment is the growth rate of non-farm payrolls from FRED. The noise factor is Gaussian with mean and variance equal to that of consumption growth, 0.00321 and 0.00782, respectively.
Table 6: Risk-Premia of Non-Traded Factors

Panel A: ET Estimated Risk Premia

| Risk Premium (%, Annualized) | S.E (%, Annualized) | T-Stat | Obs. |
|-----------------------------|---------------------|--------|------|
| Cons. Growth                | 0.2                 | 0.069  | 2.909| 281  |
| Leverage                    | 2.211               | 1.022  | 2.164| 198  |
| Emp. Growth                 | 0.036               | 0.04   | 0.902| 313  |
| Noise                       | -0.024              | 0.025  | -0.945| 362  |

Panel B: Risk Premia from Mean Excess Return of Projection

| Risk Premium (%, Annualized) | S.E (%, Annualized) | T-Stat | Obs. |
|-----------------------------|---------------------|--------|------|
| Cons. Growth                | 0.165               | 0.05   | 3.319| 281  |
| Leverage                    | 1.795               | 0.998  | 1.798| 198  |
| Emp. Growth                 | -0.006              | 0.082  | -0.068| 313  |
| Noise                       | -0.026              | 0.034  | -0.77 | 362  |

Panel C: Risk Premia from Constrained Projection

| Risk Premium (%, Annualized) | S.E (%, Annualized) | T-Stat | Obs. |
|-----------------------------|---------------------|--------|------|
| Cons. Growth                | 0.187               | 0.031  | 5.993| 281  |
| Leverage                    | 2.053               | 0.547  | 3.751| 198  |
| Emp. Growth                 | 0.028               | 0.027  | 1.056| 313  |
| Noise                       | -0.025              | 0.021  | -1.198| 362  |

This table shows the risk-premia of selected non-traded factors as estimated by ET and as the mean excess return on the projection of the factor on the Fama-French portfolios. Panel A shows results for ET. The first column shows the annualized in percentage risk premium of the factor. The second and third columns show the (annualized) standard error and t-stats, which were calculated with the bootstrap. The final column lists the number of quarterly observations for each factor. Panel B displays the same information for the projection of the factors onto the Fama-French factors. Panel C displays risk premia from a projection onto the Fama-French factors where an additional moment condition is added to bring the projection risk premium close to the ET estimated risk premium.

We choose the longest sample possible for each of the quarterly factors, and the number of observations varies from factor to factor. Table (6) Panel A shows our results on estimating the risk premia of the aforementioned factor candidates using the ET SDF. Standard errors are constructed using the same method as in previous sections. The rightmost column shows the number of quarterly observations used for each factor.

The first two factors, consumption and leverage, are priced. It is interesting to note that, though significant, the risk premium on consumption growth is economically small at 0.2% per year. Intermediary leverage delivers a larger risk premium at around 2% per year. Both
risk-premia are positive, reflecting common economic theory. When consumption growth is high, marginal utility is low, so assets which covary with consumption growth should covary negatively with the SDF. Similarly, intermediary leverage increases during “booms.” Adrian et al. (2014) show that intermediary leverage is positively related to intermediary assets, which is opposite the behavior of households. Rising intermediary leverage is a signal of rising intermediary assets, e.g., a credit boom.

Employment growth is unpriced. This is somewhat expected since employment growth is not traditionally linked to asset pricing. It is also reassuring as a test that ET does not simply assign positive risk-price to all or most macro-aggregates: ET does seem to allow us to distinguish between competing factors. Finally, as a final test, we see that the noise factor is unpriced as well.

Panel B shows results from using a projection on the Fama-French factors. That is, let $f_t$ be our factor of interest, and let $F_t$ be the vector of Fama-French returns. We estimate:

$$f_t = a + b'F_t + v_t$$

Let $\hat{a}$ and $\hat{b}$ be the fitted values from an OLS regression. We know the definition of the risk-premium of $f$ is $-R_f\mathbb{E}[m[f - \mathbb{E}f]]$. Applying this transformation to $\hat{f}_t = \hat{a} + \hat{b}'F_t$ we find:

$$\text{Risk Premium (} f_t \text{)} \approx \hat{b}'\mathbb{E}[F_t]$$

The left-side of this expression is displayed in Panel B. To calculate standard errors we estimate the risk premium jointly with OLS. That is, we append the following moment to the usual OLS moments in GMM:

$$\mathbb{E}[^{\hat{\lambda}} - b'F_t] = 0$$

where $\hat{\lambda}$ will be the estimated risk-premium from linear projection. Panel B shows that ET and linear projection agree on which factors are priced, but that the risk premium assigned
by projections is always a bit smaller than that assigned by ET.

