What is Missing in Asset-Pricing Factor Models?*

Massimo Dello Preite Raman Uppal Paolo Zaffaroni Irina Zviadadze

December 8, 2022

Abstract

Our objective is to price the cross section of asset returns. Despite considering hundreds of systematic risk factors ("factor zoo"), factor models still have a sizable pricing error. A limitation of these models is that returns compensate only for systematic risk. We allow *compensation also for unsystematic risk*. The resulting stochastic discount factor (SDF) prices the cross section of stock returns *exactly*, resolving the factor zoo. Empirically, more than half the variation of this SDF is explained by the unsystematic risk component, which is correlated with strategies reflecting market frictions and behavioral biases.

JEL classification: C52, C58, G11, G12.

Keywords: Unsystematic risk, factor models, factor zoo, model misspecification.

*Dello Preite is affiliated with Imperial College London; email: `m.dello-preite18@imperial.ac.uk`. Uppal is affiliated with EDHEC Business School and CEPR; email: `Raman.Uppal@edhec.edu`. Zaffaroni is affiliated with Imperial College London; email: `P.Zaffaroni@imperial.ac.uk`. Irina Zviadadze is affiliated with HEC Paris and CEPR; email: `zviadadze@hec.fr`. We gratefully acknowledge comments from Torben Andersen, Harjoat Bhamra, Svetlana Bryzgalova (discussant), John Campbell, Ines Chaieb, Pierre Collin-Dufresne, Magnus Dahlquist, Victor DeMiguel, René Garcia (discussant), Lars Hansen, Valentin Haddad (discussant), Ravi Jagannathan, Christian Julliard (discussant), Robert Korajczyk, Stefan Nagel, Markus Pelger, Valentina Raponi, Cesare Robotti, Simon Rottke (discussant), Olivier Scaillet, Alberto Teguia (discussant), Viktor Todorov, Fabio Trojani (discussant), Dacheng Xiu, and seminar and conference participants at Bocconi, EDHEC Business School, ESSEC, HEC Paris, HEC Montreal, Imperial College Business School, INSEAD, Kellogg School of Management, London Business School, Lund University, Stockholm Business School, Stockholm School of Economics, University of Chicago, University of Geneva, University of Lausanne and EPFL, 2nd Bristol Financial Markets Conference, 3rd Corporate Policies and Asset Prices (COAP) Conference at Cass Business School, 6th SAFE Asset Pricing Workshop, 12th Annual SoFiE Conference, 13th International Conference on Computational and Financial Econometrics, 14th Annual Meeting of Society for Financial Econometrics (SoFiE), 16th Annual Conference in Financial Economics Research at Arison School of Business, American Finance Association conference, European Finance Association conference, Financial Econometrics Conference at Toulouse School of Economics, Junior European Finance Seminar, Midwestern Finance Association Annual Meeting, SFS Cavalcade North America, and University of British Columbia Finance Summer Conference. An earlier version of this manuscript was circulated under the title "Correcting Misspecified Stochastic Discount Factors."

1 Introduction

A major challenge in asset pricing is to explain the cross section of asset returns. To address this challenge, the literature has examined a large number of systematic or common risk factors, leading to a factor zoo (Cochrane, 2011). However, virtually all models featuring factors from this zoo have sizable pricing errors, called alpha. A limitation of these models is that they allow expected returns to be related only to *systematic* sources of risk and preclude compensation for unsystematic risk, that is, asset-return shocks orthogonal to common risk factors. In our work, we allow *compensation for unsystematic risk*. This insight leads to exact pricing of the cross section of stock returns, resolving the factor zoo.

To price the cross-section of assets, we use as a foundation for our analysis the APT of Ross (1976, 1977). The APT provides an ideal framework because it allows expected returns to contain asset-specific components that are unrelated to common risk factors and satisfy a no-arbitrage restriction.

Our first contribution is to derive an *admissible* SDF, namely an SDF that prices a given cross section of assets correctly, assuming asset returns are described by the APT. In particular, we show theoretically how unsystematic risk enters into this admissible SDF, which leads to our key insight that the asset-specific components of expected returns are not pricing errors but compensation for unsystematic risk. Thus, we depart from the conventional wisdom that financial markets compensate investors only for exposure to systematic sources of risk.[1] To provide microfoundations for the idea that unsystematic risk can indeed be a component of the SDF, we demonstrate that the SDF implied by an equilibrium model such as Merton (1987), where investors are aware of only a subset of available securities, is one where unsystematic risk is priced even when the number of assets is large. Furthermore, we show that equilibrium asset returns and the SDF in Merton (1987) coincide with the ones implied by the APT.

Our second contribution is to provide empirical support for our insight that unsystematic risk is priced and to quantify its importance. To do this, we estimate the admissible SDF implied by the APT model for asset returns. Using data for monthly returns on 202

[1]The standard view in finance that unsystematic risk should be diversified away holds only because the return for bearing unsystematic risk is assumed to be zero in popular factor models. When unsystematic risk earns a nonzero reward then, instead of diversifying this risk, an investor will optimally adjust her portfolio to reap this compensation, as shown in Raponi, Uppal, and Zaffaroni (2022). As a result, the SDF will consist of not just a component based on systematic sources of risk but also a component based on unsystematic risk.

portfolios of stocks, we identify and characterize the two components of the SDF, one which reflects systematic risk (SDF-S) and the other unsystematic risk (SDF-U). A central finding from our analysis is that the asset-specific components in expected returns, which represent compensation for unsystematic risk, are nonzero in the data, and the SDF-U component explains more than half of the admissible SDF's variation. Thus, unsystematic risk plays a major role in pricing the cross-section of asset returns, despite the risk premia associated with each unsystematic shock being small on average.

We explore the nature of the SDF-U component by examining the composition of the portfolio that represents it. We find that small stocks contribute substantially to SDF-U, followed by stocks with extremely low or high values of net issuances. We also regress the SDF-U component on 457 trading strategies examined in the literature and find that strategies related to behavioral biases and financial frictions exhibit the largest correlations, in excess of thirty percent. The most prominent strategies are those based on 5-year analyst growth forecast (La Porta, 1996), betting against beta (Frazzini and Pedersen, 2014), long-term behavioral mispricing (Daniel, Hirshleifer, and Sun, 2020), ratio of book debt to market equity (Bhandari, 1988), and implied equity duration (Dechow, Sloan, and Soliman, 2004).

These high correlations imply that the expected excess returns on these trading strategies are large and reflect sizable compensation for bearing unsystematic risk. We find, for instance, that the risk premium for unsystematic risk associated with the strategy of La Porta (1996) is 8.20% per annum. There are 27 other strategies with an absolute value of risk premium earned for exposure to unsystematic risk greater than 5% per annum. Thus, one can use these strategies to construct a portfolio with high expected excess returns but zero exposure to systematic risk.

Turning next to the analysis of the systematic component of the SDF, we find that the market factor explains 95% of its variation. At first glance, this finding may seem surprising in light of the extensive literature documenting the poor performance of the market factor in explaining the cross-sectional differences in stock returns.[2] However, even though the market factor does not explain the cross-section of expected stock returns, we find, similar to Clarke (2020), among others, that it plays an important role in determining the *level* of stock returns. The systematic component of the SDF accounts for only 44% of the variation in the admissible SDF, so the overall contribution of the market factor to the total variation in the admissible SDF is only 42% (= 95% $\times$ 44%). We also find that the cross-sectional

[2]See Fama and French (2004) for a review of this literature.

differences in expected stock returns, in addition to the SDF-U component, are spanned by their exposures to nineteen traded factors. Among these traded factors, the size factor of Fama and French (1993) is the most prominent, explaining 89% of the residual systematic variation in the SDF after accounting for the market factor.

Our third contribution is to use our framework to shed light on the poor performance of popular candidate factor models used to price a cross section of stock returns. We consider: (i) a model with the market factor, as suggested by the capital asset pricing model (CAPM) of Sharpe (1964), (ii) a model with the consumption-mimicking portfolio, as implied by the consumption capital asset pricing model (C-CAPM) of Breeden (1979), and (iii) the three-factor model (FF3) of Fama and French (1993). These candidate factor models may be misspecified because they omit systematic sources of risk, which have been the focus of the existing literature. But, the candidate models may be misspecified also because they omit asset-specific components in expected returns. We identify and characterize the required correction term for each of these models to obtain an admissible SDF.

The main insight from our analysis of these three candidate models is that their implied SDFs represent less than 50% of the variation in the admissible SDF. The principal source of missing variation is unsystematic risk, which, in these models, has zero compensation. Moreover, once we use our approach to include what is missing in each of these three candidate models, we obtain admissible SDFs that are almost perfectly correlated. More generally, the quantitative importance of the SDF-U component suggests that candidate factor models with different proxies for only common risk factors will not lead to an admissible SDF. Instead, to obtain an admissible SDF one must recognize that unsystematic risk is priced, and we show how to do this.[3] This insight resolves the factor zoo.

When traditional asset pricing models, such as the CAPM, fail to explain a cross-section of stock returns, the response has typically been to search for additional systematic factors. For instance, momentum (Jegadeesh and Titman, 1993), value (Fama and French, 2015), and investment (Hou, Xue, and Zhang, 2015) have attracted attention as successful explanatory factors. We find that such factors are successful, at least partly, because they correlate more highly with the aggregate measure of unsystematic risk than with the systematic component of the SDF. These factors appear to be weak (Lettau and Pelger, 2020;

[3]Our findings emphasize the arguments of MacKinlay (1995) and Daniel and Titman (1997) about the importance of characteristics for understanding risk premia and the inability of a factor model to explain a cross section of stock returns, but with two crucial differences. First, our model ensures asymptotic no-arbitrage. Second, we demonstrate that, in our framework, the asset-specific components in expected returns represent compensation for unsystematic risk.

Giglio, Xiu, and Zhang, 2021); that is, they affect only a small number of asset returns from the cross-section of stock returns considered.

The fourth contribution of our work is methodological. The implicit accounting for unsystematic risk in factor models via weak factors suggests that perhaps adding an arbitrary combination of factors to a misspecified candidate factor model can capture the SDF-U component. However, we show formally that this is not the case. The SDF-U component is a weak factor in the cross-section of asset returns. Estimating reliably the risk premium associated with a weak factor is not feasible statistically. Therefore, one cannot recover an admissible SDF simply by using the traditional two-pass regression. Our approach, by explicitly accounting for compensation for unsystematic risk, allows us to recover an admissible SDF.

Our work is related to several strands of the literature. First, because we correct a given candidate factor model through the lens of a misspecified SDF, we contribute to the literature that studies misspecification of the SDF and develops methods to estimate the minimum-variance SDF, that is, the projection of the SDF on asset returns. The idea of misspecification of the SDF motivates the work of Hansen and Jagannathan (1991), who provide the minimum-variance bound that any admissible SDF must satisfy. Luttmer (1996) extends their analysis to economies with proportional transaction costs, short-sale constraints, and margin requirements. Korsaye, Quaini, and Trojani (2021) advance this literature substantially by allowing for more general convex pricing constraints, which then allows them to nest in a single unifying framework several asset-pricing approaches not covered by the SDF literature. In contrast to these papers, our objective is not to identify a bound on the SDF; instead, we provide the exact correction required for a proposed SDF to become admissible, and we highlight the role of unsystematic risk in this correction.

Several papers have developed a nonparametric approach to correct misspecified SDF models. Hansen and Jagannathan (1997) provide the smallest additive nonparametric adjustment (in a least-squares sense) required to make a given SDF admissible. Sandulescu and Schneider (2021) build on Almeida and Schneider (2021) to construct an SDF that is a sum of a linear part which is identical to that from Hansen and Jagannathan (1997) and a nonlinear part that ensures the positivity of the SDF and leads to an admissible SDF. Almeida and Garcia (2012) provide an additive correction term based on minimum-discrepancy projections. Ghosh, Julliard, and Taylor (2017) provide a multiplicative correction using a Kullback-Leibler entropy-minimization approach. In contrast, we ensure positivity of the

SDF by specifying it in an exponential form. More importantly, we focus on highlighting the compensation in expected returns for unsystematic risk.

To get as close as possible to the true SDF, ideally, one would like to estimate a projection of the SDF on a large number of assets. However, it is challenging to use these nonparametric approaches when the number of basis assets is large relative to the number of observations. To handle a large number of assets, Kozak, Nagel, and Santosh (2020), Lettau and Pelger (2020), and Giglio and Xiu (2021) develop methods based on Principal Component Analysis (PCA) for estimating the SDF, identifying factors that price the cross-section of expected returns, and estimating risk premia in the presence of model misspecification, respectively. However, these papers focus on systematic sources of risk and do not study the compensation for unsystematic risk. Our approach, being founded on the APT, handles a large number of assets and allows explicitly for the compensation for unsystematic risk.

Our work is also related to the literature on the idiosyncratic-volatility puzzle, which studies the relation between the compensation for asset-specific risk and the *volatility* of the asset-specific shock; see, for example, Fama and MacBeth (1973) and Ang, Hodrick, Xing, and Zhang (2006), with a comprehensive review provided by Bali, Engle, and Murray (2016). In contrast to this literature, we construct an SDF so that the compensation for unsystematic risk represents the negative covariance between this SDF and unsystematic shocks (rather than their *volatility*). Moreover, if one relied on the insights of the idiosyncratic volatility literature, one would expect an idiosyncratic-volatility factor to span the SDF-U component. However, we find that the idiosyncratic-volatility factors of Chen and Zimmermann (2021), Jensen, Kelly, and Pedersen (2021), and Kozak, Nagel, and Santosh (2020) account for at most 20% of the variation (in the R^2 sense) of our SDF-U component. Therefore, what is missing from factor models is not an idiosyncratic-volatility factor.

The rest of the paper is organized as follows. Section 2 presents our theoretical results for constructing an admissible SDF. Section 3 provides details of how to estimate an admissible SDF. Section 4 describes the data we use to illustrate our approach. Section 5 presents the empirical findings from applying our approach to this data. Section 6 provides an example of an equilibrium model in which unsystematic risk is priced. We conclude in Section 7. Proofs are collected in an appendix, with additional results in the Internet Appendix.

2 From a Candidate to an Admissible SDF

Our work is founded on the classical APT of (Ross, 1976), with Chamberlain (1983) and Chamberlain and Rothschild (1983) providing a more formal treatment. Effectively, the APT is our working assumption about the true data-generating process for asset returns. There are several advantages of using the APT. First, it is a flexible model that does not take a stand on what constitutes a pricing factor. Second, it is a no-arbitrage model; the absence of arbitrage opportunities implies the existence of an SDF. Third, and more importantly for us, it allows for asset-specific components in expected returns that are unrelated to systematic (common) risk.

In this section, we first review the classical APT. Next, we derive the closed-form expression for an SDF under the APT, thereby complementing Chamberlain (1983), who shows existence and continuity of the "cost functional" (i.e., the SDF) under the classical APT, without providing its closed-form representation. We then explain how to correct a misspecified SDF implied by an arbitrary factor model of asset returns. Finally, we address three empirical challenges that we face when estimating the admissible SDF: (i) nonnegativity of the SDF, (ii) econometric feasibility of the SDF, and (iii) weak factors (e.g., Lettau and Pelger, 2020) in the candidate factor model.

2.1 The SDF under the Arbitrage Pricing Theory (APT)

Let the N-dimensional vector $R_{t+1} = (R_{1,t+1}, R_{2,t+1}, \ldots, R_{N,t+1})'$ denotes the vector of gross returns of the N risky assets between t and $t+1$. Let R_{ft} be the gross return on the risk-free asset over the same time period.[4] Let f_{t+1} be the $K \times 1$ vector of common risk factors, with $K < N$ and a $K \times K$ positive definite covariance matrix $V_f = \mathrm{var}(f_{t+1}) > 0$.

The classical APT builds on the following two assumptions.

Assumption 1 (Linear Factor Model)**.** *The vector R_{t+1} of gross asset returns satisfies*

$$R_{t+1} = \mathbb{E}(R_{t+1}) + \beta\big(f_{t+1} - \mathbb{E}(f_{t+1})\big) + e_{t+1},$$

where $\mathbb{E}$ denotes an operator of mathematical expectation, $\beta = (\beta_1, \beta_2, \ldots, \beta_N)'$ is the $N \times K$ full-rank matrix of loadings of asset returns on the common factors, the vector

[4] If a risk-free asset does not exist, one can use instead the return on the minimum-variance portfolio or the return on the zero-beta portfolio.

of asset-specific errors e_{t+1} has zero mean and the $N \times N$ positive-definite covariance matrix $V_e = \text{var}(e_{t+1}) > 0$ with uniformly bounded eigenvalues.[5] *The asset-specific shocks e_{t+1} are uncorrelated with the K common factors f_{t+1}, implying that the covariance matrix of returns is $V_R = \text{var}(R_{t+1}) = \beta V_f \beta' + V_e$.*

Assumption 2 (Asymptotic No Arbitrage)**.** *There are no arbitrage opportunities for a sufficiently large number of assets N; that is, there is no sequence of portfolios containing a large number of risky assets with the weights $w = (w_1, w_2, \ldots, w_N)'$, for which:*

$$\text{var}(R'_{t+1}w) \to 0 \qquad \textit{and} \qquad (\mathbb{E}(R_{t+1}) - R_{ft}1_N)'w \geq \delta > 0 \quad \textit{as} \quad N \to \infty,$$

where δ denotes an arbitrary positive scalar and 1_N denotes the $N \times 1$ vector of ones.

Assumptions 1 and 2 imply that a model of asset excess returns is

$$R_{t+1} - R_{ft}1_N = a + \beta\lambda + \beta(f_{t+1} - \mathbb{E}(f_{t+1})) + e_{t+1}, \tag{1}$$

where the expected *excess* returns $\mathbb{E}(R_{t+1} - R_{ft}1_N) = a + \beta\lambda$ contain two components: a and $\beta\lambda$. The $K \times 1$ vector of risk premia λ represents the compensations for one unit of asset exposures to the factors f_{t+1}. Ingersoll (1984) derives the precise condition for λ to exist and shows that $\lambda = \lim_{N\to\infty} \left(\beta' V_e^{-1}\beta\right)^{-1} \beta' V_e^{-1}(\mathbb{E}(R_{t+1}) - R_f 1_N)$. The $N \times 1$ vector $a = (\mathbb{E}(R_{t+1}) - R_{ft}1_N) - \beta\lambda$, which is typically referred to as the vector of pricing errors, satisfies the following no-arbitrage restriction

$$a' V_e^{-1} a \leq \delta^*_{\text{apt}} < \infty, \tag{2}$$

as shown in Ross (1976), Huberman (1982), Chamberlain and Rothschild (1983), and Ingersoll (1984), where δ^*_{apt} is some arbitrary positive scalar.

Proposition 1. *An SDF M_{t+1} implied by the APT model for asset returns is*

$$M_{t+1} = M^{\beta}_{t+1} + M^{a}_{t+1},$$

where,

$$M^{\beta}_{t+1} = -\frac{\lambda' V_f^{-1}}{R_f}(f_{t+1} - \mathbb{E}(f_{t+1}))$$

and

$$M^{a}_{t+1} = -\frac{a' V_e^{-1}}{R_f} e_{t+1},$$

with $\text{cov}(M^{\beta}_{t+1}, M^{a}_{t+1}) = 0$.

[5] In the APT of Chamberlain and Rothschild (1983), the covariance matrix V_e is not restricted to be diagonal. The case of non-diagonal matrix V_e corresponds to the presence of weak factors, f^{weak}_{t+1}, in the shocks e_{t+1}.

The SDF component M^a_{t+1} is a linear function of asset-specific shocks e_{t+1} scaled by the risk-free rate. The presence of the SDF-U component in the admissible SDF leads to the main insight of our approach, which is the interpretation of the vector a. When viewed through the lens of an SDF, the following expression

$$a = -\operatorname{cov}(M_{t+1}, e_{t+1}) \times R_f = -\operatorname{cov}(M^a_{t+1}, e_{t+1}) \times R_f,$$

shows that the vector a should be interpreted as compensation for asset-specific risk e_{t+1}, rather than pricing errors. This interpretation paves the way for a quantitative assessment of asset-specific risk in financial markets that we will undertake in our empirical analysis.

2.2 Correcting Misspecified SDF Models

Relative to the APT, any standard candidate factor model with K^{can} observable risk factors f^{can}_{t+1} suffers from potentially two sources of misspecification. First, the candidate model may omit systematic risk factors, that is, $K^{\text{can}} < K$. Second, as is often the case, the candidate model may not allow for asset-specific components of expected excess returns, represented by the vector a in equation (1); however, in the true model, some components of this vector a may be nonzero. A popular example of a candidate model is the market model, in which the vector f^{can}_{t+1} includes only the market factor, $K^{\text{can}} = 1$, and $a^{\text{can}} = 0_N$.

To understand the implications of model misspecification, consider a candidate model with both sources of potential misspecification, that is, $a^{\text{can}} = 0_N$ and $K^{\text{can}} < K$. Let β^{can} denote the $N \times K^{\text{can}}$ matrix of loadings of asset returns on the candidate factors and λ^{can} the $K^{\text{can}} \times 1$ vector of risk premia for unit exposures to these factors. The candidate factor model implies

$$R_{t+1} - R_{ft}1_N = \alpha + \beta^{\text{can}}\lambda^{\text{can}} + \beta^{\text{can}}(f^{\text{can}}_{t+1} - \mathbb{E}[f^{\text{can}}_{t+1}]) + \varepsilon_{t+1}, \tag{3}$$

where $\alpha = (\mathbb{E}(R_{t+1}) - R_{ft}1_N) - \beta^{\text{can}}\lambda^{\text{can}}$ captures the residual variation in the expected excess returns left unexplained by compensation for asset exposures to common risk factors, and ε_{t+1} with covariance matrix V_ε captures the residual variation in asset returns that is not explained by the set of candidate factors f^{can}_{t+1}.

The proposition below shows that, just as the vector a in the classical APT satisfies the no-arbitrage restriction given in expression (2), the vector of pricing errors α in the candidate model satisfies a similar no-arbitrage restriction even if the candidate model omits some systematic risk factors.

Proposition 2 (APT in the presence of model misspecification)**.** *Suppose that the vector of asset returns R_{t+1} satisfies Assumptions 1 and 2. Given a candidate factor model with K^{can} factors, suppose the first $K^{\text{mis}} = K - K^{\text{can}}$ eigenvalues of the covariance matrix V_ε are unbounded when $N \to \infty$, the remaining eigenvalues are uniformly bounded, and the smallest eigenvalue is strictly positive. Then, the pricing error α in the misspecified candidate model satisfies the following no-arbitrage restriction, for some constant $\tilde{\delta}_{apt}$ possibly different from δ^*_{apt},*

$$\alpha' V_\varepsilon{}^{-1} \alpha \leq \tilde{\delta}_{\text{apt}}, \tag{4}$$

where, by no arbitrage, there exist an $N \times 1$ vector a, $K^{\text{mis}} \times 1$ vector λ^{mis}, and an $N \times K^{\text{mis}}$ matrix β^{mis} such that

$$\alpha = \beta^{\text{mis}} \lambda^{\text{mis}} + a \qquad \textit{and} \qquad V_\varepsilon = \text{var}(\varepsilon_{t+1}) = \beta^{\text{mis}} \beta^{\text{mis}\prime} + V_e. \tag{5}$$

We see from (5) that α is the sum of the vector a and the compensation for the missing systematic risk, $\beta^{\text{mis}}\lambda^{\text{mis}}$, with β^{mis} being the matrix of loadings of asset returns on the missing systematic risk factors and λ^{mis} being the vector representing the prices of the missing systematic risk factors. The covariance matrix V_ε is the variance of asset returns because of their exposure to the systematic risk factors f^{mis}_{t+1} that are missing in the candidate model and the asset-specific shocks e_{t+1}.

Without loss of generality, given that f^{mis}_{t+1} are latent factors, we can rotate them freely and normalize them in an arbitrary way, and so we assume that the factors f^{mis}_{t+1} are mutually orthogonal to the factors f^{can}_{t+1}, and f^{mis}_{t+1} has a $K^{\text{mis}} \times K^{\text{mis}}$ identity covariance matrix $V_{f^{\text{mis}}} = \text{var}(f^{\text{mis}}_{t+1}) = I_{K^{\text{mis}}}$. Therefore, the covariance matrix of asset returns can be represented as $V_R = \text{var}(R_{t+1}) = \beta^{\text{can}} V_{f^{\text{can}}} \beta^{\text{can}\prime} + \beta^{\text{mis}}\beta^{\text{mis}\prime} + V_e$, where $V_{f^{\text{can}}} = \text{var}(f^{\text{can}}_{t+1})$.

Next, we establish a class of admissible SDFs, given a misspecified candidate factor model for asset returns. To this end, we identify and construct the correction terms that transform the misspecified SDF implied by the candidate model to an admissible SDF.

Recall that the candidate factor model of asset returns has two sources of misspecification relative to the true APT model given in expression (1). First, the candidate model includes $K^{\text{can}} < K$ risk factors, thereby omitting $K^{\text{mis}} = K - K^{\text{can}}$ risk factors. Second, the candidate model omits the nonzero vector a by assuming that $a^{\text{can}} = 0_N$. The admissible linear SDF is the sum of the $M^{\beta,\text{can}}_{t+1}$, the SDF implied by the candidate factor model, and a correction term labeled the alpha-SDF M^{α}_{t+1}. The correction term has two components:

$M^{\beta,\text{mis}}_{t+1}$ and M^{a}_{t+1}. The first component accounts for the omitted systematic risk factors f^{mis}_{t+1}, whereas the second component accounts for omitted asset-specific components a in expected returns.

Proposition 3 (SDF: Linear Case)**.** *Under Assumptions 1 and 2, there exists an admissible SDF M_{t+1},*

$$M_{t+1} = M^{\beta,\text{can}}_{t+1} + M^{\alpha}_{t+1} = M^{\beta,\text{can}}_{t+1} + \underbrace{(M^{a}_{t+1} + M^{\beta,\text{mis}}_{t+1})}_{=M^{\alpha}_{t+1}}, \tag{6}$$

where,

$$M^{\beta,\text{can}}_{t+1} = \frac{1}{R_f} - \frac{(\lambda^{\text{can}})' V^{-1}_{f^{\text{can}}}}{R_f}(f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1})),$$

$$M^{\beta,\text{mis}}_{t+1} = -\frac{(\lambda^{\text{mis}})' V^{-1}_{f^{\text{mis}}}}{R_f}(f^{\text{mis}}_{t+1} - \mathbb{E}(f^{\text{mis}}_{t+1})),$$

$$M^{a}_{t+1} = -\frac{a' V_e^{-1}}{R_f} e_{t+1}, \tag{7}$$

with $\text{cov}(M^{\beta,\text{can}}_{t+1}, M^{a}_{t+1}) = 0$, $\text{cov}(M^{a}_{t+1}, M^{\beta,\text{mis}}_{t+1}) = 0$, $\text{cov}(M^{\beta,\text{can}}_{t+1}, M^{\beta,\text{mis}}_{t+1}) = 0$, *and where, without loss of generality, as indicated earlier, $V_{f^{\text{mis}}} = I_{K^{\text{mis}}}$.*

The correction components M^{a}_{t+1} and $M^{\beta,\text{mis}}_{t+1}$ are latent quantities, and therefore, the issue of econometric identification arises. At the estimation stage, we resolve this identification challenge by imposing the no-arbitrage restriction given in expression (4) that gives us precisely the condition required to identify M^{a}_{t+1} and $M^{\beta,\text{mis}}_{t+1}$.

The dependence of M^{a}_{t+1} on e_{t+1} in equation (7) implies that expanding a candidate model to include an increasing number of observable variables proxying for common risk factors is not a fruitful avenue to obtain an admissible SDF. In Appendix A.2, we show that M^{a}_{t+1} is a weak factor in the cross-section of asset returns. Therefore, even if it were possible to add to a candidate factor model an observable variable that was perfectly correlated with M^{a}_{t+1}, it would not lead to an admissible SDF because the risk premia associated with a weak factor cannot be estimated accurately (Anatolyev and Mikusheva, 2021). We show below how to account for unsystematic risk and construct an accurate estimator of M^{a}_{t+1}, and hence, of an admissible SDF.

2.3 Accounting for Time-Variation in Risk Premia

In our discussion above, to identify an admissible SDF we have used the classical APT setting where the prices of risk and the asset exposures to systematic risk factors are constant.

However, one may wonder whether the vector a represents, instead of compensation for asset-specific risk, time-variation in asset factor exposures and/or prices of risk. We show in Appendix A.3 that this is not the case and that, even if the exposures and price of risk are time varying, our method allows us to construct an admissible unconditional SDF. In particular, if the omitted source of time-variation arises from a common state variable, then its effect is captured by $M_{t+1}^{\beta,\text{mis}}$ via the scaled factors of Gagliardini, Ossola, and Scaillet (2016) and Gagliardini, Ossola, and Scaillet (2019). Alternatively, if the omitted source of time-variation arises from asset-specific state variables, then its effect is captured by M_{t+1}^{a}.[6]

2.4 Constructing an Admissible SDF in Practice

There are three problems in constructing an admissible SDF in practice. First, the linear SDF characterized in Proposition 3 may not always be strictly positive, which could result in negative asset prices. Second, the components $M_{t+1}^{\beta,\text{mis}}$ and M_{t+1}^{a} depend on unobservable quantities, f_{t+1}^{mis} and e_{t+1}, respectively. Finally, a candidate factor model may omit not only strong but also weak factors. We explain below how to address these three challenges.

2.4.1 Exponential SDF

There are at least two approaches for ensuring that the SDF is positive. The first approach is to express the SDF as the payoff to an option (Hansen and Jagannathan, 1997, eq. (24)). The second approach is to specify the SDF as an *exponential* function of the payoffs (Gourieroux and Monfort, 2007; Ghosh, Julliard, and Taylor, 2017). For obtaining closed-form solutions, we assume that asset returns are Gaussian and follow the second approach.

Proposition 4 (SDF: Exponential Case)**.** *Under Assumptions 1 and 2 and the assumption that returns R_{t+1} are Gaussian, there exists an admissible SDF $M_{\exp,t+1}$*

$$M_{\exp,t+1} = M_{\exp,t+1}^{\beta,\text{can}} \times M_{\exp,t+1}^{a} \times M_{\exp,t+1}^{\beta,\text{mis}}, \quad \textit{where}$$

$$M_{\exp,t+1}^{\beta,\text{can}} = \frac{1}{R_f} \cdot \exp\Big[- \lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} (f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) - \frac{1}{2} \lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} \lambda^{\text{can}} \Big],$$

$$M_{\exp,t+1}^{\beta,\text{mis}} = \exp\Big[- \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} (f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) - \frac{1}{2} \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \Big],$$

$$M_{\exp,t+1}^{a} = \exp\Big[- a' V_e^{-1} e_{t+1} - \frac{1}{2} a' V_e^{-1} a \Big],$$

where $\text{cov}(M_{\exp,t+1}^{a}, M_{\exp,t+1}^{\beta,\text{can}}) = 0,$ $\text{cov}(M_{\exp,t+1}^{\beta,\text{mis}}, M_{\exp,t+1}^{\beta,\text{can}}) = 0,$ *and* $V_{f^{\text{mis}}} = I_{K^{\text{mis}}}$*.*

[6] If the compensation for asset-specific risk a_i itself is time-varying, our method recovers the average a_i.

