Model-Free Factor Risk Premia

Mazi Kazemi

Arizona State University

October 2025

Abstract

I develop a semi-parametric estimator for factor risk premia that is robust to model misspecification. Standard two-pass and GMM estimators produce biased estimates when priced factors are omitted, irrelevant factors are included, or factors are weakly identified. The proposed method avoids these problems by separating the construction of the stochastic discount factor from any factor model specification. The SDF is recovered by minimizing Kullback–Leibler divergence subject only to no-arbitrage restrictions, and each factor's premium is computed through the marginal moment $-R_f\mathbb{E}[mf]$. This approach ensures that the estimate for any given factor depends only on that factor and the recovered SDF—not on which other factors the researcher specifies. Monte Carlo simulations calibrated to Fama–French–Carhart factors show that the estimator matches conventional methods under correct specification and delivers root mean squared errors 90-99% smaller when priced factors are omitted. Empirically, exponential-tilting estimates for traded factors align closely with time-series means, while Fama-MacBeth estimates for non-traded macro factors exhibit wild specification sensitivity—with sign reversals and swings exceeding 30 percentage points—that disappear under the robust approach.

JEL Codes: G12, C12, C58

Keywords: Linear factor models, Risk premia, Entropy,

1 Introduction

Factor risk premia—the compensation investors demand for exposure to systematic sources of risk—are fundamental building blocks of asset pricing. They determine the cross-section of expected returns, inform portfolio construction, and provide empirical discipline for structural

mmkazemi@asu.edu

models of the macroeconomy. Yet standard estimators such as the two-pass procedure of Fama and MacBeth (1973) and the generalized method of moments (GMM) of Hansen (1982) identify these premia only when the econometrician's linear factor model is correctly specified. When that specification is violated—because priced factors are omitted, redundant, or weak—the resulting point estimates and t-statistics are unreliable. This paper develops a semi-parametric, misspecification-robust estimator that recovers factor risk premia directly from the stochastic discount factor (SDF) implied by no-arbitrage. The SDF is obtained by minimizing relative entropy of the empirical distribution subject to the Euler restrictions, producing an always-positive kernel that prices each factor separately through its interaction with the SDF. By construction, the method remains valid even when the econometrician's linear specification is incorrect.

My starting point is the definition of a factor's risk premium as

$$\lambda = -R_f \mathbb{E}[mf],$$

where m is a stochastic discount factor (SDF), f is the factor innovation, and R_f is the gross risk-free rate. The key conceptual innovation is to separate the estimation of the SDF from the specification of any particular factor model. Rather than imposing a linear factor model and estimating loadings and premia jointly, I recover the SDF itself directly from observed asset returns. Specifically, I select m to minimize relative entropy (Kullback–Leibler divergence) between the empirical and risk-neutral distributions, subject only to the no-arbitrage Euler equations $\mathbb{E}[mR^e] = 0$ and $\mathbb{E}[m] = 1/R_f$. This yields the exponential-tilting SDF,

$$m_t(\gamma, \alpha) = \exp(\gamma^\top R_t^e + \alpha - 1),$$

which is strictly positive, nests variance- and likelihood-based constructions as special cases, and depends on the data only through the Lagrange multipliers (γ, α) associated with the pricing restrictions. Estimating factor premia through $-R_f \mathbb{E}[mf]$ then prices each factor separately via the SDF itself, without reference to any particular factor model specification.

This separation of factor pricing from model specification provides the key advantage: the premium estimate for any given factor depends only on that factor and the recovered

SDF—not on which other factors the researcher includes. This eliminates the omitted-variable biases inherent in two-pass and GMM procedures and yields estimates that are stable to the inclusion or exclusion of irrelevant factors. Traditional two-pass or GMM procedures jointly test whether a factor is priced and whether it helps explain the cross-section of returns; the present approach focuses on the more fundamental question—whether the factor commands a price of risk at all. This provides a transparent framework for evaluating both traded and non-traded risks in settings where linear completeness cannot be taken for granted.

I evaluate the performance of the exponential-tilting estimator in both simulated and empirical settings. In simulations calibrated to the Fama–French–Carhart factors, the method performs comparably to conventional two-pass procedures under correct specification and substantially outperforms them when priced factors are omitted. Table 1 shows that under misspecification, exponential tilting delivers root mean squared errors that are an order of magnitude smaller than two-pass and three-pass alternatives, in both large and small cross-sections. Whereas two-pass and three-pass estimates exhibit large bias and loss of precision under misspecification, exponential tilting continues to deliver nearly unbiased premia with stable inference. In empirical applications, exponential-tilting estimates for traded factors align closely with mean excess returns, while the method also yields economically sensible premia for non-traded macroeconomic variables such as intermediary leverage and consumption growth. Together, the simulation and empirical evidence demonstrate that exponential tilting provides a robust and practical tool for estimating factor risk premia in the presence of model misspecification.

My approach offers three advantages relative to existing SDF-based methods. First, relative to the variance-minimizing SDFs studied by Hansen and Jagannathan (1991), I move beyond second-moment efficiency and construct an SDF that incorporates higher-order moments while ensuring positivity. Second, relative to the generalized method of moments (Hansen, 1982; Cochrane, 2005), which delivers efficient estimates only under correct specification, my method remains well defined when the linear model is misspecified because it prices risk solely through the interaction of the SDF and each factor. Third, relative to information-theoretic approaches that use entropy to evaluate or bound SDFs (Stutzer, 1995; Bansal and Lehmann, 1997; Bakshi and Chabi-Yo, 2012; Backus et al., 2014; Liu, 2018), I employ entropy not as a diagnostic but as a constructive device: the SDF is chosen to minimize relative entropy subject to the no-arbitrage restrictions. This transformation turns entropy methods from tools for model evaluation into a practical estimation procedure for factor risk premia.

A large literature documents that standard linear estimators are highly sensitive to model misspecification. Jagannathan and Wang (1998) and Kan and Zhang (1999b,a) show that two-pass t-statistics can diverge even when the true premia are zero, while Kleibergen (2009) emphasizes the role of weak factors and the resulting invalidity of conventional inference. Subsequent work formalizes how redundant or irrelevant variables generate spurious inference in reduced-rank or over-factored models (Bryzgalova, 2015; Gospodinov et al., 2017) and develops specification-robust tests within the GMM and cross-sectional regression frameworks (Kan and Robotti, 2008; Kan et al., 2013). Non-traded factors pose additional challenges: measurement error complicates estimation (Hou and Kimmel, 2010), and Giglio and Xiu (2021) develop an alternative approach that recovers risk premia without imposing a complete factor model specification. Collectively, this evidence shows that conventional two-pass and GMM estimators perform well only under correct specification and that robustness to misspecification—whether due to omitted, weak, or redundant factors—is essential for credible inference in modern asset-pricing tests.

2 Minimum Discrepancy and Motivation

I construct the stochastic discount factor (SDF) directly from observed excess returns, taking the no-arbitrage restrictions as primitive. Let $R_t^e \in \mathbb{R}^N$ denote the vector of excess returns at date t, and let m_t be a strictly positive SDF. Absence of arbitrage requires

$$\mathbb{E}[m_t R_t^e] = 0, \qquad \mathbb{E}[m_t] = \frac{1}{R_f}.$$

Rather than postulating a parametric form for m_t or a correctly specified linear factor model, I choose m_t to minimize a measure of statistical discrepancy between the empirical and risk-neutral distributions subject to these restrictions. This *minimum-discrepancy* approach nests the variance-minimizing SDF of Hansen and Jagannathan (1991) as a special case but allows strictly positive and economically interpretable kernels that remain well defined under model misspecification.

Formally, I choose m_t to minimize

$$\min_{\{m_t\}} \mathbb{E}[\varphi(m_t)] \quad \text{s.t.} \quad \mathbb{E}[m_t R_t^e] = 0, \qquad \mathbb{E}[m_t] = \frac{1}{R_f},$$

where $\varphi(\cdot)$ is a convex discrepancy function. The quadratic choice $\varphi(m) = m^2$ reproduces the Hansen–Jagannathan variance–minimizing SDF. To ensure positivity and incorporate higher–order moments, I adopt the Cressie–Read family of discrepancy measures (Cressie and Read, 1984):

$$\varphi_\xi(m) = \frac{m^{\xi+1} - a^{\xi+1}}{\xi(\xi+1)}, \qquad a = \frac{1}{R_f}, \quad \xi \in \mathbb{R} \setminus \{0, -1\},$$

with limiting cases obtained by L'Hôpital's rule,

$$\lim_{\xi \to 0} \varphi_\xi(m) = m \ln m, \qquad \lim_{\xi \to -1} \varphi_\xi(m) = \ln m.$$

The first limit yields the *exponential–tilting* estimator based on relative entropy; the second yields the *empirical–likelihood* estimator. For $\xi \leq 0$, a Taylor expansion around a shows that all central moments of m_t enter the objective, naturally extending the variance–based formulation of Newey and Smith (2004) and Kitamura (2006).

2.1 Dual Formulation

Let $\gamma \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$ denote the Lagrange multipliers on the Euler and mean constraints. For exponential tilting, the dual problem is

$$\max_{\gamma,\alpha} \left\{ \frac{\alpha}{R_f} - \mathbb{E}\big[\exp(\gamma^\top R_t^e + \alpha - 1)\big] \right\}, \tag{1}$$

a strictly concave and globally well-behaved program whose solution uniquely determines the SDF

$$m_t(\gamma, \alpha) = \exp(\gamma^\top R_t^e + \alpha - 1).$$

For the empirical-likelihood case, the analogous derivation yields

$$m_t(\gamma, \alpha) = \frac{1}{\gamma^\top R_t^e + \alpha},$$

with feasibility requiring a positive denominator in sample. In both cases, the parameters (γ, α) are identified by the sample pricing and mean restrictions, and standard extremum-estimator arguments ensure consistency and asymptotic normality under mild regularity conditions (Newey and Smith, 2004).

2.2 Why Exponential Tilting

This construction has a key advantage for factor pricing: the parameters (γ, α) depend only on test asset returns $\{R_t^e\}$, not on any factor model specification. Factor risk premia are computed via $\lambda_k = -R_f \mathbb{E}[m f_k]$ separately for each factor, decoupling each premium from the researcher's treatment of other factors. This property—which I formalize in the next section—is the central robustness feature that distinguishes exponential tilting from conventional two-pass and GMM methods.

Having presented both ET and EL formulations, I focus on exponential tilting for the remainder of the paper. Relative to variance–based constructions, ET offers several advantages. It guarantees $m_t > 0$ for all realizations of returns, avoids explicit inequality constraints in the optimization, and embeds many existing SDF constructions as special cases (Snow, 1991; Černỳ, 2003; Stutzer, 1995, 1996; Almeida and Garcia, 2012, 2016; Ghosh et al., 2016). Moreover, since $m_t/E[m_t]$ is a nonnegative martingale with unit mean, it coincides with the Radon–Nikodým derivative that maps the physical measure into the risk–neutral measure. ET therefore selects the risk–neutral distribution closest—in the Kullback–Leibler (relative–entropy) sense—to the empirical distribution while exactly satisfying the Euler restrictions.

Economically, the interpretation is transparent. If each observation initially receives probability $1/T$, exponential tilting says: *"move the probabilities away from $1/T$ only insofar as doing so helps satisfy the Euler equations."* Thus ET modifies the empirical measure as little as necessary to restore no–arbitrage.

This construction provides both economic transparency and econometric tractability and motivates my focus on exponential tilting in what follows. Additional discussion in Appendix C.1 and Appendix C.2 interpret the multipliers (γ, α) as elasticities of marginal utility with respect to excess returns and connects the ET SDF to good–deal bounds (Cochrane and Saa-Requejo, 2000). These results provide intuition for the economic meaning of the dual parameters and for the admissibility of the ET SDF but are not required for the estimation and inference procedures developed in Section 5.

This construction has a key advantage for factor pricing: the parameters (γ, α) depend only on test asset returns $\{R_t^e\}$, not on any factor model specification. Factor risk premia are computed via $\lambda_k = -R_f \mathbb{E}[mf_k]$ separately for each factor, decoupling each premium from the researcher's treatment of other factors. The next section demonstrates why this robustness property matters by analyzing the specification-dependence problems that plague standard linear methods.

3 Specification dependence in linear factor models

Standard methods for estimating factor risk premia rely on linear factor models in which each asset's expected return is determined by its exposures to a set of risk factors. Two-pass regressions (Fama and MacBeth, 1973) and GMM estimators (Hansen, 1982) estimate these risk premia through joint systems that combine time-series factor loadings with cross-sectional pricing restrictions. This joint estimation creates specification dependence: the estimated premium for any given factor depends not only on that factor's data but also on which other factors are included, how those factors are measured, and whether the specified model is complete. This section analyzes four manifestations of specification dependence—omitted priced factors, redundant factors, weak identification, and weak pricing—and shows that each produces biases or inference failures that persist even as sample size grows. These problems arise because linear methods conflate the pricing of individual factors with the specification of the overall factor model, whereas the ET approach in Section 2 separates these two aspects by constructing the pricing kernel independently of any factor structure.

3.1 Setup

Consider the standard linear factor model

$$R^e_{i,t+1} = \beta_i' f_{t+1} + \varepsilon_{i,t+1}, \qquad \mathbb{E}[\varepsilon_{i,t+1}] = 0, \tag{2}$$

where $R^e_{i,t+1}$ is the excess return on asset i, β_i is its $K \times 1$ loading vector, and f_{t+1} is a $K \times 1$ vector of demeaned factor innovations. The arbitrage-pricing-theory arguments of Chamberlain and Rothschild (1983) and Ross (1976) permit weak cross-sectional dependence in $\varepsilon_{i,t+1}$ and imply that idiosyncratic risk is not priced. Under no-arbitrage, expected returns satisfy

$$\mathbb{E}[R^e_{i,t+1}] = \beta_i' \lambda, \tag{3}$$

where $\lambda = (\lambda_1, \ldots, \lambda_K)'$ is the vector of factor risk premia. For any valid stochastic discount factor m_{t+1}, these premia satisfy $\lambda_k = -R_f \mathbb{E}[m_{t+1} f_{k,t+1}]$.[1]

Two-pass (Fama–MacBeth) regressions (Black et al., 1972; Fama and MacBeth, 1973) estimate β_i from time-series regressions of each asset's returns on the factors, then estimate λ from a cross-sectional regression of average returns on the estimated loadings $\hat{\beta}_i$. GMM estimators (Hansen, 1982; Cochrane, 2005) combine the time-series and cross-sectional moment conditions into a single system. Both approaches estimate all risk premia jointly: the estimate $\hat{\lambda}_k$ for any factor k depends on the loadings and premia of all other factors in the model. When the model specification is correct, these methods are efficient. When specification is violated, they produce biased estimates or invalid inference that persist as population features.