As an interesting application to bring new information into the linear projection, we take the risk-premium as given by ET. Thus, instead of appending the above moment condition to OLS, we append:

$$\mathbb{E}[\lambda_{ET} - b'F_t] = 0$$

where $\lambda_{ET}$ is the risk-premium estimate from ET, e.g., from Panel A. Importantly, we do not jointly estimate ET and the projection. We do not want our desire to minimize projection errors to influence our estimate of the risk premium. The results of this experiment are presented in Panel C. Not surprisingly, the risk premia estimates are in between the pure ET and pure projection ones. However, more importantly, the J-test of overidentifying restrictions fails to reject the null ($J = 0.0054$, #overidentifying restrictions = 1). What this tells us is that we do not have to sacrifice the pricing properties of the un-spanned factor when we use projections (at least for set of factors considered here).

7 Conclusion

This paper has developed and evaluated a new method for estimating factor risk premia. The theory is based around finding an SDF which solves a minimum discrepancy problem, using only the test assets (i.e., equity portfolios) in its construction. We call this the ET SDF. The ET SDF can then be used to estimate the risk-premium of a proposed factor by exploiting the definition of the risk-premium: $-R_f\mathbb{E}[mf]$. These moment conditions are valid whether or not the model is missing certain factors. 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 close to their observed average.

---

39 One could do everything jointly with appropriately partitioned weighting matrices.

40 Of course, betas can be quite different: see Kleibergen and Zhan (2018).
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.

References

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

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

Almeida, C. and Garcia, R. (2012). Assessing misspecified asset pricing models with empirical likelihood estimators. *Journal of Econometrics*, 170(2):519–537.

Almeida, C. and Garcia, R. (2016). Economic implications of nonlinear pricing kernels. *Management Science*, 63(10):3361–3380.

Back, K. (2017). *Asset pricing and portfolio choice theory*. Oxford University Press.

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

Bakshi, G. and Chabi-Yo, F. (2012). New entropy restrictions and the quest for better specified asset pricing models. *Journal of Financial and Quantitative Analysis*, pages 1–47.

Balduzzi, P. and Robotti, C. (2008). Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business & Economic Statistics*, 26(3):354–368.
Bansal, R. and Lehmann, B. N. (1997). Growth-optimal portfolio restrictions on asset pricing models. *Macroeconomic dynamics*, 1(2):333–354.

Black, F., Jensen, M. C., and Scholes, M. (1972). The capital asset pricing model: Some empirical tests. *Studies in the theory of capital markets*, 81(3):79–121.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7(3).

Breeden, D. T., Gibbons, M. R., and Litzenberger, R. H. (1989). Empirical tests of the consumption-oriented capm. *The Journal of Finance*, 44(2):231–262.

Bryzgalova, S. (2015). Spurious factors in linear asset pricing models. *LSE manuscript*.

Burnside, C. (2015). Identification and inference in linear stochastic discount factor models with excess returns. *Jnl of Financial Econometrics*, 14(2):295–330.

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

Černý, A. (2003). Generalised sharpe ratios and asset pricing in incomplete markets. *Review of Finance*, 7(2):191–233.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, pages 1281–1304.

Cochrane, J. H. (2005). Asset pricing (revised edition).

Cochrane, J. H. and Saa-Requejo, J. (2000). Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of political economy*, 108(1):79–119.

Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*, pages 967–972.
Cressie, N. and Read, T. R. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):440–464.

Dybvig, P. H. and Ingersoll, J. E. (1982). Mean-variance theory in complete markets. *Journal of Business*, pages 233–251.

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

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*.

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

Gagliardini, P., Ossola, E., and Scaillet, O. (2018). A diagnostic criterion for approximate factor structure. Technical report, Swiss Finance Institute Research Paper.

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

Ghosh, A., Julliard, C., and Taylor, A. P. (2018). An information-theoretic asset pricing model. Technical report, Working Paper, London School of Economics.

Giglio, S. and Xiu, D. (2017). Inference on risk premia in the presence of omitted factors. Technical report, National Bureau of Economic Research.

Gospodinov, N., Kan, R., and Robotti, C. (2017). Spurious inference in reduced-rank asset-pricing models. *Econometrica*, 85(5):1613–1628.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054.
Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of political economy*, 99(2):225–262.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Hou, K. and Kimmel, R. (2010). On the estimation of risk premia in linear factor models. Technical report, Working Paper, Ohio State University.

Huberman, G., Kandel, S., and Stambaugh, R. F. (1987). Mimicking portfolios and exact arbitrage pricing. *The Journal of Finance*, 42(1):1–9.

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

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

Jagannathan, R. and Wang, Z. (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *The Journal of Finance*, 53(4):1285–1309.

Kan, R. and Robotti, C. (2008). Specification tests of asset pricing models using excess returns. *Journal of Empirical Finance*, 15(5):816–838.

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

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

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

Kitamura, Y. (2006). Empirical likelihood methods in econometrics: Theory and practice. Technical report, Yale University.
Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65:861–874.

Kitamura, Y. and Stutzer, M. (2002). Connections between entropic and linear projections in asset pricing estimation. *Journal of Econometrics*, 107(1-2):159–174.

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, 149(2):149–173.