2.4.2 Projection SDF

Even if the values of the parameters of the data-generating process (3) are known, the admissible SDF $M_{\exp,t+1}$ depends on the unobservable quantities f^{mis}_{t+1} and e_{t+1}, which means $M_{\exp,t+1}$ is not feasible empirically. To overcome this challenge, we rely on a projection version of the SDF, $\hat{M}_{\exp,t+1}$, with $\hat{\cdot}$ used to indicate projected quantities. In particular, we take the exponential function of the linear projections of M^a_{t+1} and $M^{\beta,\text{mis}}_{t+1}$ on the set of the risk-free and risky assets to obtain[7]

$$\hat{M}^a_{\exp,t+1} = \exp\Big[- a'V_R^{-1}(R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2}a'V_R^{-1}a\Big], \tag{8}$$

$$\hat{M}^{\beta,\text{mis}}_{\exp,t+1} = \exp\Big[- (\beta^{\text{mis}}\lambda^{\text{mis}})'V_R^{-1}(R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2}(\beta^{\text{mis}}\lambda^{\text{mis}})'V_R^{-1}\beta^{\text{mis}}\lambda^{\text{mis}}\Big], \tag{9}$$

where, from Proposition 2,

$$\mathbb{E}[R_{t+1}] - R_{ft} = a + \beta^{\text{mis}}\lambda^{\text{mis}} + \beta^{\text{can}}\lambda^{\text{can}},$$

and

$$V_R = \beta^{\text{can}}V_{f^{\text{can}}}\beta^{\text{can}\prime} + \beta^{\text{mis}\prime}\beta^{\text{mis}} + V_e.$$

Note that the component $M^{\beta,\text{can}}_{\exp,t+1}$ depends on only observable quantities, so that the projection nonnegative admissible SDF takes the form

$$\hat{M}_{\exp,t+1} = M^{\beta,\text{can}}_{\exp,t+1} \times \hat{M}^{\beta,\text{mis}}_{\exp,t+1} \times \hat{M}^a_{\exp,t+1}.$$

The next proposition shows that, as $N \to \infty$, the SDF components, $M^{\beta,\text{mis}}_{\exp,t+1}$ and $M^a_{\exp,t+1}$, and their corresponding projection versions, $\hat{M}^{\beta,\text{mis}}_{\exp,t+1}$, and $\hat{M}^a_{\exp,t+1}$, have the same limits. We denote the matrix of the loadings of returns on the candidate and missing factors by $\beta = (\beta^{\text{can}}\ , \beta^{\text{mis}})$ and an arbitrary $K \times K$ positive-definite matrix by $A > 0$. We use $\xrightarrow{p}$ to denote convergence in probability and $a = o(b)$ with $b > 0$ to denote that $|a|/b \to 0$, as $N \to \infty$, where the dependence of a and b on N is implicit; for additional details about this notation, see Appendix A.1.[8]

Proposition 5 (Limiting properties of SDF projections)**.** *Under Assumptions 1 and 2 and the conditions* $N^{-1}\beta^{\text{can}\prime}V_\varepsilon^{-1}\beta^{\text{can}} \longrightarrow D > 0$, $N^{-1}\beta^{\text{mis}\prime}V_e^{-1}\beta^{\text{mis}} \longrightarrow E > 0$, $\beta^{\text{can}\prime}V_e^{-1}a =$

[7]The formulae (8) and (9) indicate that the assumption that asset returns are Gaussian is mild in practice. By the arguments of the Central Limit Theorem, the projection version of our feasible SDF, being an exponential function of the sum of N asset returns that are non-Gaussian, is approximately log-normal as $N \to \infty$.

[8]Strictly speaking, the matrix of loadings β from the data-generating process given in expression (1) can be different from the matrix $\beta = (\beta^{\text{can}}\ , \beta^{\text{mis}})$ because missing factors are identified only up to a rotation. This difference, however, does not have any economic bearing.

$o(N^{1/2})$, *and* $\beta^{\text{mis}\prime} V_e^{-1} a = o(N^{\frac{1}{2}})$, *as* $N \to \infty$,

$$\hat{M}^{a}_{\text{exp},t+1} - M^{a}_{\text{exp},t+1} \xrightarrow{p} 0, \qquad \hat{M}^{\beta,\text{mis}}_{\text{exp},t+1} - M^{\beta,\text{mis}}_{\text{exp},t+1} \xrightarrow{p} 0,$$

$$\text{cov}(M^{\beta,\text{can}}_{\text{exp},t+1}, \hat{M}^{\beta,\text{mis}}_{\text{exp},t+1}) \to 0, \quad \text{cov}(M^{\beta,\text{can}}_{\text{exp},t+1}, \hat{M}^{a}_{\text{exp},t+1}) \to 0, \; \text{cov}(\hat{M}^{\beta,\text{mis}}_{\text{exp},t+1}, \hat{M}^{a}_{\text{exp},t+1}) \to 0.$$

The above proposition implies that in order to construct the admissible SDF we do not need to pre-estimate the missing factors and asset-specific risk that may have been omitted in the candidate factor model.

3 Estimation Details

In this section, we describe our approach for estimating the admissible SDF and the role played by the no-arbitrage restriction (4). We also discuss how to identify the number of missing factors in a candidate factor model and choose the no-arbitrage bound δ_{apt}. Finally, we provide a diagnostic tool for detecting missing factors in a candidate factor model.

3.1 Our Estimation Approach

We recover the admissible SDF in two steps. In the first step, we use a (pseudo) Gaussian maximum-likelihood estimator (MLE) subject to the no-arbitrage restriction formulated in expression (4) to estimate the model of asset returns given in expression (3), in which the candidate factor model consists of $K^{\text{can}} \geq 0$ factors. Without loss of generality, we consider candidate factor models that include tradable factors represented by either factor returns (for example, the market factor) in excess of the risk-free rate, long-minus-short strategies, or factor-mimicking portfolios. In the second step, we use the results in Propositions 4 and 5 to recover the nonnegative feasible admissible SDF. Section IA.2 of the Internet Appendix contains a more general case, in which (i) the candidate model for asset returns includes both tradable and nontradable factors and (ii) the risk factors in the candidate model are allowed to be correlated with the missing systematic risk factors.

3.1.1 The Objective Function

For a generic vector Θ that collects all the elements of the matrices β^{can}, β^{mis}, V_e, $V_{f^{\text{can}}}$, and vectors λ^{can}, λ^{mis}, and a, the (up to a constant) Gaussian joint log-likelihood of the vectors of asset returns in excess of the risk-free rate, $R_{t+1} - R_{ft}1_N$, and observable factors f^{can}_{t+1} scaled by T is

$$\log(L(\Theta)) = -\frac{1}{2}\log(|V_\varepsilon|) - \frac{1}{2}\log(|V_f^{\text{can}}|)$$
$$- \frac{1}{2T}\sum_{t=0}^{T-1} \varepsilon'_{t+1} V_\varepsilon^{-1} \varepsilon_{t+1} - \frac{1}{2T}\sum_{t=0}^{T-1} (f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1}))' V^{-1}_{f^{\text{can}}} (f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1})),$$

where $\varepsilon_{t+1} = R_{t+1} - R_{ft}1_N - a - \beta^{\text{mis}}\lambda^{\text{mis}} - \beta^{\text{can}}\lambda^{\text{can}} - \beta^{\text{can}}(f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1}))$ and $V_\varepsilon = \beta^{\text{mis}}\beta^{\text{mis}\prime} + V_e$. Because it is not possible to obtain consistent estimates of the weak factors (Lettau and Pelger, 2020, prop. 2), we assume that V_e is diagonal.

We maximize this log-likelihood function subject to the no-arbitrage restriction. We substitute the expression in (4) with

$$a' V_\varepsilon^{-1} a \leq \delta_{\text{apt}}, \tag{10}$$

which is computationally more convenient.

We use the the Karush-Kuhn-Tucker multiplier method to solve the resulting constrained optimization problem:

$$\hat{\Theta} = \underset{\Theta}{\text{argmax}}\{\log(L(\Theta)) - \kappa(a' V_\varepsilon^{-1} a - \delta_{\text{apt}})\}, \tag{11}$$

where κ is the Karush-Kuhn-Tucker multiplier on the no-arbitrage restriction. The quantities K^{mis} and δ_{apt} are obtained using a cross-validation procedure that is explained in Section 3.2 below. Appendix A2 provides the solution to the optimization problem.

3.1.2 The Role of the No-Arbitrage Restriction

The no-arbitrage restriction on the vector a serves several purposes. Economically, it rules out asymptotic arbitrage. For example, if the elements of the vector a were left unconstrained, the Hansen and Jagannathan (1997) (HJ) distance would explode, as we show in Section 5. Moreover, the no-arbitrage restriction constrains the Sharpe ratio of the so-called alpha portfolio of Raponi, Uppal, and Zaffaroni (2022).[9] In the same spirit, Kozak, Nagel, and Santosh (2020) rule out near-arbitrage opportunities by restricting the maximum squared Sharpe ratio implied by the entire SDF.

[9] This alpha portfolio is an inefficient portfolio that, when combined with a portfolio invested in the candidate factors (the so-called beta portfolio), delivers a portfolio on the efficient frontier.

Statistically, the no-arbitrage restriction (when binding) leads to identification of the vectors a and λ^{mis}. Specifically, at the estimation stage, the no-arbitrage restriction provides N conditions that allow us to split the estimate of α into the estimates of a and $\beta^{\text{mis}}\lambda^{\text{mis}}$. Identification of a and λ^{mis} is a necessary step for constructing the missing systematic and asset-specific components of the admissible SDF, $M_{t+1}^{\beta,\text{mis}}$ and M_{t+1}^{a}, respectively. Even in population, the no-arbitrage restriction binds and is further influenced by the presence of financial frictions (Korsaye, Quaini, and Trojani, 2019, sec. 2).

Finally, the estimator of a under the no-arbitrage restriction has the form of a ridge estimator, as can be seen from Proposition A2. The ridge estimator has the appealing property of mitigating estimation noise that, in our case, affects estimates of asset-specific risk premia. The estimation noise can be significant because the vector a is an N-dimensional object representing a component of expected returns.

3.2 Identifying the Number of Missing Systematic Risk Factors and δ_{apt}

Given that a candidate factor model for asset returns may feature $0 \leq K^{\text{can}} < K$ risk factors, we need to determine the number K^{mis} of missing systematic risk factors f_{t+1}^{mis}. We estimate K^{mis} and the bound δ_{apt} on the no-arbitrage restriction (10), employing cross-validation and using as a selection metric the HJ distance. The choice of the HJ distance is natural, given our objective of identifying what is missing in asset-pricing factor models.

Our cross-validation procedure uses 20 folds. We split the entire sample into 20 folds and estimate the model on all but one fold, which is used for validation. We repeat this procedure 20 times and compute the HJ distance on the validation folds. We fix a grid of δ_{apt} from 0 to 0.1 that corresponds to Sharpe ratios ranging from 0 to 0.32 per month for the portfolio associated with purely asset-specific risk.[10] We vary the number of systematic factors missing in the candidate model, K^{mis}, from 0 to 10. We pick K^{mis} and the value of δ_{apt} that deliver the smallest HJ distance in the validation step. Our procedure never selects the boundary values of K^{mis} and δ_{apt}. Finally, using the optimal K^{mis} and δ_{apt}, we reestimate the model on the entire sample.

In the literature, other methods have been used for selecting the number of systematic risk factors in SDF models. For example, Giglio and Xiu (2021) use an information criterion similar to Bai and Ng (2002). Lettau and Pelger (2020) and Kozak, Nagel, and Santosh

[10] Ross (1977) suggests using a bound that is a multiple of the Sharpe ratio for the market portfolio, which is about 0.4 per annum.

(2020) use economic restrictions relating expected returns to the covariance of returns with factors, in addition to time-series information on the variation in asset returns.[11] Because none of these approaches directly applies to a model with asset-specific components in expected returns, we face a choice: either to use a two-stage estimation that pins down K^{mis} in the first step and δ_{apt} in the second step or to design our own method. We choose the latter and optimize the objective function that explicitly incorporates a no-arbitrage restriction while selecting K^{mis} and δ_{apt} simultaneously to minimize the HJ-distance.

3.3 Detecting Missing Factors

Propositions 4 and 5 imply that, as $N \to \infty$, the estimated SDF component $\log(\hat{M}^{\beta,\text{mis}}_{\text{exp},t+1})$ converges to a linear function of the missing factors. The following proposition shows that a simple time-series regression applied to this component reveals if a set of observable variables g_t explains the variation in asset returns that is left unexplained by a candidate factor model and, if so, delivers the prices of risk associated with these missing factors.

Proposition 6 (Detecting missing factors)**.** *Consider the regression of* $\log(\hat{M}^{\beta,\text{mis}}_{\text{exp},t})$ *on an intercept and the vector* g_t*,*

$$\log(\hat{M}^{\beta,\text{mis}}_{\text{exp},t}) = \gamma_0 + \gamma_1' g_t + u_t.$$

Under the assumptions of Proposition 5 and if $g_t = Q f_t^{\text{mis}}$ *for some nonsingular* Q*, as* $N \to \infty$ *we have*

$$\hat{\gamma}_1 \xrightarrow{p} -(Q')^{-1} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \quad \textit{and} \quad R_g^2 \xrightarrow{p} 1.$$

On the other hand, if g_t *is orthogonal to* f_t^{mis} *then*

$$\hat{\gamma}_1 \xrightarrow{p} 0_{K^{\text{mis}}} \quad \textit{and} \quad R_g^2 \xrightarrow{p} 0.$$

Proposition 6 does not require large T but only large N; that is, $R_g^2 \xrightarrow{p} 1$ as long as T exceeds the number of factors in the vector g_t. Our approach for identifying factors that are correlated with the missing component of the candidate SDF is robust with respect to weak factors and also to factors with asset exposures that are highly correlated. The latter

[11] Even though our objective function is similar to that of Lettau and Pelger (2020), there are several important differences in the two approaches. First, our goal is not to compress α as much as possible, but rather to ensure that the no-arbitrage restriction holds. From the perspective of the corrected model, α is not a pricing error, and therefore does not need to be the null vector. Second, our objective function, does not explicitly include a pricing metric measuring goodness of fit. If we were to include such a pricing metric in the objective function, we would have to augment our log-likelihood function with an additional penalty term represented by the HJ distance.

would lead to multicollinearity in the second pass of the traditional two-pass cross-sectional regression method. Our method does not suffer from this limitation because we do not need to estimate the exposures of basis assets to these factors but only to verify that these factors have explanatory power for the SDF.

4 Data

In this section, we describe the data that we use for our empirical analysis. In the first subsection, we describe the set of basis assets that we use to estimate the SDF. In the second subsection, we describe the set of factors that could potentially be related to the estimated SDF and its components.

4.1 Basis Assets

We construct a projection of the SDF on a large set of standard characteristics-based portfolios of U.S. stocks. As in Giglio and Xiu (2017), we use monthly return data for 202 portfolios from July 1963 to August 2019 from Kenneth French's website. The data includes returns on 25 portfolios sorted by size and book-to-market ratio (ME & BM), 17 industry portfolios (Ind), 25 portfolios sorted by operating profitability and investment (OP & INV), 25 portfolios sorted by size and variance (ME & VAR), 35 portfolios sorted by size and net issuance (ME & NetISS), 25 portfolios sorted by size and accruals (ME & ACCR), 25 portfolios sorted by size and beta (ME & BETA), and 25 portfolios sorted by size and momentum (ME & MOM). Instead of individual assets, we use portfolios because they exhibit a more stable factor structure (Lettau and Pelger, 2020; Giglio and Xiu, 2021).

4.2 Factors Potentially Spanning the SDF

To examine which economic variables may explain the variation in the SDF, we collect a comprehensive set of variables available at a monthly frequency that include both traded and non-traded factors (e.g., macroeconomic variables, and uncertainty indices). We briefly describe these factors below, with details about the data sources and construction of these variables provided in Section IA.3 of the Internet Appendix.

Our set of 457 traded factors includes returns on the trading strategies analyzed in Kozak, Nagel, and Santosh (2020), Bryzgalova, Huang, and Julliard (2020), Chen and

Zimmermann (2021), and Jensen, Kelly, and Pedersen (2021). We also include factors from Novy-Marx (2013) and Hou, Mo, Xue, and Zhang (2021), and the momentum up-minus-down (UMD) factor from the AQR data library.

Our set of non-traded factors includes 103 variables. We include the factors considered in Bryzgalova, Huang, and Julliard (2020). We augment these factors with the first three principal components of 279 macro variables from Jurado, Ludvigson, and Ng (2015) and the first eight principal components of 128 macro variables from the FRED-MD dataset of McCracken and Ng (2015). We include consumption growth and inflation constructed from real per capita consumption data on nondurables and services and the corresponding price index from the Bureau of Economic Analysis (BEA). We also include the market-dislocation index (Pasquariello, 2014), the disagreement index (Huang, Li, and Wang, 2021), the Chicago Board Options Exchange (CBOE) volatility index (VIX), the U.S. economic-policy-uncertainty (EPU) index (Baker, Bloom, and Davis, 2016), the equity-market-volatility (EMV) tracker (Baker, Bloom, Davis, and Kost, 2019), the credit-spread index (Gilchrist and Zakrajšek, 2012), the Chicago Fed National Financial Condition Index from FRED, the consumer-sentiment index, the U.S. business-confidence index, the U.S. consumer-confidence index, the U.S. composite-leading indicator, the coincident-economic-activity index, the NBER recession index, the TED spread, the effective federal-funds rate, and the real federal-funds rate. For variables that are persistent, we include their levels and first-order differences, as well as, where appropriate, the AR(1) or VAR(1) innovations.

5 Empirical Analysis

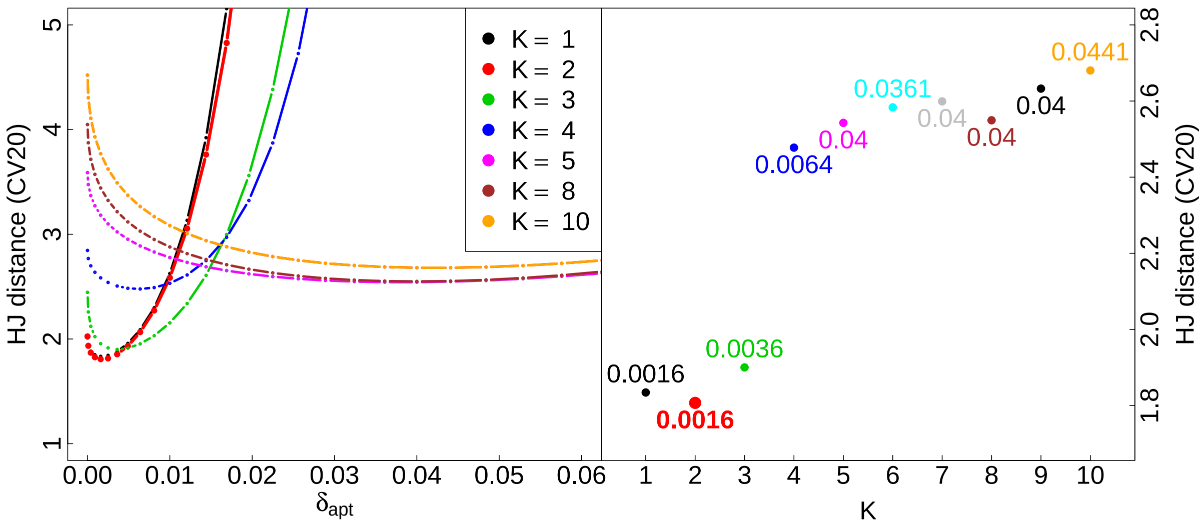
In this section, first we estimate the admissible SDF under the APT and characterize its components, thereby establishing the relative quantitative importance of systematic versus unsystematic risk. Then, we examine commonly used candidate factor models of asset returns, such as the market model, the consumption-CAPM model, and the Fama and French (1993) three-factor model. For each of these models, we characterize the missing systematic and unsystematic components of the corresponding SDFs.

5.1 The SDF under the APT Model for Asset Returns

We start our analysis by constructing the SDF implied by the APT model for asset returns in (1); that is, for the true model of returns. To determine the number of latent factors K

Figure 1: Model selection using the HJ distance

This figure illustrates how the HJ distance changes with K and δ_{apt}. The two panels show the estimation results based on cross validation.The panel on the left plots the HJ distance for a given choice of K as one varies δ_{apt}. The panel on the right displays the δ_{apt} (numbers inside the box) that minimizes the HJ distance for a given choice of K.



and the no-arbitrage bound δ_{apt}, we use the HJ distance and employ 20-fold cross validation, as described above.

The two panels of Figure 1 illustrate how the HJ distance changes as we vary K and δ_{apt}. We see that the combination of $K = 2$ latent factors (see left panel) and $\delta_{\text{apt}} = 0.0016$ (see right panel) achieves the smallest HJ distance, consistent with the evidence on low-dimensional factor pricing models in Kozak, Nagel, and Santosh (2018).[12] The nonzero value for the optimal δ_{apt} indicates that unsystematic risk is priced, that is, the vector $a \neq 0_N$. This constitutes our first main finding because it challenges the conventional view that asset returns compensate only for systematic (common) risk factors.

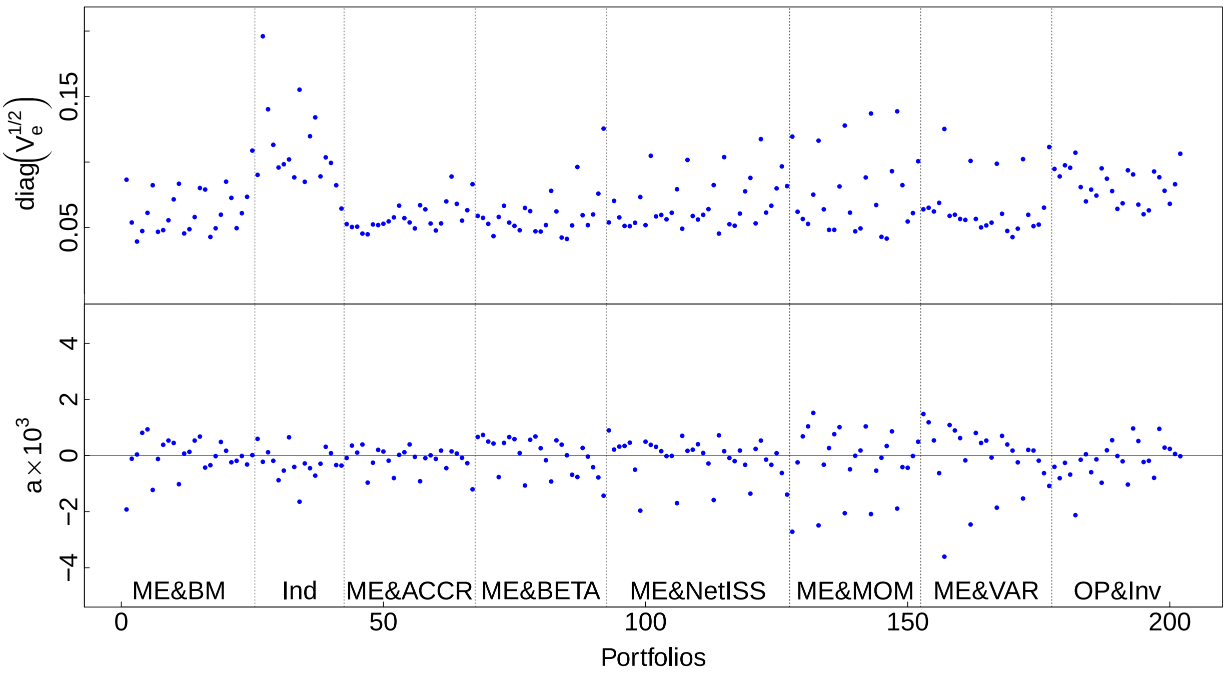
To understand the importance of accounting for compensation for asset-specific risk, we explore how the HJ distance changes if we set $a = 0_N$. We find that a model with $K = 2$ and $a = 0_N$ has significantly higher HJ distance—about 15% higher than when $a \neq 0_N$. When studying the pricing errors across the 202 basis assets, we observe that the largest increase in pricing errors from setting $a = 0_N$ is for the portfolios sorted by size and variance.

In similar vein, to understand the importance of unsystematic risk, one could also consider models based on principal components of demeaned asset returns. Again, we find that

[12]The difference in performance of the optimal model relative to models with alternative K and δ_{apt} is statistically significant. Throughout the paper, for statistical inference, we use bootstrap methods.

Figure 2: Estimated asset-specific risk and its compensation

This figure illustrates the estimated elements of the vector a and diagonal matrix V_e for the 202 basis assets, which we split into eight groups based on characteristics by which stocks are sorted into portfolios. The top panel shows the asset-specific volatilities $\mathrm{diag}(V_e^{1/2})$ and the bottom panel shows the compensation a for asset-specific risk.



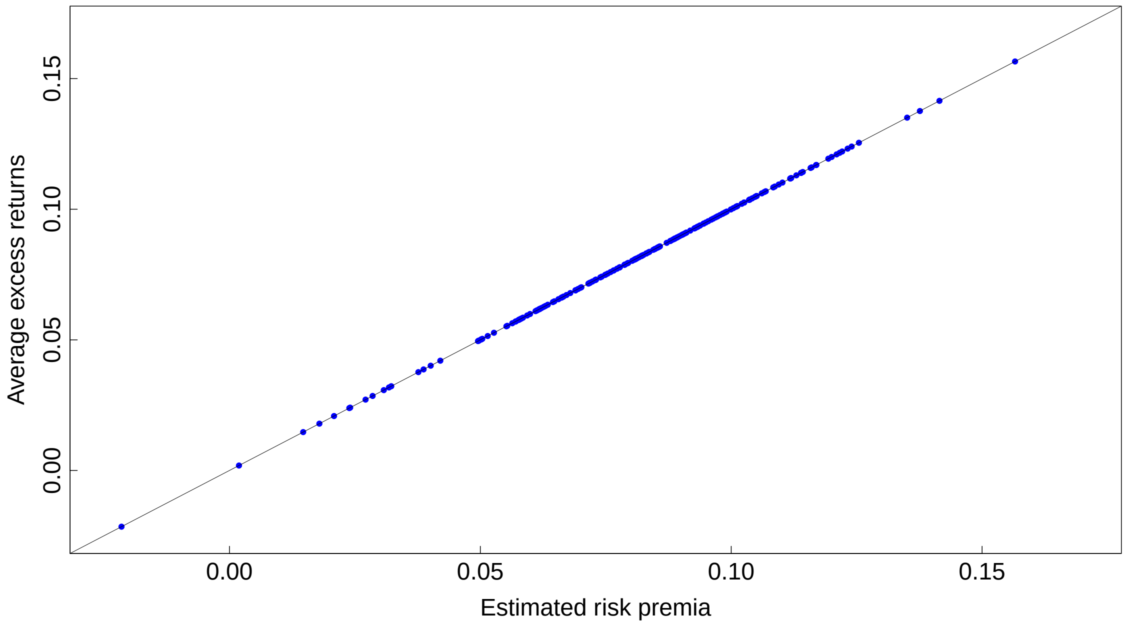
ignoring compensation for asset-specific risk leads to a statistically significant increase in the HJ distance. For example, a candidate model with the first two principal components (PCA2) has an HJ distance 14.91% larger than that implied by the APT-model. We also find that a naive strategy of including a larger number of principal components in the factor model of asset returns leads to inferior pricing performance of the implied SDF because improving the fit of the covariance matrix of asset returns by using a larger number of principal components worsens the fit of expected excess returns.[13] This tradeoff is consistent with the evidence in Lettau and Pelger (2020) and Kozak, Nagel, and Santosh (2020).

Below we will shed light on which basis assets contribute most to the SDF-U component. We will also explore which tradable strategies reflect risk premia for the unsystematic risk.

Figure 2 illustrates the estimated elements of the vector a and diagonal matrix V_e for the 202 basis assets. The top panel shows that these assets have different asset-specific volatilities, so our assumption that V_e is a diagonal rather than spherical matrix is war-

[13] The HJ distances of the models based on the first one to five principal components, relative to that of the APT model, are 14.76%, 14.91%, 17.42%, 35.11%, 34.98%, respectively.

Figure 3: Comparing average excess returns with model-implied risk premia
This figure overlays the average excess returns on the test assets with the model-implied risk premia after undoing the effect of shrinkage for the ridge estimator of the vector a.



ranted. The bottom panel indicates that the compensation for exposures of basis assets to individual asset-specific shocks is small relative to the premia of conventional risk factors.

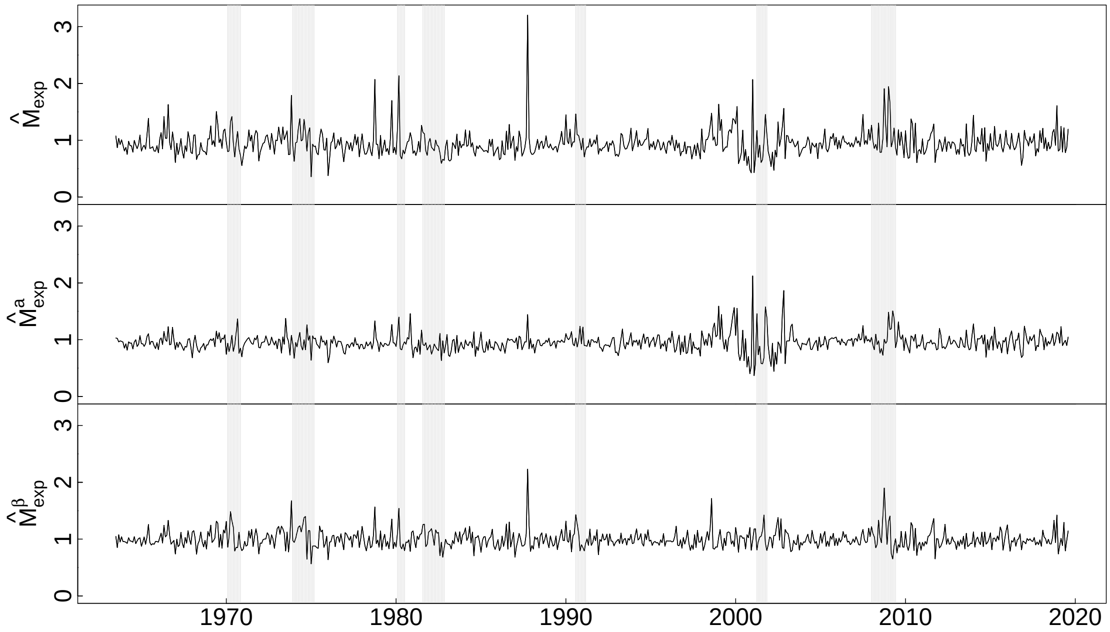
A common approach for evaluating asset-pricing factor models is to plot average excess returns on the test assets against the model-implied risk premia. We follow this approach in Figure 3, which shows that, as expected, our model exhibits a perfect fit. This is a consequence of our main insight, namely to interpret elements of the vector a as compensation for exposure to individual asset-specific shocks, as opposed to pricing errors.[14]

Next, we study the time-series properties of the estimated SDF, $\hat{M}_{\exp,t+1}$, and its components, $\hat{M}^{a}_{\exp,t+1}$ and $\hat{M}^{\beta}_{\exp,t+1}$. Figure 4 shows that both $\hat{M}^{a}_{\exp,t+1}$ and $\hat{M}^{\beta}_{\exp,t+1}$, exhibit sizable volatility during recessions and also during normal times. We see that different components of the SDF dominate its variation in different time periods. For example, in the Fall of 1987, common systematic shocks in asset returns are responsible for a dramatic increase in the level and volatility of the SDF. On the other hand, in the early 2000s, it

[14] Recall that the estimated vector a is a ridge estimator of the asset-specific risk premia (see the formula given in expression (A5) in Appendix A.6). For this exercise, we undo the shrinkage by multiplying each element of the estimated vector a by 1 plus the estimated value of the Karush-Kuhn-Tucker multiplier κ in equation (11). The estimated value of κ is 19.16.