[1]This is true if the matrix of betas has full rank K. If it does not, then risk-premia are not identified. In that case, we can simply define the risk-premium to be $\lambda_k = -R_f \mathbb{E}[m_{t+1} f_{k,t+1}]$. This is, after all, simply the negative risk-neutral expectation of a source of risk, a natural "premium" definition.

3.2 Omitted priced factors

Suppose the true data-generating process includes K priced factors but the econometrician observes only a subset. Partition the factors as $f_t = (f_t^{obs'}, f_t^{omit'})'$ with corresponding loadings $\beta_i = (\beta_i^{obs'}, \beta_i^{omit'})'$ and premia $\lambda = (\lambda^{obs'}, \lambda^{omit'})'$, where the researcher observes only f_t^{obs} containing $K_1 < K$ factors.

In the two-pass procedure, regressing returns on observed factors alone yields biased loading estimates. As $T \to \infty$,

$$\text{plim}\,\hat{\beta}_i^{obs} = \beta_i^{obs} + (\Sigma_f^{obs})^{-1}\Sigma_{obs,omit}\beta_i^{omit}, \tag{4}$$

where $\Sigma_f^{obs} = \mathbb{E}[f_t^{obs}(f_t^{obs})']$ and $\Sigma_{obs,omit} = \mathbb{E}[f_t^{obs}(f_t^{omit})']$. The bias term $(\Sigma_f^{obs})^{-1}\Sigma_{obs,omit}\beta_i^{omit}$ reflects the projection of omitted factors onto observed ones.

The second-stage cross-sectional regression compounds this bias. As $T, N \to \infty$, the estimated risk premia converge to

$$\text{plim}\,\hat{\lambda}^{obs,FM} = (B^{obs'}WB^{obs})^{-1}B^{obs'}W(B^{obs}\lambda^{obs} + B^{omit}\lambda^{omit}), \tag{5}$$

where B^{obs} and B^{omit} are $N \times K_1$ and $N \times (K - K_1)$ matrices stacking the true loadings across assets, and W depends on cross-sectional moments. The estimated premium for any observed factor conflates its true value λ^{obs} with contamination from omitted factors' premia λ^{omit}, weighted by the cross-sectional covariance structure of loadings. This omitted-variable bias persists as a population-level phenomenon. Unless omitted factors are orthogonal to observed ones ($\Sigma_{obs,omit} = 0$) or carry zero premia ($\lambda^{omit} = 0$), the estimator $\hat{\lambda}^{obs,FM}$ converges to a biased limit.

By contrast, the ET estimator computes each premium through the moment $\hat{\lambda}_{ET,k} = -R_f T^{-1}\sum_t m_t(\hat{\gamma}, \hat{\alpha}) f_{k,t}$, where the pricing kernel parameters $(\hat{\gamma}, \hat{\alpha})$ are estimated from test asset returns alone. The estimate for factor k depends only on the joint realization of $(R_t^e, f_{k,t})$ and is invariant to which other factors are included or omitted from auxiliary analyses.

3.3 Redundant factors

Including factors with zero risk premia—whether they are truly unpriced or merely irrelevant to the cross-section—distorts inference on other factors' premia even when those other factors are correctly specified. Consider a model with K factors where factor j has $\lambda_j = 0$. In the two-pass procedure, the estimator for all premia is

$$\hat{\lambda}^{FM} = \left(\frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \hat{\beta}_i' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \bar{R}_i^e, \tag{6}$$

where $\hat{\beta}_i$ includes the loading on the useless factor j.

Kan and Zhang (1999b,a) demonstrate that even though $\lambda_j = 0$ in population, including factor j causes the t-statistics for other factors' premia to diverge and confidence intervals to become unreliable as sample size grows. The asymptotic variance of $\hat{\lambda}_k$ for any factor $k \neq j$ satisfies

$$\text{Avar}(\hat{\lambda}_k^{FM}) = [(\Sigma_{\beta\beta})^{-1} \Sigma_{\beta\beta R} \Sigma_{R\beta} (\Sigma_{\beta\beta})^{-1}]_{kk}, \tag{7}$$

where $\Sigma_{\beta\beta} = \text{plim}\, N^{-1} \sum_i \hat{\beta}_i \hat{\beta}_i'$ includes the cross-sectional variance and covariances involving the useless factor's loadings. The presence of $\beta_{i,j}$ in this expression means that even factors with $\lambda_j = 0$ affect the precision and distribution of all other premium estimates.

The ET estimator avoids this problem through its marginal moment structure. The estimate $\hat{\lambda}_{ET,k}$ for factor k depends only on $(R_t^e, f_{k,t})$ and is invariant to whether factor j is included in the researcher's analysis. Including or excluding redundant factors does not affect the point estimate or standard error of any other factor's premium, because each premium is computed through an independent moment condition rather than through a joint system.

3.4 Weak identification

When factors have low correlations with test asset returns—meaning the factor loadings $\beta_{i,k}$ are small in magnitude across assets—conventional inference breaks down even if the risk premium λ_k is nonzero. Consider the cross-sectional regression

$$\mathbb{E}[R_i^e] = \beta_i' \lambda, \quad i = 1, \ldots, N. \tag{8}$$

Weak identification occurs when the matrix of factor loadings $B = (\beta_1, \ldots, \beta_N)'$ has small singular values, making this system nearly rank-deficient. In the limit, as the loadings approach zero uniformly, the concentration parameter

$$\mu_N^2 = N \cdot \lambda' \left(\frac{1}{N} B' M_R B \right) \lambda \tag{9}$$

fails to diverge, where M_R projects onto the orthogonal complement of the constant.

Kleibergen (2009) shows that under weak identification, two-pass t-statistics do not converge to a normal distribution and standard confidence intervals fail to achieve correct coverage. The finite-sample distribution of $\hat{\lambda}^{FM}$ is poorly approximated by its asymptotic normal limit, and t-statistics can diverge even when the true premium is zero. The usual sandwich formula for standard errors produces confidence intervals with incorrect coverage probabilities. The same pathology affects GMM estimators: estimation uncertainty from the first-stage loadings propagates into second-stage premium estimates in ways that violate standard asymptotic approximations.

The ET estimator handles weak identification in a different way. When a factor has low correlation with test assets, it typically corresponds to a situation where the factor is not fully spanned by the test asset returns, as discussed in Proposition A.1(ii). In such cases, the risk premium is not uniquely determined by no-arbitrage alone, and the feasible set forms an interval $[\lambda_k^{\min}, \lambda_k^{\max}]$. The ET estimator converges to a point within this interval. Importantly, standard asymptotic theory continues to apply—the estimator remains asymptotically normal, with standard errors that appropriately reflect the degree of identification. Large standard errors signal limited identifying information rather than a breakdown of inference, providing an honest assessment of what the data can identify.

3.5 Summary

The three forms of misspecification analyzed in this section—omitted priced factors, redundant factors, and weak identification—share a common root cause: two-pass and GMM methods estimate all risk premia jointly through systems that tie each factor's premium to the full model specification. When priced factors are omitted, the estimated premia converge to biased population limits that conflate true premia with contamination from omitted factors. When redundant factors are included, the asymptotic variance of all premium estimates becomes contaminated by the spurious loadings, distorting inference even on correctly specified factors. When factors are weakly identified, standard asymptotic approximations break down entirely, producing t-statistics with non-standard distributions and confidence intervals with incorrect coverage.

The ET estimator avoids these specification-dependence problems through its construction. The pricing kernel $m_t(\gamma, \alpha)$ is estimated from test asset returns $\{R_t^e\}$ alone, independently of any factor model. Each factor's premium is then computed through the marginal moment $\hat{\lambda}_{ET,k} = -R_f T^{-1} \sum_t m_t f_{k,t}$, which depends only on the joint distribution of $(R_t^e, f_{k,t})$. This separation—estimating the SDF from returns, then pricing factors through independent moments—eliminates omitted-variable bias (the estimate for factor k is invariant to other factors), avoids contamination from redundant factors (including factor j does not affect $\hat{\lambda}_{ET,k}$ for $k \neq j$), and handles weak identification through the spanning framework (incomplete spanning produces larger standard errors but preserves valid asymptotic inference). The next section establishes these robustness properties formally.

4 Theoretical properties

This section develops the large-sample foundations of the exponential-tilting (ET) estimator for factor risk premia. I establish consistency and asymptotic normality under standard regularity conditions and show that the estimator remains well defined when the econometrician's linear factor model is misspecified. The key result, formalized in Proposition A.1, is that each factor's estimated premium depends only on that factor and on the semi-parametrically recovered SDF, making the estimate invariant to the inclusion or exclusion of other factors. Complete assumptions and proofs appear in Appendix A.

4.1 Setup and identification

Let $R_t^e \in \mathbb{R}^N$ denote excess returns on N test assets and $f_{k,t}$ the innovation of factor k. The joint process $\{(R_t^e, f_{k,t})\}$ is assumed strictly stationary and ergodic with moments of order $2+\delta$

for some $\delta > 0$. Absence of arbitrage implies the existence of a strictly positive SDF m_t satisfying $E[m_t R_t^e] = 0$ and $E[m_t] = 1/R_f$, and the second-moment matrix $\Sigma_R = E[R_t^e (R_t^e)']$ is positive definite so that no test asset is redundant.

The ET SDF is parameterized by

$$m_t^{ET}(\gamma, \alpha) = \exp\bigl(\gamma' R_t^e + \alpha - 1\bigr), \tag{10}$$

where the population parameters (γ^*, α^*) solve the dual problem

$$\max_{\gamma,\alpha} \; \frac{\alpha}{R_f} - E[\exp(\gamma' R_t^e + \alpha - 1)]\,, \tag{11}$$

subject to $E[m_t^{ET}(\gamma, \alpha) R_t^e] = 0$ and $E[m_t^{ET}(\gamma, \alpha)] = 1/R_f$. The objective is strictly concave and has a unique optimum under mild moment conditions (uniform exponential integrability). Replacing expectations with time–series averages yields the sample analogue

$$(\hat{\gamma}, \hat{\alpha}) = \arg\max_{\gamma,\alpha} \Bigl\{ \frac{\alpha}{R_f} - \frac{1}{T} \sum_{t=1}^{T} \exp(\gamma' R_t^e + \alpha - 1) \Bigr\}.$$

Given $(\hat{\gamma}, \hat{\alpha})$, the estimated factor premium is

$$\hat{\lambda}_{ET,k} = -R_f \, \frac{1}{T} \sum_{t=1}^{T} m_t^{ET}(\hat{\gamma}, \hat{\alpha}) \, f_{k,t}. \tag{12}$$

Equation (12) computes each premium through a single marginal moment involving $(m_t^{ET}, f_{k,t})$ and does not require specifying a linear factor model or jointly estimating loadings and premia.

4.2 Main result

Proposition 4.1 (No-arbitrage consistent risk premia)**.** *Under the regularity conditions in Appendix A, the ET estimator* (12) *satisfies:*

(i) ***Full spanning.*** *If factor f_k is fully spanned by test assets—i.e., there exist $w \in \mathbb{R}^N$ and $c \in \mathbb{R}$ such that $f_{k,t} = c + w' R_t^e$ almost surely—then all valid SDFs imply the same premium $\lambda_k^* = w' E[R_t^e]$. Moreover, $\hat{\lambda}_{ET,k} \xrightarrow{p} \lambda_k^*$ and $\sqrt{T}(\hat{\lambda}_{ET,k} - \lambda_k^*) \xrightarrow{d} N(0, \Sigma_k)$, where*

Σ_k can be estimated with a heteroskedasticity- and autocorrelation-consistent (HAC) covariance matrix (Newey and West, 1987).

(ii) ***Incomplete spanning.*** *If f_k is not fully spanned, different valid SDFs may imply different premia. The set of all no-arbitrage-consistent premia forms a compact interval $\Lambda_k = [\lambda_k^{\min}, \lambda_k^{\max}]$, where*

$$\lambda_k^{\max} = \sup\{-R_f E[m_t f_{k,t}] : E[m_t R_t^e] = 0,\ E[m_t] = 1/R_f,\ m_t > 0\},$$
$$\lambda_k^{\min} = \inf\{-R_f E[m_t f_{k,t}] : E[m_t R_t^e] = 0,\ E[m_t] = 1/R_f,\ m_t > 0\}.$$

The ET estimator converges to a value inside this interval, $\hat{\lambda}_{ET,k} \xrightarrow{p} \lambda_k^{ET} \in [\lambda_k^{\min}, \lambda_k^{\max}]$, where λ_k^{ET} corresponds to the SDF that minimizes Kullback–Leibler divergence from the uniform probability measure subject to the pricing constraints. The width $\lambda_k^{\max} - \lambda_k^{\min}$ quantifies market incompleteness for factor f_k.

(iii) ***Misspecification robustness.*** *Conditional on a fixed set of test assets, the point estimate $\hat{\lambda}_{ET,k}$ depends only on $\{R_t^e, f_{k,t}\}$ and is invariant to which other factors the researcher includes or excludes in auxiliary analyses. Adding or omitting priced or irrelevant factors does not affect the consistency result in part (i).*

When a factor is spanned by traded returns, the law of one price fixes a unique premium $w'E[R_t^e]$ that all valid SDFs must deliver (Hansen and Jagannathan, 1991). If the factor is not spanned, unpriced residual risk introduces a continuum of feasible SDFs and an interval of admissible premia; ET selects the one closest to the empirical distribution in relative-entropy distance, consistent with the good-deal-bound logic of Cochrane and Saa-Requejo (2000). Because the ET kernel is estimated solely from returns, the resulting premium for each factor is immune to misspecification of any larger linear model.

