

What is Missing in Asset-Pricing Factor Models?*

Massimo Dello Preite Raman Uppal Paolo Zaffaroni Irina Zviadadze

March 29, 2022

Abstract

Our objective is to develop a methodology to price the cross section of asset returns. Despite the hundreds of *systematic* risk factors considered in the literature (“factor zoo”), there is still a sizable pricing error. Our key insight is that the pricing error is compensation not just for omitted systematic risk but also *asset-specific risk*. We use this insight to construct a stochastic discount factor (SDF) that prices the cross section of stock returns *exactly*, and therefore, resolves the factor zoo. Empirically, we demonstrate that more than half of the variation in this SDF is explained by an aggregate measure of asset-specific risk that reflects market frictions and behavioral biases.

JEL classification: C52, C58, G11, G12.

Keywords: Asset-specific 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.

1 Introduction

A major challenge in asset pricing is to explain the cross section of asset returns. To achieve this goal, the literature has searched for alternative systematic risk factors, leading to a factor zoo (Cochrane, 2011). However, virtually all examined factor models featuring different factors from this zoo imply sizable pricing errors in asset returns, called alpha. In our work, we develop a methodology that resolves the factor zoo by demonstrating that the alpha is compensation not just for sources of systematic risk that are missing in a candidate factor model but also for asset-specific risk. This insight allows us to price the cross-section of stock returns exactly.

The foundation for our analysis is the Arbitrage Pricing Theory (APT) of Ross (1976, 1977), with a more formal treatment in Chamberlain (1983) and Chamberlain and Rothschild (1983). The APT postulates a linear factor model for the deviations of asset returns from their means and allows for asset-specific components in expected returns. These asset-specific components are unrelated to common factors, also known as sources of systematic risk, and satisfy a no-arbitrage restriction. We exploit the APT to evaluate and correct misspecification in popular candidate linear factor models that are used by financial economists to price the cross-section of assets.

A candidate factor model may be misspecified for at least two reasons: it may omit (i) systematic sources of risk and (ii) asset-specific components in expected returns that should be present, in line, for example, with the argument of Daniel and Titman (1997). To identify what is missing in a candidate asset-pricing factor model, we evaluate misspecification in this model through the lens of the SDF. Our methodology shows how to address both sources of misspecification to construct an *admissible* SDF, namely an SDF that prices correctly a given set of assets.

The first main insight of our analysis is that the asset-specific components of expected returns should be interpreted not as pricing errors but as compensation for asset-specific risk represented by shocks to asset returns that are orthogonal to common risk factors. This insight is a major departure from the conventional wisdom that financial markets compensate investors only for exposure to systematic sources of risk. The standard view in finance, that one should diversify away all asset-specific risk, holds only because the return for bearing asset-specific risk is zero in popular factor models. However, if asset-specific

risk earns a non-zero reward then instead of diversifying this risk, an investor will optimally adjust her portfolio to reap this compensation.

To demonstrate empirical support for our insight, we examine what is missing in popular candidate factor models. To set the stage, we start our analysis with a model without any observed risk factors, so that the candidate SDF is just the inverse of the risk-free rate. Then, 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 [Breedon \(1979\)](#), and (iii) the three-factor model (FF3) of [Fama and French \(1993\)](#). Using data for 202 portfolios of monthly stock returns, for each of these models we identify and characterize the required correction term to move from the candidate to admissible SDF.

The correction term consists of two terms: a measure of systematic risk and an aggregate measure of asset-specific risk. The measure of systematic risk is a function of risk factors that are common in asset returns but missing in the candidate factor model. The aggregate measure of asset-specific risk is a portfolio of asset returns orthogonal to all common risk factors, including those missing in the candidate model. Despite starting from different candidate factor models, once we include the correction term for each of the candidate models listed above, we obtain admissible SDFs that are almost perfectly correlated.

A key finding that emerges from our empirical analysis is that the asset-specific components in expected returns are non-zero and represent reward for asset-specific risk. This risk plays a major role in pricing the cross-section of asset returns, despite the risk premia associated with each individual asset-specific shock being small on average. Specifically, the aggregate measure of asset-specific risk explains more than half of the variation in the admissible SDF. The quantitative role of the aggregate measure of asset-specific risk suggests that candidate factor models with different proxies for only common risk factors will never imply an admissible SDF. Instead, the only way to obtain an admissible SDF is to recognize that asset-specific risk is priced and we show how to do this. This insight resolves the factor zoo.

We explore the nature of the aggregate measure of asset-specific risk by examining the composition of the portfolio that represents it. We find that small stocks contribute substantially to the aggregate measure of asset-specific risk, followed by stocks with extremely low or high values of net issuances. We also measure the correlations of the aggregate measure of asset-specific risk with the returns on 472 trading strategies examined in the literature.

The strategies related to behavioral biases and financial frictions exhibit the largest correlations: the R^2 in the regressions of these strategies' returns on the aggregate measure of asset-specific risk exceeds 30 percent. The most prominent examples of these strategies are those based on 5-year analyst growth forecast (La Porta, 1996), betting-against-beta factor (Frazzini and Pedersen, 2014), long-term behavioral mispricing factor (Daniel, Hirshleifer, and Sun, 2020), the ratio of book debt to market equity (Bhandari, 1988), and implied equity duration (Dechow, Sloan, and Soliman, 2004).

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

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, Value (Fama and French, 2015), Investment (Hou, Xue, and Zhang, 2015), and Momentum (Jegadeesh and Titman, 1993) 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 asset-specific risk than with the systematic component of the SDF. These factors appear to be weak (Lettau and Pelger, 2020; Giglio, Xiu, and Zhang, 2021); that is, they affect only a small number of asset returns out of the cross-section of stock returns considered.

The implicit account for asset-specific risk in factor models via factors such as Value, Investment, or Momentum, suggests that perhaps adding an arbitrary combination of tradable factors to the candidate factor model can capture the aggregate-measure of asset specific risk. However, it is not the case. The aggregate measure of asset-specific risk is a weak factor in the cross-section of asset returns. Recovering this weak factor and its associated risk premia through the lens of a candidate factor model is not feasible statistically. We formally prove this result that constitutes the second main insight of our analysis.