Figure 4: Time series behavior of the SDF and its components

This figure has three panels. The top, middle, and bottom panels show the dynamics of the SDF $\hat{M}_{\exp,t+1}$, its asset-specific component $\hat{M}^{a}_{\exp,t+1}$, and its systematic component $\hat{M}^{\beta}_{\exp,t+1}$, respectively.



is the SDF-U component that generates a spike in the volatility of the SDF. Thus both common and asset-specific shocks contribute to explaining asset valuations.

We find that, of the total variation of the SDF, the SDF-S component contributes only 44%, while the SDF-U component contributes 56%, which highlights the importance of accounting for asset-specific risk in factor models of asset returns. These results are consistent with Daniel and Titman (1997) and Chaieb, Langlois, and Scaillet (2021), among others, who document that a substantial portion of expected excess returns is left unexplained by factor risk premia. Our finding also speaks to the puzzling evidence reported in Herskovic, Moreira, and Muir (2019) and Lopez-Lira and Roussanov (2022), who document that the portfolios of stock returns that hedge factor risk exposure exhibit high positive expected returns. These high expected returns could reflect compensation for unsystematic risk.

Furthermore, regression results (see Figure IA.1 of the Internet Appendix) indicate that $\hat{M}^{a}_{\exp,t+1}$ is acyclical: $\log(\hat{M}^{a}_{\exp,t+1})$ does not significantly correlate with any business-cycle indicator.[15] In contrast to $\log(\hat{M}^{a}_{\exp,t+1})$, the systematic component $\log(\hat{M}^{\beta}_{\exp,t+1})$ has a significant correlation with the NBER recession indicators (see Figure IA.2 of the Internet

[15] In regressions, we use the log SDF because it is linear in common risk factors and asset-specific shocks.

Appendix). Given that $\hat{M}^{\beta}_{\exp,t+1}$ and $\hat{M}^{a}_{\exp,t+1}$ reflect the composition of the systematic and asset-specific components of the SDF, our approach provides a procedure for constructing trading strategies with and without exposures to systematic risk.

Having established the quantitative importance of the SDF-U component, $\hat{M}^{a}_{\exp,t+1}$, we examine which trading strategies reflect exposures to this component of the SDF. As a first step, we run individual regressions of $\log(\hat{M}^{a}_{\exp,t+1})$ on the excess returns of 457 strategies. We find 27 strategies with an R^2 larger than 30%. As a second step, we compute the risk premia associated with the exposures of these 457 trading strategies to the SDF-U component as the negative covariance of the return on the strategy and $\hat{M}^{a}_{\exp,t+1}$:

$$\mathrm{RP}^{a}_{\text{strategy}} = -\operatorname{cov}(R_{\text{strategy},t+1}, \hat{M}^{a}_{\exp,t+1}) \times R_f.$$

As expected, we find that many of the strategies that are highly correlated with $\log(\hat{M}^{a}_{\exp,t+1})$ are associated with large risk premia. However, there are also some strategies that are not highly correlated with $\log(\hat{M}^{a}_{\exp,t+1})$ but that still command sizable risk premia; for example, momentum strategies. Table 1 lists strategies that have high correlations with and high compensation for exposure to the SDF-U component.

Examining the strategies from Table 1 closely, we find that there is large overlap across these strategies and that, adopting the classification of Jensen, Kelly, and Pedersen (2021), they fall into the following clusters: Investment, Leverage, Low Risk, Profitability, and Value. In the literature, some of these strategies have been interpreted as being behavioral—for example, the management factor of Stambaugh and Yuan (2017) and the long-horizon financial factor of Daniel, Hirshleifer, and Sun (2020))—while others as reflecting market frictions—for example, the betting-against-beta factor of Frazzini and Pedersen (2014) and constraints return relation among high R&D firms in Li (2011).

Examining the composition of $\hat{M}^{a}_{\exp,t+1}$, we identify a substantial contribution of small stocks. Specifically, of the 34 basis assets with the highest contribution to the variation of $\hat{M}^{a}_{\exp,t+1}$, fifteen represent various portfolios of small stocks, such as small stocks with low and high book-to-market, small stocks with high accruals, and small stocks with high prior returns. In addition, sixteen basis assets represent a range of portfolios of stocks sorted by size and market beta or size and variance. Finally, seven of these 34 basis assets are portfolios of stocks with extremely high or low values of net issuances.

The special role of small stocks in the SDF-U component is an estimation result, not an assumption hardwired into the corrected candidate model. Proposition 4 shows that the

Table 1: Strategies highly correlated with $\log(\hat{M}^a_{\exp,t+1})$ and with high asset-specific risk premium RP^a

This table reports trading strategies that are highly correlated with $\log(\hat{M}^a_{\exp,t+1})$ and that have a large asset-specific risk premium. The first column, using the classification scheme in Jensen, Kelly, and Pedersen (2021), gives the name of the cluster to which the strategy belongs. If a strategy is not in the list of Jensen, Kelly, and Pedersen (2021), we assign it to the cluster Unclassified. The second column gives the source. The third column gives the name of the variable, as in Chen and Zimmermann (2021) or Jensen, Kelly, and Pedersen (2021). The penultimate column reports the R^2 of the univariate regressions of $\log(\hat{M}^a_{\exp,t+1})$ on the return of each individual strategy. The last column reports the risk premium. The clusters, and within each cluster the sources, are listed in alphabetical order.

Cluster name	Source	Variable name	R^2(%)	RP^a(%)
Investment	Daniel, Hirshleifer, and Sun (2019)	beh_fin	44.85	5.33
	Fama and French (2015)	cma	30.21	2.08
	Hou, Xue, and Zhang (2015)	r_ia	30.48	2.04
	Ortiz-Molina and Phillips (2014)	aliq_at	32.21	−3.89
	Ritter (1991)	ageipo	29.31	5.83
	Stambaugh and Yuan (2016)	mgmt	33.59	3.22
	Xing (2008)	invcap	39.26	6.06
Leverage	Bhandari (1988)	leverage	43.21	5.54
	Fama and French (1992)	am	38.42	5.16
	Fama and French (1992)	bookleverage	39.58	−4.25
	Palazzo (2012)	cash	32.26	−5.98
	Palazzo (2012)	cash_at	36.66	−3.72
	Penman Richardson and Tuna (2007)	netdebt_me	34.61	3.84
Low risk	Ali, Hwang, and Trombley (2003)	idiovolaht	15.68	5.88
	Ang Chen and Xing (2006)	betadown_252d	32.11	−5.03
	Ang, Hodrick, Xing, Zhang (2006)	rvol_21d	25.92	−5.25
	Ang, Hodrick, Xing, Zhang (2006)	ivol	18.14	6.14
	Bali, Cakici, and Whitelaw (2010)	maxret	20.67	6.73
	Bradshaw, Richardson, Sloan (2006)	netequityfinance	37.50	5.24
	Bradshaw, Richardson, Sloan (2006)	xfin	39.68	6.54
	Fama and MacBeth (1973)	beta	17.37	−6.05
	Frazzini and Pedersen (2014)	betafp	17.60	−6.93
	Frazzini and Pedersen (2014)	bab	46.65	4.48
	Pontiff and Woodgate (2008)	shareiss1y	37.48	2.90
Momentum	Jegadeesh and Titman (1993)	mom12m	9.78	5.36
	Jegadeesh and Titman (1993)	mom6m	11.98	5.51
Profitability	Chen, Novy-Marx, Zhang (2011)	rome	25.22	5.11
	Diether, Malloy and Scherbina (2002)	forecastdispersion	26.83	5.05
	Frankel and Lee (1998)	predictedfe	39.73	5.46
	Frankel and Lee (1998)	analystvalue	38.00	4.86
	La Porta (1996)	fgr5yrlag	50.71	8.19
Value	Barbee, Mukherji and Raines (1996)	sp	33.09	4.46
	Basu (1977)	ep	32.65	5.33
	Boudoukh, Michaely, Richardson, Roberts (2007)	eqnpo_me	34.34	4.74
	Daniel and Titman (2006)	eqnpo_12m	30.31	3.39
	Dechow, Sloan and Soliman (2004)	equityduration	40.83	5.11
	Fama and French (1992)	hml	37.32	3.30
	Litzenberger and Ramaswamy (1979)	div12m_me	31.50	4.61
Unclassified	Cen, Wei, and Zhang (2006)	feps	24.87	6.15
	Cooper, Gulen, Schill (2008)	betaarb	24.90	6.19
	Datar, Naik, Radcliffe (1998)	shvol	25.10	5.88
	Easley, Hvidkjaer and O'Hara (2002)	probinformedtrading	21.36	6.20
	Elgers, Lo and Pfeiffer (2001)	sfe	29.24	7.00
	Li (2011)	rdcap	33.24	−4.47
	Ritter (1991)	indipo	36.42	4.30

relative weight of a basis asset in $M^{a}_{\exp,t+1}$ depends on the *ratio* of the asset's compensation for asset-specific risk, represented by the corresponding element of the vector a, and asset-specific volatility. While it is natural to expect that small stocks have larger firm-specific components of expected returns, it is also known that they have higher volatility. Therefore, ex-ante, it is not clear if small stocks will feature prominently in the SDF-U component.

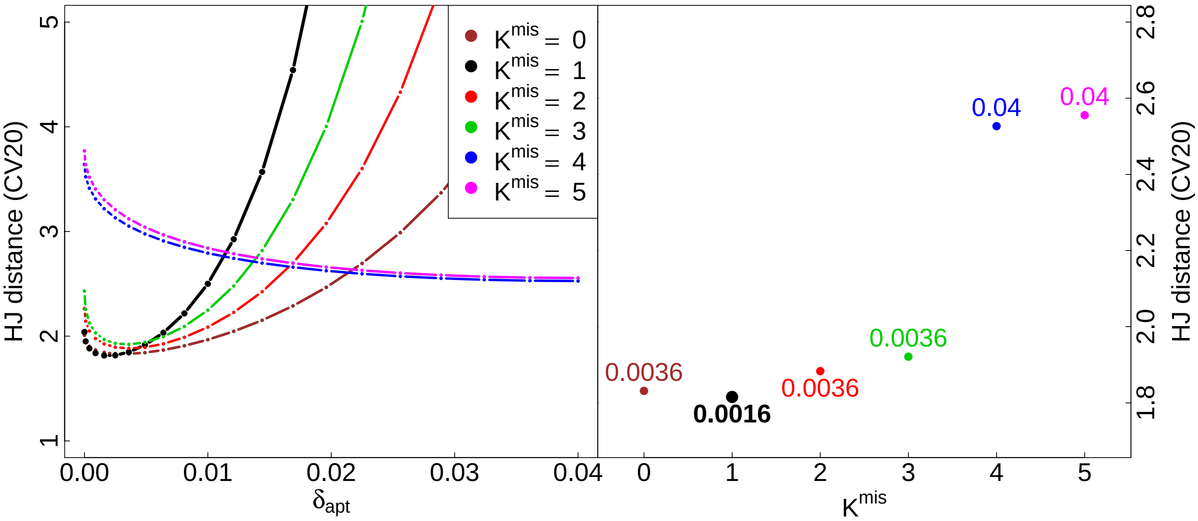
Our finding regarding the role of small stocks in the SDF-U component complements the literature on granular origins of aggregate fluctuations (Gabaix, 2011). In this literature, idiosyncratic shocks to fundamentals of *large* firms can lead to nontrivial aggregate effects, that is, these shocks explain a substantial part of variation in aggregate fundamentals, or equivalently, in $\hat{M}^{\beta}_{\exp,t+1}$. In contrast, our empirical finding is about the importance of idiosyncratic shocks to the returns of small companies that drive the acyclical component of the SDF, $\hat{M}^{a}_{\exp,t+1}$.

We now turn our attention to the component of the SDF related to systematic risk factors, $\hat{M}^{\beta}_{\exp,t+1}$. We find that the market factor of Sharpe (1964) exhibits the highest explanatory power for $\log(\hat{M}^{\beta}_{\exp,t+1})$, with an $R^2 = 0.95$. It is remarkable that, despite all the criticism of the market model, when we consider only the systematic component of the SDF, the market factor explains such a large proportion of its variation. Besides the market factor, we find that there are 23 trading strategies and 3 nontraded factors (shocks in VIX, intermediary capital (He, Kelly, and Manela, 2017), and dividend yield) that each individually explain more than 30% of variation in $\log(\hat{M}^{\beta}_{\exp,t+1})$. Because the market factor already explains a large proportion of the variation in $\log(\hat{M}^{\beta}_{\exp,t+1})$, the other factors explain only a small proportion of the variation not explained by the market factor. A combination of nineteen trading strategies is needed to explain 99% of variation in $\log(\hat{M}^{\beta}_{\exp,t+1})$.[16] These results are in line with the findings of Kozak, Nagel, and Santosh (2020) and Bryzgalova, Huang, and Julliard (2020) about the nonsparsity of the SDF in characteristics and the existence of several combinations of trading strategies that deliver a similar cross-sectional fit.

[16]These strategies include: market, size, betting-against-beta, sales-to-market (Barbee Jr, Mukherji, and Raines, 1996), change in current operating working capital and change in noncurrent operating liabilities (Richardson, Sloan, Soliman, and Tuna, 2005), Kaplan-Zingales index (Lamont, Polk, and Saa-Requejo, 2001), cash-to-assets (Palazzo, 2012), dollar trading volume Brennan, Chordia, and Subrahmanyam (1998), highest 5 days of return scaled by volatility (Asness, Frazzini, Gormsen, and Pedersen, 2020), quality-minus-junk growth (Asness, Frazzini, Israel, Moskowitz, and Pedersen, 2018), and short interest (Dechow, Hutton, Meulbroek, and Sloan, 2001).

Figure 5: Correction of the CAPM model using the HJ distance

This figure illustrates how the HJ distance changes with K^{mis} and δ_{apt}, when the candidate model includes only the market factor. The two panels show the estimation results based on cross validation. The panel on the left plots the HJ distance for a given choice of K^{mis} as one varies δ_{apt}. The panels on the right display the optimal values of the HJ distance for a given choice of K^{mis}.



5.2 Candidate Factor Models and Principal-Component-Based Models

To illustrate how our method brings new insights about the SDF corresponding to existing factor models for returns, we now consider three traditional candidate models—those implied by the CAPM of Sharpe (1964), the Consumption-CAPM (C-CAPM) of Breeden (1979), and the three-factor model of Fama and French (1993). For the SDF $\hat{M}^{\beta,\text{can}}_{\text{exp},t+1}$ implied by each of these candidate factor models, we estimate the required correction terms $\hat{M}^{a}_{\text{exp},t+1}$ and $\hat{M}^{\beta,\text{mis}}_{\text{exp},t+1}$, which allows us to identify what is missing in each of these models.

5.2.1 The CAPM

We consider a candidate model with the market return as its sole factor ($K^{\text{can}} = 1$) and the vector $a^{\text{can}} = 0_N$, which we refer to as the CAPM. Figure 5 shows that when the candidate model is the CAPM, then the estimation procedure selects $K^{\text{mis}} = 1$ and $\delta_{\text{apt}} = 0.0016$. Thus, the CAPM suffers from *both* sources of misspecification: K^{can} is less than the true number of factors K and the vector $a \neq 0_N$. The obtained number of missing factors to correct the market model is in line with our earlier finding that two latent factors summarize the common variation in asset returns, with one factor being a proxy for the market factor.

Table 2: Analysis of models before and after correction for misspecification
The first column of the table lists the models considered. Then, the table reports three sets of quantities: (1) The HJ distances of alternative models, relative to the HJ distance of the APT model, $(\text{HJ}^{\text{model}}/\text{HJ}^{\text{APT}} - 1) \times 100\%$, before and after the model is corrected for misspecification; (2) The Sharpe ratio of the corrected SDF for each of the models along with its components; and, (3) The variance decomposition of the SDF components.

	Relative HJ (%)		Sharpe ratio (p.a.)				Variance decomp. (%)			
Model	Before correction	After correction	$\hat{M}_{\text{exp}}$	$\hat{M}^{a}_{\text{exp}}$	$\hat{M}^{\beta,\text{can}}_{\text{exp}}$	$\hat{M}^{\beta,\text{mis}}_{\text{exp}}$	$\hat{M}_{\text{exp}}$	$\hat{M}^{a}_{\text{exp}}$	$\hat{M}^{\beta,\text{can}}_{\text{exp}}$	$\hat{M}^{\beta,\text{mis}}_{\text{exp}}$
APT	—	—	0.84	0.59	0.55	—	100.00	56.21	43.79	—
CAPM	14.72	0.77	0.81	0.63	0.44	0.28	100.00	57.60	31.37	11.03
CCAPM	14.83	-0.16	0.91	0.72	0.36	0.44	100.00	58.69	21.45	19.85
FF3	15.45	-3.37	0.98	0.69	0.70	0.00	100.00	42.16	57.84	0.00

The importance of the SDF-U component for correcting the CAPM is evident from Table 2. The second and third columns of this table show that after we correct the CAPM for misspecification, the HJ distance drops by a statistically significant 14%. Analyzing the pricing errors before and after correcting the CAPM, we find that the largest improvement in pricing is for the portfolios formed by sorting stocks by size and value, size and beta, size and net issuance, and size and variance. The last set of columns of this table and Figure 6 illustrate that the SDF-U component explains most of the variation in the corrected SDF. Specifically, 57.6% of the variation in $\log(\hat{M}_{\text{exp},t+1})$ is due to the SDF-U component, while only 11.03% is due to missing systematic risk in the CAPM. Thus, we conclude that the improvement in the pricing performance of the corrected CAPM is mainly due to the inclusion of the SDF-U component.

We now characterize the variation in $\log(\hat{M}_{\text{exp},t+1})$ that is due to systematic risk factors. The remaining 42% of the variation in the log SDF not explained by the SDF-U component is due to the combination of $\hat{M}^{\beta,\text{can}}_{\text{exp},t+1}$ and $\hat{M}^{\beta,\text{mis}}_{\text{exp},t+1}$, with $31.37/(31.37 + 11.03) \approx 73.99\%$ of this variation attributable to the market factor. Recall that for the APT model, we find that the market factor explains 95% of the systematic component of the admissible SDF.[17]

In contrast to the case of the APT model, we find that the *systematic* component, $\log(\hat{M}^{\beta,\text{mis}}_{\text{exp},t+1})$ is uncorrelated with nontraded factors. The reason for this zero correlation is that the candidate SDF $\log(M^{\beta,\text{can}}_{\text{exp},t+1})$ based on the market model subsumes the explana-

[17]The quantitative difference in the role of the observable market factor is because the market factor is only a proxy for the latent risk factor recovered when considering the APT model. Specifically, the ratio of the standard deviations of the systematic component of the SDF explained by the first systematic component in the case of the APT model to that explained by the market factor in the case of the CAPM candidate model is 1.2.

Figure 6: Time series of SDF and its components for CAPM candidate model
This figure has four panels, which show the dynamics of the admissible SDF, $M_{\text{exp},t+1}$ and its three components: the asset-specific component $M^{a}_{\text{exp},t+1}$, the component $M^{\beta,\text{can}}_{\text{exp},t+1}$ corresponding to the candidate model with the market factor, and the missing systematic component $M^{\beta,\text{mis}}_{\text{exp},t+1}$.

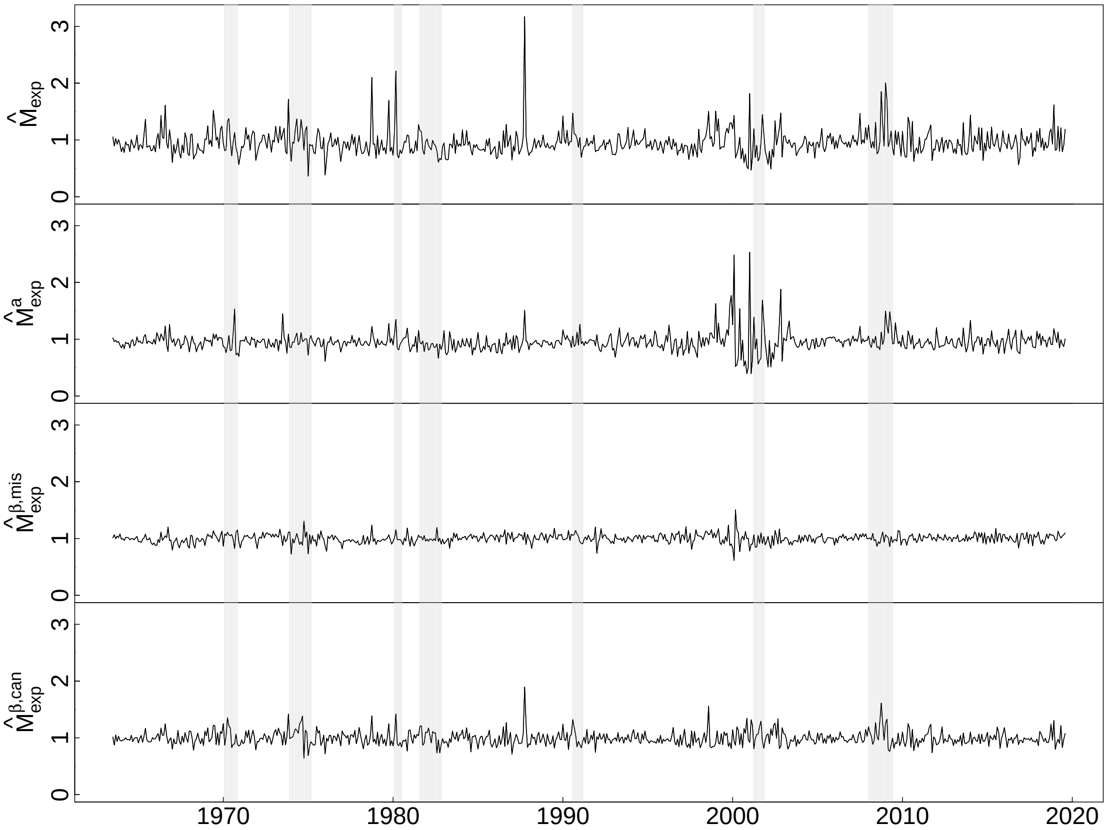


Table 3: Correlation matrix of corrected SDFs
This table reports the correlation matrix of admissible SDFs obtained after correcting different candidate models: APT, CAPM, C-CAPM, and FF3.

	APT	CAPM corrected	C-CAPM corrected	FF3 corrected
APT	1.00			
CAPM corrected	0.99	1.00		
C-CAPM corrected	0.94	0.94	1.00	
FF3 corrected	0.94	0.94	0.89	1.00

tory power of innovations in VIX, intermediary capital, and dividend yield. We find 27 trading strategies that individually explain more than 30% of variation in $\log(\hat{M}^{\beta,\text{mis}}_{\text{exp},t+1})$. The size factor (Fama and French, 1993) is one of the most prominent among them with an explanatory power of about 89%, which explains the success of the models developed in Fama and French (1993, 2015).

We conclude by highlighting that our approach is successful in correcting the CAPM to obtain an admissible SDF, which, we see from Table 3, is almost perfectly correlated with the SDF previously obtained from the APT model.

5.2.2 The C-CAPM

We now consider a candidate model with the return on a consumption-mimicking portfolio as its sole factor and the vector $a^{\text{can}} = 0_N$, which we refer to as the C-CAPM. We follow the standard approach of Breeden, Gibbons, and Litzenberger (1989) to construct the consumption-mimicking portfolio.[18]

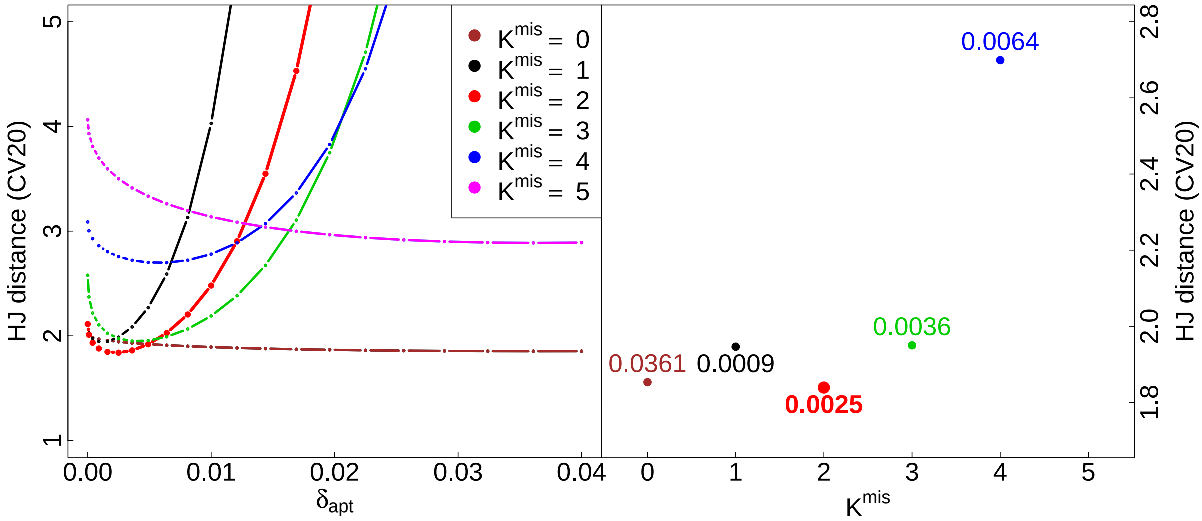
Figure 7 shows that if one starts from the C-CAPM, then the estimation procedure selects $K^{\text{mis}} = 2$ and $\delta_{\text{apt}} = 0.0025$. The consumption-mimicking portfolio is not highly correlated with either of the latent factors estimated when correcting the APT model—the correlations are 0.3 and 0—and therefore, we still require two additional factors to capture the common variation in asset returns. The value of $\delta_{\text{apt}} = 0.0025$ implies an annual Sharpe ratio associated with pure asset-specific risk being of $(0.0025 \cdot 12)^{1/2} = 0.17$.

Table 2 shows that augmenting the consumption-mimicking-portfolio factor by two latent factors and the vector of the asset-specific components in expected returns, a, leads to a substantial and statistically significant drop in the HJ distance by 14.99%. Most of this drop is accounted for by the SDF-U component, which contributes 58.69% of the time-variation in $\log(\hat{M}_{\text{exp},t+1})$. On the other hand, examining the pricing errors of the candidate model (see Figure IA.4 in the Internet Appendix), we see that they are centered around 0.06, whereas they are centered around zero in the corrected model. This observation indicates that a missing level factor in the candidate C-CAPM is also a substantial source of misspecification. However, the variance decomposition in Table 2 shows that the missing

[18] As outlined in Giglio and Xiu (2021), construction of factor mimicking portfolios can be sensitive to the choice of basis assets. They propose a three-stage procedure, which is insensitive to the choice of basis assets. However, their procedure does not allow for asset-specific risk, which we document plays a major role in the risk-return trade-off.

Figure 7: Correction of the C-CAPM model using the HJ distance

This figure illustrates how the HJ distance changes with K^{mis} and δ_{apt}, when the candidate model includes only the consumption-mimicking portfolio of Breeden, Gibbons, and Litzenberger (1989). The two panels show the estimation results based on cross validation. The panel on the left plots the HJ distance for a given choice of K^{mis} as one varies δ_{apt}. The panel on the right displays the optimal values of the HJ distance for a given choice of K^{mis}.



systematic risk factor explains a much smaller proportion of the variation in the admissible SDF, compared to the SDF-U component.[19]

5.2.3 The Three-Factor Model of Fama and French (1993)

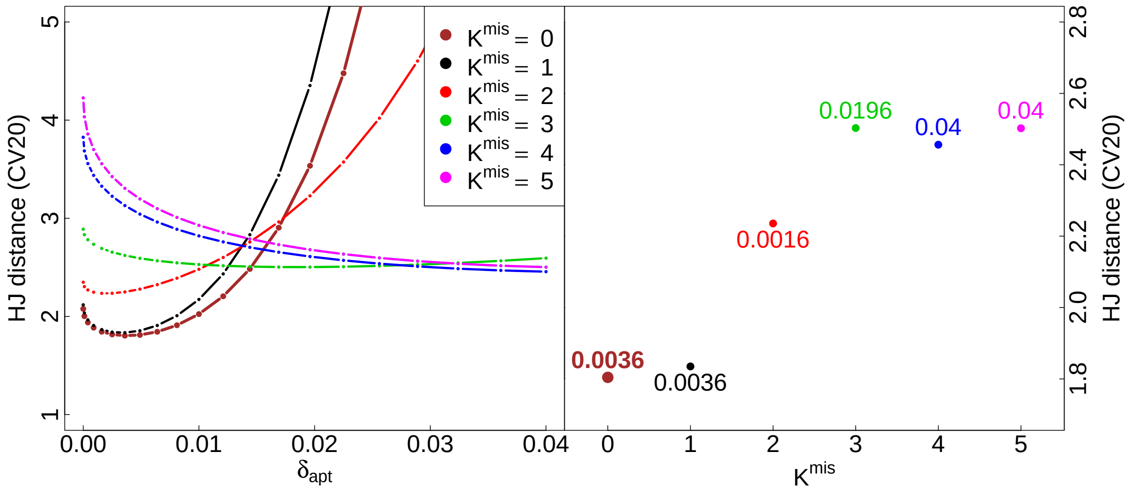
We consider a candidate model with the three factors of Fama and French (1993), market, size, and value, and the vector $a^{\text{can}} = 0_N$, and we refer to this model as FF3. Figure 8 shows that if we start with FF3 as the candidate model for asset returns, our method selects *zero* missing sources of systematic risk and an optimal $\delta_{\text{apt}} = 0.0036$. At first glance, it may seem surprising that this candidate factor model incorporates all systematic variation in asset returns, given that about 19 strategies are necessary to capture the systematic SDF, as explained in the case of the APT model in Section 5.1. However, the FF3 model already includes the market and size factors that jointly explain more than 96% of the variation in the systematic component of the SDF.