4.3 Robustness to misspecification

Proposition A.1 establishes that the ET estimator remains consistent under conditions that cause standard linear methods to fail. The three forms of misspecification analyzed in

Section 3—omitted factors, redundant factors, and weak identification—are directly addressed by the structure of the estimator.

First, omitted priced factors do not bias ET estimates because each premium is computed through the marginal moment $\hat{\lambda}_{ET,k} = -R_f T^{-1} \sum_t m_t f_{k,t}$, which depends only on $(R_t^e, f_{k,t})$ and not on which other factors are included in auxiliary analyses. This is the content of Proposition A.1(iii). By contrast, two-pass and GMM methods estimate all premia jointly, so omitting a priced factor biases every other premium estimate (Section 3).

Second, redundant or irrelevant factors—those with zero risk premia—do not affect ET estimates of other factors' premia. Adding factor f_j to the analysis does not alter the estimate $\hat{\lambda}_{ET,k}$ for any other factor $k \neq j$, again by the invariance property in Proposition A.1(iii). Linear methods lack this property: including useless factors distorts the joint covariance structure and affects all premium estimates (Kan and Zhang, 1999b,a).

Third, weak factors are handled through the spanning framework. If a factor has low correlation with test assets (weak identification through low β), this corresponds to incomplete spanning as in Proposition A.1(ii). The estimator remains consistent for a point in the feasible interval $[\lambda_k^{\min}, \lambda_k^{\max}]$, with standard errors reflecting the genuine identification limits. If a factor has a small risk premium (weak pricing through low λ), standard asymptotic theory applies: the estimate $\hat{\lambda}_{ET,k}$ will be close to zero with standard errors scaled appropriately. In neither case does the asymptotic distribution break down, contrasting with the divergent t-statistics documented by Kleibergen (2009) for linear methods under weak identification.

4.4 Practical implications

Proposition A.1 has three implications for empirical practice.

First, estimating $\hat{\lambda}_{ET,k}$ with alternative test-asset sets provides a diagnostic of spanning. Tight clustering of estimates across different test-asset configurations suggests that factor f_k is approximately spanned and that the premium is uniquely identified by no-arbitrage. Wide variation across test-asset sets signals incomplete markets and highlights the dependence of the estimated premium on the choice of pricing kernel.

Second, inference on whether a proposed factor is priced can proceed directly through $\hat{\lambda}_{ET,k}$ and its associated t-statistic without specifying a full factor model. A researcher can test $H_0 : \lambda_k = 0$ using the asymptotic distribution from part (i) of Proposition A.1, requiring only the factor data $\{f_{k,t}\}$ and a set of test assets. This contrasts with two-pass or GMM methods (Fama and MacBeth, 1973; Hansen, 1982; Kan and Robotti, 2008; Kan et al., 2013), where the specification of control factors affects every estimated premium and t-statistic. The ET approach separates the question of whether factor k commands a risk premium from the question of whether it helps explain the cross-section conditional on other factors. The former concerns risk pricing; the latter conflates pricing with explanatory power in a potentially misspecified linear model.

Third, while traditional methods are dependent on both a well-spanning test asset set and a correct model, my method only requires a well-spanning test asset set. While this is a restriction, it is a restriction shared by all methods.

5 Estimation

The minimum-discrepancy formulation yields an SDF indexed by the dual parameters (γ, α), which must be estimated from the data. The sample counterpart of the exponential-tilting program in Section 2 is

$$\max_{\gamma,\alpha} \left\{ \frac{\alpha}{R_f} - \frac{1}{T} \sum_{t=1}^{T} \exp(\gamma^\top R_t^e + \alpha - 1) \right\},$$

a strictly concave problem with a unique maximizer under mild moment conditions on R_t^e. Denote this maximizer by $(\hat{\gamma}, \hat{\alpha})$ and the associated estimated SDF by $\hat{m}_t = \exp(\hat{\gamma}^\top R_t^e + \hat{\alpha} - 1)$. The concavity of the objective and the linearity of the constraints ensure stable numerical solutions, and standard extremum-estimator arguments imply consistency and asymptotic normality under the regularity conditions stated in Appendix A.

Consistency and asymptotic normality of $(\hat{\gamma}, \hat{\alpha})$ follow from standard extremum–estimator arguments. Let $\theta = (\gamma, \alpha)$ and define the objective

$$Q_T(\theta) = \frac{\alpha}{R_f} - \frac{1}{T} \sum_{t=1}^{T} \exp\bigl(\gamma' R_t^e + \alpha - 1\bigr).$$

If $Q_T(\theta)$ converges uniformly to a deterministic limit $Q(\theta)$ with unique maximizer θ_0, then $\hat{\theta} \equiv (\hat{\gamma}, \hat{\alpha}) \to_p \theta_0$. Moreover,

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega), \tag{13}$$

where $\Omega = H^{-1} V H^{-1}$, with $H = -\mathbb{E}[\partial^2 Q(\theta_0)/\partial\theta\partial\theta']$ and V the long–run covariance matrix of the score. Consistent estimators of V are obtained using the heteroskedasticity–and–autocorrelation–robust (HAC) method of Newey and West (1987).

Given $(\hat{\gamma}, \hat{\alpha})$, the risk premium of factor f_t is estimated using the defining moment

$$\hat{\lambda} = -\frac{R_f}{T} \sum_{t=1}^{T} \hat{m}_t f_t. \tag{14}$$

Equation (14) highlights a key feature of the method: the point estimate of λ depends only on the factor of interest and on the estimated SDF $\hat{m}_t$. In contrast to two–pass regressions where inference on λ is joint with the assumed specification of the linear factor model. The separation of SDF estimation from factor pricing underlies the robustness of the approach to model misspecification.

Define

$$\mu(\theta) \equiv -R_f\, \mathbb{E}\big[g_t(\theta) f_t\big], \qquad \phi_t(\theta) \equiv -R_f\, g_t(\theta) f_t, \qquad g_t(\theta) = \exp(\gamma' R_t^e + \alpha - 1),$$

so that $\hat{\lambda} = \mathbb{E}_T[\phi_t(\hat{\theta})]$. A first–order expansion around θ_0 gives

$$\mathrm{IF}_t(\hat{\lambda}) = \underbrace{\phi_t(\theta_0) - \mu(\theta_0)}_{\text{empirical average}} + \underbrace{\nabla_\theta \mu(\theta_0)'\, \mathrm{IF}_t(\hat{\theta})}_{\text{plug–in via score}},$$

with $\mathrm{IF}_t(\hat{\theta}) = -G^{-1}\psi_t(\theta_0)$, where

$$\psi_t(\theta) = \begin{bmatrix} -\, R_t^e\, g_t(\theta) \\ \frac{1}{R_f} - g_t(\theta) \end{bmatrix}, \qquad G = \mathbb{E}\Big[\frac{\partial \psi_t(\theta_0)}{\partial \theta'}\Big].$$

Hence

$$\sqrt{T}\big(\hat{\lambda} - \mu(\theta_0)\big) \;\Rightarrow\; N(0, \sigma^2), \qquad \sigma^2 = \mathrm{Var}\big(\mathrm{IF}_t(\hat{\lambda})\big),$$

which expands to

$$\sigma^2 = \mathrm{Var}\Big(\phi_t(\theta_0) - \mu(\theta_0)\Big) + \nabla_\theta\mu(\theta_0)' G^{-1}\,\mathrm{Var}\Big(\psi_t(\theta_0)\Big)\, G^{-1'} \nabla_\theta\mu(\theta_0)$$
$$+ 2\,\mathrm{Cov}\Big(\phi_t(\theta_0) - \mu(\theta_0),\ \nabla_\theta\mu(\theta_0)' G^{-1}\psi_t(\theta_0)\Big).$$

In practice, it is convenient to estimate σ^2 by stacking the moments into a single system and applying the usual sandwich formula. Define

$$\Phi_t(\theta, \lambda) = \begin{bmatrix} \psi_t(\theta) \\ \phi_t(\theta) - \lambda \end{bmatrix},$$

and consider the just–identified system $\mathbb{E}_T[\Phi_t(\theta, \lambda)] = 0$. The Jacobian is

$$J = \mathbb{E}\Big[\frac{\partial \Phi_t(\theta_0, \lambda_0)}{\partial(\theta', \lambda)'}\Big] = \begin{bmatrix} G & 0 \\ \nabla_\theta\mu(\theta_0)' & -1 \end{bmatrix},$$

and the long–run covariance is $S = \mathrm{HAC}(\Phi_t(\theta_0, \lambda_0))$. Then

$$\sqrt{T}\begin{bmatrix} \hat\theta - \theta_0 \\ \hat\lambda - \lambda_0 \end{bmatrix} \ \Rightarrow\ N\Big(0,\ J^{-1} S\, J^{-1'}\Big),$$

so the asymptotic variance of $\hat\lambda$ is the (λ, λ) element of $J^{-1}SJ^{-1'}$, consistently estimated by plugging in $(\hat\theta, \hat\lambda)$ and a HAC estimator of S. This stacked–moment approach automatically incorporates the empirical–average term, the plug–in term, and their covariance, while preserving the two–step estimation of $\hat\theta$ used to construct the SDF.

The estimation procedure is therefore as follows. I first solve the dual program (1) to obtain $(\hat\gamma, \hat\alpha)$ and construct $\hat m_t$. I then compute $\hat\lambda$ using (14). Standard errors are obtained by stacking the moments and applying the sandwich formula with HAC covariance estimation, which correctly captures all components of sampling variability.

6 Simulation Results

This section evaluates the finite-sample behavior of the exponential–tilting (ET) estimator relative to the Fama–MacBeth two–pass (FM) and three–pass estimators. The design is calibrated directly to the empirical properties of the Fama–French–Carhart four–factor model. Simulations examine both efficiency under correct specification and robustness when priced factors are omitted from the econometrician's model.

The data-generating process is

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

where $u_{i,t} \sim N(0, \sigma_i^2)$ represents idiosyncratic risk. Factor realizations $\tilde{f}_t$ are drawn from a multivariate normal with mean and covariance matching the historical Fama–French–Carhart factors. Portfolio betas are sampled from the empirical cross-section of 80 portfolios (sorted on size, book-to-market, momentum, and industry), and residual variances σ_i^2 are set equal to the median residual variance from the time-series regressions. True premia λ_k equal the historical monthly means of the four factors. Two designs are considered: a large cross-section ($N = 200$, $T = 600$) and a small cross-section ($N = 25$, $T = 1000$). Each experiment is repeated 1,000 times.

Table 1 reports root mean squared errors (RMSEs) for the three estimators. When all four factors are observed (Panel A), the methods perform similarly when N and T are large, and differences in RMSE are economically negligible. With a smaller cross-section, sampling error increases for all methods, but ET deteriorates less than the alternatives, outperforming Fama–MacBeth for at least two factors and three–pass for several. When priced factors are omitted from the econometrician's model (Panel B), the RMSEs of Fama–MacBeth and three-pass increase sharply, while ET's RMSEs remain virtually unchanged. The error rates are anywhere from 90-99% smaller for ET.

Table 2 reports the average bias of estimated factor premia across Monte Carlo replications. The results reinforce the RMSE patterns in Table 1. When all four factors are observed (Panel A), biases are small for all estimators, indicating that differences in RMSEs primarily

Table 1: Simulation Results: Root Mean Squared Error (RMSE)

	N=200, T=600				N=25, T=1000			
	RMRF	SMB	HML	UMD	RMRF	SMB	HML	UMD
	Panel A: All Factors Observed							
Three-Pass	0.168	0.256	0.254	0.395	0.663	0.853	0.564	1.528
Fama-MacBeth	0.778	0.230	0.262	0.377	1.771	0.570	0.658	0.882
Exponential Tilting	0.186	0.303	0.326	0.441	0.649	0.841	0.563	1.493
	Panel B: Two Factors Unobserved							
Three-Pass	2.541	3.163			2.588	3.150		
Fama-MacBeth	18.502	8.393			19.017	8.380		
Exponential Tilting	0.185	0.303			0.649	0.842		

This table reports the annualized RMSE (in percent) of estimated factor risk premia across 1000 simulations. The true DGP includes four priced factors: $RMRF$, SMB, HML, and UMD. Panel A reports results when all four factors are observed. Panel B reports results when only $RMRF$ and SMB are observed, omitting HML and UMD.

reflect sampling variability. When priced factors are omitted (Panel B), however, the biases of the Fama–MacBeth estimator become large and persistent, reflecting the population-level omitted-variable problem inherent in two–pass regressions. Three-pass biases also increase in absolute value, though less than Fama-MacBeth. Exponential tilting delivers the smallest and most stable biases across all experiments, confirming that the estimator prices each factor correctly even when the econometrician's model is misspecified. Overall, the bias results highlight the robustness advantage of ET and complement the RMSE evidence in Table 1.

Taken together, the simulation evidence establishes two central points. When the econometrician's specification is correct, exponential tilting is competitive with the standard alternatives, with efficiency similar to Fama–MacBeth and three–pass. When priced factors are omitted, exponential tilting shows no change in error rate, while the other two methods deteriorate, and delivers reliable premia even when the number of test assets is small. These findings underscore the robustness advantages of the method in environments where misspecification is a first–order concern.