Turning next to the analysis of the systematic component of the SDF, we find that 95% of its variation is explained by the market factor. 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.¹ 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. In fact, the systematic component of the SDF accounts for only 44% of the variation in the admissible SDF, so that 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 differences in expected stock returns are spanned by their exposures to nineteen traded factors, in addition to the aggregate measure of asset-specific risk. 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 adjusting for the market factor.

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 the 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 \(1991\)](#), in which they provide the minimum-variance bound that must be satisfied by any admissible SDF. [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 asset-specific risk in this correction.

A number of papers develop a non-parametric approach to correct misspecified SDF models. [Hansen and Jagannathan \(1997\)](#) provide the pricing factor that is the smallest additive nonparametric adjustment (in a least-squares sense) required to make a given SDF admissible. [Almeida and Garcia \(2012\)](#) provide an additive correction term that is based on minimum-discrepancy projections. [Ghosh, Julliard, and Taylor \(2017\)](#) provide a multiplicative correction using a Kullback-Leibler entropy-minimization approach. In order to get as close as possible to the true SDF, ideally one would like to estimate a projection of

¹See [Fama and French \(2004\)](#) for a review of this literature.

the SDF on a large number of assets. However, it is challenging to use these non-parametric approaches when the number of basis assets is large relative to the number of observations. Furthermore, these methodologies are silent about the role of common versus idiosyncratic shocks in the admissible SDF. In contrast, our approach, because it is founded on the APT, is designed to handle a large number of assets and to explicitly allow for asset-specific risk to matter in pricing asset returns.

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. We complement this literature by developing a methodology that allows us to study the importance of asset-specific risk when correcting a candidate SDF for misspecification.

From the perspective of financial economics, our work is related to the literature on the role of idiosyncratic risk in asset pricing. Early contributions include [Douglas \(1969\)](#), [Fama and MacBeth \(1973\)](#), [Levy \(1978\)](#), and [Lehmann \(1990\)](#), with more recent work by, among others, [Goyal and Santa-Clara \(2003\)](#), [Herskovic, Kelly, Lustig, and Van Nieuwerburgh \(2016\)](#), and [Mehra, Wahal, and Xie \(2021\)](#). These papers study the effect on risk premia of the volatilities of asset residuals that are obtained after risk-adjusting asset returns for the market or other commonly used risk factors. We complement this literature by establishing, through the lens of the SDF, a formal risk-return relation between asset-specific risk, represented by direct shocks to systematic risk-adjusted returns rather than their volatilities, and expected asset-specific returns. Our factorization of the SDF into systematic and asset-specific components is in line with a body of theoretical work, including [Levy \(1978\)](#), [Merton \(1987\)](#), and [Malkiel and Xu \(2006\)](#).

Our findings also emphasize the arguments of [MacKinlay \(1995\)](#) and [Daniel and Titman \(1997\)](#) about the importance of characteristics for understanding risk premia and inability of a factor model to explain a cross section of stock returns, but with two important 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 asset-specific risk.

The rest of the paper is organized as follows. Section 2 contains our main theoretical results underpinning the methodology that we develop to construct an admissible SDF in the presence of model misspecification. Section 3 provides details of how to estimate a correction

term for a candidate SDF in order to obtain the admissible SDF. Section 4 describes the data that we use to illustrate our methodology. Section 5 presents the empirical findings from applying our methodology to this data. We collect our conclusions in Section 6. Proofs for all the propositions along with additional results are collected in appendices.

2 From a Candidate to an Admissible SDF

The starting point for our analysis is a candidate asset pricing linear factor model, represented by a vector of K^{can} observable risk factors f_{t+1}^{can} . Potentially, any candidate model suffers from misspecification. We evaluate misspecification by reviewing the candidate model’s pricing performance measured by the covariance of the stochastic discount factor and asset returns. Because asset returns are given, effectively we evaluate misspecification in the SDF implied by the candidate factor model.

Our work is founded on the classical APT (Ross, 1976) to correct candidate misspecified SDFs.² 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, the APT is a flexible model that does not take a stand on what constitutes a pricing factor. Second, the APT is a no-arbitrage model. The absence of arbitrage opportunities implies the existence of an SDF.

In this section, we first review the classical APT. Next, we explain how through the lens of the APT we correct a misspecified SDF. We derive the closed-form expression for an admissible SDF and identify what the candidate SDF is missing in order to be an admissible SDF. 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 Arbitrage Pricing Theory (APT)

Let the N -dimensional vector $R_{t+1} = (R_{1,t+1}, R_{2,t+1}, \dots, R_{N,t+1})'$ denotes the vector of gross returns on the N risky assets. Let R_{ft} be the gross return on the risk-free asset. If a risk-free asset does not exist, one can use instead the return on the minimum-variance

²Chamberlain (1983), and Chamberlain and Rothschild (1983) provide a formal analysis of the APT.

portfolio or the return on the zero-beta portfolio. 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 $V_f = \text{var}(f_{t+1}) > 0$.

The classical APT builds on two assumptions.

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

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

where $\beta = (\beta_1, \beta_2, \dots, \beta_N)'$ is the $N \times K$ full-rank matrix of loadings of asset returns on the common factors, the vector 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. 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 $w = (w_1, w_2, \dots, w_N)'$, for which:*

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

where δ denotes an arbitrary positive scalar and 1_N denotes the $N \times 1$ vector of ones.³

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

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

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 $\lambda = \lim_{N \rightarrow \infty} (\beta' V_e^{-1} \beta)^{-1} \beta' V_e^{-1} (\mathbb{E}(R_{t+1}) - R_{ft}1_N)$. 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

$$a' V_e^{-1} a \leq \delta_{\text{apt}}^* < \infty, \quad (2)$$

as shown in Ross (1976), Huberman (1982), Chamberlain and Rothschild (1983), and Ingersoll (1984), where δ_{apt}^* is some arbitrary positive scalar. The main insight of our methodology, which we will describe shortly, is a different interpretation of the vector a . We show

³Throughout the paper, we use δ to denote an arbitrary positive scalar, not always taking the same value.

that through the lens of an SDF the vector a should be interpreted as compensation for asset-specific risk e_{t+1} , rather than the vector of pricing errors.