Furthermore, we find that the value factor correlates more strongly with the SDF-U component (the correlation is -0.61) than with the systematic component of the SDF (the

[19] Figure IA.3 in the Internet Appendix shows the estimated time series of the admissible SDF and its components obtained after correcting the candidate C-CAPM.

Figure 8: Correction of FF3 model using HJ distance

This figure illustrates how the HJ distance changes with K^{mis} and δ_{apt}, when the candidate model is the three-factor model (Fama and French, 1993). The two panels show the estimation results based on cross validation. The panel on the left plots the HJ distance for a given choice of K^{mis} as one varies δ_{apt}. The panel on the right displays the optimal values of the HJ distance for a given K^{mis}.



correlation is 0.14). Thus, FF3 implicitly incorporates some asset-specific risk and/or weak factors.

Table 2 shows that augmenting the FF3 model with the vector of asset-specific components in expected returns, a, leads to a substantial and statistically significant improvement in pricing performance: the HJ distance drops by 18.82%. The large drop in the HJ distance indicates the quantitative importance of unsystematic risk for pricing. Thus, similar to Stambaugh and Yuan (2017), Bryzgalova, Huang, and Julliard (2020), and Clarke (2020) among others, we document sizable misspecification in the FF3 model, but in contrast to these papers, we attribute the misspecification to compensation for unsystematic risk, i.e., $a \neq 0_N$. Analyzing the pricing errors before and after correcting the FF3 model, we find the largest improvement in pricing is for the portfolios formed by sorting stocks by size and momentum and size and variance.[20]

Table 3 shows that our approach for correcting misspecification in the original FF3 model leads to an admissible SDF that is highly correlated with that implied by the APT model and those obtained after correcting the other candidate factor models.

[20]The time-series behavior of the admissible SDF obtained from correcting the original FF3 model is displayed in Figure IA.5 in the Internet Appendix.

5.3 Robustness of estimation results

To illustrate the robustness of our estimation approach, we undertake two exercises. First, we show that if one were to estimate K and δ_{apt} *in sample*, instead of estimating them using cross-validation (as described in Section 3.2 and shown in Section 5.1), the pricing errors would have been much larger. Second, we examine the pricing performance of the SDF estimated on one set of return data but evaluated on two other datasets.

In Figure 9, we display the results from estimating the APT model in sample. The two panels show that a naive in-sample analysis leads to a choice of $K = 10$ and $\delta_{\text{apt}} = 0.04$. These estimated parameters are substantially larger than the $K = 2$ and $\delta_{\text{apt}} = 0.0016$ obtained from estimating the APT model using cross-validation. The larger number of in-sample factors, K, is to fit as best as possible the covariance matrix for returns, while the larger value of δ_{apt} is to fit as best as possible the cross-sectional variation in expected excess returns. However, the right-hand panel of Figure 1 (on page 20) shows that the in-sample combination of $K = 10$ and $\delta_{\text{apt}} = 0.04$ performs extremely poorly in the cross-validation exercise because of overfitting.

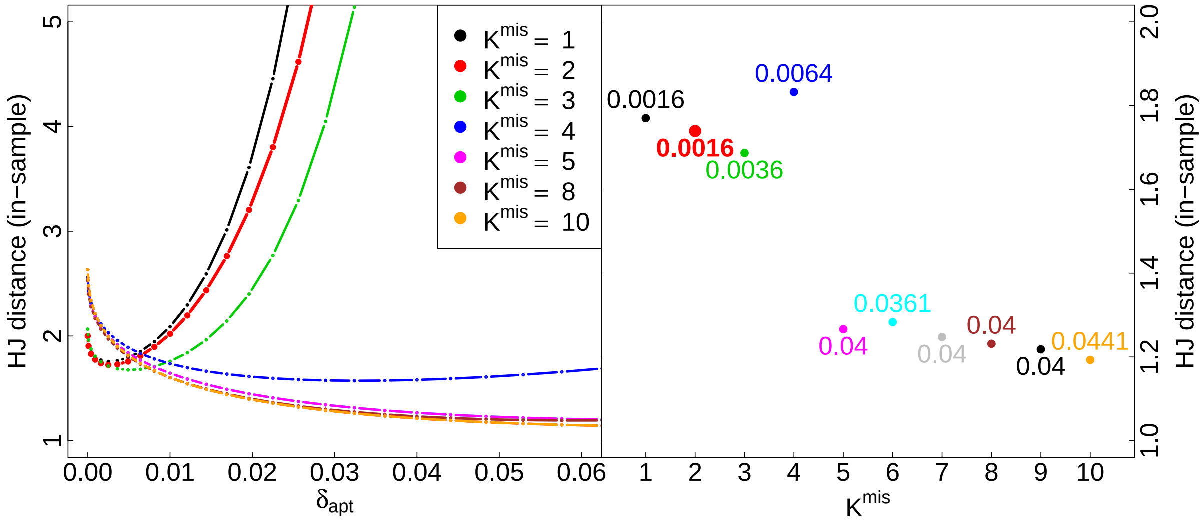
The second exercise we undertake to illustrate the robustness of our approach is to run an out-of-sample analysis: we evaluate how the candidate and corrected CAPM, C-CAPM, and FF3 models estimated from the first dataset (described in Section 4.1) price two different cross-sections of stock returns. The second dataset we consider, also used in Korsaye, Quaini, and Trojani (2021), includes 100 portfolios sorted by size and book-to-market, 25 portfolios sorted by size and long-term reversal, 25 portfolios sorted by size and short-term reversal, and 49 industry portfolios.[21] The third dataset we consider includes 100 portfolios sorted by size and book-to-market, 100 portfolios sorted by size and operating profitability, 100 portfolios sorted by size and investment, and 49 industry portfolios.

The two alternative datasets we consider include portfolios formed by sorting stocks on the same or similar characteristics as those used to form the set of basis assets, but the sort has a different level of granularity or different order. For example, the set of basis assets in our first dataset includes portfolios formed by sorting stocks by size and momentum, whereas the second dataset, used also by Korsaye, Quaini, and Trojani (2021), includes portfolios formed by sorting on size and long- or short-term reversal. Also, while the set of basis assets includes portfolios formed by sorting stocks on operating profitability and investment, the

[21] The dataset of Korsaye, Quaini, and Trojani (2021) also includes twenty five momentum portfolios that we exclude because they are present in our 202 basis assets.

Figure 9: Model selection using the HJ distance in sample

This figure illustrates how the HJ distance changes with K^{mis} and δ_{apt} when these parameters are estimated *in sample*. The panel on the left plots the HJ distance for a given choice of K^{mis} as one varies δ_{apt}. The panel on the right displays δ_{apt} (numbers inside the box) that minimizes the HJ distance for a given choice of K^{mis}.



third dataset includes portfolios formed by sorting stocks by size and investment and size and operating profitability. Similarly, the compositions of industry portfolios in the two out-of-sample datasets differ from those in the first dataset.

This choice of test assets for our out-of-sample analysis is reasonable because the estimated admissible SDF represents not the marginal utility but a projection of the marginal utility on a set of basis assets. Thus, by construction, the estimated SDF would not price test assets whose returns are orthogonal to those of the basis assets (Cochrane, 2005), and hence, we use test assets that are not orthogonal to the basis assets. Our choice of test assets allows us to evaluate if the estimated SDF is subject to overfitting or captures successfully the risks associated with selected characteristics.

Table 4 reports results for the three datasets. For the first dataset, we have already discussed the two key insights: (a) the APT model exhibits much better pricing performance than the three candidate factor models (see column (1)) and (b) after correction, the candidate factor models perform as well as the APT model (see column (2)). From the second and third datasets there are three main insights: (a) the pricing performance of the APT model continues to be significantly better than that of the candidate factor models (see columns (3) and (5)); (b) the relative edge in pricing performance of the APT model declines out-of-sample (compare columns (1) and (3) and columns (1) and (5)); and

Table 4: Cross-sectional Out-of-Sample Pricing Performance

This table reports the HJ distances of alternative models, relative to the HJ distance of the APT model, $(\text{HJ}^{\text{model}}/\text{HJ}^{\text{APT}} - 1) \times 100\%$, before a candidate model is corrected for misspecification, and after it has been corrected. All the models are estimated on our first dataset, the set of 202 basis assets described in Section 4.1. The performance of these models is then evaluated for the set of 202 basis assets ("First dataset") and for two additional datasets not used at the estimation ("Second dataset" and "Third dataset").

	First dataset		Second dataset		Third dataset	
	(1)	(2)	(3)	(4)	(5)	(6)
	Before correction	After correction	Before correction	After correction	Before correction	After correction
CAPM	14.72	0.77	9.68	0.34	8.92	0.42
C-CAPM	14.83	−0.16	10.07	1.06	8.57	0.60
FF3	15.45	−3.37	11.57	−1.81	11.03	−1.65

(c) the *corrected* candidate factor models continue to perform as well as the APT model (see columns (4) and (6)).

6 Microfoundations for Priced Unsystematic Risk

In the previous section, we have shown empirically the need to include an SDF-U component in the SDF. We could repeat our empirical analysis for other candidate factor models. However, our main conclusion is not going to change—the SDF-U component accounts for the lion's share of pricing of the cross-section of asset returns.[22] This is consistent with the empirical finding in Bryzgalova, Huang, and Julliard (2020), who undertake a large-scale search for a factor model that prices a cross-section of asset returns but find none. We have also shown that, given a candidate asset-pricing model, adding extra common risk factors to this model cannot proxy for the SDF-U component. The SDF-U component is a weak factor in the cross-section of asset returns, and therefore, its risk premia cannot be estimated accurately. We show this result explicitly in Appendix A.2.

At this point, one may wonder in what kind of economic environment unsystematic risk will be priced. Below we present an example of an equilibrium model that provides microfoundations for the notion that unsystematic risk is priced. Our example relies on the

[22]If by chance a candidate factor model contains a factor that is correlated with the SDF-U component, then one may find that the role of missing unsystematic risk is biased down, as we saw in the case of the corrected FF3 model.

well-known static model of Merton (1987). We show that the equilibrium asset returns and SDF in this model have the same functional forms as those we have for our APT model.

In Merton (1987), investors are aware about only a subset of the available securities. This type of "incomplete information" then implies that not only common risk factor but also shocks specific to each security are priced. While this kind of incomplete information may not be the only reason why the SDF-U component plays a dominant role in the pricing assets, it is an appealing argument given the large empirical evidence documenting that both retail (Polkovnichenko, 2005; Campbell, 2006; Goetzmann and Kumar, 2008) and institutional investors (Koijen and Yogo, 2019, table 2) invest in only a small number of available stocks.[23]

Below we summarize the main assumptions of the model and then analyze its equilibrium implications for the SDF. For details of the model, we refer the reader to Merton (1987).

Assume that there are N firms in the economy whose end-of-period cash flows are technologically given by[24]:

$$C_i = I_i \left[\mu_i + \eta_i Y + s_i \epsilon_i\right],$$

where, for simplicity, it is assumed that there is a single random common factor Y with $E(Y) = 0$ and $E(Y^2) = 1$, with $E(\varepsilon_i) = E(\varepsilon_i \mid \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_{i+1}, \ldots, \varepsilon_N, Y) = 0$, for $i = \{1, \ldots, N\}$, where ε_i are asset-specific shocks. Here, I_i is the amount of physical investment in firm i and μ_i, η_i, and s_i represent parameters of firm i 's production technology.

Let V_i denote the equilibrium value of firm i at the beginning of the period. If R_i is the equilibrium return per dollar from investing in firm i over the period, then $R_i = C_i/V_i$, and

$$R_i = \mathbb{E}(R_i) + b_i Y + \sigma_i \epsilon_i, \tag{12}$$

where b_i and σ_i are functions of the parameters of firm i's production technology.

There are two additional securities in the economy, both in zero net supply: a security that is risk-free with return R_f and the $(N + 1)$st risky security, which combines the risk-free security and a forward contract with cash settlements on the factor Y. Without loss

[23] Other mechanisms, such as market segmentation, institutional restrictions, transaction costs, illiquidity, imperfect divisibility of securities, may lead to the same observable behavior. That is, the modeling framework of Merton (1987) can be viewed as a reduced-form representation of all these microfoundations leading investors to invest in only a subset of available securities.

[24] We have made the following changes to the notation used in Merton (1987) so that it is consistent with the notation in our paper. We denote an investor's risk aversion by γ instead of δ; we denote the total number of assets by N instead of n; we index individual assets by i instead of k; and we denote the unsystematic risk premium by a_i instead of λ_k.

of generality, the forward price of the contract is assumed to be such that the standard deviation of the equilibrium returns on the security is unity. As a result, its return is

$$R_{N+1} = \mathbb{E}(R_{N+1}) + Y. \tag{13}$$

There is a sufficiently large number of investors with a sufficiently disperse distribution of wealth so that each investor acts as a price taker. Each investor is risk averse and exhibits mean-variance preferences over the end-of-period wealth:

$$U^j = E(R^j W^j) - \frac{\gamma^j}{2W^j}\,\mathrm{var}(R^j W^j),$$

where W^j denotes the value of the initial endowment of investor j evaluated at equilibrium prices, R^j denotes the return per dollar on investor j's optimal portfolio, and $\gamma^j > 0$ is the risk-aversion of investor j.

Investors differ in their information sets. The common part of investors' information sets includes: (i) the return on the risk-free security, (ii) the structure of securities' return given in expression (12), and (iii) the expected return and variance of the forward-contract security given in (13). However, different investors have knowledge about the parameters b_i and σ_i for different *subsets* of securities. The investors who know about security i agree on its characteristics. To simplify the analysis, investors are assumed to have identical risk aversion $\gamma^j = \gamma$ and identical initial wealth $W^j = W$.

The optimal solution of the each investor's portfolio problem allows us to obtain the aggregate demand for each security. Equating this to the aggregate supply for each security leads to the equilibrium expected return for asset i (Merton, 1987, eq. (16)):

$$\mathbb{E}(R_i) = R_f + \gamma b_i b + \gamma x_i \sigma_i^2 / q_i, \quad \text{for} \quad i = \{1, \ldots, N\}, \tag{14}$$

where x_i is the fraction of the market portfolio invested in asset i, $b = \sum_{i=1}^{N} x_i b_i$, and q_i is the fraction of investors who know about security i.

Denoting the return on the market as $R_m = \sum_{i=1}^{N} x_i R_i$, Merton (1987, eq. (24)) obtains the equilibrium expected excess return on the market:

$$\mathbb{E}(R_m) - R_f = \gamma\,\mathrm{var}(R_m) + a_m, \tag{15}$$

where $a_m = \sum_{i=1}^{N} x_i a_i$,

$$a_i = (1 - q_i)\Delta_i,$$
$$\Delta_i = \mathbb{E}(R_i) - R_f - b_i(\mathbb{E}(R_{N+1}) - R_f).$$

Equations (12) and (15) then imply

$$R_i - R_f = \beta_i(\mathbb{E}(R_m) - R_f) + a_i - \beta_i a_m + b_i Y + \sigma_i \epsilon_i, \tag{16}$$

where β_i denotes the covariance of the return on security i with the return on the market portfolio, divided by the variance of the market return. Equation (16) contains Y on the right-hand side. We substitute out Y by using the definition of the market portfolio return along with equations (12) and (14), to obtain

$$R_i - R_f = a_i - \beta_i a_m + \beta_i(\mathbb{E}(R_m) - R_f) + \frac{b_i}{b}(R_m - \mathbb{E}(R_m)) + \sigma_i \epsilon_i.$$

We now derive the SDF in this economy. In particular, we consider the case where the number of available assets is large, that is, $N \to \infty$. In this case, as we show in the proof for Proposition 7: (i) $\beta_i \to b_i/b$, (ii) $a_m \to 0$, and (iii) the market return is asymptotically orthogonal to all asset-specific shocks, ϵ_i. The proposition below then shows that this leads to equilibrium asset returns and an SDF that have the same functional form as those for the APT model, as specified in equations (1) and (6), respectively.

Proposition 7. *When the number of assets is large, $N \to \infty$, equilibrium asset returns are*

$$\begin{aligned} R_i - R_f &= a_i + \beta_i(\mathbb{E}(R_m) - R_f) + \beta_i(R_m - \mathbb{E}(R_m)) + \sigma_i \epsilon_i, \qquad (17) \\ &= a_i + \beta_i(R_m - R_f) + \sigma_i \epsilon_i, \end{aligned}$$

and the equilibrium SDF is

$$M = \underbrace{-\frac{1}{R_f}\sum_{i=1}^{N}\left(\frac{a_i}{\sigma_i}\epsilon_i\right)}_{M^a} + \underbrace{\frac{1}{R_f} - \frac{\mathbb{E}(R_m) - R_f}{R_f \cdot \mathrm{var}(R_m)}(R_m - \mathbb{E}(R_m))}_{M^\beta}. \tag{18}$$

The SDF in (18) consists of two components representing adjustments for risk: the first one for unsystematic risk, M^a, and the second for systematic risk, M^β, exactly as prescribed by the SDF under the APT. Note that a_i in (17) represents the compensation for unsystematic risk, because

$$a_i = -\operatorname{cov}\left(R_i - R_f, -\frac{1}{R_f}\sum_{i=1}^{N}\left(\frac{a_i}{\sigma_i}\epsilon_i\right)\right) \times R_f,$$

which coincides with the elements of the vector a in the APT. Naturally, the other part of the risk premium in (17), $\beta_i(\bar{R}_m - R_f)$, is compensation for exposure to systematic risk, represented by market risk because of the assumption of a single common factor:

$$\beta_i(\mathbb{E}(R_m) - R_f) = -\operatorname{cov}\left(R_i - R_f, -\frac{\mathbb{E}(R_m) - R_f}{R_f \cdot \mathrm{var}(R_m)}(R_m - \mathbb{E}(R_m))\right) \times R_f.$$

If all investors were fully informed about all N assets, that is, $q_i = 1$, then $a_i = 0$, and the results in (17) and (18) simplify to the expressions for security returns and the SDF under the CAPM, respectively. Thus, in Merton (1987), the no-arbitrage APT restriction is equivalent to stating that there are only a handful of assets that do not belong to the common information set of investors.

Thus, the above discussion shows that there are equilibrium models that support the notion that unsystematic risk is priced. Moreover, Proposition 7 shows that the result that unsystematic risk is priced is not limited to an economy with a finite number of assets.[25]

7 Conclusion

A fundamental challenge in finance is to price the cross section of assets. The main difficulty when pricing assets is to determine how exactly to adjust their returns for risk. The literature has proposed a large number of alternative factor models to accomplish this task. Despite the proliferation of systematic risk factors, referred to as the factor zoo (Cochrane, 2011), there is still a sizable pricing error, called alpha. This leads one to the question posed in the title of this paper: "What is missing in asset-pricing factor models?"

We challenge the conventional wisdom that only systematic sources of risk receive compensation in financial markets by showing that unsystematic risk is also compensated. That is, the pricing error alpha implied by factor models includes compensation not only for missing common risk factors but also for unsystematic risk. Theoretically, we demonstrate this key insight through the lens of the SDF under the assumptions of the APT and support it by demonstrating that an equilibrium model such as Merton (1987) is consistent with our insight. Empirically, we show that the component of the admissible SDF reflecting unsystematic risk, which is represented by a linear combination of unsystematic shocks, accounts for more than half (56%) of the variation in the admissible SDF. What is missing in virtually all factor models is compensation for this unsystematic risk.

[25] Often, relying on the expression in Merton (1987)

$$a_i = (1 - q_i)\Delta_i = \gamma \sigma_i^2 \left(\frac{1}{q_i} - 1 \right) \frac{V_i}{V_m},$$

where V_m is the value of the market, the empirical literature has implicitly assumed that a_i is cross-sectionally perfectly correlated with σ_i^2. However, the above expression shows that a_i depends not just on σ_i^2 but also on q_i and V_i.

The approach we develop in this paper applies widely—to reduced-form factor models, but also to partial- and general-equilibrium asset pricing models—without needing to identify which factors (strong or weak) are missing. In terms of estimation, the approach is designed and feasible for a large number of assets; in fact, its performance improves with the number of assets considered. Our novel insight, which establishes the importance of compensation for unsystematic risk, is crucial both for empiricists wanting to resolve the factor zoo and for theorists wishing to develop microfounded models of asset pricing.

A Appendix

A.1 Notation

The following notation is adopted in the manuscript and appendix. $\mathbb{E}(\cdot)$ denotes the expectation operator and $\bar{X}$ denotes the sample average of the variable X. Capital letters denote matrices, while lowercase letters denote scalars or vectors. A matrix $A > 0$ denotes a positive-definite matrix A and $\|A\|$ denotes the matrix norm $\|A\| = (\text{tr}(A'A))^{\frac{1}{2}}$, where $\text{tr}(\cdot)$ is the trace operator. For deterministic sequences $\{a_N\}$ and $\{b_N\}$, the notation $a_N = O(b_N)$ means that $|a_N|/b_N < \delta$, where $\delta > 0$ is some finite number, and $a_N = o(b_N)$ means that $|a_N|/b_N \to 0$ as $N \to \infty$. The notation $A_N = O(b_N)$ and $A_N = o(b_N)$ for a sequence of matrices $\{A_N\}$ of constant dimensions $a_1 \times a_2$ means that the previous statements hold for every element of A_N. Finally, the notation $a_N = O_p(b_N), a_N = o_p(b_N), A_N = O_p(b_N)$, and $A_N = o_p(b_N)$ means that the previous statements hold in probability.

A.2 Can one recover M^a_{t+1} using observable variables?

In this appendix, we show that M^a_{t+1} is a weak factor in the cross-section of asset returns. Therefore, even if it were possible to add to a candidate factor model an observable variable that was perfectly correlated with M^a_{t+1}, it would not lead to an admissible SDF. The risk premia associated with a weak factor cannot be estimated accurately (Anatolyev and Mikusheva, 2021), which leads to the problem of recovering the admissible SDF.

Proposition A1. *Under Assumptions 1 and 2, assume, without loss of generality, that* $K^{\text{can}} = 1$, *and there are no missing systematic risk factors, that is,* $K^{\text{mis}} = 0$, *implying*

$$R_{t+1} - R_{ft} = a + \beta^{\text{can}} f^{\text{can}}_{t+1} + e_{t+1}, \quad \textit{where}$$

$$T^{-1} \sum_{t=1}^{T} (e_t - \bar{e})(e_t - \bar{e})' \xrightarrow{p} V_e \quad \textit{and}$$

$$T^{-1} \sum_{t=1}^{T} (f^{\text{can}}_t - \bar{f}^{\text{can}})(f^{\text{can}}_t - \bar{f}^{\text{can}})' \xrightarrow{p} V^{\text{can}}_f > 0, \quad \textit{as } T \to \infty.$$

If an observable variable f^{idio}_{t+1} *is such that* $f^{\text{idio}}_{t+1} - \mathbb{E}(f^{\text{idio}}_{t+1}) = a' V_e^{-1} e_{t+1}$, *then* f^{idio}_{t+1} *must be a weak factor.*

Proof: Without loss of generality, given that the vector e_t is uncorrelated with f^{can}_t, consider that f^{can}_t and f^{idio}_t are uncorrelated.

In a time-series regression

$$R_{t+1} - R_{ft} = \beta_0 + \beta^{\text{can}} f^{\text{can}}_{t+1} + \beta^{\text{idio}} f^{\text{idio}}_{t+1} + u_t,$$

the parameter β^{idio} in population is

$$\begin{aligned}
\beta_i^{\text{idio}} &= \frac{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))(R_{it} - \mathbb{E}(R_{it}))\right]}{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))^2\right]} \\
&= \frac{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))(\beta_i^{\text{can}}(f_t^{\text{can}} - \mathbb{E}(f_t^{\text{can}})) + e_{it})\right]}{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))^2\right]} \\
&= \frac{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))\beta_i^{\text{can}}(f_t^{\text{can}} - \mathbb{E}(f_t^{\text{can}}))\right]}{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))^2\right]} + \frac{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))e_{it}\right]}{\mathbb{E}\left[(f_t^{\text{idio}} - \mathbb{E}(f_t^{\text{idio}}))^2\right]} \\
&= \frac{\mathbb{E}\left[(a'V_e^{-1}e_t)e_{it}\right]}{\mathbb{E}\left[(a'V_e^{-1}e_t)^2\right]} \\
&= \frac{a'V_e^{-1}V_e\iota_i}{a'V_e^{-1}V_eV_e^{-1}a} \\
&= \frac{a_i}{a'V_e^{-1}a} \\
&= \frac{a_i}{\delta_{\text{apt}}},
\end{aligned}$$

where ι_i is the ith row/column of the matrix I_N, and assume, without loss of generality, that the APT constraint binds, i.e., $a'V_e^{-1}a = \delta_{\text{apt}}$. Given that $\beta^{\text{idio}\prime}\beta^{\text{idio}} < \delta < \infty$ for any N, f_t^{idio} satisfies the definition of a weak factor (Lettau and Pelger, 2020). □

Estimation of the price of risk associated with the factor f_t^{idio} is problematic, as is the case for any weak factor; for a formal analysis see Anatolyev and Mikusheva (2021). Thus, the traditional two-pass regression approach does not permit one to estimate M^a_{t+1} accurately. In contrast, the method described in this paper explains how to construct an accurate estimate of M^a_{t+1}, which does not rely on the existence of an observable f^{idio}_{t+1} that is perfectly correlated with M^a_{t+1}.

Proposition A1 extends to the multivariate case, that is, when f^{idio}_{t+1} is a vector. It also extends to the case in which f^{idio}_{t+1} spans unsystematic risk *imperfectly*, that is,

$$f^{\text{idio}}_{t+1} - \mathbb{E}(f^{\text{idio}}_{t+1}) = \gamma a'V_e^{-1}e_{t+1} + \eta_{t+1},$$

where $\mathbb{E}(\eta_{t+1}) = 0, \text{corr}(a'V_e^{-1}e_{t+1}, \eta_{t+1}) = 0, \text{var}(\eta_{t+1}) = \sigma^2_\eta = \text{var}(f^{\text{idio}}_{t+1})(1 - \rho^2_{f^{\text{idio}},M^a})$.

A.3 Robustness to omitted sources of time-variation in risk premia

In this section, we show that our approach delivers an admissible SDF even if the true data-generating process for asset returns features time-variation in risk premia. One qualification applies: our method is not designed to capture conditional cross-sectional differences in excess returns but it is to capture unconditional differences in excess returns. And it does so, even if the misspecification of the candidate model is about omitted sources of time-variation in risk premia. In turn, we consider two different sources of time-variation in risk premia: (i) time-varying risk exposures and (ii) time-varying prices of risk.

A.3.1 A model with time-varying risk exposures

Without loss of generality, assume that the true model for asset returns is a conditional one-factor model without compensation for unsystematic risk ($a = 0$):

$$R_{t+1} - \mathbb{E}_t(R_{t+1}) = \beta_t f_{t+1} + e_{t+1},$$

where f_{t+1} is an observable factor with unconditional risk premium λ, $\mathbb{E}_t(f_{t+1}) = 0$, β_t is an $N \times 1$ vector of risk exposures of asset returns R_{t+1} to the factor f_{t+1}, and e_{t+1} is an $N \times 1$ vector of unsystematic shocks with a diagonal covariance matrix V_e.

Case 1: Common source of variation in risk exposures

Furthermore, assume that

$$\beta_t = \beta_0 + \beta_1 g_t,$$

where g_t is a common source of time-variation in assets' exposures β_t to the risk factor f_{t+1}. Without loss of generality, assume that $\mathbb{E}(g_t) = 0$. Given these assumptions, the true data generating process for asset returns is

$$R_{t+1} - R_{ft} = (\beta_0 + \beta_1 g_t)\lambda + (\beta_0 + \beta_1 g_t) f_{t+1} + e_{t+1}$$

If a candidate model is a one-factor model with the risk factor $f^{\text{can}}_{t+1} = f_{t+1}$ and constant risk exposures β^{can}, our method recognizes that there are two extra common sources of risk g_t and $g_t f_{t+1}$ that are omitted in the candidate model. The component $M^{\beta,\text{mis}}_{t+1}$ of the admissible SDF M_{t+1} captures the pricing impact of these omitted factors. Relatedly, our method recovers the pricing implications of these omitted factors via the component $\beta^{\text{mis}}\lambda^{\text{mis}}$ of the unconditional risk premia. As a result, a one-factor model with time-variation in risk premia driven by one common variable is observationally equivalent to a model with one observable and two unobservable common factors and constant risk premia. This equivalence holds with respect to capturing the unconditional risk premia.

Case 2: Asset-specific source of variation in risk exposures

Now, assume that

$$\beta_t = \beta_0 + \beta_1 \circ G_t,$$

where $G_t = (g_{1t},\ g_{2t},\ \cdots,\ g_{Nt})'$ is a vector of asset-specific sources of time-variation in risk exposures β_t to the risk factor f_{t+1} and a symbol $\circ$ denotes the Hadamard product. Without loss of generality, assume that $\mathbb{E}(g_{it}) = 0$ for each $1 \leq i \leq N$. Given these assumptions, the true data-generating process for asset returns is

$$R_{t+1} - R_{ft} = (\beta_0 + \beta_1 \circ G_t)\lambda + (\beta_0 + \beta_1 \circ G_t)f_{t+1} + e_{t+1}.$$

If a candidate model is a one-factor model with the risk factor $f_{t+1}^{\text{can}} = f_{t+1}$ and constant risk exposures β^{can}, then our method recognizes that there are N unsystematic shocks $\eta_{t+1} = G_t\lambda + G_t f_{t+1}$, which if being priced, are captured by the component M_{t+1}^a of the admissible SDF M_{t+1}. The pricing implications of these shocks are reflected in a vector a. Thus, a one-factor model with time-variation in risk premia driven by asset-specific variables is equivalent to a model with one observable common risk factor, in which unsystematic shocks are priced. This equivalence holds with respect to capturing the unconditional risk premia.

A.3.2 A model with time-varying prices of risk

Now assume that the true data-generating process for asset returns features constant risk exposures but time-varying prices of risk. Without loss of generality, we assume that there is only one risk factor f_{t+1} and that unsystematic risk is not priced (i.e., $a = 0$):

$$R_{t+1} - R_{ft} = \beta(\lambda_0 + \lambda_1 g_t) + \beta f_{t+1} + e_{t+1},$$

where $\mathbb{E}_t(f_{t+1}) = 0$, $\mathbb{E}(g_t) = 0$.