Table 2: Simulation Results: Average Bias

	N=200, T=600				N=25, T=1000			
	RMRF	SMB	HML	UMD	RMRF	SMB	HML	UMD
Panel A: All Factors Observed								
Three-Pass	-0.107	0.137	0.030	-0.231	-0.593	0.719	0.251	-1.362
Fama-MacBeth	-0.351	0.050	0.003	-0.120	-0.313	0.031	-0.009	-0.071
Exponential Tilting	-0.083	0.105	0.024	-0.179	-0.577	0.701	0.241	-1.322
Panel B: Two Factors Unobserved								
Three-Pass	-2.531	3.154	-2.539	3.143				
Fama-MacBeth	-18.427	8.354	-18.645	8.333				
Exponential Tilting	-0.084	0.106	-0.577	0.702				

This table reports the average bias (in percent per year) of estimated factor risk premia across 1000 simulations. The true DGP includes four priced factors: $RMRF$, SMB, HML, and UMD. Panel A reports results when all four factors are observed. Panel B reports results when only $RMRF$ and SMB are observed, omitting HML and UMD.

7 Empirical Results

I now turn to the empirical application of the exponential–tilting (ET) estimator. The analysis proceeds in two parts. I first study monthly traded factors, including the Fama–French–Carhart factors and extensions, where the risk premia can be benchmarked against mean excess returns. I then examine non-traded monthly macroeconomic factors, including industrial production growth, the term spread, and the credit spread. These results illustrate both the practical implementation of ET and the interpretive advantages of estimating factor premia directly from the SDF.

7.1 Traded Factors

The first set of results concerns seven monthly factors: the four Fama–French–Carhart factors ($RMRF$, SMB, HML, and UMD), the profitability and investment factors proposed by Fama and French (2015), and the square of the market return, which serves as a nonlinear transformation that could in principle capture skewness. Six of these are traded factor returns, and one is constructed. Because the premia of traded factors are, in population, equal to

Table 3: Risk Premia of Monthly Factors

	Time–Series Mean			Fama–MacBeth			Exponential Tilting		
Factor	RP	S.E.	t	RP	S.E.	t	RP	S.E.	t
Panel A: All Seven Factors									
Market	7.017	2.142	3.275	6.989	2.112	3.309	7.240	0.117	61.775
Size	1.560	1.486	1.049	1.730	1.539	1.124	1.578	0.096	16.394
Value	3.335	1.722	1.936	3.561	1.777	2.004	3.198	0.200	15.980
Momentum	6.939	1.922	3.610	7.741	2.023	3.827	7.012	0.289	24.287
Profitability	3.450	1.208	2.857	3.008	1.402	2.145	3.093	0.166	18.675
Investment	3.383	1.135	2.979	3.368	1.255	2.684	3.180	0.202	15.735
Market^2				42.826	74.918	0.572	-22.546	20.359	-1.107
Panel B: Fama–French Three Factors and Market^2									
Market	7.017	2.142	3.275	7.644	2.073	3.688	7.240	0.117	61.775
Size	1.560	1.486	1.049	0.829	1.540	0.538	1.578	0.096	16.394
Value	3.335	1.722	1.936	4.645	1.972	2.356	3.198	0.200	15.980
Market^2				-522.764	94.254	-5.546	-22.546	20.359	-1.107

This table reports estimated annualized factor risk premia (in percent) for monthly factors, 1967M7–2025M6. Panel A includes seven factors; Panel B restricts attention to the three Fama–French factors and the squared market return. Reported statistics are the mean excess return, the two–pass Fama–MacBeth estimate, and the exponential–tilting estimate, with Newey–West standard errors and t–statistics.

their mean excess returns, these provide a natural benchmark against which to judge the ET estimates.

The sample runs from July 1967 to June 2025. Test assets comprise 100 portfolios from Kenneth French's data library: 25 portfolios sorted on size and book-to-market, 25 on size and operating profitability, 25 on size and momentum, and 25 on size and investment. For ET, these portfolios provide the cross-section for recovering the SDF parameters $(\hat{\gamma}, \hat{\alpha})$. For Fama–MacBeth (FM), they provide the cross-section in the two-pass procedure.

Table 3 reports estimated premia for the seven monthly factors. Panel A includes all factors jointly, while Panel B restricts attention to the original three Fama–French factors and the squared market return. For each factor I report its mean excess return, the FM estimate, and the ET estimate, together with standard errors and t–statistics.

The results show that ET estimates for the four benchmark factors are close to their mean excess returns and are statistically significant at conventional levels. The market premium is estimated at 7.24% annually with a t-statistic above 60, while size, value, and momentum premia are 1.58%, 3.20%, and 7.01% respectively, all highly significant. FM also recovers significant premia for these factors, though with notably different point estimates and lower precision—for instance, the FM market premium is 6.99% versus the time-series mean of 7.02%.

For the profitability and investment factors added by Fama and French (2015), both methods find positive and statistically significant premia. FM estimates 3.01% for profitability ($t = 2.145$) and 3.37% for investment ($t = 2.684$), while ET estimates 3.09% ($t = 18.675$) and 3.18% ($t = 15.735$) respectively. These estimates are economically similar but ET delivers substantially tighter standard errors. Neither method finds evidence of a premium on the squared market return in either specification.

The key advantage of ET becomes apparent when comparing Panels A and B. When I reduce the factor set to the original Fama–French three plus MKT^2, ET estimates for the overlapping factors remain exactly unchanged (Market=7.240, Size=1.578, Value=3.198 in both panels), precisely as the theory in Proposition A.1 predicts. In contrast, FM estimates shift materially across specifications, underscoring the dependence of two-pass inference on model specification. This invariance property—that each factor's premium depends only on that factor and the SDF, not on which other factors are included—is the central robustness advantage of the exponential-tilting approach.

7.2 Non–Traded Factors

I next examine three monthly macroeconomic factors that are not directly tradable: industrial production growth, the term spread (10-year minus 3-month Treasury yield), and the credit spread (BAA minus AAA corporate bond yield). These factors appear frequently in asset pricing studies but pose challenges for traditional methods because they require constructing mimicking portfolios or making strong assumptions about market completeness. The ET estimator allows me to estimate premia directly from the definition $\lambda_k = -R_f \mathbb{E}[m f_k]$ without such auxiliary constructions.

The sample runs from July 1967 to June 2025. Test assets comprise 100 portfolios from Kenneth French's data library: 25 portfolios sorted on size and book-to-market, 25 on size and operating profitability, 25 on size and momentum, and 25 on size and investment. These portfolios provide a rich cross-section for recovering the SDF parameters $(\hat{\gamma}, \hat{\alpha})$. For comparison, I also report Fama-MacBeth (FM) estimates under four specifications—estimating each macro factor's premium alone, controlling for the market factor, controlling for the Fama-French three factors, and controlling for the Fama-French five factors—as well as three-pass (3P) estimates using the FF5 specification.

Table 4 reports the results. The Fama-MacBeth estimates exhibit extreme specification sensitivity. Industrial production growth has an estimated premium of -29.0% per year when estimated alone, but $+12.3\%$ when the market factor is included, $+11.0\%$ with FF3 controls, and $+7.5\%$ with FF5 controls. The term spread premium ranges from $+10.1\%$ (alone) to -6.3% (FF3 controls). The credit spread shows similar instability. These wild swings—including sign reversals—illustrate the specification dependence problem analyzed in Section 3: FM estimates conflate each factor's true premium with contamination from the included control factors.

By contrast, the ET estimates are stable and economically small. Industrial production growth has an estimated premium of 0.17% per year ($t = 0.37$), the term spread 0.10% ($t = 0.60$), and the credit spread 0.06% ($t = 0.96$). None of these estimates is statistically distinguishable from zero. The three-pass estimates are similarly small: 0.17% for IP growth, 0.01% for term spread, and -0.01% for credit spread, with only IP growth marginally significant ($t = 0.92$). The broad agreement between ET and three-pass—both of which are designed to be robust to misspecification—suggests these macro factors do not command substantial risk premia once specification biases are removed.

These findings reinforce two lessons. First, the specification sensitivity of FM estimates is not merely a theoretical concern but produces economically implausible and unstable results in practice. Adding or removing control factors generates premium estimates that differ by tens of percentage points and even flip signs. Second, methods designed for robustness—ET through its marginal moment structure, three-pass through latent factor extraction—deliver consistent messages: the macro factors examined here do not appear to be strongly priced in the cross-section of equity returns. The simulations in Section 6 demonstrated that ET maintains accuracy under misspecification where FM deteriorates; the present results show that this robustness matters for empirical inference.

Table 4: Non-Traded Factors

Statistic	I.P. Growth	Term Spread	Credit Spread
FM: Alone	−28.971 (9.437) [−3.070]	10.141 (3.137) [3.233]	−1.885 (0.591) [−3.191]
FM: +RMRF	12.336 (2.488) [4.959]	−0.061 (1.334) [−0.045]	0.123 (0.370) [0.332]
FM: +FF3	11.027 (2.360) [4.673]	−6.257 (1.234) [−5.070]	0.294 (0.243) [1.209]
FM: +FF5	7.503 (2.121) [3.538]	−4.051 (1.089) [−3.719]	0.030 (0.213) [0.140]
ET	0.166 (0.452) [0.366]	0.099 (0.166) [0.600]	0.055 (0.057) [0.959]
3-Pass	0.165 (0.180) [0.918]	0.005 (0.071) [0.074]	−0.011 (0.023) [−0.463]

This table reports annualized risk premia on non-traded macro factors estimated with Fama–MacBeth (FM), Exponential-Tilting (ET), and Three-Pass (GX) procedures. For each factor, FM estimates are shown under four specifications: Alone, +MKT, +FF3, and +FF5. ET and 3-Pass rows use the FF5 + macro-factor specification. Entries are λ (standard error) [t-statistic]; all are annualized. FM and ET standard errors are heteroskedasticity- and autocorrelation-consistent (Newey–West, 12 lags); 3-Pass standard errors are based on a time-series bootstrap.

8 Conclusion

This paper develops a semi-parametric methodology for estimating factor risk premia without imposing a correctly specified linear factor model. By constructing the stochastic discount factor through exponential tilting of observed returns, the estimator produces an always-positive kernel that enforces the Euler restrictions implied by no-arbitrage while remaining agnostic about the functional form of the pricing relation. The central feature of the method is that the estimated premium for any factor depends only on that factor and the recovered SDF, and is invariant to the inclusion or exclusion of other factors. Simulations calibrated to the Fama–French–Carhart model show that exponential tilting matches conventional two-pass efficiency under correct specification and remains unbiased when priced factors are omitted. Empirically, the method delivers traded-factor premia close to time-series means and economically interpretable premia for non-traded macro variables such as intermediary leverage and consumption growth. Together, these results demonstrate that exponential tilting offers a practical and theoretically coherent approach for estimating factor risk premia when model misspecification is a first-order concern.

References

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.

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.

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.

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

Černỳ, 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.

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.

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.

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.

Giglio, S. and Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, 129(7):1947–1990.

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.

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

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.

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

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

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.

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 Theoretical Properties

This section establishes the theoretical properties of the exponential tilting estimator. We show that the estimator delivers risk premia consistent with no-arbitrage conditions even when the researcher's linear factor model is misspecified. Under standard regularity conditions, the exponential tilting estimator is consistent and asymptotically normal when factors are spanned by test assets, and remains a valid no-arbitrage pricing operator under incomplete spanning. The key advantage relative to conventional methods is that the estimator computes each factor's risk premium through a marginal moment condition rather than through joint cross-sectional regression, thereby achieving robustness to model misspecification.

A.1 Setup and Assumptions

Let $R^e_t \in \mathbb{R}^N$ denote a vector of excess returns on N test assets at date t, and let $f_{k,t}$ denote the innovation of factor k at date t. Throughout this section we maintain the following regularity conditions, which are standard in the asset pricing literature.

Assumption A.1 (Stationarity and Ergodicity). *The joint sequence* $\{(R^e_t, f_{k,t})\}_{t=1}^{\infty}$ *is strictly stationary and ergodic.*

Assumption A.2 (Moment Conditions). *For some* $\delta > 0$*, we have* $\sup_t \mathbb{E}[\|R^e_t\|^{2+\delta}] < \infty$ *and* $\mathbb{E}[|f_{k,t}|^{2+\delta}] < \infty$*.*

Assumption A.3 (No Arbitrage). *There exists at least one strictly positive random variable* m_t *such that* $\mathbb{E}[m_t R^e_t] = 0$ *and* $\mathbb{E}[m_t] = 1/R_f$*, where* $R_f > 0$ *denotes the gross risk-free rate.*

Assumption A.4 (Identification). *The second-moment matrix* $\Sigma_R = \mathbb{E}[R^e_t (R^e_t)']$ *is positive definite.*

Assumption A.5 (Regularity for Extremum Estimation). *There exists an open neighborhood* $\mathcal{N}$ *of the population parameter* (γ^*, α^*) *such that* $\mathbb{E}\big[\sup_{(\gamma,\alpha)\in\mathcal{N}} \exp(\gamma' R^e_t + \alpha)\big] < \infty$*. Moreover, the Jacobian matrix of the constraint mapping* $(\gamma, \alpha) \mapsto \big(\mathbb{E}[m_t(\gamma,\alpha) R^e_t], \mathbb{E}[m_t(\gamma,\alpha)]\big)$ *has full row rank at* (γ^*, α^*)*, where* $m_t(\gamma, \alpha) = \exp(\gamma' R^e_t + \alpha - 1)$*.*

Assumption A.6 (Bounded Admissible SDF Set). *The set of admissible stochastic discount factors* $\mathcal{M}$ *satisfies* $\sup_{m\in\mathcal{M}} \mathbb{E}[m_t^2] \leq M_0 < \infty$ *for some finite constant* M_0*. This bound can be motivated in two equivalent ways: either through good-deal bounds that restrict the maximum Sharpe ratio to rule out economically implausible trading opportunities (Cochrane and Saa-Requejo, 2000), or through an entropy penalty that limits the Kullback-Leibler divergence* *between the risk-neutral and physical measures (Hansen and Jagannathan, 1997; Bansal and Lehmann, 1997). In practice, we verify that the estimated exponential tilting stochastic discount factor satisfies $\mathbb{E}[(\hat{m}_t^{ET})^2] < \infty$ in our empirical applications.*

Assumption A.7 (Asymptotic Regularity)**.** *The process $(R_t^e, f_{k,t})$ is strongly mixing with summable mixing coefficients, and the Jacobian matrix $J = \mathbb{E}[\partial \Phi_t(\theta^*, \lambda_k^*)/\partial(\theta', \lambda)']$ is nonsingular, where Φ_t denotes the stacked moment vector defined below and $\theta = (\gamma, \alpha)$.*

Assumptions A.1 through A.4 are standard in asset pricing econometrics. Assumption A.1 ensures that time-series averages converge to their population counterparts under the appropriate laws of large numbers. Assumption A.2 imposes sufficient integrability to guarantee the existence of relevant moments and to support asymptotic normality via central limit theorems. The moment bound exceeding second order accommodates the nonlinear nature of the exponential tilting objective function. Assumption A.3 formalizes the absence of arbitrage opportunities in the cross-section of test assets and ensures the existence of at least one valid stochastic discount factor. Assumption A.4 rules out redundant or linearly dependent test assets and guarantees that the Euler equations uniquely identify a stochastic discount factor within the linear span of excess returns.