Relative to the APT, any standard candidate factor model with K^{can} observable risk factors f_{t+1}^{can} potentially suffers from two sources of misspecification. First, the candidate model may omit systematic risk factors, that is, $K^{\text{can}} < K$. Second, the candidate model does not allow for the asset-specific components of expected excess returns represented by the vector a in equation (1). In the data, some components of this vector a may be non-zero, and therefore the candidate factor model that assumes that $a = 0_N$, where 0_N denotes the $N \times 1$ vector of zeros, is misspecified. A popular example of a candidate model is the market model, in which $a^{\text{can}} = 0_N$ and the vector f_{t+1}^{can} includes only the market factor, with $K^{\text{can}} = 1$.

To understand the implications of model misspecification, consider a candidate model with $a^{\text{can}} = 0_N$ 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

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

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_{t+1}^{can} .

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 1 (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 $N \rightarrow \infty$, 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 $\tilde{\delta}_{\text{apt}}$ possibly different from δ_{apt}^* ,*

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

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

$$\alpha = \beta^{\text{mis}} \lambda^{\text{mis}} + a \quad \text{and} \quad V_\varepsilon = \text{var}(\varepsilon_{t+1}) = \beta^{\text{mis}} \beta^{\text{mis}'} + V_e. \quad (5)$$

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 due to their exposure to the systematic risk factors f_{t+1}^{mis} that are missing in the candidate model and the asset-specific shocks e_{t+1} .

Without loss of generality, given that f_{t+1}^{mis} are latent factors, and therefore we can rotate them freely and normalize in an arbitrary way, we assume that the factors f_{t+1}^{mis} are mutually orthogonal to the factors f_{t+1}^{can} , and f_{t+1}^{mis} has a $K^{\text{mis}} \times K^{\text{mis}}$ identity covariance matrix $V_{f^{\text{mis}}} = \text{var}(f_{t+1}^{\text{mis}}) = I_{K^{\text{mis}} \times K^{\text{mis}}}$. Therefore, the covariance matrix of asset returns can be represented as $V_R = \text{var}(R_{t+1}) = \beta^{\text{can}} V_{f^{\text{can}}} \beta^{\text{can}'} + \beta^{\text{mis}} \beta^{\text{mis}'} + V_e$, where $V_{f^{\text{can}}} = \text{var}(f_{t+1}^{\text{can}})$. Because it is difficult to empirically distinguish an exact from approximate factor structure of asset returns, we assume that the covariance matrix of shocks e_{t+1} , $V_e = \text{var}(e_{t+1})$, is diagonal.

2.2 The SDF in the Presence of Misspecification

Below, we derive the closed-form expression for the SDF in the presence of misspecification. This result complements [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. More importantly, we establish a class of admissible SDFs in the presence of misspecification in a 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.

2.2.1 A Candidate Model that is Correctly Specified

To set the stage for our analysis, we start by considering the case in which the *true* APT model given in expression (1) has K factors and $a = 0_N$ with the APT restriction (2) guaranteed to hold. Suppose the candidate model coincides with the true APT model, that

is, $K^{\text{can}} = K$ and thus there is no misspecification. Denote $M_{t+1}^{\beta, \text{can}}$ the SDF implied by the candidate model

$$M_{t+1}^{\beta, \text{can}} = \frac{1}{R_{ft}} - \frac{1}{R_{ft}} \lambda' V_f^{-1} (f_{t+1} - \lambda).$$

In the absence of misspecification, $M_{t+1}^{\beta, \text{can}}$ prices all the risky assets and the risk-free asset exactly:

$$\mathbb{E}[M_{t+1}^{\beta, \text{can}}(R_{it+1} - R_{ft})] = 0 \quad \text{for each asset } i \quad \text{and} \quad \mathbb{E}[M_{t+1}^{\beta, \text{can}} R_{ft}] = 1,$$

that is, $M_{t+1}^{\beta, \text{can}}$ coincides with the linear admissible SDF M_{t+1} under the APT

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

Because the admissible SDF M_{t+1} is spanned by the K -dimensional vector f_{t+1} , the asset-specific shocks e_{it+1} are not priced:

$$\text{cov}[M_{t+1} e_{it+1}] = 0 \quad \text{for each asset } i.$$

2.2.2 A Candidate Model with Omitted Factors

Consider again the case, in which the true APT model given in expression (1) has K factors and $a = 0_N$. However, the candidate model now has $K^{\text{can}} < K$ factors, that is, there are $K^{\text{mis}} = K - K^{\text{can}}$ omitted systematic risk factors in the candidate model. In this case, under the APT the admissible linear SDF M_{t+1} can be written as the sum of two components:

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

The component $M_{t+1}^{\beta, \text{can}}$, which is based on the candidate factor model, is

$$M_{t+1}^{\beta, \text{can}} = \frac{1}{R_{ft}} - \frac{1}{R_{ft}} \lambda^{\text{can}'} V_{f^{\text{can}}}^{-1} (f_{t+1}^{\text{can}} - \lambda^{\text{can}}),$$

and the correction term $M_{t+1}^{\beta, \text{mis}}$, which accounts for the missing systematic risk factors, is

$$M_{t+1}^{\beta, \text{mis}} = -\frac{1}{R_{ft}} \lambda^{\text{mis}'} V_{f^{\text{mis}}}^{-1} (f_{t+1}^{\text{mis}} - \lambda^{\text{mis}}),$$

where, as discussed earlier, we normalize $V_{f^{\text{mis}}} = I_{K^{\text{mis}} \times K^{\text{mis}}}$.⁴

From the perspective of the candidate model, the vector α from equation (3) is the vector of pricing errors

$$\mathbb{E}[M_{t+1}^{\beta, \text{can}}(R_{t+1} - R_{ft})] \times R_f = \alpha.$$

⁴Our normalization assumption about mutual orthogonality of the candidate risk factors f_{t+1}^{can} with the missing risk factors f_{t+1}^{mis} has no bearing for the admissible SDF.