If a candidate model is a one-factor model with the risk factor $f_{t+1}^{\text{can}} = f_{t+1}$ and constant prices of and exposures to risk, then our method recognizes that there is an extra common source of risk g_t that is omitted in the candidate model. The component $M_{t+1}^{\beta,\text{mis}}$ of the admissible SDF M_{t+1} captures the pricing impact of this omitted common source of variation in asset returns. Relatedly, our method recovers the pricing implications of this extra factor via the component $\beta^{\text{mis}}\lambda^{\text{mis}}$ of the unconditional risk premia. Thus, a one-factor model with time-varying price of risk driven by one variable is equivalent to a model with one observable and one unobservable common risk factor. This equivalence holds with respect to capturing the unconditional risk premia.

A.4 Assumptions

In this section, we provide a set of assumptions that we use in the lemmas and propositions of Sections A.5 and A.6, respectively. Whereas the assumptions about β^{can} are unique (Assumptions A1 and A2), the assumptions about β^{mis} differ, depending on whether missing factors f_{t+1}^{mis} are strong (Assumptions A3 and A4) or weak (Assumptions A5 and A6).

Assumption A1 (Strong candidate factors)**.** *We assume that a candidate model contains only strong factors* f_t^{can}*, that is,* $\frac{1}{N}\beta^{\text{can}\prime}V_e{}^{-1}\beta^{\text{can}} \to D$ *, where* $D > 0$ *is some* $K^{\text{can}} \times K^{\text{can}}$ *positive-definite matrix.*

Assumption A2. *We assume that* $\beta^{\text{can}\prime}V_e{}^{-1}a = o(N^{\frac{1}{2}})$*.*[26]

Assumption A3 (Strong missing factors)**.** *We assume that a candidate factor model misses only strong factors* f_t^{mis}*, that is,* $\frac{1}{N}\beta^{\text{mis}\prime}V_e{}^{-1}\beta^{\text{mis}} \to E$*, where* $E > 0$ *is some* $K^{\text{mis}} \times K^{\text{mis}}$ *positive-definite matrix.*

Assumption A4. *We assume that* $\beta^{\text{mis}\prime}V_e{}^{-1}a = o(N^{\frac{1}{2}})$*.*[27]

Assumption A5 (Weak missing factors)**.** *We assume that a candidate factor model misses only weak factors* f_t^{mis}*, that is,* $\beta^{\text{mis}\prime}V_e{}^{-1}\beta^{\text{mis}} \to E$*, where* $E > 0$ *is some* $K^{\text{mis}} \times K^{\text{mis}}$ *positive-definite matrix.*

Assumption A6. *We assume that* $\beta^{\text{mis}\prime}V_e{}^{-1}\beta^{\text{can}} = o(N^{\frac{1}{2}})$*.*[28]

Note that Assumptions 1 and 2 imply that $a'V_e{}^{-1}a = O(1)$.

A.5 Lemmas

We now provide a set of lemmas that will be useful for proving our propositions.

[26]Assumption A1 and asymptotic no arbitrage, along with the Cauchy-Schwarz inequality, imply that $\beta^{\text{can}\prime}V_e{}^{-1}a = O(N^{\frac{1}{2}})$, but we need a slightly faster convergence rate.

[27]Assumption A3 and asymptotic no arbitrage, along with the Cauchy-Schwarz inequality, imply that $\beta^{\text{mis}\prime}V_e{}^{-1}a = O(N^{\frac{1}{2}})$, but we need a slightly faster convergence rate.

[28]Assumptions A1 and A5, along with the Cauchy-Schwarz inequality, imply that $\beta^{\text{mis}\prime}V_e{}^{-1}\beta^{\text{can}} = O(N^{\frac{1}{2}})$, but we need a slightly faster rate.

Lemma A.1. *For a random vector $z \sim N(\mu_z, \Sigma_z)$, and any constant vector d:*

$$\mathbb{E}(ze^{d'z}) = \mu^* e^{\frac{1}{2}(\mu^{*\prime}\Sigma_z^{-1}\mu^* - \mu_z'\Sigma_z^{-1}\mu_z)}, \quad \textit{where} \;\; \mu^* = (\mu_z + \Sigma_z d).$$

Proof: Denote by n_z the dimension of the vector z. Then, using the definition of the expectation,

$$\mathbb{E}(ze^{d'z}) = \frac{1}{(\sqrt{2\pi})^{n_z}|\Sigma_z|^{\frac{1}{2}}} \int_{-\infty}^{\infty} ze^{d'z} e^{-\frac{1}{2}(z-\mu_z)'\Sigma_z^{-1}(z-\mu_z)} dz.$$

Note that

$$\begin{aligned} e^{d'z} e^{-\frac{1}{2}(z-\mu_z)'\Sigma_z^{-1}(z-\mu_z)} &= e^{d'z - \frac{1}{2}z'\Sigma_z^{-1}z - \frac{1}{2}\mu_z'\Sigma_z^{-1}\mu_z + \mu_z'\Sigma_z^{-1}z} \\ &= e^{-\frac{1}{2}z'\Sigma_z^{-1}z - \frac{1}{2}\mu_z'\Sigma_z^{-1}\mu_z + (\Sigma_z d + \mu_z)'\Sigma_z^{-1}z} \\ &= e^{-\frac{1}{2}z'\Sigma_z^{-1}z - \frac{1}{2}\mu_z'\Sigma_z^{-1}\mu_z + \mu^{*\prime}\Sigma_z^{-1}z} \\ &= e^{-\frac{1}{2}\mu_z'\Sigma_z^{-1}\mu_z + \frac{1}{2}\mu^{*\prime}\Sigma_z^{-1}\mu^*} e^{-\frac{1}{2}z'\Sigma_z^{-1}z + \mu^{*\prime}\Sigma_z^{-1}z - \frac{1}{2}\mu^{*\prime}\Sigma_z^{-1}\mu^*} \\ &= e^{-\frac{1}{2}\mu_z'\Sigma_z^{-1}\mu_z + \frac{1}{2}\mu^{*\prime}\Sigma_z^{-1}\mu^*} e^{-\frac{1}{2}(z-\mu^*)'\Sigma_z^{-1}(z-\mu^*)}, \end{aligned}$$

implying that

$$\mathbb{E}(ze^{d'z}) = e^{-\frac{1}{2}\mu_z'\Sigma_z^{-1}\mu_z + \frac{1}{2}\mu^{*\prime}\Sigma_z^{-1}\mu^*} \left(\frac{1}{(\sqrt{2\pi})^{n_z}|\Sigma_z|^{\frac{1}{2}}} \int_{-\infty}^{\infty} ze^{-\frac{1}{2}(z-\mu^*)'\Sigma_z^{-1}(z-\mu^*)} dz \right).$$

Note that by the definition of the expectation, the component in brackets is equal to μ^*. □

Lemma A.2. *Under Assumptions A1 and A3:*

$$\beta^{\text{mis}\,\prime} V_e^{-1} \beta^{\text{can}} = O(N).$$

Proof: Applying Cauchy-Schwarz inequality for matrices we get

$$0 \le ||\beta^{\text{mis}\,\prime} V_e^{-1} \beta^{\text{can}}|| \le ||\beta^{\text{mis}\,\prime} V_e^{-1} \beta^{\text{mis}}||^{\frac{1}{2}} \cdot ||\beta^{\text{can}\,\prime} V_e^{-1} \beta^{\text{can}}||^{\frac{1}{2}} = O(N). \qquad \square$$

Lemma A.3. *Under Assumptions A1 and A3:*

$$\beta^{\text{can}\,\prime} V_\varepsilon^{-1} \beta^{\text{can}} = O(N).$$

Proof: Applying the Sherman-Morrison-Woodbury formula to V_ε^{-1} and using Lemma A.2 gives

$$\begin{aligned} \beta^{\text{can}\,\prime} V_\varepsilon^{-1} \beta^{\text{can}} &= \beta^{\text{can}\,\prime} V_e^{-1} \beta^{\text{can}} - \beta^{\text{can}\,\prime} V_e^{-1} \beta^{\text{mis}} (V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\,\prime} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}\,\prime} V_e^{-1} \beta^{\text{can}} \\ &= O(N) + O(N) \cdot [O(1) + O(N)]^{-1} \cdot O(N) \\ &= O(N). \end{aligned} \qquad \square$$

Lemma A.4. *Under Assumptions A1, A5 and A6:*

$$\beta^{\text{can}\prime}{V_\varepsilon}^{-1}\beta^{\text{can}} = O(N).$$

Proof: Follow the same steps as those in the proof of Lemma A.3. □

Lemma A.5. *Under Assumption A3:*

$$\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}\beta^{\text{mis}} = O(1) \quad \textit{and} \quad \beta^{\text{mis}\prime}{V_\varepsilon}^{-1}\beta^{\text{mis}} \to V_{f^{\text{mis}}}^{-1} \quad \textit{as} \quad N \to \infty.$$

Proof: Applying the Sherman-Morrison-Woodbury formula to ${V_\varepsilon}^{-1}$ leads to

$$\begin{aligned}
\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}\beta^{\text{mis}} &= \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}} - \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}} \\
&= V_{f^{\text{mis}}}^{-1}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}} \\
&= V_{f^{\text{mis}}}^{-1} \cdot [O(1) + O(N)]^{-1} \cdot O(N) \\
&\to V_{f^{\text{mis}}}^{-1}
\end{aligned}$$

□

Lemma A.6. *Under Assumptions A1 and A3:*

$$\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}\beta^{\text{can}} = O(1).$$

Proof: Applying the Sherman-Morrison-Woodbury formula to ${V_\varepsilon}^{-1}$ and using Lemma A.2 leads to

$$\begin{aligned}
\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}\beta^{\text{can}} &= \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{can}} - \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{can}} \\
&= V_{f^{\text{mis}}}^{-1}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{can}} \\
&= O(1) \cdot [O(1) + O(N)]^{-1} \cdot O(N) \\
&= O(1).
\end{aligned}$$

□

Lemma A.7. *Under Assumptions A1, A5 and A6:*

$$\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}\beta^{\text{can}} = O(N^{\frac{1}{2}}).$$

Proof: Follow the same steps as those in the proof of Lemma A.6. □

Lemma A.8. *Under Assumptions A2, A3 and A4:*

$$a'{V_\varepsilon}^{-1}a - a'{V_e}^{-1}a \to 0 \quad \textit{as} \quad N \to \infty.$$

Proof: Applying the Sherman-Morrison-Woodbury formula to ${V_\varepsilon}^{-1}$ leads to

$$
\begin{aligned}
a'{V_\varepsilon}^{-1}a &= a'{V_e}^{-1}a - a'{V_e}^{-1}\beta^{\text{mis}}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}a \\
&= a'{V_e}^{-1}a + o(N^{\frac{1}{2}}) \cdot [O(1) + O(N)]^{-1} \cdot o(N^{\frac{1}{2}}) \\
&= a'{V_e}^{-1}a + o(1)
\end{aligned}
$$

where $a'{V_e}^{-1}a = O(1)$. □

Lemma A.9. *Under Assumptions A1, A2, A3 and A4:*

$$\beta^{\text{can}\prime}{V_\varepsilon}^{-1}a = o(N^{\frac{1}{2}}).$$

Proof: Applying the Sherman-Morrison-Woodbury formula to ${V_\varepsilon}^{-1}$ and using Lemma A.2 leads to

$$
\begin{aligned}
\beta^{\text{can}\prime}{V_\varepsilon}^{-1}a &= \beta^{\text{can}\prime}{V_e}^{-1}a - \beta^{\text{can}\prime}{V_e}^{-1}\beta^{\text{mis}}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}a \\
&= o(N^{\frac{1}{2}}) + O(1) \cdot [O(1) + O(N)]^{-1} \cdot o(N^{\frac{1}{2}}) \\
&= o(N^{\frac{1}{2}}).
\end{aligned}
$$

□

Lemma A.10. *Under Assumptions A1, A2, A5 and A6:*

$$\beta^{\text{can}\prime}{V_\varepsilon}^{-1}a = o(N^{\frac{1}{2}}).$$

Proof: Follow the same steps as those in the proof of Lemma A.9. □

Lemma A.11. *Under Assumptions A3 and A4:*

$$\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}a = o(N^{-\frac{1}{2}}).$$

Proof: Applying the Sherman-Morrison-Woodbury formula to ${V_\varepsilon}^{-1}$ leads to

$$
\begin{aligned}
\beta^{\text{mis}\prime}{V_\varepsilon}^{-1}a &= \beta^{\text{mis}\prime}{V_e}^{-1}a - \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}a \\
&= V_{f^{\text{mis}}}^{-1}(V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime}{V_e}^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\prime}{V_e}^{-1}a \\
&= O(1) \cdot [O(1) + O(N)]^{-1} \cdot o(N^{\frac{1}{2}}) \\
&= o(N^{-\frac{1}{2}}).
\end{aligned}
$$

□

Lemma A.12. *Let e be a $N \times 1$ random vector with zero mean and covariance matrix V_e. Under Assumptions A1 and A3:*

$$\beta^{\text{can}\prime}{V_\varepsilon}^{-1}e = O_p(N^{\frac{1}{2}}).$$

Proof: For any random variable X with a finite second moment, we have that $X = O_p((\mathbb{E}(X^2))^{\frac{1}{2}})$. If $X = \beta^{\text{can}\prime} V_e^{-1} e$, then

$$\mathbb{E}(\beta^{\text{can}\prime} V_e^{-1} e e' V_e^{-1} \beta^{\text{can}}) = \beta^{\text{can}\prime} V_e^{-1} \beta^{\text{can}} = O(N),$$

and therefore, $\beta^{\text{can}\prime} V_e^{-1} e = O_p(N^{\frac{1}{2}})$. Similarly, we can show that $\beta^{\text{mis}\prime} V_e^{-1} e = O_p(N^{\frac{1}{2}})$. Applying the Sherman-Morrison-Woodbury formula to V_ε^{-1} and using Lemma A.2 gives

$$\begin{aligned}
\beta^{\text{can}\prime} V_\varepsilon^{-1} e &= \beta^{\text{can}\prime} V_e^{-1} e - \beta^{\text{can}\prime} V_e^{-1} \beta^{\text{mis}} (V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}\prime} V_e^{-1} e \\
&= O_p(N^{\frac{1}{2}}) + O(N) \cdot [O(1) + O(N)]^{-1} \cdot O_p(N^{\frac{1}{2}}) \\
&= O_p(N^{\frac{1}{2}}). \qquad \square
\end{aligned}$$

Lemma A.13. *Under Assumption A3:*

$$\beta^{\text{mis}\prime} V_\varepsilon^{-1} e = O_p(N^{-\frac{1}{2}}).$$

Proof: From the proof of Lemma A.12, $\beta^{\text{mis}\prime} V_e^{-1} e = O_p(N^{\frac{1}{2}})$.
Applying the Sherman-Morrison-Woodbury formula to V_ε^{-1} and using Lemma A.2 gives

$$\begin{aligned}
\beta^{\text{mis}\prime} V_\varepsilon^{-1} e &= \beta^{\text{mis}\prime} V_e^{-1} e - \beta^{\text{mis}\prime} V_e^{-1} \beta^{\text{mis}} (V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}\prime} V_e^{-1} e \\
&= V_{f^{\text{mis}}}^{-1} (V_{f^{\text{mis}}}^{-1} + \beta^{\text{mis}\prime} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}\prime} V_e^{-1} e \\
&= O(1) \cdot [O(1) + O(N)]^{-1} \cdot O_p(N^{\frac{1}{2}}) \\
&= O_p(N^{-\frac{1}{2}}). \qquad \square
\end{aligned}$$

A.6 Proofs of Propositions

In this section, we provide the proofs for the propositions in the manuscript.

Proof of Proposition 2

Note that $R_{t+1} - R_{ft} 1_N - \beta^{\text{can}} \big(f_{t+1}^{\text{can}} + \lambda^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}) \big) = \alpha + \varepsilon_{t+1}$. By Chamberlain and Rothschild (1983, thm. 4) the covariance matrix of ε_{t+1} has an *approximate* factor structure, and satisfies

$$\text{var}(\varepsilon) = V_\varepsilon = \beta^{\text{mis}} V_{f^{\text{mis}}} \beta^{\text{mis}\prime} + V_e,$$

where $V_e > 0$ with uniformly bounded eigenvalues, and by Chamberlain and Rothschild (1983, cor. 2) there exists a vector λ^{mis} such that $(\alpha - \beta^{\text{mis}} \lambda^{\text{mis}})' V_e^{-1} (\alpha - \beta^{\text{mis}} \lambda^{\text{mis}})$ is bounded for any N, where β^{mis} is the $N \times K^{\text{mis}}$ matrix consisting of the K^{mis} dominant eigenvectors of V_ε (the eigenvectors associated with the largest K^{mis} eigenvalues of the matrix V_ε), each multiplied by the square-root of the corresponding eigenvalues. We set $a = \alpha - \beta^{\text{mis}}\lambda^{\text{mis}}$.

We need to show that $\alpha' V_\varepsilon^{-1}\alpha = O(1)$ under Assumptions 1, 2, A1, and A3. We use the definition of α to express $\alpha' V_\varepsilon^{-1}\alpha$ as

$$\begin{aligned}\alpha' V_\varepsilon^{-1}\alpha &= (a + \beta^{\text{mis}}\lambda^{\text{mis}})' V_\varepsilon^{-1}(a + \beta^{\text{mis}}\lambda^{\text{mis}}) \\ &= a' V_\varepsilon^{-1} a + \lambda^{\text{mis}\prime}\beta^{\text{mis}\prime} V_\varepsilon^{-1}\beta^{\text{mis}}\lambda^{\text{mis}} + 2a' V_\epsilon^{-1}\beta^{\text{mis}}\lambda^{\text{mis}}.\end{aligned}$$

The result then follows from Lemmas A.5, A.8, and A.11. □

Note that Proposition 2 assumes the presence of at least one omitted systematic risk factor, that is, $K^{\text{mis}} > 0$. If instead $K^{\text{mis}} = 0$, that is, all eigenvalues of V_ε are bounded, then the data-generating process of asset returns with K^{can} factors given in expression (3) satisfies the assumptions of the classical APT.

Proof of Proposition 3

We use a guess-and-verify method to derive the SDF. We guess that the SDF has the following functional form

$$M_{t+1} = \mathbb{E}(M_{t+1}) + b^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c' e_{t+1},$$

where b^{can} is a $K^{\text{can}} \times 1$ vector , b^{mis} is a $K^{\text{mis}} \times 1$ vector, and c is an $N \times 1$ vector. We identify the unknown vectors b^{can}, b^{mis}, and c by using the Law of One Price. Specifically, because we assume the existence of the risk-free asset, to determine the mean of the SDF we use the condition:

$$\mathbb{E}(M_{t+1}) = \frac{1}{R_f}.$$

Next, because λ^{can} represents a vector of prices of risk of f_{t+1}^{can}, we have that

$$-\operatorname{cov}(M_{t+1}, f_{t+1}^{\text{can}}) \cdot R_f = \lambda^{\text{can}\prime}.$$

These K^{can} conditions identify b^{can}:

$$b^{\text{can}\prime} = -\frac{1}{R_f}\lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1}.$$

Similarly, λ^{mis} is the price of risk associated with factors f_{t+1}^{mis}, or equivalently,

$$-\operatorname{cov}(M_{t+1}, f_{t+1}^{\text{mis}}) \cdot R_f = \lambda^{\text{mis}\,\prime}.$$

These K^{mis} conditions identify b^{mis}:

$$b^{\text{mis}\,\prime} = -\frac{1}{R_f}\lambda^{\text{mis}\,\prime} V_{f^{\text{mis}}}^{-1}.$$

Finally, it must be the case that the SDF prices the N assets

$$\mathbb{E}\big(M_{t+1}(R_{t+1} - R_{ft}1_N)\big) = 0_N.$$

These N equations identify c:

$$c' = -\frac{1}{R_f} a' V_e^{-1}.$$

Taken together

$$M_{t+1} = M_{t+1}^{\beta,\text{can}} + M_{t+1}^{\beta,\text{mis}} + M_{t+1}^{a},$$

where

$$M_{t+1}^{\beta,\text{can}} = \frac{1}{R_f} - \frac{1}{R_f}\lambda^{\text{can}\,\prime} V_{f^{\text{can}}}^{-1}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})),$$

$$M_{t+1}^{\beta,\text{mis}} = -\frac{1}{R_f}\lambda^{\text{mis}\,\prime} V_{f^{\text{mis}}}^{-1}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})),$$

$$M_{t+1}^{a} = -\frac{1}{R_f} a' V_e^{-1} e_{t+1}.$$

Pairwise uncorrelatedness (and independence by Gaussianity) of f_t^{can}, f_t^{mis}, and e_t implies that the pairwise covariances between $M_{t+1}^{\beta,\text{can}}$, $M_{t+1}^{\beta,\text{mis}}$, and M_{t+1}^{a} are all zero. □

Proof of Proposition 4

We use a guess and verify method to derive a nonnegative SDF. We guess that the SDF has the following functional form:

$$M_{\exp,t+1} = \exp\left[\mu_+ + b_+^{\text{can}\,\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b_+^{\text{mis}\,\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c_+' e_{t+1}\right],$$

with unknown vectors b_+^{can}, b_+^{mis}, and c_+, as well as an unknown scalar μ_+.

To identify the unknowns and verify our guess we use the following $K^{\text{can}} + K^{\text{mis}} + N + 1$ equations, which are implications of the Law of One Price:

$$-\operatorname{cov}(M_{\exp,t+1}, f_{t+1}^{\text{can}}) \cdot R_f = \lambda^{\text{can}},$$

$$
\begin{aligned}
-\text{cov}(M_{\exp,t+1}, f_{t+1}^{\text{mis}}) \cdot R_f &= \lambda^{\text{mis}}, \\
\mathbb{E}(M_{\exp,t+1}(R_{t+1} - R_{ft}1_N)) &= 0, \\
\mathbb{E}(M_{\exp,t+1}) &= R_f^{-1},
\end{aligned}
$$

The first K^{can} equations imply that

$$-\mathbb{E}(M_{\exp,t+1}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}))) = \mathbb{E}(M_{\exp,t+1}) \cdot \lambda^{\text{can}},$$

which, along with Lemma A.1, give:

$$b_+^{\text{can}} = -V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}.$$

Similarly, the next K^{mis} equations imply that

$$-\mathbb{E}(M_{\exp,t+1}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}))) = \mathbb{E}(M_{\exp,t+1}) \cdot \lambda^{\text{mis}},$$

which, along with Lemma A.1, lead to:

$$b_+^{\text{mis}} = -V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}}.$$

Using the next N equations and Lemma A.1 gives:

$$
\begin{aligned}
0_N &= \mathbb{E}(M_{\exp,t+1}(R_{t+1} - R_{ft}1_N)) \\
&= \mathbb{E}(M_{\exp,t+1}(a + \beta^{\text{mis}}\lambda^{\text{mis}} + \beta^{\text{can}}\lambda^{\text{can}} + \beta^{\text{can}}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) \\
&\quad + \beta^{\text{mis}}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + e_{t+1})) \\
&= (a + \beta^{\text{mis}}\lambda^{\text{mis}} + \beta^{\text{can}}\lambda^{\text{can}})\mathbb{E}(M_{\exp,t+1}) + \mathbb{E}(M_{\exp,t+1}e_{t+1}) \\
&\quad + \mathbb{E}(M_{\exp,t+1}\beta^{\text{can}}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}))) + \mathbb{E}(M_{\exp,t+1}\beta^{\text{mis}}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}))) \\
&= (a + \beta^{\text{mis}}\lambda^{\text{mis}} + \beta^{\text{can}}\lambda^{\text{can}})\mathbb{E}(M_{\exp,t+1}) + V_e c_+ \mathbb{E}(M_{\exp,t+1}) \\
&\quad - \beta^{\text{can}}\lambda^{\text{can}}\mathbb{E}(M_{\exp,t+1}) - \beta^{\text{mis}}\lambda^{\text{mis}}\mathbb{E}(M_{\exp,t+1}) \\
&= (a + V_e c_+)\mathbb{E}(M_{\exp,t+1}).
\end{aligned}
$$

As a result,

$$c_+ = -V_e^{-1} a.$$

Finally, the last identifying condition implies

$$
\begin{aligned}
\mathbb{E}(M_{\exp,t+1}) &= R_f^{-1} \\
&= \mathbb{E}(\exp[\mu_+ + b_+^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b_+^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c_+' e_{t+1}]) \\
&= \exp[\mu_+ + b_+^{\text{can}\prime} V_{f^{\text{can}}} b_+^{\text{can}}/2 + b_+^{\text{mis}\prime} V_{f^{\text{mis}}} b_+^{\text{mis}}/2 + c_+' V_e c_+/2].
\end{aligned}
$$

Thus,

$$\exp(\mu_+) = R_f^{-1} \cdot \exp\big[-\lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}/2 - \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}}/2 - a' V_e^{-1} a/2\big].$$

Collecting all these results, we obtain

$$M_{\exp,t+1} = M_{\exp,t+1}^{\beta,\text{can}} \cdot M_{\exp,t+1}^{\beta,\text{mis}} \cdot M_{\exp,t+1}^{a},$$

where

$$\begin{aligned}
M_{\exp,t+1}^{\beta,\text{can}} &= R_f^{-1} \cdot \exp\,[-\lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} (f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) - \lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}/2)],\\
M_{\exp,t+1}^{\beta,\text{mis}} &= \exp\,[-\lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} (f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) - \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}}/2)],\\
M_{\exp,t+1}^{a} &= \exp\,[-a' V_e^{-1} e_{t+1} - a' V_e^{-1} a/2].
\end{aligned}$$

Pairwise uncorrelatedness (and independence by Gaussianity) of f_t^{can}, f_t^{mis}, and e_t implies that the pairwise covariances between $M_{\exp,t+1}^{\beta,\text{can}}$, $M_{\exp,t+1}^{\beta,\text{mis}}$, and $M_{\exp,t+1}^{a}$ are all zero. □

Proof of Proposition 5

We start by analyzing the exponent of $\hat{M}_{\exp,t+1}^{a}$. Our goal is to show that

$$-a' V_R^{-1} (R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2} a' V_R^{-1} a = -a' V_e^{-1} e_{t+1} - \frac{1}{2} a' V_e^{-1} a + o_p(1).$$

First, we note that

$$\begin{aligned}
-a' V_R^{-1} (R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2} a' V_R^{-1} a = & -a' V_R^{-1} \beta^{\text{can}} (f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}))\\
& - a' V_R^{-1} \beta^{\text{mis}} (f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}))\\
& - a' V_R^{-1} e_{t+1}\\
& - \frac{1}{2} a' V_R^{-1} a.
\end{aligned}$$

We analyze the four right-hand-side terms one-by-one. Applying the Sherman-Morrison-Woodbury formula to V_R^{-1} and using Lemmas A.3, A.6, A.8, A.9, A.11, and A.12 gives

$$\begin{aligned}
a' V_R^{-1} \beta^{\text{can}} &= a' V_\varepsilon^{-1} \beta^{\text{can}} - a' V_\varepsilon^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon^{-1} \beta^{\text{can}}\\
&= a' V_\varepsilon^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon^{-1} \beta^{\text{can}})^{-1} V_{f^{\text{can}}}^{-1}\\
&= o(N^{\frac{1}{2}}) \cdot [O(1) + O(N)]^{-1} \cdot O(1)\\
&= o(N^{-1/2}),\\
a' V_R^{-1} \beta^{\text{mis}} &= a' V_\varepsilon^{-1} \beta^{\text{mis}} - a' V_\varepsilon^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon^{-1} \beta^{\text{mis}}
\end{aligned}$$

$$= o(N^{-\frac{1}{2}}) + o(N^{\frac{1}{2}}) \cdot [O(1) + O(N)]^{-1} \cdot O(1) =$$

$$= o(N^{-1/2}),$$

$$a' V_R^{-1} e_{t+1} = a' V_\varepsilon{}^{-1} e_{t+1} - a' V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon{}^{-1} e_{t+1}$$

$$= a' V_\varepsilon{}^{-1} e_{t+1} + o(N^{\frac{1}{2}}) \cdot [O(1) + O(N)]^{-1} \cdot O_p(N^{\frac{1}{2}})$$

$$= a' V_\varepsilon{}^{-1} e_{t+1} + o_p(1),$$

$$a' V_R^{-1} a = a' V_\varepsilon{}^{-1} a - a' V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon{}^{-1} a$$

$$= (a' V_e{}^{-1} a + o(1)) + o(N^{\frac{1}{2}}) \cdot [O(1) + O(N)]^{-1} \cdot o(N^{\frac{1}{2}})$$

$$= a' V_e^{-1} a + o(1).$$

We use these results to show that

$$- a' V_R^{-1} (R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2} a' V_R^{-1} a = -a' V_e{}^{-1} e_{t+1} - \frac{1}{2} a' V_e{}^{-1} a + o_p(1),$$

and subsequently obtain

$$\hat{M}^a_{\exp,t+1} - M^a_{\exp,t+1} \xrightarrow{p} 0 \qquad \text{as} \quad N \to \infty.$$

Next, we analyze the exponent of $\hat{M}^{\beta,\text{mis}}_{\exp,t+1}$:

$$- (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} (R_{t+1} - \mathbb{E}(R_{t+1})) - \frac{1}{2} (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} \beta^{\text{mis}} \lambda^{\text{mis}}$$

$$= -(\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} \beta^{\text{can}} (f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1}))$$

$$- (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} \beta^{\text{mis}} (f^{\text{mis}}_{t+1} - \mathbb{E}(f^{\text{mis}}_{t+1}))$$

$$- (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} e_{t+1}$$

$$- \frac{1}{2} (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} \beta^{\text{mis}} \lambda^{\text{mis}}.$$

We apply the Sherman-Morrison-Woodbury formula and Lemmas A.3, A.5, A.6, A.12, and A.13 to each of the four terms above.

$$\beta^{\text{mis}\prime} V_R^{-1} \beta^{\text{can}} = \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}} - \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}}$$

$$= \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} V_{f^{\text{can}}}^{-1}$$

$$= O(1) \cdot [O(1) + O(N)]^{-1} \cdot O(1)$$

$$= O(N^{-1}),$$

$$\beta^{\text{mis}\prime} V_R^{-1} \beta^{\text{mis}} = \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{mis}} - \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{mis}}$$

$$
\begin{aligned}
&= (V_{f^{\text{mis}}}^{-1} + o(1)) + O(1) \cdot [O(1) + O(N)]^{-1} \cdot O(1) \\
&= V_{f^{\text{mis}}}^{-1} + o(1),
\end{aligned}
$$

$$
\begin{aligned}
\beta^{\text{mis}\prime} V_R^{-1} e_{t+1} &= \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} e_{t+1} - \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}\prime} V_\varepsilon{}^{-1} e_{t+1} \\
&= O_p(N^{-\frac{1}{2}}) + O(1) \cdot [O(1) + O(N)]^{-1} \cdot O_p(N^{\frac{1}{2}}) \\
&= O_p(N^{-\frac{1}{2}}),
\end{aligned}
$$

$$
\begin{aligned}
\beta^{\text{mis}\prime} V_R^{-1} \beta^{\text{mis}} &= \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{mis}} - \beta^{\text{mis}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}} (V_{f^{\text{can}}}^{-1} + \beta^{\text{can}\prime} V_\varepsilon{}^{-1} \beta^{\text{can}})^{-1} \beta^{\text{can}} V_\varepsilon{}^{-1} \beta^{\text{mis}} \\
&= (V_{f^{\text{mis}}}^{-1} + o(1)) + O(1) \cdot [O(1) + O(N)]^{-1} O(1) \\
&= V_{f^{\text{mis}}}^{-1} + o(1).
\end{aligned}
$$

We use these results to show that

$$
\begin{aligned}
&- (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} (R_{t+1} - \mathbb{E}(R_{t+1})) - \frac{1}{2} (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} \beta^{\text{mis}} \lambda^{\text{mis}} \\
&\quad = -\lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} (f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) - \frac{1}{2} \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} + o_p(1)
\end{aligned}
$$

and subsequently obtain

$$
\hat{M}_{\exp,t+1}^{\beta,\text{mis}} - M_{\exp,t+1}^{\beta,\text{mis}} \xrightarrow{p} 0 \quad \text{as} \quad N \to \infty.
$$

Pairwise uncorrelatedness (and independence by Gaussianity) of $f_t^{\text{can}}, f_t^{\text{mis}}$, and e_t implies that the pairwise covariances between $M_{\exp,t+1}^{\beta,\text{can}}, M_{\exp,t+1}^{\beta,\text{mis}}$, and $M_{\exp,t+1}^{a}$ are all zero. The same remains true for the projected versions, $\hat{M}_{\exp,t+1}^{a}$ and $\hat{M}_{\exp,t+1}^{\beta,\text{mis}}$, thanks to the asymptotic-in-N equivalences proven above. □

Proof of Proposition 6

Define g_t as the vector of some observable variables representing missing factors in the candidate model and collect its values for each t in a matrix $G = (g_1 \cdots g_T)'$. Likewise, define $F^{\text{mis}} = (f_1^{\text{mis}} \cdots f_T^{\text{mis}})'$. For each t, collect the values of the systematic component $\log(\hat{M}_{\exp,t+1}^{\beta,\text{mis}})$ of the admissible SDF in a vector $\log(\hat{M}_{\exp}^{\beta,\text{mis}}) = (\log(\hat{M}_{\exp,1}^{\beta,\text{mis}}) \cdots \log(\hat{M}_{\exp,T}^{\beta,\text{mis}}))'$. Then, the R^2 of the regression of $\log(\hat{M}_{\exp,t}^{\beta,\text{mis}})$ on an intercept and the vector g_t,

$$
\log(\hat{M}_{\exp,t}^{\beta,\text{mis}}) = \gamma_0 + \gamma_1' g_t + u_t,
$$

is

$$R_g^2 = \frac{\hat{\gamma}_1' G'(I_T - 1_T 1_T'/T) G \hat{\gamma}_1}{\log(\hat{M}_{\exp}^{\beta,\text{mis}})'(I_T - 1_T 1_T'/T) \log(\hat{M}_{\exp}^{\beta,\text{mis}})},$$

where $\hat{\gamma}_1 = (G'(I_T - 1_T 1_T'/T)G)^{-1} G'(I_T - 1_T 1_T'/T) \log(\hat{M}_{\exp}^{\beta,\text{mis}})$.