Assumption A.5 provides the regularity conditions necessary for standard extremum estimator theory to apply to the exponential tilting dual problem. The uniform exponential moment condition ensures that the sample objective function converges uniformly to the population objective over a neighborhood of the true parameter value, which is required for consistency of the maximizer. The full-rank condition on the constraint Jacobian is a standard constraint qualification that ensures the first-order conditions uniquely characterize the optimum. Assumption A.6 bounds the second moments of admissible stochastic discount factors uniformly, which ensures that the set of feasible risk premia forms a compact interval in the incomplete spanning case. This assumption can be motivated by appealing to good-deal bounds in the sense of Cochrane and Saa-Requejo (2000), which rule out economically implausible Sharpe ratios. Finally, Assumption A.7 provides the mixing and non-singularity conditions necessary for asymptotic normality of the estimator via standard central limit theorems for dependent processes.

We now introduce several definitions that formalize the key economic concepts underlying our theoretical analysis.

Definition A.1 (Valid Stochastic Discount Factor)**.** *A random variable m_t is called a valid stochastic discount factor if it satisfies the no-arbitrage conditions stated in Assumption A.3. That is, m_t must satisfy $\mathbb{E}[m_t R_t^e] = 0$ and $\mathbb{E}[m_t] = 1/R_f$, and must be strictly positive almost surely.*

In incomplete markets, there exist multiple valid stochastic discount factors. Any two valid discount factors differ by a component that is orthogonal to the payoff space spanned by traded assets. This multiplicity reflects the fundamental incompleteness of real-world financial markets and cannot be resolved without imposing additional economic structure or making auxiliary completeness assumptions. Our analysis acknowledges this multiplicity and characterizes how it affects the estimation of factor risk premia.

Definition A.2 (Factor Spanning)**.** *Factor f_k is said to be* ***fully spanned*** *by test assets R_t^e if there exist portfolio weights $w \in \mathbb{R}^N$ and a constant $c \in \mathbb{R}$ such that*

$$f_{k,t} = c + w' R_t^e \quad \textit{almost surely.} \tag{15}$$

Equivalently, the demeaned factor $f_{k,t} - \mathbb{E}[f_{k,t}]$ lies in the linear span of the test asset returns span$\{R_{1,t}^e, \ldots, R_{N,t}^e\}$ almost surely.

Factor f_k is said to be ***approximately spanned*** *if*

$$f_{k,t} = c + w' R_t^e + \eta_t \tag{16}$$

where the residual η_t satisfies $\mathbb{E}[\eta_t | R_t^e] = 0$ and the ***spanning*** *R^2, defined as*

$$R^2_{span} \equiv 1 - \frac{Var(\eta_t)}{Var(f_{k,t})}, \tag{17}$$

is close to unity.

The spanning condition formalizes the notion that test assets contain economically relevant information about the factor. For traded factors that are themselves portfolios constructed from the test assets, full spanning holds by construction with w being the portfolio weights and $R^2_{\text{span}} = 1$. For non-traded factors such as consumption growth, labor income growth, or other macroeconomic variables, approximate spanning holds when the test assets exhibit sufficient statistical exposure to the underlying source of risk.

Spanning and identification. The spanning condition is necessary for identification of the risk premium and applies equally to all return-based estimation methods, not uniquely to exponential tilting. To see why, note that under full spanning, the law of one price immediately implies a unique risk premium $\lambda_k^* = w'\mathbb{E}[R_t^e]$ that is agreed upon by all valid stochastic discount factors. Under approximate spanning with high R^2_{span}, different valid SDFs can assign different prices to the unspanned residual η_t, but the magnitude of this variation is bounded by $\text{Var}(\eta_t)$, which becomes negligible as $R^2_{\text{span}} \to 1$.

Empirical implementation. The spanning portfolio weights w can be estimated via the population projection

$$w = \Sigma_R^{-1}\text{Cov}(R_t^e, f_{k,t}), \tag{18}$$

where $\Sigma_R = \mathbb{E}[R_t^e(R_t^e)']$ is the second-moment matrix of test asset returns. In practice, we implement this projection by regressing the factor on the test assets and computing the resulting R^2 as a measure of spanning quality. An $R^2_{\text{span}} > 0.90$ generally indicates that the factor is well-spanned, while $R^2_{\text{span}} < 0.50$ suggests substantial incompleteness that may lead to identification challenges.

Definition A.3 (Risk Premium)**.** *For any valid stochastic discount factor m_t and factor innovation $f_{k,t}$, the risk premium is defined as*

$$\lambda_k(m) \equiv \mathbb{E}[f_{k,t}] - \mathbb{E}^{Q_m}[f_{k,t}] = -R_f\mathbb{E}[m_t(f_{k,t} - \mathbb{E}[f_{k,t}])], \tag{19}$$

where Q_m denotes the risk-neutral probability measure induced by the stochastic discount factor m_t through the Radon-Nikodym derivative $dQ_m/dP = R_f m_t$.

The risk premium thus measures the difference between the physical expectation and the risk-neutral expectation of the factor, appropriately scaled by the gross risk-free rate. When factor innovations are demeaned so that $\mathbb{E}[f_{k,t}] = 0$, the risk premium simplifies to

$\lambda_k(m) = -R_f\mathbb{E}[m_t f_{k,t}]$. This formulation makes transparent that the risk premium represents the compensation investors require for bearing exposure to factor f_k, as measured by the covariance between the factor and the pricing kernel.

A.2 Main Theoretical Result

We now state and prove our main theoretical result, which establishes the consistency, asymptotic normality, and misspecification robustness of the exponential tilting estimator for factor risk premia.

Let $m_t^{ET}(\gamma, \alpha) = \exp(\gamma' R_t^e + \alpha - 1)$ denote the exponential tilting stochastic discount factor constructed from test assets R_t^e. The population parameters (γ^*, α^*) solve the dual optimization problem

$$\max_{\gamma,\alpha} \frac{\alpha}{R_f} - \mathbb{E}\left[\exp(\gamma' R_t^e + \alpha - 1)\right] \tag{20}$$

subject to the constraints $\mathbb{E}[m_t^{ET}(\gamma, \alpha) R_t^e] = 0$ and $\mathbb{E}[m_t^{ET}(\gamma, \alpha)] = 1/R_f$, which ensure that the resulting stochastic discount factor satisfies the no-arbitrage restrictions. The sample analog replaces population expectations with time-series averages:

$$\max_{\gamma,\alpha} \frac{\alpha}{R_f} - \frac{1}{T}\sum_{t=1}^{T} \exp(\gamma' R_t^e + \alpha - 1). \tag{21}$$

Given the estimated parameters $(\hat{\gamma}, \hat{\alpha})$, we define the exponential tilting estimator of the risk premium for factor k as

$$\hat{\lambda}_k^{ET} = -R_f \frac{1}{T}\sum_{t=1}^{T} m_t^{ET}(\hat{\gamma}, \hat{\alpha}) f_{k,t}. \tag{22}$$

This estimator computes the risk premium through a marginal moment condition involving only the factor of interest and the estimated stochastic discount factor. The key feature distinguishing this approach from conventional methods is that each factor's risk premium is estimated independently, conditional on a common pricing kernel but without any cross-factor restrictions imposed through joint optimization.

Proposition A.1 (No-Arbitrage Consistent Risk Premia)**.** *Under Assumptions A.1 through*

A.7, the exponential tilting estimator defined in equation (22) *satisfies the following properties.*

Part (i): Full Spanning. *If factor f_k is fully spanned by test assets R_t^e in the sense of Definition A.2, then all valid stochastic discount factors imply the same risk premium. Specifically, for any two valid stochastic discount factors $m_t^{(1)}$ and $m_t^{(2)}$, we have $\lambda_k(m^{(1)}) = \lambda_k(m^{(2)}) \equiv \lambda_k^*$, where the common value is given by $\lambda_k^* = w'\mathbb{E}[R_t^e]$ for the portfolio weights w appearing in the spanning representation. Moreover, the exponential tilting estimator is consistent for this unique risk premium:*

$$\hat{\lambda}_k^{ET} \xrightarrow{p} \lambda_k^* \quad \text{as } T \to \infty, \tag{23}$$

and is asymptotically normally distributed:

$$\sqrt{T}(\hat{\lambda}_k^{ET} - \lambda_k^*) \xrightarrow{d} \mathcal{N}(0, \Sigma_k) \quad \text{as } T \to \infty, \tag{24}$$

where the asymptotic variance Σ_k is derived in Appendix B and can be consistently estimated using heteroskedasticity and autocorrelation consistent methods as described in Newey and West (1987).

Part (ii): Incomplete Spanning. *If factor f_k is not fully spanned by test assets R_t^e, then different valid stochastic discount factors may imply different risk premia. The set of all risk premia consistent with no-arbitrage conditions forms a compact interval:*

$$\Lambda_k = \{\lambda_k(m) : m \text{ is a valid SDF}\} = [\lambda_k^{\min}, \lambda_k^{\max}], \tag{25}$$

where the bounds are characterized by the optimization problems

$$\lambda_k^{\max} = \sup\{-R_f\mathbb{E}[m_t f_{k,t}] : \mathbb{E}[m_t R_t^e] = 0, \mathbb{E}[m_t] = 1/R_f, m_t > 0\}, \tag{26}$$

$$\lambda_k^{\min} = \inf\{-R_f\mathbb{E}[m_t f_{k,t}] : \mathbb{E}[m_t R_t^e] = 0, \mathbb{E}[m_t] = 1/R_f, m_t > 0\}. \tag{27}$$

The exponential tilting estimator converges in probability to a value within this interval:

$$\hat{\lambda}_k^{ET} \xrightarrow{p} \lambda_k^{ET} \in [\lambda_k^{\min}, \lambda_k^{\max}] \quad \text{as } T \to \infty, \tag{28}$$

where the limiting value λ_k^{ET} corresponds to the stochastic discount factor that minimizes the Kullback-Leibler divergence from the uniform probability measure subject to the pricing constraints. The specific location of λ_k^{ET} within the feasible interval depends on the choice of test assets, but the estimate remains a valid no-arbitrage price. The width $\lambda_k^{\max} - \lambda_k^{\min}$ quantifies the degree to which factor f_k is unspanned by the test assets.

Part (iii): Misspecification Robustness. *Conditional on a fixed set of test assets R_t^e, the point estimate $\hat{\lambda}_k^{ET}$ depends only on the factor innovation $f_{k,t}$ and the estimated stochastic discount factor parameters $(\hat{\gamma}, \hat{\alpha})$ obtained from solving equation (21). The estimator does not depend on which other factors the researcher observes, includes in auxiliary analyses, or specifies in any linear factor model. In particular, if the researcher omits priced factors or includes irrelevant factors when conducting separate analyses, these choices do not affect the consistency result stated in equation (23) for factors satisfying the full spanning condition.*

Proof. We establish each part of the proposition through a sequence of steps, beginning with preliminary results on the properties of the exponential tilting stochastic discount factor.

Preliminary Lemmas

We first establish that the sample solution to the exponential tilting dual problem converges to the population solution, and that the limiting stochastic discount factor satisfies the no-arbitrage conditions.

Lemma A.1 (Consistency of Exponential Tilting Parameters)**.** *Under Assumptions A.1 through A.5, the sample solution $(\hat{\gamma}, \hat{\alpha})$ to the optimization problem in equation (21) converges in probability to the unique solution (γ^*, α^*) of the population problem in equation (20). That is, $(\hat{\gamma}, \hat{\alpha}) \xrightarrow{p} (\gamma^*, \alpha^*)$ as $T \to \infty$.*

Proof. The objective function in equation (20) is strictly concave in (γ, α) because the exponential function is strictly convex and the expectation operator preserves strict convexity. Under Assumption A.5, the full-rank condition on the constraint Jacobian combined with strict concavity of the objective ensures that the population optimization problem has a unique solution. By Slater's condition, there exists a feasible interior point for the constraint set, and the Karush-Kuhn-Tucker conditions are necessary and sufficient for optimality. Therefore, the first-order conditions uniquely characterize the population parameter (γ^*, α^*).

To establish consistency, we apply the general theory for extremum estimators developed in Newey and McFadden (1994, Theorem 2.1). The uniform law of large numbers ensures that the sample objective function converges uniformly to the population objective over compact parameter sets. Specifically, Assumption A.5 provides the uniform integrability condition necessary to apply Newey and McFadden (1994, Theorem 2.4). For any compact set Θ containing the true parameter value in its interior, we have

$$\sup_{(\gamma,\alpha)\in\Theta} \left| \frac{1}{T}\sum_{t=1}^{T} \exp(\gamma' R_t^e + \alpha - 1) - \mathbb{E}[\exp(\gamma' R_t^e + \alpha - 1)] \right| \xrightarrow{p} 0 \tag{29}$$

as $T \to \infty$ by Assumption A.1 and the uniform exponential moment bound in Assumption A.5. The continuous mapping theorem then implies that the maximizer of the sample problem converges to the unique maximizer of the population problem. □

Lemma A.2 (Validity of the Exponential Tilting Stochastic Discount Factor)**.** *The stochastic discount factor* $m_t^{ET}(\gamma^*, \alpha^*) = \exp(\gamma^{*\prime} R_t^e + \alpha^* - 1)$ *constructed using the population parameters is a valid stochastic discount factor in the sense of Definition A.1. That is, it satisfies* $\mathbb{E}[m_t^{ET}(\gamma^*, \alpha^*) R_t^e] = 0$*,* $\mathbb{E}[m_t^{ET}(\gamma^*, \alpha^*)] = 1/R_f$*, and* $m_t^{ET}(\gamma^*, \alpha^*) > 0$ *almost surely.*

Proof. Strict positivity of the stochastic discount factor follows immediately from the exponential functional form, which is strictly positive for all real arguments. The satisfaction of the Euler equation and mean constraint follows from the Karush-Kuhn-Tucker first-order conditions of the constrained optimization problem in equation (20).