However, from the point of view of the true model given in expression (1), the vector α represents the compensation for assets' exposures to f_{t+1}^{mis} , because it satisfies

$$-\text{cov}(M_{t+1}^{\beta, \text{mis}}, (R_{t+1} - R_{ft})) \times R_{ft} = \alpha = \beta^{\text{mis}} \lambda^{\text{mis}}.$$

The *admissible* SDF, just as in the previous case of the correctly specified candidate model, is fully spanned by the K risk factors f_{t+1} , and therefore, asset-specific shocks are not priced:

$$\text{cov}[M_{t+1} e_{it+1}] = 0 \quad \text{for each asset } i.$$

2.2.3 A Candidate Model with Omitted Factors and Asset-Specific Components in Expected Excess Returns

We now consider the general case, in which the true APT model given in expression (1) has K factors and a non-zero vector $a \neq 0_N$ that satisfies the no-arbitrage restriction (2), while the candidate model features *two* sources of misspecification. 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 non-zero vector a by assuming that $a^{\text{can}} = 0_N$. The admissible linear SDF is now the sum of $M_{t+1}^{\beta, \text{can}}$, which is the SDF based on the candidate factor model, and a correction term labeled the alpha-SDF M_{t+1}^{α} . The correction term has two components: $M_{t+1}^{\beta, \text{mis}}$ and M_{t+1}^a . The first component accounts for the omitted systematic risk factors f_{t+1}^{mis} , whereas the second component accounts for omitted asset-specific components a in expected returns.

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

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

$$M_{t+1}^{\beta, \text{can}} = \frac{1}{R_{ft}} - \frac{(\lambda^{\text{can}})' V_{f^{\text{can}}}^{-1}}{R_{ft}} (f_{t+1}^{\text{can}} - \lambda^{\text{can}}),$$

$$M_{t+1}^{\beta, \text{mis}} = -\frac{(\lambda^{\text{mis}})' V_{f^{\text{mis}}}^{-1}}{R_{ft}} (f_{t+1}^{\text{mis}} - \lambda^{\text{mis}}),$$

$$M_{t+1}^a = -\frac{a' V_e^{-1}}{R_{ft}} e_{t+1},$$

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

The SDF component M_{t+1}^a is a linear function (scaled by the risk-free rate) of asset-specific shocks e_{t+1} . As M_{t+1}^a constitutes part of the admissible SDF, we refer to it as an *aggregate* measure of asset-specific risk to emphasize that it represents an aggregate source of risk even though it consists of asset-specific shocks.

The correction components M_{t+1}^a and $M_{t+1}^{\beta, \text{mis}}$ 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_{t+1}^a and $M_{t+1}^{\beta, \text{mis}}$.

A central insight from the above proposition is that the interpretation of α is different in the candidate and corrected models. From the perspective of the candidate model, α represents the vector of pricing errors

$$\alpha = \mathbb{E}[M_{t+1}^{\beta, \text{can}}(R_{t+1} - R_{ft})] \times R_f,$$

whereas in the corrected model α represents the compensation for assets' exposures to the risk factors f_{t+1}^{mis} and asset-specific risk e_{t+1} :

$$\alpha = -\text{cov}(M_{t+1}^a, (R_{t+1} - R_{ft})) \times R_f.$$

In particular, in the corrected model the elements of the vector a represent the compensation for exposure to the asset-specific shocks e_{t+1} :

$$a = -\text{cov}(M_{t+1}^a, e_{t+1}) \times R_f = -\text{cov}(M_{t+1}^a, e_{t+1}) \times R_f.$$

This result paves the way towards a quantitative assessment of asset-specific risk in financial markets that we will explore in our empirical analysis.

The dependence of M_{t+1}^a on e_{t+1} 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 build an admissible SDF. In Appendix A, we show that M_{t+1}^a is a weak factor in the cross-section of asset returns, and therefore, even if it were possible to add a perfectly correlated with M_{t+1}^a observable variable as a factor to a candidate factor model, it would not have led to the admissible SDF. The risk premia associated with any weak factor can not be estimated accurately (Anatolyev and Mikusheva, 2021). We show below how to account for asset-specific risk and construct an accurate estimator of M_{t+1}^a , and thus, of an admissible SDF.

2.3 Nonnegative Feasible SDF

There are three problems in constructing an admissible SDF in practice. First, the linear SDF characterized in the previous section may not always be strictly positive, thus possibly leading to negative asset prices. Second, the components $M_{t+1}^{\beta, \text{mis}}$ and M_{t+1}^a depend on unobservable quantities, such as, 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.

2.3.1 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 (Ghosh, Julliard, and Taylor, 2017; Gourioux and Monfort, 2007). For obtaining closed-form solutions, we assume that asset returns are Gaussian and follow the second approach.

Proposition 3 (SDF: Exponential Case). *Under Assumptions 1 and 2 and the assumption that returns R_{t+1} are Gaussian, there exists an admissible SDF $M_{\text{exp}, t+1}$*

$$\begin{aligned} M_{\text{exp}, t+1} &= M_{\text{exp}, t+1}^{\beta, \text{can}} \times M_{\text{exp}, t+1}^a \times M_{\text{exp}, t+1}^{\beta, \text{mis}}, \quad \text{where} \\ M_{\text{exp}, t+1}^{\beta, \text{can}} &= \frac{1}{R_{ft}} \exp \left[-(\lambda^{\text{can}})' V_{f^{\text{can}}}^{-1} (f_{t+1}^{\text{can}} - \lambda^{\text{can}}) - \frac{1}{2} (\lambda^{\text{can}})' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}} \right], \\ M_{\text{exp}, t+1}^{\beta, \text{mis}} &= \exp \left[-(\lambda^{\text{mis}})' V_{f^{\text{mis}}}^{-1} (f_{t+1}^{\text{mis}} - \lambda^{\text{mis}}) - \frac{1}{2} (\lambda^{\text{mis}})' V_{f^{\text{mis}}}^{-1} \lambda^{\text{mis}} \right], \\ M_{\text{exp}, t+1}^a &= \exp \left[-a' V_e^{-1} e_{t+1} - \frac{1}{2} a' V_e^{-1} a \right], \end{aligned}$$