In Proposition 5, we showed that

$$\log(\hat{M}_{\exp,t+1}^{\beta,\text{mis}}) \xrightarrow{p} -\lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) - \frac{1}{2}\lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}}.$$

For simplicity, we set $M_{1_T} = I_T - 1_T 1_T'/T$ and, given that $M_{1_T} 1_T = 0_{T\times T}$, we obtain

$$\begin{aligned}
\hat{\gamma}_1 &\xrightarrow{p} -(G' M_{1_T} G)^{-1} G' M_{1_T} (F^{\text{mis}} - 1_T \mathbb{E}(f_{t+1}^{\text{mis}\prime})) V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \\
&= -(Q F^{\text{mis}\prime} M_{1_T} F^{\text{mis}} Q')^{-1} Q F^{\text{mis}\prime} M_{1_T} F^{\text{mis}} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \\
&= -(Q')^{-1} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \\
&= \gamma_1.
\end{aligned}$$

The limiting behavior of the numerator of R_g^2 is as follows

$$\begin{aligned}
\gamma_1'(G' M_{1_T} G)\gamma_1 &\xrightarrow{p} \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} Q^{-1} Q (F^{\text{mis}\prime} M_{1_T} F^{\text{mis}}) Q'(Q')^{-1} V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \\
&= \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} (F^{\text{mis}\prime} M_{1_T} F^{\text{mis}}) V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}}, \quad \text{when } N \to \infty.
\end{aligned}$$

The limiting behavior of the denominator of R_g^2 is as follows

$$\begin{aligned}
&\log(\hat{M}_{\exp}^{\beta,\text{mis}})'(I_T - 1_T 1_T'/T) \log(\hat{M}_{\exp}^{\beta,\text{mis}}) \\
&\quad \xrightarrow{p} \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} (F^{\text{mis}} - 1_T \mathbb{E}(f_{t+1}^{\text{mis}\prime}))' M_{1_T} (F^{\text{mis}} - 1_T \mathbb{E}(f_{t+1}^{\text{mis}\prime})) V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \\
&\quad = \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} (F^{\text{mis}\prime} M_{1_T} F^{\text{mis}}) V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}}, \quad \text{when } N \to \infty.
\end{aligned}$$

Given that the limit of the numerator equals the limit of the denominator, $R_g^2 \xrightarrow{p} 1$. The proof of the case of G being orthogonal to F^{mis}, that is, when $G'(I_T - 1_T 1_T'/T) F^{\text{mis}} = 0_{K^{\text{mis}} \times K^{\text{mis}}}$, is straightforward, and therefore, omitted. □

Proof of Proposition 7

The equilibrium process for asset returns, given by (2) and (24) is

$$R_i - R_f = \beta_i(\mathbb{E}(R_m) - R_f) + a_i - \beta_i a_m + b_i Y + \sigma_i \epsilon_i.$$

We posit that the SDF M has the following form,

$$M = \xi + \chi Y + \sum_{i=1}^{N} \zeta_i \epsilon_i,$$

where ξ, χ, and ζ_i, $i = \{1, \ldots, N\}$, are determined using the $N + 2$ equations for the Law of One Price:

$$\mathbb{E}[M] = \frac{1}{R_f}, \tag{A1}$$

$$\mathbb{E}[M(R_{N+1} - R_f)] = 0 \tag{A2}$$

$$\mathbb{E}[M(R_i - R_f)] = 0, \quad \text{for} \quad i = \{1, \ldots, N\}, \tag{A3}$$

where, from (3) and (11) in the manuscript,

$$R_{N+1} = R_f + \gamma b + Y.$$

From expression (A1), we get

$$\xi = \frac{1}{R_f}.$$

From expression (A2), we get

$$\chi = -\frac{\gamma b}{R_f}.$$

From expression (A3), for each $i = \{1, \ldots, N\}$ we have

$$\xi \beta_i (\mathbb{E}(R_m) - R_f) + \xi (a_i - \beta_i a_m) + \chi \beta_i + \zeta \sigma_i = 0.$$

As a result,

$$\zeta = -\frac{1}{R_f} \frac{\beta_i (\mathbb{E}(R_m) - R_f) + a_i - \beta_i a_m - b_i \gamma b}{\sigma_i}.$$

Recalling that

$$R_m = \sum_{i=1}^{N} x_i R_i$$

and using (2) and (16) from the manuscript, we obtain

$$\begin{aligned} R_m - R_f &= \sum_{i=1}^{N} x_i (\gamma b_i b + \gamma x_i \sigma_i^2 / q_i) + \sum_{i=1}^{N} x_i b_i Y + \sum_{i=1}^{N} x_i \sigma_i \epsilon_i \\ &= \gamma b^2 + \gamma \sum_{i=1}^{N} x_i^2 \sigma_i^2 / q_i + bY + \sum_{i=1}^{N} x_i \sigma_i \epsilon_i. \end{aligned}$$

From the last expression, we obtain

$$bY = (R_m - R_f) - \gamma b^2 - \gamma \sum_{i=1}^{N} x_i^2 \sigma_i^2 / q_i - \sum_{i=1}^{N} x_i \sigma_i \epsilon_i.$$

As a result, the SDF is

$$M = \frac{1}{R_f} - \frac{\gamma}{R_f} \left((R_m - R_f) - b^2\gamma - \gamma \sum_{i=1}^{N} x_i^2 \sigma_i^2 / q_i - \sum_{i=1}^{N} x_i \sigma_i \epsilon_i \right)$$
$$- \frac{1}{R_f} \sum_{t=1}^{N} \frac{\beta_i(\mathbb{E}(R_m) - R_f) + a_i - \beta_i a_m - b_i \gamma b}{\sigma_i} \epsilon_i.$$

Grouping together similar terms, we obtain

$$M = \frac{1}{R_f} + \frac{\gamma^2 b^2}{R_f} + \frac{\gamma^2 \sum_{i=1}^{N} x_i^2 \sigma_i^2 / q_i}{R_f} - \frac{\gamma}{R_f}(R_m - R_f)$$
$$- \frac{1}{R_f} \sum_{i=1}^{N} \left(\frac{\beta_i(\mathbb{E}(R_m) - R_f) + a_i - \beta_i a_m - b_i \gamma b - \gamma x_i \sigma_i^2}{\sigma_i} \epsilon_i \right).$$

Finally, we use expressions (22) and (24) in Merton (1987) to simplify the loading of M on ϵ_i and obtain

$$-\frac{1}{R_f} \sum_{i=1}^{N} \left(\frac{\beta_i(\mathbb{E}(R_m) - R_f) + a_i - \beta_i a_m - b_i \gamma b - \gamma x_i \sigma_i^2}{\sigma_i} \right) = -\frac{1}{R_f} \sum_{i=1}^{N} \frac{a_i}{\sigma_i}.$$

Using demeaned returns on the market portfolio as a factor in the SDF, along with expressions (15), (19), and (24), we obtain

$$M = \underbrace{-\frac{1}{R_f} \sum_{i=1}^{N} \left(\frac{a_i}{\sigma_i} \epsilon_i \right)}_{M^a} + \underbrace{\frac{1}{R_f} - \frac{(\mathbb{E}(R_m) - R_f)}{R_f \operatorname{var}(R_m)} (R_m - \mathbb{E}(R_m))}_{M^\beta}.$$

As the number of available assets increases, that is, $N \to \infty$, then

$$\beta_i = \frac{b_i b + x_i \sigma_i^2}{b^2 + \sum_{i=1}^{N} x_i^2 \sigma_i^2} \to \frac{b_i b}{b^2} = \frac{b_i}{b},$$

$$a_m = \sum_{i=1}^{N} x_i a_i = \sum_{i=1}^{N} x_i (1 - q_i) \Delta_i = \sum_{i=1}^{N} \gamma x_i^2 \sigma_i^2 \frac{(1 - q_i)}{q_i} \to 0,$$

$$\operatorname{cov}\left(\sum_{i=1}^{N} x_i \sigma_i \epsilon_i, \epsilon_i \right) = \sum_{i=1}^{N} x_i \sigma_i \to 0.$$

Thus, given $N \to \infty$, we have: (i) $\beta_i \to b_i / b$, (ii) $a_m \to 0$, and (iii) the market return is asymptotically orthogonal to all unsystematic shocks, ϵ_i. Making these substitutions gives the results in (17) and (18). □

A.7 Estimation

We start this section, by listing a number of identifying restrictions that we use to fix the rotation of latent factors f_{t+1}^{mis} omitted in an arbitrary candidate model. Next, we show how to estimate the model of asset returns.

A.7.1 Identification conditions

The identification of the loadings of asset returns on the missing factors in a candidate model is unique up to a rotation. Thus, at the estimation stage, we need to impose identifying restrictions. Denote f^{can} (f^{mis}) a matrix $T \times K^{\text{can}}$ ($T \times K^{\text{mis}}$) that collects candidate (missing) factors column by column. Combine these matrices in a $T \times (K^{\text{can}} + K^{\text{mis}})$ matrix $f = [f^{\text{can}},\ f^{\text{mis}}]$. Note that the rotation of this matrix is defined by a squared invertible matrix of a dimension $(K^{\text{can}} + K^{\text{mis}}) \times (K^{\text{can}} + K^{\text{mis}})$, and therefore, the rotation is pinned down by $(K^{\text{can}} + K^{\text{mis}})^2$ parameters.

At the estimation stage, we impose the following $(K^{\text{can}} + K^{\text{mis}})^2$ identifying restrictions to fix the rotation:

- The first K^{can} columns of the rotation matrix are fixed because f^{can} includes factors that are observable. This is equivalent to $K^{\text{can}}(K^{\text{can}} + K^{\text{mis}})$ restrictions being imposed already.
- $V_{f^{\text{mis}}} = I_{K^{\text{mis}}}$ introduces $K^{\text{mis}}(K^{\text{mis}} + 1)/2$ restrictions. We also assume that the latent factors f_{t+1}^{mis} have positive means to identify the latent factors uniquely rather than up to a sign.
- $\beta^{\text{mis}\prime}\beta^{\text{mis}}$ is a diagonal matrix that is equivalent to imposing $(K^{\text{mis}} - 1)K^{\text{mis}}/2$ restrictions. We also introduce an order restriction that requires that the diagonal elements of the matrix $\beta^{\text{mis}\prime}\beta^{\text{mis}}$ are in decreasing order.
- Candidate factors f_{t+1}^{can} are pairwise uncorrelated with missing factors f_{t+1}^{mis}, which is equivalent to imposing $K^{\text{can}}K^{\text{mis}}$ additional restrictions.

A.7.2 Parameter estimates

Proposition A2 (Parameter estimates)**.** *Suppose that the vector of asset returns R_{t+1} satisfies the data-generating process in equations* (1) *and* (10)*. Without loss of generality,* *assume that f_{t+1}^{can} are tradable factors in the form of excess returns on investment strategies (if any candidate factor is not tradable, we use its factor-mimicking portfolio, as in Breeden, Gibbons, and Litzenberger (1989)). Assume that the number of missing factors in the candidate model, K^{mis}, and the no-arbitrage bound δ_{apt} are known, and that the sample covariance matrix of candidate factors $\hat{V}_{f^{\text{can}}} = \hat{M}_{f^{\text{can}}} - \bar{f}^{\text{can}}\bar{f}^{\text{can}\prime}$ is nonsingular, where $\hat{M}_{f^{\text{can}}} = T^{-1}\sum_{t=1}^{T} f_t^{\text{can}} f_t^{\text{can}\prime}$, and $\bar{f}^{\text{can}} = T^{-1}\sum_{t=1}^{T} f_t^{\text{can}}$. Then:*

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier $\hat{\kappa}$ is greater than zero, the estimators of β^{can}, λ^{mis}, and a are

$$\text{vec}(\hat{\beta}^{\text{can}}) = \left(\hat{M}_{f^{\text{can}}} \otimes I_{N\times N} - \bar{f}^{\text{can}}\bar{f}^{\text{can}\prime} \otimes \hat{G}\right)^{-1} \cdot \text{vec}\left(\hat{M}_{Rf^{\text{can}}} - \hat{G}(\bar{R} - R_f 1_N)\bar{f}^{\text{can}\prime}\right), \tag{A4}$$

$$\hat{\lambda}^{\text{mis}} = (\hat{\beta}^{\text{mis}\prime}\hat{V}_{\varepsilon}^{-1}\hat{\beta}^{\text{mis}})^{-1}\hat{\beta}^{\text{mis}\prime}\hat{V}_{\varepsilon}^{-1}\left(\bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}}\hat{\lambda}^{\text{can}}\right), \quad \textit{and}$$

$$\hat{a} = \frac{1}{\hat{\kappa}+1}\left(\bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}}\hat{\lambda}^{\text{can}} - \hat{\beta}^{\text{mis}}\hat{\lambda}^{\text{mis}}\right),$$

where

$$\hat{G} = \frac{1}{(\hat{\kappa}+1)} I_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)}\hat{\beta}^{\text{mis}}(\hat{\beta}^{\text{mis}\prime}\,\hat{V}_{\varepsilon}^{-1}\,\hat{\beta}^{\text{mis}})^{-1}\hat{\beta}^{\text{mis}\prime}\,\hat{V}_{\varepsilon}^{-1},$$

$$\hat{V}_{\varepsilon} = \hat{\beta}^{\text{mis}}\hat{\beta}^{\text{mis}\prime} + \hat{V}_e,$$

$$\hat{M}_{Rf^{\text{can}}} = \frac{1}{T}\sum_{t=1}^{T}(R_t - R_{ft-1}1_N)f_t^{\text{can}\prime},$$

$$\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t, \quad \textit{and} \quad \bar{R}_f = \frac{1}{T}\sum_{t=1}^{T} R_{ft}.$$

The estimators $\hat{\lambda}^{\text{can}}$ and $\hat{V}_f^{\text{can}}$ coincide with the sample mean and sample covariance of the factors f_t^{can}. The estimators $\hat{\beta}^{\text{mis}}$ and $\hat{V}_e$ do not admit a closed-form solution.

(ii) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} = 0$, it is possible to estimate only vector $\alpha = \beta^{\text{mis}}\lambda^{\text{mis}} + a$ but not its components, a and $\beta^{\text{mis}}\lambda^{\text{mis}}$, separately. The estimator of α is

$$\hat{\alpha} = \bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}}\hat{\lambda}^{\text{can}}. \tag{A5}$$

The estimator of $\text{vec}(\hat{\beta}^{\text{can}})$ is given by expression (A4) with $\hat{\kappa} = 0$. As in case (i), the estimators $\hat{\lambda}^{\text{can}}$ and $\hat{V}_{f^{\text{can}}}$ coincide with the sample mean and covariance of the factors f_t^{can}, while the estimators $\hat{\beta}^{\text{mis}}$ and $\hat{V}_e$ do not admit closed-form expressions.

Proof. We consider the two cases, $\kappa > 0$ and $\kappa = 0$, separately. First, suppose that $\kappa > 0$. Differentiating the penalized (scaled) log-likelihood function,

$$\log L_p(\Theta) = -\frac{\kappa}{2}(a' V_\varepsilon^{-1} a - \delta_{\text{apt}}) - \frac{1}{2}\log(|V_\varepsilon|) - \frac{1}{2}\log(|V_{f^{\text{can}}}|)$$
$$- \frac{1}{2T}\sum_{t=0}^{T-1} \varepsilon'_{t+1} V_\varepsilon^{-1} \varepsilon_{t+1} - \frac{1}{2T}\sum_{t=0}^{T-1} (f_{t+1}^{\text{can}} - \mathbb{E}(f_t^{\text{can}}))' V_{f^{\text{can}}}^{-1} (f_{t+1}^{\text{can}} - \mathbb{E}(f_t^{\text{can}})),$$

where $\varepsilon_{t+1} = R_{t+1} - R_{ft} 1_N - a - \beta^{\text{mis}} \lambda^{\text{mis}} - \beta^{\text{can}} \lambda^{\text{can}} - \beta^{\text{can}}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}))$. For λ^{mis} and a, we obtain the following $K^{\text{mis}} + N$ equations:

$$\begin{pmatrix} \beta^{\text{mis}\prime} V_\varepsilon^{-1} \\ I_N \end{pmatrix} \left(\bar{R} - \bar{R}_f 1_N - \beta^{\text{can}} \lambda^{\text{can}}\right) = \begin{pmatrix} \beta^{\text{mis}\prime} V_\varepsilon^{-1} \beta^{\text{mis}} & \beta^{\text{mis}\prime} V_\varepsilon^{-1} \\ \beta^{\text{mis}} & (1+\kappa) I_N \end{pmatrix} \begin{pmatrix} \hat{\lambda}^{\text{mis}} \\ \hat{a} \end{pmatrix}.$$

From the APT no-arbitrage restriction, $\hat{\lambda}^{\text{mis}}$ and $\hat{a}$ are identified separately (given that $\kappa > 0$). In fact, the above system of linear equations can be solved because the matrix premultiplying the vector $(\hat{\lambda}^{\text{mis}\prime},\ \hat{a}')'$ is nonsingular for every $\kappa > 0$, leading to the closed-form solutions:

$$\hat{\lambda}^{\text{mis}} = (\beta^{\text{mis}\prime} V_\varepsilon^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}\prime} V_\varepsilon^{-1} \left(\bar{R} - \bar{R}_f 1_N - \beta^{\text{can}} \lambda^{\text{can}}\right), \tag{A6}$$

$$\hat{a} = \frac{1}{\hat{\kappa} + 1}\left(\bar{R} - \bar{R}_f 1_N - \beta^{\text{can}} \lambda^{\text{can}} - \beta^{\text{mis}} \hat{\lambda}^{\text{mis}}\right). \tag{A7}$$

Recall that the candidate factors f_t^{can} represent excess returns on tradable investment strategies, that is, $\mathbb{E}(f_t^{\text{can}}) = \lambda^{\text{can}}$. Thus, taking the first derivative of the log-likelihood with respect to λ^{can} results in

$$\hat{\lambda}^{\text{can}} = \frac{1}{T}\sum_{t=1}^{T} f_{t^{\text{can}}}.$$

Similarly, the first derivative of the log-likelihood with respect to V_f^{can} gives

$$\hat{V}_{f^{\text{can}}} = \frac{1}{T}\sum_{t=1}^{T} (f_t^{\text{can}} - \hat{\lambda}^{\text{can}})(f_t^{\text{can}} - \hat{\lambda}^{\text{can}})'.$$

Next, we take the derivative of the penalized (scaled) log-likelihood function with respect to κ and obtain:

$$\hat{\kappa} = \left(\frac{(\bar{R} - \bar{R}_f 1_N - \beta^{\text{can}} \lambda^{\text{can}} - \beta^{\text{mis}} \lambda^{\text{mis}})' V_\varepsilon^{-1} (\bar{R} - \bar{R}_f 1_N - \beta^{\text{can}} \lambda^{\text{can}} - \beta^{\text{mis}} \lambda^{\text{mis}})}{\delta_{\text{apt}}} - 1\right)^{1/2}. \tag{A8}$$

Note that when $\kappa > 0$, the first-order condition implies that $a' V_\varepsilon^{-1} a = \delta_{\text{apt}}$.

Finally, we consider the first-order condition with respect to the generic (a, b)th element of β^{can}, denoted by β^{can}_{ab} with $1 \le a \le N,\ 1 \le b \le K^{\text{can}}$ and obtain

$$-\frac{1}{T}\sum_{t=1}^{T}\Big(R_t - R_{ft-1}1_N - \beta^{\text{mis}}\lambda^{\text{mis}} - a - \hat{\beta}^{\text{can}} f_t^{\text{can}}\Big)' V_\varepsilon{}^{-1}\Big(-\frac{\partial \beta^{\text{can}}}{\partial \beta^{\text{can}}_{ab}} f_t^{\text{can}}\Big) = 0,$$

which can be rearranged by stacking together the first-order conditions as

$$\hat{M}_{Rf^{\text{can}}} - (a + \beta^{\text{mis}}\lambda^{\text{mis}})\bar{f}^{\text{can}\,\prime} - \hat{\beta}^{\text{can}}\hat{M}_{f^{\text{can}}} = 0_{N\times K^{\text{can}}}. \tag{A9}$$

Next, define

$$G = \frac{1}{(\hat{\kappa}+1)} I_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)}\beta^{\text{mis}}(\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\,\prime} V_\varepsilon^{-1},$$

and use the formulas (A6), and (A7) to rewrite equation (A9) as follows

$$\hat{\beta}^{\text{can}}\hat{M}_{f^{\text{can}}} - G\hat{\beta}^{\text{can}}\bar{f}^{\text{can}}\bar{f}^{\text{can}\,\prime} = \hat{M}_{Rf^{\text{can}}} - G(\bar{R} - \bar{R}_f 1_N)\bar{f}^{\text{can}\,\prime}.$$

Then, we take the vec operator and solve for $\hat{\beta}^{\text{can}}$ to obtain

$$\text{vec}(\hat{\beta}^{\text{can}}) = \big(\hat{M}_{f^{\text{can}}} \otimes I_N - \bar{f}^{\text{can}}\bar{f}^{\text{can}\,\prime} \otimes G\big)^{-1} \cdot \text{vec}\left(\hat{M}_{Rf^{\text{can}}} - G(\bar{R} - R_f 1_N)\bar{f}^{\text{can}\,\prime}\right). \tag{A10}$$

It is important to note that the solution for $\hat{\beta}^{\text{can}}$ exists because the matrix

$$\hat{M}_{f^{\text{can}}} \otimes I_N - \bar{f}^{\text{can}}\,\bar{f}^{\text{can}\,\prime} \otimes G = (\hat{M}_{f^{\text{can}}} - \bar{f}^{\text{can}}\,\bar{f}^{\text{can}\,\prime}) \otimes I_N + \bar{f}^{\text{can}}\,\bar{f}^{\text{can}\,\prime} \otimes (I_N - G)$$

is nonsingular. The nonsingularity follows because the matrix $(\hat{M}_{f^{\text{can}}} - \bar{f}^{\text{can}}\,\bar{f}^{\text{can}\,\prime}) \otimes I_N$ is positive semi-definite, given that $\hat{M}_{f^{\text{can}}} - \bar{f}^{\text{can}}\bar{f}^{\text{can}\,\prime}$ is the covariance matrix of f_t^{can}, while the matrix $\bar{f}^{\text{can}}\,\bar{f}^{\text{can}\,\prime} \otimes (I_N - G)$ is positive semi-definite, because $I_N - G$ is semi-definite as shown below. Specifically,

$$\begin{aligned}
I_N - G &= I_N - \frac{1}{(\hat{\kappa}+1)} I_N - \Big(\frac{\hat{\kappa}}{1+\hat{\kappa}}\Big)\beta^{\text{mis}}(\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\,\prime} V_\varepsilon^{-1} \\
&= \Big(\frac{\hat{\kappa}}{1+\hat{\kappa}}\Big)(I_N - \beta^{\text{mis}}(\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}) \\
&= \Big(\frac{\hat{\kappa}}{1+\hat{\kappa}}\Big)V_\varepsilon(V_\varepsilon^{-1} - V_\varepsilon^{-1}\beta^{\text{mis}}(\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}) \\
&= \Big(\frac{\hat{\kappa}}{1+\hat{\kappa}}\Big)V_\varepsilon(V_\varepsilon^{-1})^{\frac{1}{2}}(I_N - (V_\varepsilon^{-1})^{\frac{1}{2}}\beta^{\text{mis}}(\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\,\prime}(V_\varepsilon^{-1})^{\frac{1}{2}})(V_\varepsilon^{-1})^{\frac{1}{2}}
\end{aligned}$$

is the product of the positive-definite matrices $I_N - (V_\varepsilon^{-1})^{\frac{1}{2}}\beta^{\text{mis}}(\beta^{\text{mis}\,\prime} V_\varepsilon^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}\,\prime}(V_\varepsilon^{-1})^{\frac{1}{2}}$, V_ε, and $(V_\varepsilon^{-1})^{\frac{1}{2}}$. The first matrix is positive semi-definite because it is a projection matrix.

Next, we plug expression (A10) for $\hat{\beta}^{\text{can}}$ into expressions (A6), (A7), and (A8) for $\hat{\lambda}^{\text{mis}}$ $\hat{a}$, and $\hat{\kappa}$, respectively, and obtain

$$\hat{\lambda}^{\text{mis}} = \hat{\lambda}^{\text{mis}}(\beta^{\text{mis}}, V_e), \quad \hat{a} = \hat{a}(\beta^{\text{mis}}, V_e), \quad \text{and} \quad \hat{\kappa} = \hat{\kappa}(\beta^{\text{mis}}, V_e).$$

We substitute these expressions together with $\hat{\beta}^{\text{can}} = \hat{\beta}^{\text{can}}(\beta^{\text{mis}}, V_e)$, into $L_p(\Theta)$ to obtain the concentrated log-likelihood function, which is a function of only β^{mis} and V_e. We maximize the concentrated log-likelihood numerically, thereby obtaining $\hat{\beta}^{\text{mis}}$ and $\hat{V}_e$, which also imply the optimal values of the other parameters.

Now consider the case in which the Karush-Kuhn-Tucker multiplier is zero: $\kappa = 0$. In this case, a feasible solution to the optimization problem satisfies $a' V_\varepsilon^{-1} a < \delta_{\text{apt}}$.

The first-order conditions with respect to λ^{mis} and a imply the following system of $K^{\text{mis}} + N$ equations

$$\begin{pmatrix} \beta^{\text{mis}\prime} V_\varepsilon^{-1} \\ I_N \end{pmatrix} \left(\bar{R} - \bar{R}_f 1_N - \beta^{\text{can}} \lambda^{\text{can}} \right) = \begin{pmatrix} \beta^{\text{mis}\prime} V_\varepsilon^{-1} \beta^{\text{mis}} & \beta^{\text{mis}\prime} V_\varepsilon^{-1} \\ \beta^{\text{mis}} & I_N \end{pmatrix} \begin{pmatrix} \hat{\lambda}^{\text{mis}} \\ \hat{a} \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} \beta^{\text{mis}\prime} V_\varepsilon^{-1} \beta^{\text{mis}} & \beta^{\text{mis}\prime} V_\varepsilon^{-1} \\ \beta^{\text{mis}} & I_N \end{pmatrix}$$

is of dimension $(N + K^{\text{mis}}) \times (N + K^{\text{mis}})$ but of rank N, and therefore it is noninvertible. As a result, we cannot separately identify a and λ^{mis}, implying that if $\kappa = 0$, we can only identify the sum: $a + \beta^{\text{mis}} \lambda^{\text{mis}}$. All the other parameters of the vector Θ are identified separately, and their expressions follow from differentiating the log likelihood and solving the resulting first-order conditions. For instance, the formula for $\hat{\beta}^{\text{can}}$ follows by setting $\hat{G} = I_N$ into (A4).

When both cases, $\kappa > 0$ and $\kappa = 0$, are feasible, we choose the one under which the log-likelihood $L_p(\Theta)$ is larger. □

References

Almeida, C., and R. Garcia (2012): "Assessing Misspecified Asset Pricing Models With Empirical Likelihood Estimators," *Journal of Econometrics*, 170, 519–537.

Almeida, C., and P. Schneider (2021): "Constrained Polynomial Likelihood," Swiss Finance Institute Research Paper.

Anatolyev, S., and A. Mikusheva (2021): "Factor models with many assets: Strong factors, weak factors, and the two-pass procedure," *Journal of Econometrics*.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61(1).