Kleibergen, F. and Zhan, Z. (2018). Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics*, 16(2):155–190.

Liu, Y. (2018). Index option returns and generalized entropy bounds. *Available at SSRN 2149265*.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, pages 1429–1445.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41(5):867–887.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

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

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

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

Snow, K. N. (1991). Diagnosing asset pricing models using the distribution of asset returns. *The Journal of Finance*, 46(3):955–983.
Stutzer, M. (1995). A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics*, 68(2):367–397.

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

A Alternate Estimation Methods

In this section of the Appendix, we show results for our monthly factors using four different methods, including the main method from the body of the paper. We consider the following methods:

1. We use all portfolios in the construction of the SDF. That is, we do not do check and retain only portfolios with significant gammas. The risk-premia are then estimated with the Delta Method as before.

2. We use all portfolios in the construction of the SDF, and we estimate the risk-premia jointly with the parameters in the SDF as one large method of moments system.

3. We estimate the SDF separately, retain only significant portfolios, and then re-estimate the SDF, but this time jointly with the risk premia.

4. This last method is the same as in the main body of the text: Retain only significant portfolios, re-estimate the SDF with only these, and use the Delta Method to estimate risk-premia.
Table 7: Alternate Methods of Estimation for Risk-Premia

| Factor | Time-Series Mean | All Portfolios & Delta Method | All Portfolios & Joint Estimation |
|--------|------------------|------------------------------|-----------------------------------|
|        | Risk-Premium     | S.E.  | T-Stats | Risk-Premium | S.E.  | T-Stats | Risk-Premium | S.E.  | T-Stats |
| RMRF   | 6.089            | 2.282 | 2.668    | 6.36         | 3.445 | 1.846   | 6.36         | 0.043 | 149.136 |
| SMB    | 2.772            | 1.61  | 1.722    | 2.808        | 1.934 | 1.451   | 2.808        | 0.055 | 51.173  |
| HML    | 4.455            | 1.633 | 2.729    | 4.525        | 2.58  | 1.754   | 4.525        | 0.127 | 35.689  |
| MOM    | 8.246            | 2.096 | 3.934    | 8.372        | 4.621 | 1.812   | 8.372        | 0.142 | 59.134  |
| RMW    | 3.009            | 1.292 | 2.33     | 0.232        | 1.827 | 0.127   | 0.232        | 0.537 | 0.432   |
| CMA    | 3.846            | 1.121 | 3.429    | 2.517        | 1.625 | 1.549   | 2.517        | 0.54  | 4.664   |

| Factor | Sig. Gammas & Joint Estimation |
|--------|--------------------------------|
|        | Risk-Premium | S.E.  | T-Stats | Risk-Premium | S.E.  | T-Stats |
| RMRF   | 6.244         | 0.177 | 35.216  | 6.448        | 4.066 | 1.586   |
| SMB    | 3.069         | 0.365 | 8.407   | 3.214        | 2.201 | 1.46    |
| HML    | 2.644         | 0.882 | 2.999   | 5.098        | 2.726 | 1.87    |
| MOM    | 8.399         | 0.316 | 26.586  | 8.495        | 5.157 | 1.647   |
| RMW    | 0.769         | 0.854 | 0.9     | 0.524        | 1.828 | 0.287   |
| CMA    | 1.761         | 0.866 | 2.034   | 2.494        | 1.62  | 1.539   |

This table shows the risk-premia of our monthly factors (excluding the market squared, for space) using alternate methods of estimation. The first column displays the factor names: market (RMRF), size (SMB), value (HML), momentum (MOM), profitability (RMW), and investment (CMA). The next three columns simply display the time-series means, standard errors, and t-statistics for the factors. We consider the following methods: We use all portfolios in the construction of the SDF. That is, we do not do check and retain only portfolios with significant gammas. The risk-premia are then estimated with the Delta Method as before. We use all portfolios in the construction of the SDF, and we estimate the risk-premia jointly with the parameters in the SDF as one large method of moments system. We estimate the SDF separately, retain only significant portfolios, and then re-estimate the SDF, but this time jointly with the risk premia. This last method is the same as in the main body of the text: Retain only significant portfolios, re-estimate the SDF with only these, and use the Delta Method to estimate risk-premia. All risk-premia and standard errors have been annualized in percentage terms.

Table (7) displays the results of these different methods. For comparison, we calculate risk-premia and standard errors using the time-series means as well. The main take-aways are as follows: First, the point estimates are similar across all methods. One notable exception is the HML risk-premium for the method using significant portfolios in joint estimation. This seems to be an outlier though. Second, the standard errors are much smaller with joint estimation (and hence the t-statistics are larger). This is not wholly surprising as joint estimation should be more asymptotically efficient. However, the sample size is not that large, and we are cautious about taking the extremely large t-statistics at face value. Lastly, even though the t-statistics and standard errors do change across methods, the results are consistent vis a vis the question of whether a certain factor is “priced” or not.