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

2.3.2 Projection SDF

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

and risky assets and obtain⁵

$$\hat{M}_{\text{exp},t+1}^a = \exp \left[-a' V_R^{-1} (R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2} a' V_R^{-1} a \right], \quad (6)$$

$$\hat{M}_{\text{exp},t+1}^{\beta, \text{mis}} = \exp \left[-(\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} (R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2} (\beta^{\text{mis}} \lambda^{\text{mis}})' V_R^{-1} \beta^{\text{mis}} \lambda^{\text{mis}} \right], \quad (7)$$

where

$$\begin{aligned} \mathbb{E}[R_{t+1}] - R_{ft} &= a + \beta^{\text{mis}} \lambda^{\text{mis}} + \beta^{\text{can}} \lambda^{\text{can}}, \\ V_R &= \beta^{\text{can}} V_{f\text{can}} \beta^{\text{can}'} + \beta^{\text{mis}'} \beta^{\text{mis}} + V_e. \end{aligned}$$

The component $M_{\text{exp},t+1}^{\beta, \text{can}}$ depends on the observable quantities, so that the projection non-negative admissible SDF takes the form

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

The next proposition shows that, as $N \rightarrow \infty$, the SDF components, $M_{\text{exp},t+1}^{\beta, \text{mis}}$ and $M_{\text{exp},t+1}^a$, and their corresponding projection versions, $\hat{M}_{\text{exp},t+1}^{\beta, \text{mis}}$, and $\hat{M}_{\text{exp},t+1}^a$, have the same limits. We denote the matrix of the loadings of returns on the candidate and missing factors by $\beta = (\beta^{\text{can}} \ \beta^{\text{mis}})$, an arbitrary $K \times K$ positive-definite matrix by $A > 0$, and use $\xrightarrow{p}$ to denote convergence in probability.⁶

Proposition 4 (Limiting properties of SDF projections). *Under Assumptions 1 and 2 and the conditions $N^{-1} \beta' V_\varepsilon^{-1} \beta \xrightarrow{p} A$ and $\beta' V_\varepsilon^{-1} a = o(N^{\frac{1}{2}})$, as $N \rightarrow \infty$,*

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

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

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

⁵The formulae (6) and (7) 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 an N asset returns that are non-Gaussian, is approximately log-normal as $N \rightarrow \infty$.

⁶Strictly 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.

2.4 Weak Factors

Finally, we consider the case, in which the candidate model includes all strong systematic risk factors but omits weak factors collected in a vector f_{t+1}^{mis} . Following [Lettau and Pelger \(2020\)](#), we define weak factors as factors that affect only a subset of the underlying assets or affect all assets only marginally. Mathematically, the loadings of asset returns on weak factors satisfy

$$\beta^{\text{mis}'} \beta^{\text{mis}} \xrightarrow{p} A,$$

where $A > 0$ is some arbitrary positive-definite matrix. This property implies that $\beta^{\text{mis}'} \beta^{\text{mis}} / N \xrightarrow{p} 0_{K^{\text{mis}} \times K^{\text{mis}}}$, where $0_{K^{\text{mis}} \times K^{\text{mis}}}$ is a $K^{\text{mis}} \times K^{\text{mis}}$ matrix of zeros. The proposition below shows that also in this case we can correct the candidate SDF to obtain the admissible SDF, even though it is not possible to estimate consistently missing weak factors ([Lettau and Pelger \(2020, proposition 2\)](#)).

Proposition 5 (Properties of $\hat{M}_{\text{exp},t+1}^\alpha$, when missing factors are weak and $N \rightarrow \infty$).

Under Assumptions 1 and 2 and conditions $N^{-1} \beta' V_\varepsilon^{-1} \beta \xrightarrow{p} A$, $\beta^{\text{mis}'} V_\varepsilon^{-1} \beta^{\text{mis}} = O(1)$, $\beta^{\text{mis}'} V_\varepsilon^{-1} \beta = o(N^{\frac{1}{2}})$, $\beta' V_\varepsilon^{-1} a = o(N^{\frac{1}{2}})$, $\beta^{\text{mis}'} V_\varepsilon^{-1} a = O(1)$, and β^{mis} and β are not asymptotically collinear, then as $N \rightarrow \infty$,

$$\hat{M}_{\text{exp},t+1}^\alpha - M_{\text{exp},t+1}^\alpha \xrightarrow{p} 0 \quad \text{with} \quad M_{\text{exp},t+1}^\alpha = \exp \left[-\alpha' V_\varepsilon^{-1} \varepsilon_{t+1} - \frac{1}{2} \alpha' V_\varepsilon^{-1} \alpha \right].$$

This proposition highlights that weak factors and pure asset-specific risk (shocks uncorrelated across asset returns) cannot be identified separately. The presence of weak factors is compatible with the approximate factor structure in the APT of [Chamberlain and Rothschild \(1983\)](#). The approximate factor structure requires that the covariance matrix of returns adjusted for strong factors has uniformly bounded eigenvalues, limiting the degree of cross-sectional dependence. Proposition 5 clarifies that, although latent weak factors cannot be estimated, let alone their corresponding risk premia, we can still estimate the component M_{t+1}^a , which is a function of the priced weak factors and pure asset-specific risk.

3 Estimation Details

In this section, we describe our approach for estimating the admissible SDF, explain the role of the no-arbitrage restriction (4), and discuss how to identify the number of missing

factors in a candidate factor model and how to choose the no-arbitrage bound δ'_{apt} . Also, we provide a diagnostic tool for detecting missing factors in the candidate factor model.