Asness, C., A. Frazzini, N. J. Gormsen, and L. H. Pedersen (2020): "Betting against correlation: Testing theories of the low-risk effect," *Journal of Financial Economics*, 135(3), 629–652.

Asness, C., A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen (2018): "Size matters, if you control your junk," *Journal of Financial Economics*, 129(3), 479–509.

Bai, J., and S. Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

Baker, M., and J. Wurgler (2006): "Investor sentiment and the cross-section of stock returns," *Journal of Finance*, 61(4), 1645–1680.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring economic policy uncertainty," *The quarterly journal of economics*, 131(4), 1593–1636.

Baker, S. R., N. Bloom, S. J. Davis, and K. J. Kost (2019): "Policy news and stock market volatility," Discussion paper, National Bureau of Economic Research.

Bali, T. G., R. F. Engle, and S. Murray (2016): *Empirical Asset Pricing: The Cross Section of Stock Returns*. John Wiley & Sons, Hoboken, New Jersey.

Barbee Jr, W. C., S. Mukherji, and G. A. Raines (1996): "Do sales-price and debt-equity explain stock returns better than book-market and firm size?," *Financial Analysts Journal*, 52(2), 56–60.

Bhandari, L. C. (1988): "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence," *Journal of Finance*, 43(2), 507–528.

Breeden, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989): "Empirical Test of the Consumption-Oriented CAPM," *Journal of Finance*, 44, 231–262.

Brennan, M. J., T. Chordia, and A. Subrahmanyam (1998): "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, 49(3), 345–373.

Bryzgalova, S., J. Huang, and C. Julliard (2020): "Bayesian solutions for the factor zoo: We just ran two quadrillion models," *Available at SSRN 3481736*.

Campbell, J. Y. (2006): "Household Finance," *Journal of Finance*, 61(4), 1553–1604.

Chaieb, I., H. Langlois, and O. Scaillet (2021): "Factors and risk premia in individual international stock returns," *Journal of Financial Economics*, 141(2), 669–692.

Chamberlain, G. (1983): "Funds, Factors and Diversification in Arbitrage Pricing Models," *Econometrica*, 51, 1305–1324.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

Chen, A. Y., and T. Zimmermann (2021): "Open Source Cross-Sectional Asset Pricing," Working Paper.

Clarke, C. (2020): "The Level, Slope and Curve Factor Model for Stocks," forthcoming in *Journal of Financial Economics*.

Cochrane, J. H. (2005): *Asset Pricing (revised edition)*. Princeton University Press, Princeton, New Jersey.

——— (2011): "Presidential address: Discount rates," *Journal of Finance*, 66(4), 1047–1108.

Daniel, K., D. Hirshleifer, and L. Sun (2020): "Short-and Long-Horizon Behavioral Factors," *Review of Financial Studies*, 33(4), 1673–1736.

Daniel, K., and S. Titman (1997): "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance*, 52(1), 1–33.

Dechow, P. M., A. P. Hutton, L. Meulbroek, and R. G. Sloan (2001): "Short-sellers, fundamental analysis, and stock returns," *Journal of Financial Economics*, 61(1), 77–106.

Dechow, P. M., R. G. Sloan, and M. T. Soliman (2004): "Implied Equity Duration: A New Measure of Equity Risk," *Review of Accounting Studies*, 9(2), 197–228.

Fama, E. F., and K. French (2015): "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116(1-22).

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

——— (2004): "The Capital Asset Pricing Model: Theory and Evidence," *Journal of Economic Perspectives*, 18(3), 25–46.

Fama, E. F., and J. D. MacBeth (1973): "Risk, return, and equilibrium: Empirical tests," *The Journal of Political Economy*, 81, 607–636.

Frazzini, A., and L. H. Pedersen (2014): "Betting against Beta," *Journal of Financial Economics*, 111(1), 1–25.

Gabaix, X. (2011): "The granular origins of aggregate fluctuations," *Econometrica*, 79(3), 733–772.

Gagliardini, P., E. Ossola, and O. Scaillet (2016): "Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets," *Econometrica*, 84(3), 985–1046.

——— (2019): "A diagnostic criterion for approximate factor structure," forthcoming in *Journal of Econometrics*.

Ghosh, A., C. Julliard, and A. P. Taylor (2017): "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 D. Xiu (2017): "Inference on Risk Premia in the Presence of Omitted Factors," Working Paper, University of Chicago.

——— (2021): "Asset Pricing with Omitted Factors," forthcoming, Journal of Political Economy.

Giglio, S., D. Xiu, and D. Zhang (2021): "Test Assets and Weak Factors," Chicago Booth Research Paper.

Gilchrist, S., and E. Zakrajšek (2012): "Credit spreads and business cycle fluctuations," *American Economic Review*, 102(4), 1692–1720.

Goetzmann, W. N., and A. Kumar (2008): "Equity Portfolio Diversification," *Review of Finance*, 12, 433–463.

Gourieroux, C., and A. Monfort (2007): "Econometric specification of stochastic discount factor models," *Journal of Econometrics*, 136(2), 509–530.

Hansen, L. P., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

——— (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52(2), 557–590.

He, Z., B. Kelly, and A. Manela (2017): "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 126(1), 1–35.

Herskovic, B., A. Moreira, and T. Muir (2019): "Hedging risk factors," Available at SSRN 3148693.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021): "An augmented q-factor model with expected growth," *Review of Finance*, 25(1), 1–41.

Hou, K., C. Xue, and L. Zhang (2015): "Digesting anomalies: An investment approach," *Review of Financial Studies*, 28(3), 650–705.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015): "Investor sentiment aligned: A powerful predictor of stock returns," *Review of Financial Studies*, 28(3), 791–837.

Huang, D., J. Li, and L. Wang (2021): "Are disagreements agreeable? Evidence from information aggregation," *Journal of Financial Economics*.

Huberman, G. (1982): "A Simple Approach to the Arbitrage Pricing Theory," *Journal of Economic Theory*, 28, 183–191.

Ingersoll, J. (1984): "Some Results in the Theory of Arbitrage Pricing," *Journal of Finance*, 39, 1021–1039.

Jegadeesh, N., and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance*, 48, 65–91.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (2021): "Is There a Replication Crisis in Finance?," Working Paper.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring uncertainty," *American Economic Review*, 105(3), 1177–1216.

Koijen, R. S. J., and M. Yogo (2019): "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127(4), 1475–1515.

Korsaye, S., A. Quaini, and F. Trojani (2019): "Smart SDFs," University of Geneva, Working Paper.

——— (2021): "Smart Stochastic Discount Factors," University of Geneva, Working Paper.

Kozak, S., S. Nagel, and S. Santosh (2018): "Interpreting Factor Models," *The Journal of Finance*, 73(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the Cross-section," *Journal of Financial Economics*, 135, 271–292.

La Porta, R. (1996): "Expectations and the Cross-Section of Stock Returns," *Journal of Finance*, 51(5), 1715–1742.

Lamont, O., C. Polk, and J. Saa-Requejo (2001): "Financial Constraints and Stock Returns," *Review of Financial Studies*, 14(2), 529–54.

Lettau, M., and M. Pelger (2020): "Factors that fit the time-series and cross-section of stock returns," *Review of Financial Studies*, 33, 2274–2325.

Li, D. (2011): "Financial Constraints, R& D Investment, and Stock Returns," *Review of Financial Studies*, 24(9), 2974–3007.

Lopez-Lira, A., and N. Roussanov (2022): "Do Common Factors Really Explain the Cross-Section of Stock Returns?," Working Paper.

Luttmer, E. G. (1996): "Asset Pricing in Economies with Frictions," *Econometrica*, 64, 1439–1467.

MacKinlay, A. C. (1995): "Multifactor Models Do Not Explain Deviations from the CAPM," *Journal of Financial Economics*, 38, 3–28.

McCracken, M. W., and S. Ng (2015): "FRED-MD: A Monthly Database for Macroeconomic Research," Working Paper.

MERTON, R. C. (1987): "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, 42(3), 483–510.

NOVY-MARX, R. (2013): "The other side of value: The gross profitability premium," *Journal of financial economics*, 108(1), 1–28.

PALAZZO, B. (2012): "Cash holdings, risk, and expected returns," *Journal of Financial Economics*, 104(1), 162–185.

PASQUARIELLO, P. (2014): "Financial market dislocations," *The Review of Financial Studies*, 27(6), 1868–1914.

POLKOVNICHENKO, V. (2005): "Household Portfolio Diversification: A Case for Rank Dependent Preferences," *Review of Financial Studies*, 18, 1467–1502.

RAPONI, V., R. UPPAL, AND P. ZAFFARONI (2022): "Robust Portfolio Choice," Working paper, Imperial College.

RICHARDSON, S. A., R. G. SLOAN, M. T. SOLIMAN, AND I. TUNA (2005): "Accrual reliability, earnings persistence and stock prices," *Journal of Accounting and Economics*, 39(3), 437–485.

ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

ROSS, S. A. (1977): "Return, Risk, and Arbitrage," in *Risk and Return in Finance*, ed. by I. Friend, and J. Bicksler. Ballinger, Cambridge, MA.

SANDULESCU, M., AND P. SCHNEIDER (2021): "International Mispricing and Arbitrage Premia," Swiss Finance Institute Research Paper.

SHARPE, W. (1964): "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, 1919, 425–442.

STAMBAUGH, R. F., AND Y. YUAN (2017): "Mispricing factors," *Review of Financial Studies*, 30(4), 1270–1315.

Internet Appendix

This appendix generalizes a number of the results reported in the main text. It also provides detailed information about the data used in our empirical analysis.

IA.1 The SDF with Nonorthogonal Components

In the main text of the manuscript, we assume that the candidate risk factors f_{t+1}^{can} are orthogonal to the missing sources of systematic risk f_{t+1}^{mis} and unsystematic shocks e_{t+1}. This assumption is without loss of generality because if the (observable) systematic risk factors f_{t+1}^{mis} that the candidate model omits are in fact correlated with f_{t+1}^{can}, there exists an *observationally equivalent* representation of the SDF M_{t+1}, such that factors f_{t+1}^{can} are orthogonal to some latent sources of systematic risk (residuals from an orthogonal projection of omitted observable risk factors onto the candidate factors).

In particular,

$$\begin{aligned} M_{t+1} &= \frac{1}{R_f} + b^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c' e_{t+1} \\ &= \frac{1}{R_f} + \tilde{b}^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b^{\text{mis}\prime}(\tilde{f}_{t+1}^{\text{mis}} - \mathbb{E}(\tilde{f}_{t+1}^{\text{mis}})) + c' e_{t+1} \end{aligned} \tag{IA1}$$

where $Q = cov(f_{t+1}^{\text{can}},\ f_{t+1}^{\text{mis}\prime})$ is a $K^{\text{can}} \times K^{\text{mis}}$ matrix of covariances and

$$\tilde{b}^{\text{can}} = b^{\text{can}} + V_{f^{\text{can}}}^{-1} Q b^{\text{mis}},$$

$$\tilde{f}_{t+1}^{\text{mis}} - \mathbb{E}(\tilde{f}_{t+1}^{\text{mis}}) = (f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) - Q' V_{f^{\text{can}}}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})).$$

Notice that by construction $\text{cov}(f_{t+1}^{\text{can}}, \tilde{f}_{t+1}^{\text{mis}\prime}) = 0_{K^{\text{can}} \times K^{\text{mis}}}$, because $\tilde{f}_{t+1}^{\text{mis}}$ represent the linear-projection residual from projecting $f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})$ on $f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})$. For some applications, the representation (IA1) of the SDF M_{t+1} that is based on correlated systematic factors f_{t+1}^{can} and f_{t+1}^{mis} is more useful. For example, this would be the case in an exercise of quantifying prices of risk associated with non-tradable risk factors.

We now show how all our results can be generalized to allow for the case of correlated observed and missing factors.

Proposition IA.1 (SDF: Correlated case)**.** *Under Assumptions 1 and 2 of the APT, there exists an admissible SDF of the form*

$$M_{t+1} = \frac{1}{R_f} + b^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c' e_{t+1}, \quad \textit{where}$$

$$b^{\text{can}\prime} = \left(-\frac{1}{R_f}\lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} + \frac{1}{R_f}\lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} Q' V_{f^{\text{can}}}^{-1}\right) \cdot \left(I_{K^{\text{can}}} - Q V_{f^{\text{mis}}}^{-1} Q' V_{f^{\text{can}}}^{-1}\right)^{-1},$$

$$b^{\text{mis}\prime} = \left(-\frac{1}{R_f}\lambda^{\text{mis}}V_{f^{\text{mis}}}^{-1} + \frac{1}{R_f}\lambda^{\text{can}\prime}V_{f^{\text{can}}}^{-1}QV_{f^{\text{mis}}}^{-1}\right)\cdot\left(I_{K^{\text{mis}}} - Q'V_{f^{\text{can}}}^{-1}QV_{f^{\text{mis}}}^{-1}\right)^{-1},$$

$$c' = -\frac{1}{R_f}aV_e^{-1}.$$

Proof. We guess that the SDF has the following functional form

$$M_{t+1} = \mathbb{E}_t(M_{t+1}) + b^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c'e_{t+1},$$

where b^{can} is a $K^{\text{can}} \times 1$ vector, b^{mis} is a $K^{\text{mis}} \times 1$ vector, and c is an $N \times 1$ vector. We identify the unknown vectors b^{can}, b^{mis}, and c by using the Law of One Price. Specifically, because we assume the existence of the risk-free asset, to determine the mean of the SDF we use the condition

$$\mathbb{E}(M_{t+1}) = \frac{1}{R_f}.$$

Next, because λ^{can} represents a vector of prices of risk of f_{t+1}^{can} we have that

$$-cov(M_{t+1}, f_{t+1}^{\text{can}}) \cdot R_f = \lambda^{\text{can}\prime}.$$

These K^{can} conditions identify b^{can}:

$$b^{\text{can}\prime} = -\frac{1}{R_f}\lambda^{\text{can}\prime}V_{f^{\text{can}}}^{-1} - b^{\text{mis}\prime}Q'V_{f^{\text{can}}}^{-1}. \tag{IA2}$$

Similarly, λ^{mis} is the price of risk associated with factors f_{t+1}^{mis}, or equivalently,

$$-cov(M_{t+1}, f_{t+1}^{\text{mis}}) \cdot R_f = \lambda^{\text{mis}\prime}.$$

These K^{mis} conditions identify b^{mis}:

$$b^{\text{mis}\prime} = -\frac{1}{R_f}\lambda^{\text{mis}\prime}V_{f^{\text{mis}}}^{-1} - b^{\text{can}\prime}QV_{f^{\text{mis}}}^{-1}. \tag{IA3}$$

Putting together expressions (IA2) and (IA3), we obtain

$$b^{\text{can}\prime} = \left(-\frac{1}{R_f}\lambda^{\text{can}\prime}V_{f^{\text{can}}}^{-1} + \frac{1}{R_f}\lambda^{\text{mis}\prime}V_{f^{\text{mis}}}^{-1}Q'V_{f^{\text{can}}}^{-1}\right)\cdot\left(I_{K^{\text{can}}} - QV_{f^{\text{mis}}}^{-1}Q'V_{f^{\text{can}}}^{-1}\right)^{-1},$$

$$b^{\text{mis}\prime} = \left(-\frac{1}{R_f}\lambda^{\text{mis}}V_{f^{\text{mis}}}^{-1} + \frac{1}{R_f}\lambda^{\text{can}\prime}V_{f^{\text{can}}}^{-1}QV_{f^{\text{mis}}}^{-1}\right)\cdot\left(I_{K^{\text{mis}}} - Q'V_{f^{\text{can}}}^{-1}QV_{f^{\text{mis}}}^{-1}\right)^{-1}.$$

Finally, it must be the case that the SDF prices the N basis assets:

$$\mathbb{E}(M_{t+1}(R_{t+1} - R_{ft}1_N)) = 0_N.$$

These N equations identify c:

$$c' = -\frac{1}{R_f}aV_e^{-1}.$$

To derive c, we used expressions (IA2) and (IA3). □

Next, we provide a non-negative SDF.

Proposition IA.2 (Nonnegative SDF: Correlated case)**.** *Under Assumptions 1 and 2 of the APT and the assumption that returns R_{t+1} are Gaussian, there exists an admissible SDF $M_{\exp,t+1}$*

$$
\begin{aligned}
M_{\exp,t+1} &= M^{\beta,\text{can}}_{\exp,t+1} \times M^{a}_{\exp,t+1} \times M^{\beta,\text{mis}}_{\exp,t+1} \quad \textit{where} \\
M^{\beta,\text{can}}_{\exp,t+1} &= \frac{1}{R_f} \exp\left(b^{\text{can}}_{+}(f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1})) - \frac{1}{2} b^{\text{can}\prime}_{+} V_{f^{\text{can}}} b^{\text{can}\prime}_{+} - \frac{1}{2} b^{\text{can}\prime}_{+} Q b^{\text{mis}}_{+} \right) \\
M^{\beta,\text{mis}}_{\exp,t+1} &= \exp\left(b^{\text{mis}}_{+}(f^{\text{mis}}_{t+1} - \mathbb{E}(f^{\text{mis}}_{t+1})) - \frac{1}{2} b^{\text{mis}\prime}_{+} V_{f^{\text{mis}}} b^{\text{mis}\prime}_{+} - \frac{1}{2} b^{\text{can}\prime}_{+} Q b^{\text{mis}}_{+} \right) \\
M^{a}_{\exp,t+1} &= \exp\left(-a' V_e^{-1} e_{t+1} - \frac{1}{2} a' V_e^{-1} a \right), \quad \textit{where} \\
b^{\text{can}\prime}_{+} &= (-\lambda^{\text{can}\prime} V_{f^{\text{can}}} + \lambda^{\text{mis}\prime} V^{-1}_{f^{\text{mis}}} Q' V^{-1}_{f^{\text{can}}}) \cdot (I_{K^{\text{can}}} - Q V^{-1}_{f^{\text{mis}}} Q' V^{-1}_{f^{\text{can}}})^{-1}, \\
b^{\text{mis}\prime}_{+} &= (-\lambda^{\text{mis}\prime} V_{f^{\text{mis}}} + \lambda^{\text{can}\prime} V^{-1}_{f^{\text{can}}} Q V^{-1}_{f^{\text{mis}}}) \cdot (I_{K^{\text{mis}}} - Q' V^{-1}_{f^{\text{can}}} Q V^{-1}_{f^{\text{mis}}})^{-1}.
\end{aligned}
$$

Proof. We use a guess and verify method to derive a nonnegative SDF. We guess that the SDF has the following functional form:

$$M_{\exp,t+1} = \exp\left[\mu_{+} + b^{\text{can}\prime}_{+}(f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1})) + b^{\text{mis}\prime}_{+}(f^{\text{mis}}_{t+1} - \mathbb{E}(f^{\text{mis}}_{t+1})) + c'_{+} e_{t+1}\right],$$

with unknown vectors b^{can}_{+}, b^{mis}_{+}, and c_{+}, as well as an unknown scalar μ_{+}. To identify the unknowns and verify our guess, we use the following $K^{\text{can}} + K^{\text{mis}} + N + 1$ equations, which are implications of the Law of One Price:

$$
\begin{aligned}
-\operatorname{cov}(M_{\exp,t+1}, f^{\text{can}}_{t+1}) \cdot R_f &= \lambda^{\text{can}}, \\
-\operatorname{cov}(M_{\exp,t+1}, f^{\text{mis}}_{t+1}) \cdot R_f &= \lambda^{\text{mis}}, \\
\mathbb{E}(M_{\exp,t+1}(R_{t+1 - R_{ft} 1_N})) &= 0 \\
\mathbb{E}(M_{\exp,t+1}) &= R_f^{-1}.
\end{aligned}
$$

The first K^{can} equations imply that

$$-\mathbb{E}(M_{\exp,t+1}(f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1}))) = \mathbb{E}(M_{\exp,t+1}) \cdot \lambda^{\text{can}},$$

which along with Lemma A.1, give

$$V_{f^{\text{can}}}(b^{\text{can}}_{+} + V^{-1}_{f^{\text{can}}} Q b^{\text{mis}}_{+}) = -\lambda^{\text{can}}. \tag{IA4}$$

Similarly, the next K^{mis} equations imply that

$$-\mathbb{E}(M_{\exp,t+1}(f^{\text{mis}}_{t+1} - \mathbb{E}(f^{\text{mis}}_{t+1}))) = \mathbb{E}(M_{\exp,t+1}) \cdot \lambda^{\text{mis}},$$

which, along with Lemma A.1, lead to:

$$V_{f^{\text{mis}}}(b^{\text{mis}}_{+} + V^{-1}_{f^{\text{mis}}} Q' b^{\text{mis}}_{+}) = -\lambda^{\text{mis}}. \tag{IA5}$$

Taken together expressions (IA4) and (IA5), we obtain

$$b_+^{\text{can}\prime} = (-\lambda^{\text{can}\prime} V_{f^{\text{can}}} + \lambda^{\text{mis}\prime} V_{f^{\text{mis}}}^{-1} Q' V_{f^{\text{can}}}^{-1}) \cdot (I_{K^{\text{can}}} - Q V_{f^{\text{mis}}}^{-1} Q' V_{f^{\text{can}}}^{-1})^{-1},$$

$$b_+^{\text{mis}\prime} = (-\lambda^{\text{mis}\prime} V_{f^{\text{mis}}} + \lambda^{\text{can}\prime} V_{f^{\text{can}}}^{-1} Q V_{f^{\text{mis}}}^{-1}) \cdot (I_{K^{\text{mis}}} - Q' V_{f^{\text{can}}}^{-1} Q V_{f^{\text{mis}}}^{-1})^{-1}.$$

Next, because the SDF prices the N basis assets, which, given Lemma A.1, leads to

$$c' = -a' V_e^{-1}.$$

Finally, the last identifying condition implies

$$\begin{aligned}
&\mathbb{E}(M_{\exp,t+1}) = R_f^{-1} \\
&= \mathbb{E}(\exp[\mu_+ + b_+^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + b_+^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + c_+' e_{t+1}]) \\
&= \exp\left(\mu_+ + (b_+^{\text{can}} + V_{f^{\text{can}}}^{-1} Q b_+^{\text{mis}})' V_{f^{\text{can}}} (b_+^{\text{can}} + V_{f^{\text{can}}}^{-1} Q b_+^{\text{mis}})/2 + b_+^{\text{mis}\prime} V_{f^{\text{mis}}} b_+^{\text{mis}}/2 + c_+' V_e c_+/2\right) \\
&= \exp(\mu_+ + b_+^{\text{can}\prime} V_{f^{\text{can}}} b_+^{\text{can}}/2 + b_+^{\text{mis}\prime} V_{f^{\text{mis}}} b_+^{\text{mis}}/2 + a' V_e^{-1} a/2 + b_+^{\text{can}\prime} Q b_+^{\text{mis}}).
\end{aligned}$$

In the last equation, we use $V_{f^{\text{mis}}} = Q' V_{f^{\text{can}}}^{-1} Q + V_{\tilde{f}^{\text{mis}}}$, where f_{t+1}^{can} and $\tilde{f}_{t+1}^{\text{mis}}$ are orthogonal, and $c' = -a' V_e^{-1}$. As a result,

$$\exp(\mu_+) = R_f^{-1} \cdot \exp[-b_+^{\text{can}\prime} V_{f^{\text{can}}} b_+^{\text{can}}/2 - b_+^{\text{mis}\prime} V_{f^{\text{mis}}} b_+^{\text{mis}}/2 - a' V_e^{-1} a/2 - b_+^{\text{can}\prime} Q b_+^{\text{mis}}]. \quad \square$$

Next, let us introduce the projection version of the SDF $M_{\exp,t+1}$. First, notice that we can express this SDF as

$$M_{\exp,t+1} = \frac{1}{R_f} \cdot \exp\left(m_{t+1} - \frac{1}{2} m\right) = \frac{1}{R_f} \cdot \exp\left(m_{t+1}^{\beta,\text{can}} + m_{t+1}^{\beta,\text{mis}} + m_{t+1}^{a} - \frac{1}{2} m^{\beta,\text{can}} - \frac{1}{2} m^{\beta,\text{mis}} - \frac{1}{2} m^{a}\right),$$

where

$$\begin{aligned}
m_{t+1}^{\beta,\text{can}} &= b_+^{\text{can}\prime}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})), \\
m_{t+1}^{\beta,\text{mis}} &= b_+^{\text{mis}\prime}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})), \\
m_{t+1}^{a} &= -a' V_e^{-1} e_{t+1}, \\
m^{\beta,\text{can}} &= b_+^{\text{can}\prime} V_{f^{\text{can}}} b_+^{\text{can}} + b_+^{\text{can}\prime} Q b_+^{\text{mis}}, \\
m^{\beta,\text{mis}} &= b_+^{\text{mis}\prime} V_{f^{\text{mis}}} b_+^{\text{mis}} + b_+^{\text{can}\prime} Q b_+^{\text{mis}}, \\
m^{a} &= a' V_e^{-1} a.
\end{aligned}$$

Second, set $X_{t+1} = (R_{t+1} - R_f 1_N - \mu)$ with $\mu = \mathbb{E}(R_{t+1} - R_f 1_N)$, $f_t = (f_t^{\text{can}\prime}, f_t^{\text{mis}\prime})'$, and $\beta = (\beta^{\text{can}}, \beta^{\text{mis}})$ and notice that $V_R = \beta V \beta' + V_e$ with $V = \begin{pmatrix} V_{f^{\text{can}}} & Q \\ Q' & V_{f^{\text{mis}}} \end{pmatrix}$.

Finally, define the projected non-negative SDF as

$$\hat{M}_{\exp,t+1} = \frac{1}{R_f} \cdot \exp\left(\hat{m}_{t+1} - \frac{1}{2}\hat{m}\right), \quad \text{where}$$

$$\begin{aligned}
\hat{m}_{t+1} &= \mathbb{E}(m_{t+1}X'_{t+1})(\mathbb{E}(X_{t+1}X'_{t+1})^{-1}X_{t+1} = \hat{m}^{\beta,\text{can}}_{t+1} + \hat{m}^{\beta,\text{mis}}_{t+1} + \hat{m}^{a}_{t+1}, \quad \text{where} \\
\hat{m}^{\beta,\text{can}}_{t+1} &= b^{\text{can}\,\prime}_{+}[V_{f^{\text{can}}}\beta^{\text{can}\,\prime} + Q\beta^{\text{mis}\,\prime}]V_R^{-1}X_{t+1} = b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},\ Q)\beta'V_R^{-1}X_{t+1}, \\
\hat{m}^{\beta,\text{mis}}_{t+1} &= b^{\text{mis}\,\prime}_{+}[Q'\beta^{\text{can}\,\prime} + V_{f^{\text{mis}}}\beta^{\text{mis}\,\prime}]V_R^{-1}X_{t+1} = b^{\text{mis}\,\prime}_{+}(Q',\ V_f^{\text{mis}})\beta'V_R^{-1}X_{t+1} \\
\hat{m}^{a}_{t+1} &= c'_{+}V_eV_R^{-1}X_{t+1}, \quad \text{and} \\
\hat{m} &= \hat{m}^{\beta,\text{can}} + \hat{m}^{\beta,\text{mis}} + \hat{m}^{a} \\
\hat{m}^{\beta,\text{can}} &= b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},Q)\beta'V_R^{-1}\beta(V_{f^{\text{can}}},Q)'b^{\text{can}}_{+} + b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},Q)\beta'V_R^{-1}\beta(Q',V_{f^{\text{mis}}})'b^{\text{mis}}_{+}, \\
\hat{m}^{\beta,\text{mis}} &= b^{\text{mis}\,\prime}_{+}(Q',V_{f^{\text{mis}}})\beta'V_R^{-1}\beta(Q',V_{f^{\text{mis}}})'b^{\text{mis}}_{+} + b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},Q)\beta'V_R^{-1}\beta(Q',V_{f^{\text{mis}}})'b^{\text{mis}}_{+}, \\
\hat{m}^{a} &= c'_{+}V_eV_R^{-1}V_ec_{+}.
\end{aligned}$$

Under the assumptions of Proposition 5, we have

$$\begin{aligned}
c'_{+}V_eV_R^{-1}e_{t+1} &\xrightarrow{p} c'_{+}e_{t+1}, \\
c'_{+}V_eV_R^{-1}\beta &\xrightarrow{p} 0'_{K^{\text{can}}+K^{\text{mis}}}, \\
\beta'V_R^{-1}e_{t+1} &\xrightarrow{p} 0_{K^{\text{can}}+K^{\text{mis}}}, \\
\beta'V_R^{-1}\beta &\xrightarrow{p} V^{-1} \qquad \text{as} \quad N \to \infty.
\end{aligned}$$

Given these results, we show that

$$\begin{aligned}
&\hat{m}_{t+1} = [b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},Q) + b^{\text{mis}\,\prime}_{+}(Q',V_{f^{\text{mis}}})]\beta'V_R^{-1}\beta(f_{t+1} - \mathbb{E}(f_{t+1})) \\
&+ [b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},Q) + b^{\text{mis}\,\prime}_{+}(Q',V_{f^{\text{mis}}})]\beta'V_R^{-1}e_{t+1} + c'_{+}V_eV_R^{-1}\beta(f_{t+1} - \mathbb{E}(f_{t+1})) + c'_{+}V_eV_R^{-1}e_{t+1} \\
&\xrightarrow{p} [b^{\text{can}\,\prime}_{+}(V_{f^{\text{can}}},Q) + b^{\text{mis}\,\prime}_{+}(Q',V_{f^{\text{mis}}})]V^{-1}(f_{t+1} - \mathbb{E}(f_{t+1})) + c'_{+}e_{t+1} \\
&= (b^{\text{can}\,\prime}_{+}, b^{\text{mis}\,\prime}_{+})\begin{pmatrix} V_{f^{\text{can}}} & Q \\ Q' & V_{f^{\text{mis}}} \end{pmatrix}V^{-1}(f_{t+1} - \mathbb{E}(f_{t+1})) + c'_{+}e_{t+1} = (b^{\text{can}\,\prime}_{+}, b^{\text{mis}\,\prime}_{+})(f_{t+1} - \mathbb{E}(f_{t+1})) \\
&= b^{\text{can}\,\prime}_{+}(f^{\text{can}}_{t+1} - \mathbb{E}(f^{\text{can}}_{t+1})) + b^{\text{mis}\,\prime}_{+}(f^{\text{mis}}_{t+1} - \mathbb{E}(f^{\text{mis}}_{t+1})) - a'V_e^{-1}e_{t+1} = m_{t+1}, \qquad \text{as} \quad N \to \infty.
\end{aligned}$$

Similarly, $\beta'V_R^{-1}\beta \xrightarrow{p} V^{-1}$ as $N \to \infty$ implies $\hat{m}^{\beta,\text{can}} \xrightarrow{p} m^{\beta,\text{can}}$, $\hat{m}^{\beta,\text{mis}} \xrightarrow{p} m^{\beta,\text{mis}}$, and $\hat{m}^{a} \xrightarrow{p} m^{a}$, as $N \to \infty$; thus, yielding $\hat{m} \xrightarrow{p} m$, as $N \to \infty$. Taken together that $\hat{m}_{t+1} \xrightarrow{p} m_{t+1}$ and $\hat{m} \xrightarrow{p} m$, as $N \to \infty$, we obtain $\hat{M}_{\text{exp},t+1} \xrightarrow{p} M_{\text{exp},t+1}$, as $N \to \infty$.