To see this formally, consider the Lagrangian function for the dual problem:

$$\mathcal{L}(\gamma, \alpha, \nu, \mu) = \frac{\alpha}{R_f} - \mathbb{E}[m_t(\gamma, \alpha)] - \nu' \mathbb{E}[m_t(\gamma, \alpha) R_t^e] - \mu \left(\mathbb{E}[m_t(\gamma, \alpha)] - \frac{1}{R_f} \right), \tag{30}$$

where $m_t(\gamma, \alpha) = \exp(\gamma' R_t^e + \alpha - 1)$, and $\nu \in \mathbb{R}^N$ and $\mu \in \mathbb{R}$ are Lagrange multipliers associated with the Euler equation constraint and mean constraint, respectively. By Slater's condition and the strict concavity established in Lemma A.1, the KKT conditions are necessary and sufficient. Taking derivatives with respect to the primal variables and evaluating at the optimum (γ^*, α^*) yields the first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \gamma} = -\mathbb{E}[m_t(\gamma^*, \alpha^*)R_t^e] - \mathbb{E}[m_t(\gamma^*, \alpha^*)R_t^e(R_t^e)']\nu = 0, \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{1}{R_f} - \mathbb{E}[m_t(\gamma^*, \alpha^*)] - \mu = 0, \tag{32}$$

together with the binding constraints $\mathbb{E}[m_t(\gamma^*, \alpha^*)R_t^e] = 0$ and $\mathbb{E}[m_t(\gamma^*, \alpha^*)] = 1/R_f$. These conditions jointly imply that the exponential tilting stochastic discount factor is valid. □

Proof of Part (i): Full Spanning

We now establish that when a factor is fully spanned, all valid stochastic discount factors agree on its risk premium, and that the exponential tilting estimator consistently identifies this unique value.

Suppose factor f_k is fully spanned by test assets R_t^e in the sense of Definition A.2. Then there exist portfolio weights $w \in \mathbb{R}^N$ and a constant $c \in \mathbb{R}$ such that $f_{k,t} = c + w' R_t^e$ holds almost surely. Let $m_t^{(1)}$ and $m_t^{(2)}$ be any two valid stochastic discount factors. We will show that they imply identical risk premia for factor f_k.

Since both stochastic discount factors are valid, they satisfy the no-arbitrage conditions:

$$\mathbb{E}[m_t^{(1)} R_t^e] = 0, \quad \mathbb{E}[m_t^{(2)} R_t^e] = 0, \quad \mathbb{E}[m_t^{(1)}] = \mathbb{E}[m_t^{(2)}] = \frac{1}{R_f}. \tag{33}$$

Consider the risk premium implied by the first stochastic discount factor. Using the spanning representation of the factor and Definition A.3, we can write:

$$
\begin{aligned}
\lambda_k(m^{(1)}) &= -R_f \mathbb{E}[m_t^{(1)}(f_{k,t} - \mathbb{E}[f_{k,t}])] \\
&= -R_f \mathbb{E}[m_t^{(1)}(c + w' R_t^e - c - w' \mathbb{E}[R_t^e])] \\
&= -R_f \mathbb{E}[m_t^{(1)} w' (R_t^e - \mathbb{E}[R_t^e])] \\
&= -R_f w' \mathbb{E}[m_t^{(1)} (R_t^e - \mathbb{E}[R_t^e])] \\
&= -R_f w' \left(\mathbb{E}[m_t^{(1)} R_t^e] - \mathbb{E}[m_t^{(1)}] \mathbb{E}[R_t^e] \right) \\
&= -R_f w' \left(0 - \frac{1}{R_f} \mathbb{E}[R_t^e] \right) \\
&= w' \mathbb{E}[R_t^e]. \qquad (34)
\end{aligned}
$$

An identical sequence of steps applied to the second stochastic discount factor yields $\lambda_k(m^{(2)}) = w'\mathbb{E}[R_t^e]$. Therefore, we have established that $\lambda_k(m^{(1)}) = \lambda_k(m^{(2)}) \equiv \lambda_k^*$ for all valid stochastic discount factors, where the common value is $\lambda_k^* = w'\mathbb{E}[R_t^e]$. This uniqueness result holds regardless of which valid stochastic discount factor is employed to compute the risk premium. To clarify the mechanism, note that the difference $\Delta m_t = m_t^{(1)} - m_t^{(2)}$ between any two valid stochastic discount factors satisfies $\mathbb{E}[\Delta m_t] = 0$ and $\mathbb{E}[\Delta m_t R_t^e] = 0$, which implies that Δm_t is orthogonal to the span of $(1, R_t^e)$. Since the spanned factor $f_{k,t} = c + w' R_t^e$ lies entirely within this span, we have $\mathbb{E}[\Delta m_t f_{k,t}] = 0$, establishing that the risk premium is pinned down solely by the no-arbitrage conditions and the spanning property.

To establish consistency of the exponential tilting estimator, we proceed as follows. By Lemma A.1, we have $(\hat{\gamma}, \hat{\alpha}) \xrightarrow{p} (\gamma^*, \alpha^*)$ as $T \to \infty$. By the continuous mapping theorem, this implies $m_t^{ET}(\hat{\gamma}, \hat{\alpha}) \xrightarrow{p} m_t^{ET}(\gamma^*, \alpha^*)$ pointwise for each t. The estimator can be written as

$$
\hat{\lambda}_k^{ET} = -R_f \frac{1}{T} \sum_{t=1}^{T} m_t^{ET}(\hat{\gamma}, \hat{\alpha}) f_{k,t}. \qquad (35)
$$

Applying the ergodic theorem under Assumption A.1 and the continuous mapping theorem, we obtain:

$$
\hat{\lambda}_k^{ET} \xrightarrow{p} -R_f \mathbb{E}[m_t^{ET}(\gamma^*, \alpha^*) f_{k,t}] = \lambda_k(m^{ET}) = \lambda_k^*, \qquad (36)
$$

where the final equality follows from the uniqueness result we established above, since $m_t^{ET}(\gamma^*, \alpha^*)$ is a valid stochastic discount factor by Lemma A.2. This establishes the consistency stated in equation (23).

For the asymptotic distribution, we apply the general theory for generalized empirical likelihood estimators developed in Newey and Smith (2004, Theorem 3.1). Define the stacked moment vector:

$$\Phi_t(\theta, \lambda) = \begin{pmatrix} \psi_t(\theta) \\ \phi_t(\theta) - \lambda \end{pmatrix}, \tag{37}$$

where $\theta = (\gamma, \alpha)$, the vector $\psi_t(\theta)$ collects the first-order conditions from the exponential tilting dual problem:

$$\psi_t(\theta) = \begin{pmatrix} -m_t(\theta) R_t^e \\ \frac{1}{R_f} - m_t(\theta) \end{pmatrix}, \tag{38}$$

and $\phi_t(\theta) = -R_f m_t(\theta) f_{k,t}$ is the moment function defining the risk premium. The population moment condition is $\mathbb{E}[\Phi_t(\theta^*, \lambda_k^*)] = 0$ where $\theta^* = (\gamma^*, \alpha^*)$.

The sample estimator $(\hat{\theta}, \hat{\lambda}_k)$ is defined implicitly by setting the sample moments equal to zero: $\frac{1}{T} \sum_t \Phi_t(\hat{\theta}, \hat{\lambda}_k) = 0$. Under Assumption A.7, the mixing condition ensures that a functional central limit theorem applies to the moment vector, and the non-singularity of the Jacobian J ensures invertibility of the asymptotic covariance matrix. Standard results for exactly identified generalized method of moments estimators as presented in Newey and Smith (2004, Theorem 3.1) yield the asymptotic distribution:

$$\sqrt{T} \begin{pmatrix} \hat{\theta} - \theta^* \\ \hat{\lambda}_k - \lambda_k^* \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, J^{-1} S (J^{-1})'\right), \tag{39}$$

where $J = \mathbb{E}[\partial \Phi_t(\theta^*, \lambda_k^*) / \partial(\theta', \lambda)']$ is the Jacobian matrix of the moment function evaluated at the true parameter values, and S denotes the long-run covariance matrix of the moment vector $\Phi_t(\theta^*, \lambda_k^*)$. The asymptotic variance of $\hat{\lambda}_k$ is the $(N+2, N+2)$ element of the matrix $J^{-1} S (J^{-1})'$. We denote this scalar variance by Σ_k. The explicit derivation of Σ_k via the influence function is provided in Appendix B. This variance can be estimated consistently using heteroskedasticity and autocorrelation consistent covariance estimators such as the Newey-West estimator described in Newey and West (1987). This completes the proof of

Part (i).

Proof of Part (ii): Incomplete Spanning

We now characterize the case in which factor f_k is not fully spanned by test assets R_t^e. In this situation, the factor can be decomposed as $f_{k,t} = c + w' R_t^e + \eta_t$ where η_t represents the component of the factor that lies outside the span of the traded returns. This unspanned component cannot be perfectly replicated using portfolios of the test assets. Different valid stochastic discount factors may price this unspanned component differently, leading to a multiplicity of risk premia consistent with the no-arbitrage conditions.

To characterize the set of feasible risk premia, let $\mathcal{M}$ denote the set of all valid stochastic discount factors:

$$\mathcal{M} = \{m_t : \mathbb{E}[m_t R_t^e] = 0, \mathbb{E}[m_t] = 1/R_f, m_t > 0 \text{ almost surely}\}. \tag{40}$$

For each valid stochastic discount factor $m \in \mathcal{M}$, the implied risk premium for factor k is $\lambda_k(m) = -R_f \mathbb{E}[m_t f_{k,t}]$. Define the set of all risk premia consistent with no-arbitrage as:

$$\Lambda_k = \{\lambda_k(m) : m \in \mathcal{M}\}. \tag{41}$$

We now establish that Λ_k forms a non-empty, closed, and bounded interval.

Non-emptiness. By Assumption A.3, at least one valid stochastic discount factor exists. Therefore, the set $\mathcal{M}$ is non-empty, which immediately implies that Λ_k is non-empty.

Boundedness. Under Assumption A.6, the second moments of all admissible stochastic discount factors are uniformly bounded by the constant M_0. By the Cauchy-Schwarz inequality, we have

$$|\lambda_k(m)| = R_f |\mathbb{E}[m_t f_{k,t}]| \leq R_f \sqrt{\mathbb{E}[m_t^2]} \sqrt{\mathbb{E}[f_{k,t}^2]} \leq R_f \sqrt{M_0} \sqrt{\mathbb{E}[f_{k,t}^2]} < \infty \tag{42}$$

uniformly over all $m \in \mathcal{M}$, where the final inequality follows from Assumption A.2. Therefore, Λ_k is bounded. We note that Assumption A.6 can be motivated by appealing to good-deal bounds in the sense of Cochrane and Saa-Requejo (2000), which impose economically motivated restrictions on the admissible set of stochastic discount factors by ruling out Sharpe ratios that are implausibly large. An alternative motivation is through entropy penalties that limit the Kullback-Leibler divergence between risk-neutral and physical measures, as discussed in Hansen and Jagannathan (1997) and Bansal and Lehmann (1997).

Closedness. Consider a sequence $\{\lambda_k^{(n)}\}_{n=1}^{\infty} \subset \Lambda_k$ converging to some limit $\lambda_k^{(\infty)}$. For each n, there exists a valid stochastic discount factor $m^{(n)} \in \mathcal{M}$ such that $\lambda_k^{(n)} = -R_f \mathbb{E}[m_t^{(n)} f_{k,t}]$. Under Assumption A.6, the sequence $\{m^{(n)}\}$ is uniformly bounded in $L^2(\mathbb{P})$, which ensures uniform integrability of the sequence. By the Banach-Alaoglu theorem (see Hansen and Jagannathan, 1991, for application to stochastic discount factors), the unit ball in the dual space of $L^2(\mathbb{P})$ is weakly compact. The pricing constraints defining $\mathcal{M}$ are linear and hence weakly closed. Therefore, the set $\mathcal{M}$ is weakly sequentially compact, and the sequence $\{m^{(n)}\}$ has a weakly convergent subsequence with weak limit $m^{(\infty)} \in \mathcal{M}$. By the definition of weak convergence, we have $\lambda_k^{(\infty)} = -R_f \mathbb{E}[m_t^{(\infty)} f_{k,t}] \in \Lambda_k$, establishing that Λ_k is closed.

Connectedness. To show that Λ_k is connected, observe that the set $\mathcal{M}$ is convex: if $m^{(1)}$ and $m^{(2)}$ are valid stochastic discount factors, then for any $\tau \in [0, 1]$, the convex combination $m^{(\tau)} = \tau m^{(1)} + (1 - \tau) m^{(2)}$ also satisfies $\mathbb{E}[m^{(\tau)} R_t^e] = 0$, $\mathbb{E}[m^{(\tau)}] = 1/R_f$, and $m^{(\tau)} > 0$ almost surely. The function $\tau \mapsto \lambda_k(m^{(\tau)}) = -R_f \mathbb{E}[m_t^{(\tau)} f_{k,t}]$ is continuous in τ, and it traces out a continuous path from $\lambda_k(m^{(1)})$ to $\lambda_k(m^{(2)})$ as τ varies from 0 to 1. This establishes that Λ_k is path-connected and hence connected.