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).⁷ We consider the case of a factor model for asset returns that includes only tradable factors represented by either factor returns in excess of the risk-free rate (for example, market factor) or long-minus-short strategies. In the second step, we use the results in Proposition 3 and Proposition 4 to recover the nonnegative feasible admissible SDF. The Online Appendix contains a more general case, in which (i) the candidate model for asset returns includes both tradable and non-tradable factors and (ii) the risk factors in the candidate model are allowed to be correlated with the missing systematic risk factors.

3.1 Formulating the Likelihood

For a generic vector Θ that collects all the elements of the matrices β^{can} , β^{mis} , V_ε , $V_{f^{\text{can}}}$, 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, $R_{t+1} - R_{ft}$, and observable factors f_{t+1}^{can} is

$$\begin{aligned} \log(L(\Theta)) = & -\frac{T}{2} \log(|V_\varepsilon|) - \frac{T}{2} \log(|V_{f^{\text{can}}}|) \\ & - \frac{1}{2} \sum_{t=0}^{T-1} \varepsilon'_{t+1} V_\varepsilon^{-1} \varepsilon_{t+1} - \frac{1}{2} \sum_{t=0}^{T-1} (f_{t+1}^{\text{can}} - \lambda^{\text{can}})' V_{f^{\text{can}}}^{-1} (f_{t+1}^{\text{can}} - \lambda^{\text{can}}), \end{aligned}$$

where $\varepsilon_{t+1} = (R_{t+1} - R_{ft}) - a - \beta^{\text{mis}} \lambda^{\text{mis}} - \beta^{\text{can}} f_{t+1}^{\text{can}}$.

We maximize this log-likelihood function subject to the no-arbitrage restriction. We substitute the no-arbitrage restriction given in expression (4) with the one that is computationally simpler to handle

$$a' V_\varepsilon^{-1} a \leq \delta_{\text{apt}} \tag{8}$$

and that ensures the no-arbitrage restriction (4), when $N \rightarrow \infty$.

We use the Lagrange multiplier method to solve the resulting constrained optimization problem:

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

⁷For the asymptotic analysis of the MLE for large dimensional latent factor models see [Bai and Li \(2012\)](#).

In the above expression, the parameter κ is the Lagrange multiplier on the APT restriction and δ_{apt} is obtained using a cross-validation procedure that is explained below in Section 3.3. Appendix B1 provides the solution of the optimization problem.

3.2 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 are left unconstrained, the Hansen and Jagannathan (1997) (HJ) distance explodes, as we demonstrate in Section 5. Moreover, the APT restriction constrains the Sharpe ratio of the so-called alpha portfolio of Raponi, Uppal, and Zaffaroni (2021). 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. In the same spirit, Kozak, Nagel, and Santosh (2020) rules out near-arbitrage opportunities by restricting the maximum squared Sharpe ratio implied by the entire SDF.

Statistically, the APT restriction (when binding) leads to identification of the vectors a and λ^{mis} . Specifically, at the estimation stage the APT restriction provides N conditions that allow one 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, section 2)).

The estimator of a under the APT restriction has the form of a ridge estimator, as can be seen from Proposition B1. The ridge estimator has the appealing property of mitigating the estimation noise that in our case affects the estimates of the asset-specific risk premia. The estimation noise can be significant because the vector a is an N -dimensional object.

3.3 Identifying the Number of Missing Systematic Risk Factors

Given that the candidate factor model for asset returns may feature $K^{\text{can}} < K$ risk factors, we need to determine the number K^{mis} of missing systematic risk factors f_{t+1}^{mis} . We estimate K^{mis} together with the bound δ_{apt} on the no-arbitrage restriction (8), using cross-validation with the HJ distance as a selection metric. The choice of the HJ distance is natural, given

our objective of identifying the correction required to obtain an admissible SDF from the candidate factor model.

Our cross-validation procedure uses 20 folds. We split the entire sample into 20 folds and estimate the model on all but one fold. 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.⁸ 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 binding 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 \(2002\)](#). Alternatively, [Lettau and Pelger \(2020\)](#) and [Kozak, Nagel, and Santosh \(2020\)](#) use economic restrictions relating expected returns to the covariance of returns with factors in addition to time series information on variation of asset returns.⁹ Because none of these approaches directly applies to a model with asset-specific components in expected returns, we face a choice: either use a two-stage estimation that pins down K^{mis} in the first step and δ_{apt} in the second step or design our own method. We choose the latter and optimize the objective function that explicitly incorporates a no-arbitrage restriction and simultaneously select K^{mis} and δ_{apt} that deliver the minimal value of the HJ-distance.

3.4 Detecting the Missing Factors

Propositions 3 and 4 imply that, as $N \rightarrow \infty$, $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$ converges to a linear function of the missing factors. As a result, a simple time-series regression approach applied to $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$ provides a diagnostic tool for detecting the missing factors in the candidate factor model for asset returns. The advantage of our approach is the absence of a cross-

⁸[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.

⁹Even though our objective function reminds that of [Lettau and Pelger \(2020\)](#), there are several principal differences in our approaches. First and foremost, our goal is not to compress α as much as possible, but rather 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. Relatedly, our objective function, in contrast to that of [Lettau and Pelger \(2020\)](#), does not explicitly include a pricing metric measuring goodness of fit. If we were to include such a pricing metric into the objective function, we would have to augment our log-likelihood function with an additional penalty term represented by the HJ distance.

sectional step, which is present in the standard two-pass procedure. The statistical inference in the cross-sectional step is fragile if a candidate factor is weak or spurious or in the presence of multiple risk factors with highly correlated asset exposures to them.