Note that because f^{can}_{t+1} are observable factors, in empirical work we may use the exact component $m^{\beta,\text{can}}_{t+1} + m^{\beta,\text{can}}$ rather than its projected counterpart $\hat{m}^{\beta,\text{can}}_{t+1} + \hat{m}^{\beta,\text{can}}$.

IA.2 Estimation of the APT: The General Case

We now explain how to estimate the APT allowing for both tradable and nontradable factors, and for both asset-specific pricing errors and pricing errors arising from omitted systematic risk factors. Assume that

$$R_{t+1} - R_{ft}1_N = a + \beta^{\text{mis}}\lambda^{\text{mis}} + \beta_1^{\text{can}}(\lambda_1^{\text{can}} + f_{1t+1}^{\text{can}} - \mathbb{E}(f_{1t+1}^{\text{can}})) + \beta_2^{\text{can}} f_{2t+1}^{\text{can}} + \varepsilon_{t+1}, \quad \text{with}$$
$$\alpha = a + \beta^{\text{mis}}\lambda^{\text{mis}}, \;\; \varepsilon_{t+1} = \beta^{\text{mis}}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + e_{t+1},$$

where $\beta^{\text{can}} = (\beta_1^{\text{can}}, \beta_2^{\text{can}})$, $V_f = \text{var}(f_{t+1})$, $f_{t+1} = (f'_{1t+1}, f^{e\prime}_{2t+1})'$, with f_{1t+1} denoting the set of K_1^{can} nontradable observed factors and f_{2t+1}^e the set of K_2^{can} tradable observed factors, expressed as excess returns, with $K^{\text{can}} = K_1^{\text{can}} + K_2^{\text{can}}$. We assume that the missing factors are uncorrelated with the observed factors.[29] Given that f_{2t}^{can} are excess returns on tradable assets, their risk premia satisfy $\lambda_2^{\text{can}} = \mathbb{E}(f_{2t}^{\text{can}})$ and, to avoid confusion with the risk premia of the nontradable assets λ_1^{can}, we will use the expectation formulation for λ_2^{can}.

The joint log-likelihood function takes the following form:

$$\begin{aligned}
L(\tilde{\theta}) &= -\frac{1}{2}\log(\det(\tilde{\beta^{\text{mis}}}\tilde{\beta^{\text{mis}}}' + \tilde{V}_e)) \qquad \text{(IA1)}\\
&- \frac{1}{2T}\sum_{t=1}^{T}\Big(R_t - R_{ft}1_N - \tilde{\beta^{\text{mis}}}\tilde{\lambda}^{\text{mis}} - \tilde{a} - \tilde{\beta}_1^{\text{can}}(\tilde{\lambda}_1^{\text{can}} + f_{1t}^{\text{can}} - \tilde{\mathbb{E}}(f_{1t}^{\text{can}})) - \tilde{\beta}_2^{\text{can}} f_{2t}^{\text{can}}\Big)'\\
&\times (\tilde{\beta^{\text{mis}}}\tilde{\beta^{\text{mis}}}' + \tilde{V}_e)^{-1}\Big(R_t - R_{ft}1_N - \tilde{\beta^{\text{mis}}}\tilde{\lambda}^{\text{mis}} - \tilde{a} - \tilde{\beta}_1^{\text{can}}(\tilde{\lambda}_1^{\text{can}} + f_{1t}^{\text{can}} - \tilde{\mathbb{E}}(f_{1t}^{\text{can}})) - \tilde{\beta}_2^{\text{can}} f_{2t}^{\text{can}}\Big)\\
&- \frac{1}{2}\log(\det(\tilde{V}_f)) - \frac{1}{2T}\sum_{t=1}^{T}(f_t - \tilde{\mathbb{E}}(f_t))'\tilde{V}_f^{-1}(f_t - \tilde{\mathbb{E}}(f_t)).
\end{aligned}$$

Without loss of generality, one can assume that the missing factors have unit variance, that is, $\text{var}(f_t^{\text{mis}}) = I_{K^{\text{mis}}}$, achieving identification of β^{mis}.

Proposition IA.1 (Parameter estimation of APT: General Case)**.** *Suppose that the vector of asset returns, R_t, satisfies Assumption 1 and that $\hat{M}_{f_2^{\text{can}}} - \bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime}$ is nonsingular, where $\hat{M}_{f_2^{\text{can}}} = T^{-1}\sum_{t=1}^{T} f_{2t}^{\text{can}} f_{2t}^{\text{can}\prime}$ and $\bar{f}_2^{\text{can}} = T^{-1}\sum_{t=1}^{T} f_{2t}^{\text{can}}$. Then*

$$\hat{\theta} = \underset{\tilde{\theta}}{\text{argmax}} \;\; L(\tilde{\theta}) \quad \textit{subject to} \quad \tilde{a}'\tilde{V}_\varepsilon^{-1}\tilde{a} \le \delta_{\text{apt}},$$

where $L(\tilde{\theta})$ is defined in (IA1), and $\hat{\theta} = (\hat{a}', \hat{\lambda^{\text{mis}}}', \hat{\lambda}_1^{\text{can}\prime}, \hat{\mathbb{E}}(f_{1t}^{\text{can}})', \hat{\mathbb{E}}(f_{2t}^{\text{can}})', \text{vec}(\hat{\beta^{\text{mis}}})', \text{vec}(\hat{\beta}^{\text{can}})', \text{vech}(\hat{V}_e)', \text{vech}(\hat{V}_f)')'$.

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} > 0$, then setting

$$D = (\beta^{\text{mis}}, \beta_1^{\text{can}}), \;\; \lambda = (\lambda^{\text{mis}\prime}, \lambda_1^{\text{can}\prime})'$$

and using $\otimes$ to denote the Kronecker product,

$$\text{vec}(\hat{B}_2^{\text{can}}) = \Big((\hat{M}_{f_2^{\text{can}}} \otimes I_N) - (\bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime} \otimes \hat{G})\Big)^{-1} \text{vec}\Big(\hat{M}_{hf_2^{\text{can}}} - \hat{G}\bar{h}\bar{f}_2^{\text{can}\prime}\Big), \qquad \text{(IA2)}$$

[29] The estimator can be extended to the case of correlated observed and omitted risk factors; details are available upon request.

$$\hat{\lambda} = (\hat{D}'\,\hat{V}_\varepsilon^{-1}\,\hat{D})^{-1}\hat{D}'\,\hat{V}_\varepsilon^{-1}\Big(\bar{h} - \hat{\beta}_2^{\text{can}}\;\bar{f}_2^{\text{can}}\Big),$$

$$\hat{a} = \frac{1}{\hat{\kappa}+1}\Big(\bar{h} - \hat{\beta}_2^{\text{can}}\;\bar{f}_2^{\text{can}} - \hat{D}\hat{\lambda}\Big),$$

where $\hat{V}_\varepsilon = \hat{\beta^{\text{mis}}}\hat{\beta^{\text{mis}}}' + \hat{V}_e$, $\hat{M}_{hf_2^{\text{can}}} = T^{-1}\sum_{t=1}^{T} h_t f_{2t}^{\text{can}\,\prime}$, $\bar{h} = T^{-1}\sum_{t=1}^{T} h_t$ *with* $h_t = R_t - R_f 1_N - \hat{\beta}_1^{\text{can}}(f_{1t}^{\text{can}} - \bar{f}_1^{\text{can}})$ *and* $\bar{f}_1^{\text{can}} = T^{-1}\sum_{t=1}^{T} f_{1t}^{\text{can}}$, *and*

$$\hat{G} = \frac{1}{(\hat{\kappa}+1)} I_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)}\hat{D}(\hat{D}'\,\hat{V}_\varepsilon^{-1}\,\hat{D})^{-1}\hat{D}'\,\hat{V}_\varepsilon^{-1}.$$

Note that $\hat{D} = (\hat{\beta^{\text{mis}}}, \hat{B}_1^{\text{can}})$ *and* $\hat{V}_e$ *do not admit closed-form solutions, and, as before,* $\hat{\mathbb{E}}(f_t)$ *and* $\hat{V}_f$ *coincide with the sample mean and sample covariance of the observed factors,* f_t*.*

(ii) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} = 0$ *one can estimate only* $\alpha_N = a + D\lambda$ *but not the two components separately, and one obtains*

$$\hat{\alpha}_{N,MLC} = \bar{R} - \bar{R}_f 1_N - \hat{\beta}_2^{\text{can}}\;\bar{f}_2^{\text{can}},$$

in which the expression for $\text{vec}(\hat{\beta}_2^{\text{can}})$ *can be obtained by setting* $\hat{\kappa} = 0$ *in the terms that appear in (IA2). The expressions for* $\hat{\mathbb{E}}(f_t)$ *and* $\hat{V}_f$ *are unchanged, and, as before, the expressions for the estimators of* $\hat{D}$ *and* $\hat{V}_e$ *do not admit a closed-form solution.*

Proof. Within this proof, for simplicity, we do not use the $\tilde{\cdot}$ notation to denote feasible parameter values.

Defining by $\tilde{\theta}$ the MLC corresponding to $\hat{\kappa} = 0$, this is unfeasible whenever we have that $\tilde{a}'\tilde{V}_\varepsilon^{-1}\tilde{a} > \delta_{\text{apt}}$. Similarly, case $\hat{\kappa} > 0$ is unfeasible whenever,

$$\Big(\bar{R} - \bar{R}_f 1_N - \hat{\beta}_2^{\text{can}}\;\bar{f}_2^{\text{can}} - \hat{D}\,\hat{\lambda}\Big)'\hat{V}_\varepsilon^{-1}\Big(\bar{R} - \bar{R}_f 1_N - \hat{\beta}_2^{\text{can}}\;\bar{f}_2^{\text{can}} - \hat{D}\,\hat{\lambda}\Big) < \delta_{\text{apt}},$$

because $(1+\hat{\kappa})^2 = \frac{\left[\bar{R}-\bar{R}_f 1_N-\hat{\beta}_2^{\text{can}}\,\bar{f}_2^{\text{can}}-\hat{D}\,\hat{\lambda}\right]'\hat{V}_\varepsilon^{-1}\left[\bar{R}-\bar{R}_f 1_N-\hat{\beta}_2^{\text{can}}\,\bar{f}_2^{\text{can}}-\hat{D}\,\hat{\lambda}\right]}{\delta_{\text{apt}}}$. When both cases are feasible, the optimal value for the Karush-Kuhn-Tucker multiplier will be greater than zero or equal to zero, depending on which case maximizes the log-likelihood, namely depending on whether $L(\hat{\theta})$ or $L(\tilde{\theta})$ is largest, respectively. Note that when $\kappa > 0$ then $\hat{a}'\hat{V}_\varepsilon^{-1}\hat{a} = \delta_{\text{apt}}$ by construction.

We now derive the expressions for the estimators. Assume for now that case $\hat{\kappa} > 0$ holds. Differentiating the penalized log-likelihood with respect to λ, a, and κ, the first $K^* + N$ equations, setting $K^* = K^{\text{mis}} + K_1^{\text{can}}$, (after some algebra) are:

$$\begin{pmatrix} D'V_\varepsilon^{-1} \\ I_N \end{pmatrix}\Big(\bar{R} - \bar{R}_f 1_N - \beta_2^{\text{can}}\bar{f}_2^{\text{can}}\Big) = \begin{pmatrix} D'V_\varepsilon^{-1}D & D'V_\varepsilon^{-1} \\ D & (1+\hat{\kappa})I_N \end{pmatrix}\begin{pmatrix} \hat{\lambda} \\ \hat{a} \end{pmatrix},$$

where, recall that $V_\varepsilon = \beta^{\text{mis}}\beta^{\text{mis}\prime} + V_e$, and noting that all the expressions above and below are left as function of the feasible values for V_e and D (as opposed to their MLC values). Because of the APT restriction, λ and a can be identified separately, as long as $\hat{\kappa} > 0$. In fact, the above system of linear equations can be solved because the matrix pre-multiplying $\hat{\lambda}$ and $\hat{a}$ is nonsingular for every $\hat{\kappa} > 0$, leading to the closed-form solution:

$$\hat{\lambda} = (D'V_\varepsilon{}^{-1}D)^{-1}D'V_\varepsilon{}^{-1}\Big(\bar{R} - \bar{R}_f 1_N - \beta_2^{\text{can}}\bar{f}_2^{\text{can}}\Big), \tag{IA3}$$

$$\hat{a} = \frac{1}{\hat{\kappa}+1}\Big(\bar{R} - \bar{R}_f 1_N - \beta_2^{\text{can}}\bar{f}_2^{\text{can}} - D\hat{\lambda}\Big). \tag{IA4}$$

Turning now to the first-order condition with respect to the generic (a,b)th element of $\beta_{2N,\text{MLC}}$, denoted by B_{2ab} with $1 \le a \le N,\ 1 \le b \le K_2^{\text{can}}$, one obtains

$$\frac{1}{T}\sum_{t=1}^{T}\Big(R_t - R_{ft}1_N - \beta^{\text{mis}}\lambda^{\text{mis}} - a - \beta_1^{\text{can}}(\lambda_1^{\text{can}} + f_{1t}^{\text{can}} - \bar{f}_1^{\text{can}}) - \hat{\beta}_2^{\text{can}} f_{2t}^{\text{can}}\Big)' V_\varepsilon{}^{-1}(-\frac{\partial \beta_2^{\text{can}}}{\partial B_{2ab}} f_{2t}^{\text{can}}) = 0.$$

Inserting (IA3) and (IA4) into the above expression, setting $\hat{M}_{f_1^{\text{can}} f_2^{\text{can}}} = 1/T\sum_{t=1}^{T} f_{1t}^{\text{can}} f_{2t}^{\text{can}\prime}$,

$$G = \frac{1}{(\hat{\kappa}+1)} I_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)} D(D'V_\varepsilon{}^{-1}D)^{-1}D'V_\varepsilon{}^{-1},$$

and rearranging terms, yields

$$\hat{\beta}_2^{\text{can}}\hat{M}_{f_2^{\text{can}}} - G\hat{\beta}_2^{\text{can}}\bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime} = \hat{M}_{Rf_2^{\text{can}}} - G(\bar{R} - \bar{R}_f 1_N)\bar{f}_2^{\text{can}\prime} - \beta_1^{\text{can}}(\hat{M}_{f_1^{\text{can}} f_2^{\text{can}}} - \bar{f}_1^{\text{can}}\bar{f}_2^{\text{can}\prime}),$$

which can be rewritten more succinctly as

$$\frac{1}{T}\sum_{t=1}^{T} f_{2t}^{\text{can}} g_t' = 0_{K_2^{\text{can}}\times N},$$

with $g_t = \Big(h_t - G\bar{h} - \hat{\beta}_2^{\text{can}} f_{2t}^{\text{can}} + G\hat{\beta}_2^{\text{can}}\bar{f}_2^{\text{can}}\Big)$. Taking the vec operator and solving for $\hat{\beta}_2^{\text{can}}$ gives the desired expression in (IA2).

We need to show that a solution for $\hat{\beta}_2^{\text{can}}$ exists. This requires one to establish that the matrix $\Big((\hat{M}_{f_2^{\text{can}}} \otimes I_N) - (\bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime} \otimes G)\Big)$ is invertible. This matrix can be written as

$$\Big((\hat{M}_{f_2^{\text{can}}} \otimes I_N) - (\bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime} \otimes G)\Big) = \Big(((\hat{M}_{f_2^{\text{can}}}\bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime}) \otimes I_N) + (\bar{f}_2^{\text{can}}\bar{f}_2^{\text{can}\prime} \otimes (I_N - G))\Big).$$

Given the assumptions made, the first matrix on the right hand side is nonsingular. One then just needs to show that the second matrix is positive semi-definitive, which follows from the proof of Theorem A2.

Therefore, plugging $\hat{\beta}_2^{\text{can}} = \hat{\beta}_2^{\text{can}}(D, V_e)$ into $\hat{\lambda}$ and $\hat{a}$, and then $\hat{\lambda}$ into $\hat{a}$, one obtains that

$$\hat{\beta}_2^{\text{can}} = \hat{\beta}_2^{\text{can}}(D, V_e), \hat{\lambda} = \hat{\lambda}(D, V_e), \quad \hat{a} = \hat{a}(D, V_e) \quad \text{and} \quad \hat{\kappa} = \hat{\kappa}(D, V_e).$$

Substituting them into $L(\theta) - \kappa(a' {V_\varepsilon}^{-1} a - \delta_{\text{apt}})$, gives the concentrated log-likelihood function, which is a function of only D and V_e and it will be maximized numerically to obtain $\hat{D}$ and $\hat{V}_e$. Observe that the penalization term vanishes for the concentrated log likelihood function for both $\hat{\kappa} = 0$ and $\hat{\kappa} > 0$.

(ii) Suppose now that $\hat{\kappa} = 0$. One can clearly obtain a unique solution for $(D, I_N)\begin{pmatrix} \tilde{\lambda} \\ \tilde{a} \end{pmatrix} = D\tilde{\lambda} + \tilde{a}$. However, to solve for $\tilde{\lambda}$ and $\tilde{a}$ separately, one needs to invert the matrix

$$\begin{pmatrix} D' {V_\varepsilon}^{-1} \\ I_N \end{pmatrix} (D, I_N) = \begin{pmatrix} D' {V_\varepsilon}^{-1} D & D' {V_\varepsilon}^{-1} \\ D & I_N \end{pmatrix},$$

which is not possible because it is of dimension $(N + K^*) \times (N + K^*)$ but of rank N, because the left-hand side shows that it is obtained from the product of two matrices of dimension $(N + K^*) \times N$. Thus, only the sum $D\tilde{\lambda} + \tilde{a}$ can be estimated. However, all the other parameters are identified and their expressions follow from differentiating $L(\theta)$ and solving the resulting first-order conditions. For instance, the formula for $\tilde{\beta}_2^{\text{can}}$ follows from setting $G = I_N$ into (IA2). □

IA.3 Data description

To examine which economic variables may explain variation in the SDF, we collect a set of 457 traded variables and 103 nontraded variables.

The set of traded variables contains:

- 205 factors from Chen and Zimmermann (2021).
- 153 factors in the Global Factor Dataset from Jensen, Kelly, and Pedersen (2021).
- 55 factors from Kozak, Nagel, and Santosh (2020).
- 35 factors from Bryzgalova, Huang, and Julliard (2020), and the sources of these factors are specified in their Internet Appendix. Their dataset contains a list of 51 variables, 34 of which are traded. We have one additional factor because we consider two versions of SMB, one from FF3 and the other from FF5.
- We add the following 9 factors:
  - Industry-adjusted value, momentum, and profitability factors; intra-industry value, momentum, and profitability factors; profitable-minus-unprofitable factor from Novy-Marx (2013), available from http://rnm.simon.rochester.edu.

  - Expected-growth factor of Hou, Mo, Xue, and Zhang (2021), available from https://global-q.org/index.html.
  - Up-minus-down (UMD) factor from the AQR data library, available from https://www.aqr.com/Insights/Datasets.

The set of nontraded variables contains:

- 53 factors from Bryzgalova, Huang, and Julliard (2020), and the sources of these are given in their Internet Appendix. Their dataset contains a list of 51 variables, 17 of which are nontraded. Below we explain how we get to 53 variables.
  - For indices of financial uncertainty, real uncertainty and macroeconomic uncertainty, we consider time horizons of 1, 3 and 12 months. We use these variables in levels and also consider the AR(1) innovations of these variables, for a total of 18 variables.
  - For the investor-sentiment measures of Baker and Wurgler (2006) and Huang, Jiang, Tu, and Zhou (2015), labeled as BW_INV_SENT and HJTZ_INV_SENT, respectively. We consider both the orthogonalized and non-orthogonalized versions. We use these variables in levels and also consider AR(1) innovations of these variables, for a total of 8 variables.
  - For other persistent variables, such as the term spread (TERM), change in the difference between a 10-year Treasury bond yield and a 3-month Treasury bill yield (DELTA_SLOPE), credit spread (CREDIT), dividend yield (DIV), price-earnings ratio (PE), unemployment rate (UNRATE), growth rate of industrial production (IND_PROD), monthly growth rate of the Producer Price Index for Crude Petroleum (OIL), we look at both levels and first order differences, for a total of 16 variables.
  - Real per capita consumption growth on nondurable goods and services both separately and jointly. We also include the 3-year consumption growth (nondurable goods and services) together with its AR(1) innovations, for a total of 5 variables.
  - Inflation, computed as the log-difference in the price index corresponding to both nondurable goods and services, and its AR(1) innovations, for a total of 2 variables.
  - The level of the intermediary capital ratio and its innovations, for a total of 2 variables.
  - The level of the aggregate liquidity factor and its innovations.
- The first 3 principal components and their VAR(1) innovations for the 279 macroeconomic variables from Jurado, Ludvigson, and Ng (2015), for a total of 6 variables.

- The first 8 principal components and their VAR(1) innovations for the 128 macroeconomic variables from the FRED-MD dataset of McCracken and Ng (2015), gives a total of 16 variables. We obtain these macro variables from `https://research.stlouisfed.org/econ/mccracken/fred-databases` and use the data vintage for February 2021. We exclude four variables, ACOGNO, ANDENOx, TWEXAFEGSMTHx, UMCSENT, which have missing observations at the start of the sample.

- Consumer sentiment and its first-order differences.

- Market-dislocation index of Pasquariello (2014), its first-order differences, and AR(1) innovations.

- The disagreement index of Huang, Li, and Wang (2021) and its first-order differences.

- The Chicago Board Options Exchange (CBOE) volatility index (VIX) available on the website of the CBOE, its first order differences, and AR(1) innovations.

- The U.S. economic policy uncertainty index (EPU) of Baker, Bloom, and Davis (2016) and the equity market volatility (EMV) tracker of Baker, Bloom, Davis, and Kost (2019), which are available from `www.policyuncertainty.com`. For both indices, we also consider their first-order differences and AR(1) innovations.

- The U.S. business-confidence index, the U.S. consumer-confidence index, and the U.S. composite leading indicator from the OECD library.

- The coincident economic-activity index and its first-order differences, downloaded from `https://fred.stlouisfed.org/series/USPHCI`.

- The NBER recession index, downloaded from `https://fred.stlouisfed.org/series/USREC`.

- The TED spread, downloaded from `https://fred.stlouisfed.org/series/TEDRATE`.

- The effective federal funds rate and the real federal funds rate, downloaded from `https://fred.stlouisfed.org/series/FEDFUNDS`.

- The credit-spread index of Gilchrist and Zakrajšek (2012) and its first order differences.

- The Chicago Fed National Financial Condition Index, downloaded from `https://fred.stlouisfed.org/series/NFCI`.

IA.4 Additional Figures

Figure IA.1: Individual regressions of $\log \hat{M}^a_{\text{exp},t+1}$ on excess returns of trading strategies and nontradable (macroeconomic) variables

This figure shows the R^2 of individual regressions of $\log \hat{M}^a_{\text{exp},t+1}$ on excess returns of trading strategies and nontradable (macroeconomic) factors. We see from the figure that while M^a has a high R^2 with the excess returns of several trading strategies, the R^2 for nontradable (macroeconomic) factors is low.

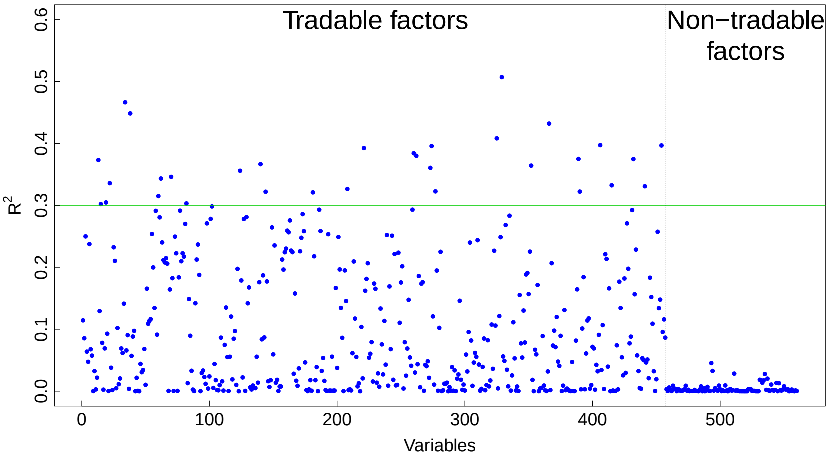


Figure IA.2: Individual regressions of $\log \hat{M}^{\beta,\text{mis}}_{\text{exp},t+1}$ on excess returns of trading strategies and nontradable (macroeconomic) variables

This figure shows the R^2 of individual regressions of $\log \hat{M}^{\beta,\text{mis}}_{\text{exp},t+1}$ on excess returns of trading strategies and nontradable (macroeconomic) factors. The (top-left corner of the) figure shows that $M^{\beta,\text{mis}}$ has a very high R^2 (95%) with one strategy, which is the market factor. It also has a high R^2 with some of the nontradable (macroeconomic) factors.

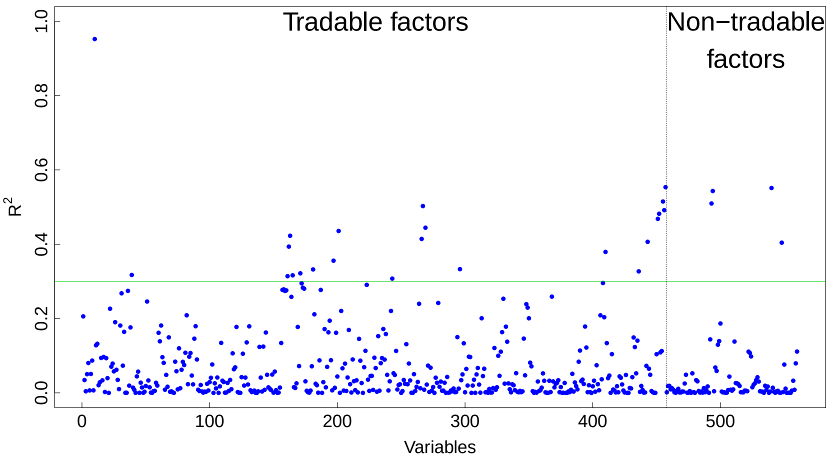


Figure IA.3: Time series of SDF and its components for the corrected C-CAPM candidate model

This figure has four panels, which show the dynamics of the SDF $M_{\exp,t+1}$ and its three components: the SDF-U, the component $M^{\beta,\text{can}}_{\exp,t+1}$ corresponding to the candidate model with a consumption-mimicking portfolio, and the missing systematic component $M^{\beta,\text{mis}}_{\exp,t+1}$.

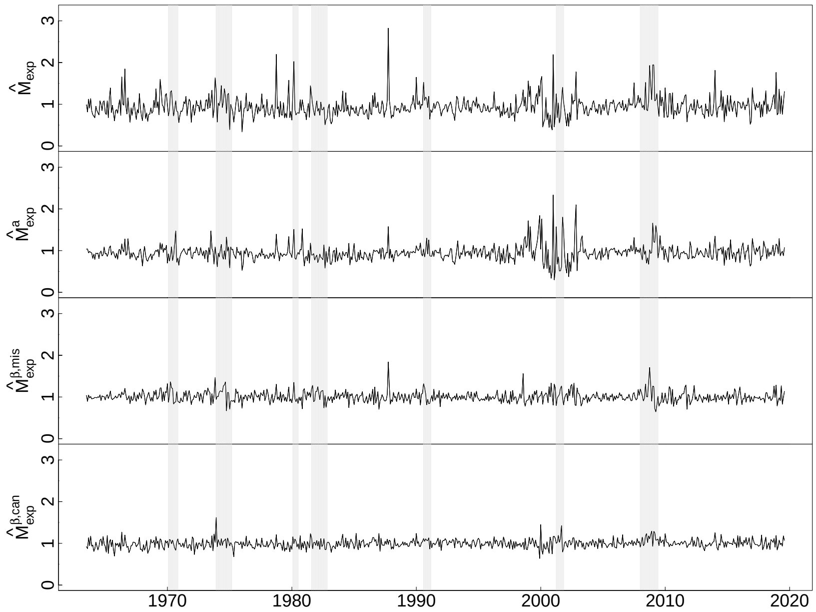


Figure IA.4: Pricing performance of the model with a consumption-mimicking portfolio with and without correction

This figure compares the pricing performance of the APT model to the model with a consumption-mimicking portfolio, where $a = 0$ and the only systematic risk factor is the consumption-mimicking portfolio of Breeden, Gibbons, and Litzenberger (1989).

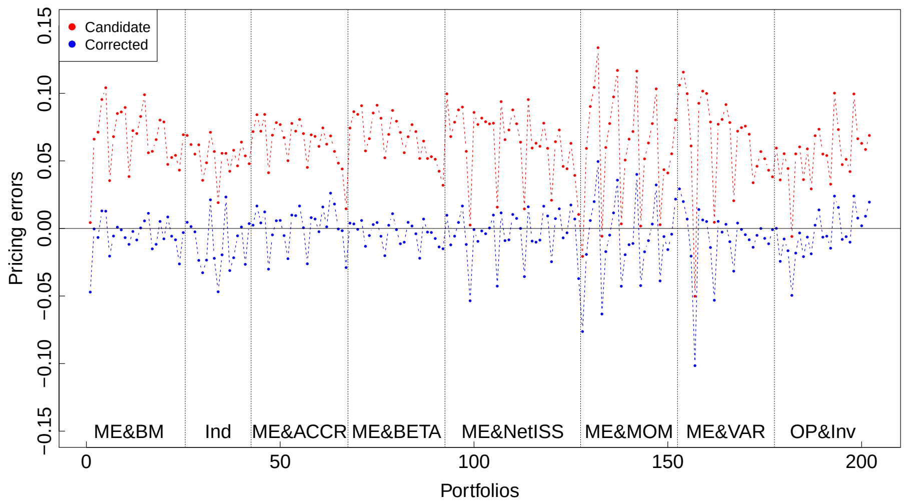


Figure IA.5: Time series of the SDF and its components for the corrected FF3 candidate model
This figure has four panels, which show the dynamics of the SDF $M_{\exp,t+1}$ and its three components: the SDF-S $M^{a}_{\exp,t+1}$, the component $M^{\beta,\mathrm{can}}_{\exp,t+1}$ corresponding to the candidate FF3 model, and the missing systematic component $M^{\beta,\mathrm{mis}}_{\exp,t+1}$.