Since Λ_k is a non-empty, closed, bounded, and connected subset of $\mathbb{R}$, it must be a closed interval. We can therefore write $\Lambda_k = [\lambda_k^{\min}, \lambda_k^{\max}]$, where the bounds are given by the supremum and infimum in equations (26) and (27). These bounds can be computed explicitly when additional structure is imposed on the admissible set, such as entropy or variance constraints, as discussed in the good-deal bounds literature.

The exponential tilting stochastic discount factor m_t^{ET} is one particular element of $\mathcal{M}$, distinguished by the property that it minimizes the Kullback-Leibler divergence (relative entropy) from the uniform probability measure subject to satisfying the pricing constraints. By Lemma A.2, we have $m_t^{ET} \in \mathcal{M}$, which implies that the corresponding risk premium $\lambda_k^{ET} = \lambda_k(m^{ET}) = -R_f \mathbb{E}[m_t^{ET} f_{k,t}]$ lies in the interval Λ_k.

By the consistency result in Lemma A.1 and the continuous mapping theorem, we have:

$$\hat{\lambda}_k^{ET} = -R_f \frac{1}{T} \sum_{t=1}^{T} m_t^{ET}(\hat{\gamma}, \hat{\alpha}) f_{k,t} \xrightarrow{p} -R_f \mathbb{E}[m_t^{ET}(\gamma^*, \alpha^*) f_{k,t}] = \lambda_k^{ET} \in [\lambda_k^{\min}, \lambda_k^{\max}], \tag{43}$$

establishing the consistency result stated in equation (28).

The specific value of λ_k^{ET} within the feasible interval depends on which test assets are used to construct the stochastic discount factor. Different choices of test assets lead to different exponential tilting stochastic discount factors, each of which is valid but may yield different risk premia for factors that are not fully spanned. This sensitivity to test asset selection is not a defect of the estimation method but rather reflects genuine economic content: it reveals and quantifies the degree of market incompleteness with respect to factor f_k.

The width of the interval $\lambda_k^{\max} - \lambda_k^{\min}$ provides an interpretable measure of how much the risk premium depends on the specific stochastic discount factor employed. When this width is small, the factor is approximately spanned and the risk premium is nearly unique across all valid stochastic discount factors. When the width is large, the factor has a substantial unspanned component and the range of no-arbitrage consistent prices is correspondingly wide. Linear factor models implicitly assume that each factor has a unique, well-defined risk premium, thereby ruling out such incompleteness by assumption. The exponential tilting framework acknowledges this incompleteness explicitly and provides a method for quantifying it empirically. This completes the proof of Part (ii).

Proof of Part (iii): Misspecification Robustness

We establish that the exponential tilting risk premium estimator for any factor k is invariant to the researcher's specification choices regarding other factors. The result follows from the separability of the estimation procedure into kernel recovery and factor pricing stages.

Step 1: Invariance of the Pricing Kernel to Factor Specifications.

The exponential tilting parameters solve:

$$(\hat{\gamma}, \hat{\alpha}) \in \arg\max_{(\gamma,\alpha)\in\Theta} \left\{ \frac{\alpha}{R_f} - \frac{1}{T} \sum_{t=1}^{T} \exp(\gamma' R_t^e + \alpha - 1) \right\}, \tag{44}$$

where $\Theta \subset \mathbb{R}^{N+1}$ denotes the parameter space. The first-order conditions characterizing the solution are:

$$\frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\gamma}, \hat{\alpha}) R_t^e = 0, \tag{45}$$

$$\frac{1}{R_f} - \frac{1}{T}\sum_{t=1}^{T} m_t(\hat{\gamma}, \hat{\alpha}) = 0, \tag{46}$$

where $m_t(\gamma, \alpha) = \exp(\gamma' R_t^e + \alpha - 1)$.

The system (45)–(46) comprises $N+1$ equations in $N+1$ unknowns and depends exclusively on the test asset excess returns $\{R_t^e\}_{t=1}^T$ and the risk-free rate R_f. Under Assumptions A.1–A.5, the solution exists, is unique, and satisfies $(\hat{\gamma}, \hat{\alpha}) \xrightarrow{p} (\gamma^*, \alpha^*)$ as $T \to \infty$ by Lemma A.1.

Observe that neither the optimization problem (44) nor the first-order conditions (45)–(46) involve any factor realizations $f_{k,t}$, factor loadings β_{ik}, or specifications of linear factor models. The pricing kernel $m_t^{ET}(\hat{\gamma}, \hat{\alpha})$ is determined entirely by the no-arbitrage restrictions on test assets. Consequently, for any two researchers using identical test asset data $\{R_t^e\}_{t=1}^T$ but potentially analyzing different sets of factors or imposing different factor model specifications, the estimated parameters $(\hat{\gamma}, \hat{\alpha})$ and the implied pricing kernel m_t^{ET} are identical.

Step 2: Risk Premium as an Independent Marginal Moment.

Given the estimated pricing kernel from Step 1, the risk premium for factor k is:

$$\hat{\lambda}_k^{ET} = -R_f \frac{1}{T}\sum_{t=1}^{T} m_t^{ET}(\hat{\gamma}, \hat{\alpha}) f_{k,t}. \tag{47}$$

Equation (47) defines a univariate sample moment involving only the sequence $\{(m_t^{ET}, f_{k,t})\}_{t=1}^T$. For distinct factors k and j, the estimators $\hat{\lambda}_k^{ET}$ and $\hat{\lambda}_j^{ET}$ are computed via separate moment evaluations and are not coupled through any system of equations, matrix inversions, or joint optimization. The value of $\hat{\lambda}_k^{ET}$ does not enter the formula for $\hat{\lambda}_j^{ET}$ and vice versa.

This separability contrasts with conventional methods. In the Fama and MacBeth (1973) two-pass procedure, risk premia are jointly estimated via the cross-sectional regression:

$$\hat{\lambda}^{FM} = (\hat{B}'\hat{B})^{-1}\hat{B}'\bar{R}^e, \tag{48}$$

where $\hat{B} \in \mathbb{R}^{N\times K}$ collects estimated factor loadings and $\bar{R}^e \in \mathbb{R}^N$ contains sample mean excess returns. Modifying the factor set alters the dimension and composition of $\hat{B}$, thereby changing the projection matrix $(\hat{B}'\hat{B})^{-1}\hat{B}'$ and affecting the estimated premium for every factor, including those whose inclusion was unchanged. In GMM estimation (Hansen, 1982), all parameters are determined by a coupled system of moment conditions, and specification changes similarly propagate across all estimates.

The exponential tilting approach avoids this coupling. Once $(\hat{\gamma}, \hat{\alpha})$ are determined from asset returns alone, computing $\hat{\lambda}_k^{ET}$ requires only the bivariate sample $\{(m_t^{ET}, f_{k,t})\}$ and imposes no restrictions on other factors.

Step 3: Formal Statement of Invariance.

Let $\mathcal{F}_A = \{f_1, \ldots, f_K\}$ and $\mathcal{F}_B = \{f_1, \ldots, f_K, f_{K+1}, \ldots, f_{K+J}\}$ denote two distinct factor sets with $\mathcal{F}_A \subset \mathcal{F}_B$, and suppose factor $k \in \mathcal{F}_A$. Consider two researchers who use identical test asset data $\{R_t^e\}_{t=1}^T$ but analyze different factor sets. Researcher A computes $\hat{\lambda}_k^{ET,A}$ using factor set $\mathcal{F}_A$, while Researcher B computes $\hat{\lambda}_k^{ET,B}$ using factor set $\mathcal{F}_B$.

By Step 1, both researchers obtain identical SDF parameters:

$$(\hat{\gamma}^A, \hat{\alpha}^A) = (\hat{\gamma}^B, \hat{\alpha}^B) \equiv (\hat{\gamma}, \hat{\alpha}), \tag{49}$$

since these parameters depend only on $\{R_t^e\}_{t=1}^T$.

By Step 2, the risk premium estimates satisfy:

$$\hat{\lambda}_k^{ET,A} = -R_f \frac{1}{T}\sum_{t=1}^{T} \exp(\hat{\gamma}' R_t^e + \hat{\alpha} - 1) f_{k,t}, \tag{50}$$

$$\hat{\lambda}_k^{ET,B} = -R_f \frac{1}{T}\sum_{t=1}^{T} \exp(\hat{\gamma}' R_t^e + \hat{\alpha} - 1) f_{k,t}. \tag{51}$$

Since the expressions are identical, we conclude:

$$\hat{\lambda}_k^{ET,A} = \hat{\lambda}_k^{ET,B}. \tag{52}$$

More generally, for any fixed test asset data $\{R_t^e\}_{t=1}^T$, the mapping from factor realizations to risk premium estimates is:

$$f_{k,t} \mapsto \hat{\lambda}_k^{ET} = h_k(R_t^e, f_{k,t}; R_f), \tag{53}$$

where the functional form h_k is independent of $\{f_j\}_{j \neq k}$. This functional independence ensures that $\hat{\lambda}_k^{ET}$ is invariant to: (i) addition or removal of factors f_j with $j \neq k$; (ii) misspecification of the factor model structure; (iii) inclusion of spurious or omission of priced factors in concurrent or subsequent analyses.

Step 4: Consistency and Asymptotic Properties.

The invariance property holds in finite samples for the point estimate $\hat{\lambda}_k^{ET}$. For the limiting behavior, note that under the spanning condition in Part (i), there exists a unique no-arbitrage risk premium $\lambda_k^* = w'\mathbb{E}[R_t^e]$ for factor k. By Lemma A.1 and the continuous mapping theorem:

$$\hat{\lambda}_k^{ET} = -R_f \frac{1}{T} \sum_{t=1}^{T} m_t^{ET}(\hat{\gamma}, \hat{\alpha}) f_{k,t} \xrightarrow{p} -R_f \mathbb{E}[m_t^{ET}(\gamma^*, \alpha^*) f_{k,t}] = \lambda_k^* \tag{54}$$

as $T \to \infty$. This consistency result obtains regardless of which other factors are included in auxiliary analyses, since (γ^*, α^*) and hence the limiting pricing kernel $m_t^{ET}(\gamma^*, \alpha^*)$ are determined solely by the population Euler equations on test assets.

The asymptotic distribution $\sqrt{T}(\hat{\lambda}_k^{ET} - \lambda_k^*) \xrightarrow{d} \mathcal{N}(0, \Sigma_k)$ derived in Part (i) similarly holds uniformly across different researcher specifications, since the influence function depends on the joint distribution of $(R_t^e, f_{k,t})$ but not on other factors.

Remark on Standard Errors. While point estimates exhibit invariance, standard errors depend on the full information set through the asymptotic covariance matrix $\Sigma_k = (J^{-1}SJ^{-1'})[N+2, N+2]$, where S is the long-run covariance of the stacked moment vector $\Phi_t(\theta^*, \lambda_k^*)$. If the researcher uses additional factors to construct test assets or modifies the test asset set, the dimension N changes, altering J and S and hence affecting precision. The robustness property concerns point estimation conditional on a fixed test asset set, not the choice of that set itself.

This completes the proof of Part (iii) and thereby establishes all three claims of Proposition A.1. $\square$

B Derivation of the Asymptotic Variance and Influence Function

This appendix derives the explicit formula for the asymptotic variance Σ_k of the exponential tilting risk premium estimator $\hat{\lambda}_k^{ET}$, following the influence function approach for generalized empirical likelihood estimators developed in Newey and Smith (2004). We provide the complete derivation to facilitate implementation and inference in empirical applications.

Setup and Notation

Recall that the exponential tilting estimator is defined by the stacked moment conditions:

$$\Phi_t(\theta, \lambda_k) = \begin{pmatrix} \psi_t(\theta) \\ \phi_t(\theta) - \lambda_k \end{pmatrix} = \begin{pmatrix} -m_t(\theta) R_t^e \\ \frac{1}{R_f} - m_t(\theta) \\ -R_f m_t(\theta) f_{k,t} - \lambda_k \end{pmatrix}, \tag{55}$$

where $\theta = (\gamma, \alpha) \in \mathbb{R}^{N+1}$ are the parameters of the exponential tilting stochastic discount factor $m_t(\theta) = \exp(\gamma' R_t^e + \alpha - 1)$, and $\lambda_k \in \mathbb{R}$ is the risk premium for factor k. The dimension of the moment vector is $\dim(\Phi_t) = N + 2$, matching the dimension of the parameter vector (θ, λ_k).