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

$$\hat{M}_{t+1}^{\beta,\text{mis}} = \gamma_0 + \gamma_1' g_{t+1} + u_{t+1},$$

is

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

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

The following proposition confirms that a simple time-series regression approach reveals if a set of observable variables 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 the missing factors). *Under the assumptions of Proposition 4 and if $g_{t+1} = Q f_{t+1}^{\text{mis}}$ for some non-singular Q , as $N \rightarrow \infty$, we have*

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

If g_{t+1} is orthogonal to f_{t+1}^{mis} , that is $G' (I_T - 1_T 1_T' / T) F^{\text{mis}} = 0_{K^{\text{mis}} \times K^{\text{mis}}}$ then

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

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 .¹⁰

4 Data

In this section, we describe the data that we use in our empirical analysis. In the first subsection, we describe the set of basis assets that we use to estimate the SDF. In the

¹⁰In practice, one uses the estimated $\hat{M}_{t+1}^{\beta,\text{mis}}$, which depends on the estimated parameters, and the result of Proposition 6 can be interpreted as holding for a sufficiently large T .

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 stochastic discount factor on a large set of standard characteristics-based portfolios of U.S. stocks. We collect monthly data on 202 portfolios from July 1963 to August 2019 from Kenneth French’s website. The dataset includes 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).

We use portfolios instead of individual assets because we posit a data generating process for asset returns with constant exposures to factors and prices of risk. [Giglio and Xiu \(2021\)](#) and [Lettau and Pelger \(2020\)](#) argue that constant factor loadings is a reasonable modeling assumption in the case of portfolio returns. We view the case of a model with constant factor loadings and prices of risk as a natural setting to illustrate how our methodology works in practice.

4.2 Factors Potentially Spanning the SDF

To examine which economic variables may explain variation in the SDF, we collect a comprehensive set of factors available at monthly frequency. We include factors used in [Chen and Zimmermann \(2021\)](#), [Jensen, Kelly, and Pedersen \(2021\)](#), and [Kozak, Nagel, and Santosh \(2020\)](#). We also include factors from [Bryzgalova, Huang, and Julliard \(2020\)](#); the Online Appendix of [Bryzgalova, Huang, and Julliard \(2020\)](#) lists the sources of these factors.

We augment the dataset of factors with vector-autoregressive (VAR) residuals of the first 3 principal components of 279 macro variables from [Jurado, Ludvigson, and Ng \(2015\)](#). We also include VAR residuals of the first 8 principal components of 128 macro variables

(McCracken and Ng, 2015) and log differences in consumer sentiment.¹¹ We use the data vintage from February 2021.

We collect real per capita consumption data on nondurables and services and the corresponding price index from the Bureau of Economic Analysis (BEA). We use a one-month and 3-year log consumption growth and an AR(1) residual in log inflation as factors. Our dataset also includes the U.S. business confidence index, U.S. consumer confidence index and U.S. composite leading indicator from the OECD library, confidence economic activity index, NBER recession index, TED spread, effective federal funds rate, real federal funds rate, and Chicago Fed National Financial Condition Index from FRED, and credit spread index of Gilchrist and Zakrajšek (2012).

We add the market-dislocations index of Pasquariello (2014), and the disagreement index of Huang, Li, and Wang (2021). We include industry-adjusted value, momentum, and profitability factors (Novy-Marx, 2013), intra-industry value, momentum, and profitability factors, and basic profitable-minus-unprofitable factor. We complement our dataset with the expected growth factor of Hou, Mo, Xue, and Zhang (2021) and the momentum Up minus Down (UMD) factor. Finally, we include the Chicago Board Options Exchange (CBOE) volatility index (VIX) available on the website of the CBOE, 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) (with the last two available on www.policyuncertainty.com). For highly persistent variables, e.g., the disagreement index of Huang, Li, and Wang (2021), VIX, etc., we define factors as the first-order log differences and residuals from the corresponding univariate autoregressive processes of order 1.

5 Empirical Analysis

In this section, we demonstrate how our methodology can be used to correct popular candidate factor models with the purpose of constructing an admissible SDF. To set the stage, we start by studying a candidate model with zero risk factors so that $M_{\text{exp},t+1}^{\beta,\text{can}}$ is just the inverse of the risk-free rate. We use this model to establish the relative quantitative importance of systematic versus asset-specific risk. This exercise is a clean experiment that is free of concerns that candidate factors may include unpriced sources of common variation

¹¹We download macroeconomic variables from <https://research.stlouisfed.org/econ/mccracken/fred-databases>. We exclude four variables, ACOGNO, ANDENOx, TWEXAFEGSMTHx, UMCSENT, which have missing observations at the start of the sample.

in asset returns (Daniel, Mota, Rottke, and Santos, 2020) or spurious factors (Kan and Zhang, 1999; Kleibergen, 2009). Furthermore, we use this experiment to examine which popular factors correlate with the systematic component of the estimated admissible SDF and which trading strategies correlate with the aggregate measure of asset-specific risk.

After our analysis of a candidate model with zero risk factors, we examine commonly used candidate factor models of asset returns, such as the market model, the model with a consumption-mimicking portfolio, and the model of Fama and French (1993) and characterize the missing systematic and asset-specific components in each of the candidate SDFs.

5.1 A Candidate Model with Zero Risk Factors

We start our analysis of a candidate model with zero risk factors ($K^{\text{can}} = 0$) by identifying how many systematic factors explain the cross-section of asset returns. That is, we define K^{mis} given that the candidate model does not include any risk factors and determine the no-arbitrage bound δ_{apt} . Figure 1 illustrates how the HJ distance changes as we vary K^{mis} and δ_{apt} . The top two panels show the estimation results based on cross validation, while the bottom two panels show the in-sample results.