The estimator $(\hat{\theta}, \hat{\lambda}_k)$ solves the sample moment condition:

$$\frac{1}{T} \sum_{t=1}^{T} \Phi_t(\hat{\theta}, \hat{\lambda}_k) = 0. \tag{56}$$

At the population level, the true parameters (θ^*, λ_k^*) satisfy:

$$\mathbb{E}[\Phi_t(\theta^*, \lambda_k^*)] = 0. \tag{57}$$

Jacobian Matrix

Define the Jacobian matrix of the moment function evaluated at the true parameters:

$$J = \mathbb{E}\left[\frac{\partial \Phi_t(\theta^*, \lambda_k^*)}{\partial(\theta', \lambda_k)'}\right] \in \mathbb{R}^{(N+2)\times(N+2)}. \tag{58}$$

To compute this explicitly, we calculate the derivatives of each component. For the first block corresponding to $\psi_t(\theta) = (-m_t R_t^e, 1/R_f - m_t)'$:

$$\frac{\partial \psi_t}{\partial \gamma} = \begin{pmatrix} -m_t R_t^e (R_t^e)' \\ -m_t (R_t^e)' \end{pmatrix} \in \mathbb{R}^{(N+1)\times N}, \tag{59}$$

$$\frac{\partial \psi_t}{\partial \alpha} = \begin{pmatrix} -m_t R_t^e \\ -m_t \end{pmatrix} \in \mathbb{R}^{N+1}, \tag{60}$$

$$\frac{\partial \psi_t}{\partial \lambda_k} = 0 \in \mathbb{R}^{N+1}. \tag{61}$$

For the second block corresponding to $\phi_t(\theta) - \lambda_k = -R_f m_t f_{k,t} - \lambda_k$:

$$\frac{\partial(\phi_t - \lambda_k)}{\partial \gamma} = -R_f m_t f_{k,t} (R_t^e)' \in \mathbb{R}^{1\times N}, \tag{62}$$

$$\frac{\partial(\phi_t - \lambda_k)}{\partial \alpha} = -R_f m_t f_{k,t} \in \mathbb{R}, \tag{63}$$

$$\frac{\partial(\phi_t - \lambda_k)}{\partial \lambda_k} = -1 \in \mathbb{R}. \tag{64}$$

Assembling these components and taking expectations (all evaluated at (θ^*, λ_k^*)) yields:

$$J = \begin{pmatrix} -\mathbb{E}[m_t(\theta^*) R_t^e (R_t^e)'] & -\mathbb{E}[m_t(\theta^*) R_t^e] & 0 \\ -\mathbb{E}[m_t(\theta^*)(R_t^e)'] & -\mathbb{E}[m_t(\theta^*)] & 0 \\ -R_f \mathbb{E}[m_t(\theta^*) f_{k,t} (R_t^e)'] & -R_f \mathbb{E}[m_t(\theta^*) f_{k,t}] & -1 \end{pmatrix}, \tag{65}$$

where we have made explicit that all expectations are evaluated at the true parameter $\theta^* = (\gamma^*, \alpha^*)$. Note that under the no-arbitrage conditions, we have $\mathbb{E}[m_t(\theta^*)R_t^e] = 0$ and $\mathbb{E}[m_t(\theta^*)] = 1/R_f$, which simplifies the middle column. The parameter vector is ordered as $(\gamma', \alpha, \lambda_k)'$ where $\gamma \in \mathbb{R}^N$, $\alpha \in \mathbb{R}$, and $\lambda_k \in \mathbb{R}$, yielding a total dimension of $N + 2$.

Long-Run Covariance Matrix

Define the long-run covariance matrix of the moment vector:

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}[\Phi_t(\theta^*, \lambda_k^*)\Phi_{t-j}(\theta^*, \lambda_k^*)']. \tag{66}$$

This matrix accounts for both heteroskedasticity and serial correlation in the moment conditions. Under Assumption A.7 (strong mixing with summable mixing coefficients), the infinite sum in equation (66) is well-defined and finite.

In practice, S is estimated using a heteroskedasticity and autocorrelation consistent (HAC) estimator such as the Newey-West estimator:

$$\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{L} w_j(\hat{\Gamma}_j + \hat{\Gamma}_j'), \tag{67}$$

where $\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} \Phi_t(\hat{\theta}, \hat{\lambda}_k)\Phi_{t-j}(\hat{\theta}, \hat{\lambda}_k)'$ is the sample autocovariance at lag j, $w_j = 1 - j/(L+1)$ is the Bartlett kernel weight, and L is the bandwidth parameter chosen to balance bias and variance (e.g., $L = \lfloor 4(T/100)^{2/9} \rfloor$ as recommended by Newey and West, 1994).

Asymptotic Distribution

By the central limit theorem for dependent processes under the mixing conditions in Assumption A.7, we have:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_t(\theta^*, \lambda_k^*) \xrightarrow{d} \mathcal{N}(0, S). \tag{68}$$

A first-order Taylor expansion of the sample moment condition around the true parameters yields:

$$0 = \frac{1}{T}\sum_{t=1}^{T}\Phi_t(\hat{\theta}, \hat{\lambda}_k) \approx \frac{1}{T}\sum_{t=1}^{T}\Phi_t(\theta^*, \lambda_k^*) + J \cdot \sqrt{T}\begin{pmatrix} \hat{\theta} - \theta^* \\ \hat{\lambda}_k - \lambda_k^* \end{pmatrix} \cdot \frac{1}{\sqrt{T}}. \tag{69}$$

Rearranging and applying equation (68) gives:

$$\sqrt{T}\begin{pmatrix} \hat{\theta} - \theta^* \\ \hat{\lambda}_k - \lambda_k^* \end{pmatrix} \approx -J^{-1} \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_t(\theta^*, \lambda_k^*) \xrightarrow{d} \mathcal{N}(0, J^{-1}S(J^{-1})'). \tag{70}$$

Asymptotic Variance of the Risk Premium Estimator

The asymptotic variance of $\hat{\lambda}_k$ is the $(N+2, N+2)$ element of the covariance matrix $J^{-1}S(J^{-1})'$. To extract this element explicitly, partition the inverse Jacobian as:

$$J^{-1} = \begin{pmatrix} J^{-1}_{11} & J^{-1}_{12} & J^{-1}_{13} \\ J^{-1}_{21} & J^{-1}_{22} & J^{-1}_{23} \\ J^{-1}_{31} & J^{-1}_{32} & J^{-1}_{33} \end{pmatrix}, \tag{71}$$

where $J^{-1}_{11} \in \mathbb{R}^{N\times N}$, $J^{-1}_{22} \in \mathbb{R}^{1\times 1}$, and $J^{-1}_{33} \in \mathbb{R}^{1\times 1}$ correspond to the γ, α, and λ_k components respectively.

The asymptotic variance is then:

$$\Sigma_k = (J^{-1}S(J^{-1})')_{N+2,N+2} = \sum_{i=1}^{N+2}\sum_{j=1}^{N+2} J^{-1}_{(N+2),i} S_{ij} J^{-1}_{(N+2),j}, \tag{72}$$

where the subscript $(N+2, N+2)$ denotes the bottom-right element of the matrix, and $J^{-1}_{(N+2),i}$ denotes the $(N+2, i)$ element of J^{-1}.

Influence Function Representation

An alternative and equivalent way to express the asymptotic variance uses the influence function. For the exponential tilting estimator, the influence function of $\hat{\lambda}_k$ at observation t is:

$$\mathrm{IF}_t(\hat{\lambda}_k) = -[J^{-1}\Phi_t(\theta^*, \lambda_k^*)]_{N+2}, \tag{73}$$

where the subscript $N+2$ extracts the final component corresponding to λ_k. The asymptotic variance can then be expressed as:

$$\Sigma_k = \mathbb{E}[\mathrm{IF}_t(\hat{\lambda}_k)^2] + 2\sum_{j=1}^{\infty} \mathbb{E}[\mathrm{IF}_t(\hat{\lambda}_k) \cdot \mathrm{IF}_{t-j}(\hat{\lambda}_k)]. \tag{74}$$

This representation makes clear that the asymptotic variance accounts for both the variance of the influence function and its serial correlation structure.

Consistent Estimation

To obtain a consistent estimate of Σ_k, we implement the following procedure:

1. Estimate $(\hat{\theta}, \hat{\lambda}_k)$ by solving equation (56).

2. Compute the sample Jacobian:

$$\hat{J} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \Phi_t(\hat{\theta}, \hat{\lambda}_k)}{\partial(\theta', \lambda_k)'}. \tag{75}$$

3. Compute the HAC estimate of the long-run covariance matrix using equation (67):

$$\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{L} \left(1 - \frac{j}{L+1}\right)(\hat{\Gamma}_j + \hat{\Gamma}_j'). \tag{76}$$

4. Compute the asymptotic variance estimate:

$$\hat{\Sigma}_k = (\hat{J}^{-1}\hat{S}(\hat{J}^{-1})')_{N+2,N+2}. \tag{77}$$

5. Construct the standard error and confidence interval:

$$\mathrm{SE}(\hat{\lambda}_k) = \sqrt{\hat{\Sigma}_k/T}, \tag{78}$$

$$\mathrm{CI}_{95\%} = \left[\hat{\lambda}_k - 1.96 \cdot \mathrm{SE}(\hat{\lambda}_k),\ \hat{\lambda}_k + 1.96 \cdot \mathrm{SE}(\hat{\lambda}_k)\right]. \tag{79}$$

Under Assumptions A.1 through A.7, this procedure yields asymptotically valid inference as $T \to \infty$. In finite samples, the coverage properties depend on the bandwidth choice L and the strength of serial correlation in the moment conditions. Monte Carlo evidence suggests that the Newey-West estimator with automatic bandwidth selection performs well for sample sizes $T \geq 500$, which is satisfied in our empirical applications with monthly data spanning 50 years.

Simplified Expression Under Conditional Homoskedasticity

If the moment conditions are conditionally homoskedastic and serially uncorrelated (i.e., $\mathbb{E}[\Phi_t \Phi'_{t-j} | \mathcal{F}_{t-1}] = 0$ for all $j > 0$), then the long-run covariance simplifies to:

$$S = \mathbb{E}[\Phi_t(\theta^*, \lambda_k^*)\Phi_t(\theta^*, \lambda_k^*)'], \tag{80}$$

and can be consistently estimated by:

$$\hat{S} = \frac{1}{T}\sum_{t=1}^{T} \Phi_t(\hat{\theta}, \hat{\lambda}_k)\Phi_t(\hat{\theta}, \hat{\lambda}_k)'. \tag{81}$$

In this case, no bandwidth choice is required and the estimator simplifies considerably. However, we do not impose this restriction in our empirical work and always use the HAC-robust procedure to allow for arbitrary forms of heteroskedasticity and serial correlation.

C Additional Results on the Minimum–Discrepancy SDF

This appendix provides two complementary perspectives on the exponential tilting SDF that are useful for interpretation but not required for the main analysis.

C.1 Interpretation of the Lagrange Multipliers

The dual parameters (γ, α) that characterize the exponential–tilting (ET) stochastic discount factor admit a natural economic interpretation. Recall that the ET SDF is given by

$$m_t(\gamma, \alpha) = \exp(\gamma' R_t^e + \alpha - 1)\,,$$

where R_t^e denotes the vector of excess returns. Because the SDF is proportional to the intertemporal marginal rate of substitution, I write $MU_{t+1} \propto m_{t+1}$. Differentiating with respect to the ith excess return yields

$$\frac{\partial MU_{t+1}}{\partial R^e_{i,t+1}} \propto \frac{\partial m_{t+1}}{\partial R^e_{i,t+1}} = \gamma_i\, m_{t+1},$$

which implies that

$$\frac{\partial MU_{t+1}}{\partial R^e_{i,t+1}} \cdot \frac{R^e_{i,t+1}}{MU_{t+1}} = \gamma_i\, R^e_{i,t+1}.$$

Hence γ_i measures the elasticity of marginal utility with respect to $R^e_{i,t+1}$, scaled by the inverse of the return. Assets that contribute more strongly to marginal utility per unit of payoff carry larger multipliers in absolute value. The sign of the multiplier is typically negative for assets that are held long in equilibrium, reflecting the fact that marginal utility declines in good states.

To anchor this interpretation, consider a one–period portfolio problem with date–1 consumption

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

where $\sum_j \pi_j = 1$ and $R_{j,1}$ are gross returns. For constant absolute risk aversion (CARA) preferences with $U(c_1) = -\exp(-\alpha c_1)$, marginal utility is $MU_1 = \alpha e^{-\alpha c_1}$ and $dc_1/dR_{i,1} = (W_0 - c_0)\pi_i$. It follows that

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

so that the implied elasticity satisfies

$$\frac{\varepsilon_{i,1}}{R_{i,1}} = -\alpha(W_0 - c_0)\pi_i = \gamma_i.$$

Thus, under CARA, γ_i is proportional to the signed dollar position in asset i and to risk aversion. For constant relative risk aversion (CRRA) preferences with $U(c_1) = c_1^{1-\rho}/(1-\rho)$, marginal utility is $MU_1 = c_1^{-\rho}$ and

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

The corresponding elasticity is

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

In both cases, the magnitude of γ_i increases with risk aversion and with the size of the position in the asset, confirming that the multipliers reflect the economic significance of each return for investors' marginal utility. This provides a direct portfolio interpretation of the dual parameters and explains the negative signs often observed empirically for assets that are held long.

C.2 Relation to Good–Deal Bounds

The exponential–tilting SDF is also related to the good–deal pricing bounds of Cochrane and Saa-Requejo (2000). Consider the payoff x^n of an asset not used in constructing the SDF and let $A > 0$ be an upper bound on the second moment of admissible SDFs. Good–deal bounds define lower and upper prices by

$$\underline{p} = \inf_m \mathbb{E}[m\, x^n], \qquad \overline{p} = \sup_m \mathbb{E}[m\, x^n],$$

subject to

$$\mathbb{E}[m] = \frac{1}{R_f}, \quad \mathbb{E}[mR^e] = \mathbf{0}, \quad m \geq 0, \quad \mathbb{E}[m^2] \leq A,$$

where R^e collects the base–asset excess returns. The ET SDF by construction satisfies the mean and Euler constraints and is strictly positive. Its feasibility under the good–deal criterion depends only on whether A is chosen greater than or equal to $\mathbb{E}[m_t^2(\hat{\gamma}, \hat{\alpha})]$.

To see this, augment the primal problem with the second–moment bound. The first–order condition for the constrained optimization takes the form

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

where $\omega \geq 0$ is the multiplier on the variance constraint and χ_t collects constants and terms in R_t^e. For $\omega > 0$, feasibility requires χ_t to lie below a finite upper bound determined by $1/\omega$. If A is sufficiently large so that the bound is slack, $\omega = 0$ and the solution reduces to the exponential tilting kernel. If A is smaller than the empirical second moment of the ET SDF, no feasible m_t exists that prices the assets and respects the bound simultaneously. Thus, for economically meaningful values of A, the ET kernel is feasible and its implied price $\mathbb{E}[m_t x^n]$ lies between the good–deal bounds.

This connection highlights that the ET SDF respects admissibility conditions in the sense of good–deal bounds, while still delivering exact pricing for the assets used in its construction. The discrepancy–minimization approach therefore yields an SDF that is both empirically tractable and economically reasonable, producing prices for new assets that remain consistent with no–arbitrage and with economically plausible Sharpe–ratio limits.