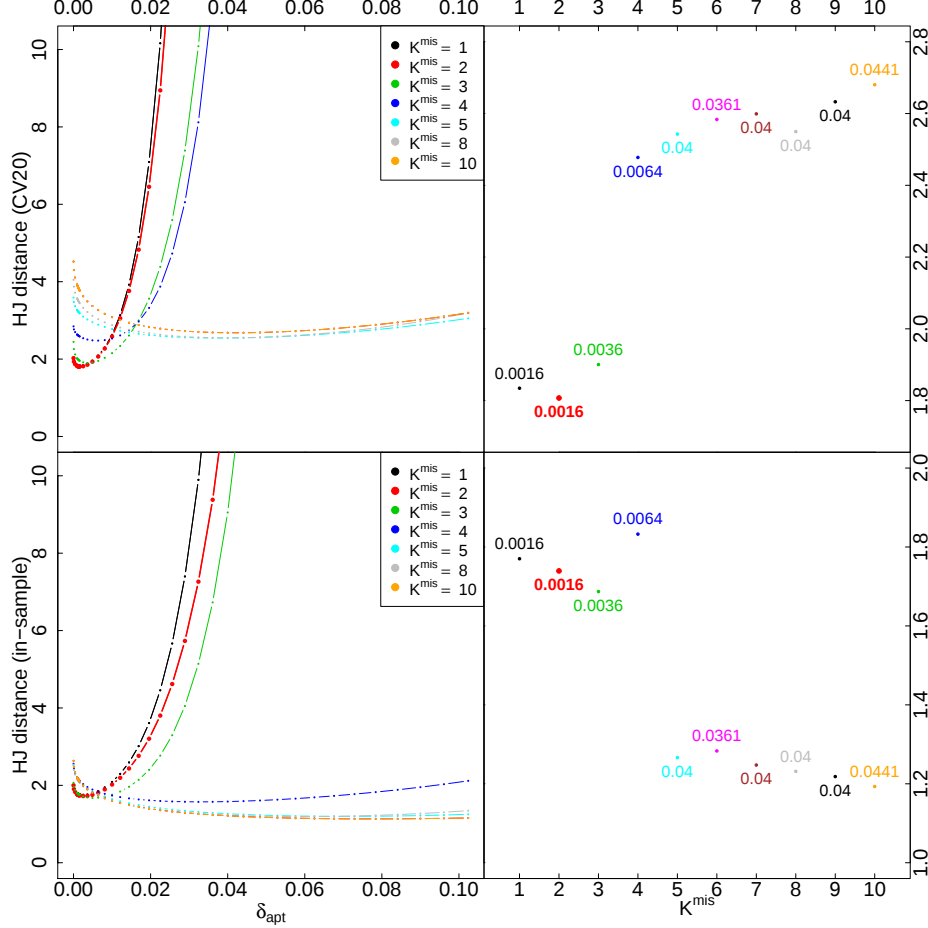
Using the HJ distance as the selection criteria, we see from Figure 1 that our estimation procedure selects $K^{\text{mis}} = 2$ systematic factors and $\delta_{\text{apt}} = 0.0016$. The two top panels demonstrate that this combination of K^{mis} and δ_{apt} achieves the smallest HJ distance, consistent with the evidence on low-dimensional factor pricing models in Kozak, Nagel, and Santosh (2018). The two bottom panels show that a naive in-sample analysis would lead to a choice of $K^{\text{mis}} = 10$ and $\delta_{\text{apt}} = 0.04$. The top right-hand panel shows that the combination of $K^{\text{mis}} = 10$ and $\delta_{\text{apt}} = 0.04$ performs poorly in the cross-validation exercise, with the poor fit in the out-of-sample analysis suggesting evidence of overfitting.

The non-zero value for the optimal δ_{apt} indicates that the asset-specific risk is priced. This result constitutes our first main finding because it challenges the conventional view in financial economics that only systematic (common) risk factors are compensated in the market. Below we will shed light on what constitutes the priced asset-specific 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-

Figure 1: Model selection using the HJ distance

This figure illustrates how the HJ distance changes with K^{mis} and δ_{apt} . The top two panels show the estimation results based on cross validation, while the bottom two panels show the in-sample results. The panels on the left plot the HJ distance for a given choice of K^{mis} as one varies δ_{apt} . The panels on the right display the δ_{apt} (numbers inside the boxes) that minimizes the HJ distance for a given choice of K^{mis} .

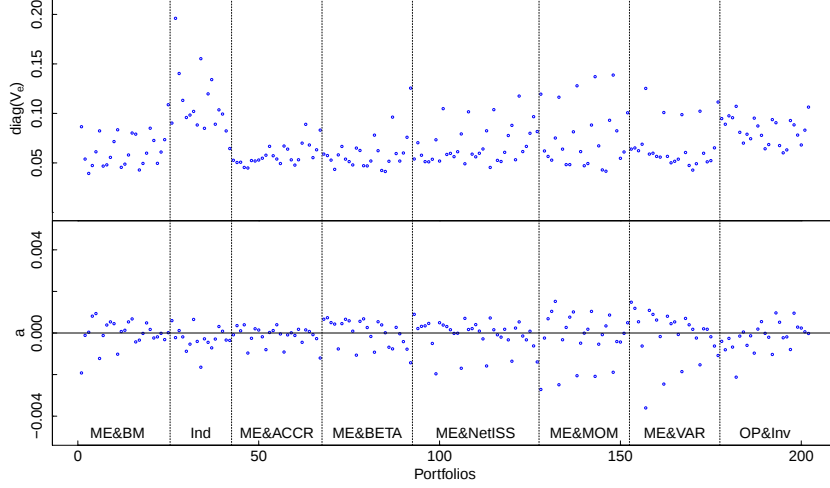


ranted. The bottom panel indicates that the compensation for exposures of basis assets to individual asset-specific shocks is small relative to the premia of conventional risk factors.

To examine the importance of asset-specific risk, we compare the pricing performance of the model selected by our methodology with a model with two systematic factors but where the asset-specific components of expected returns represented by the vector a are set equal to zero. We also compare the results of our approach with those based on principal component analysis (PCA). We choose a different number of principal components to include, ranging from one to five (increasing the number of principal components to ten does not lead to different results).

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 $\text{diag}(V_e^{1/2})$ and the bottom panel shows the compensation a for asset-specific risk.



Four empirical observations stand out from our specification analysis and Table 1. First, we find that ignoring the compensation for asset-specific risk, that is, setting $a = 0_N$, leads to a sizable increase in pricing errors, $\mathbb{E}[\hat{M}_{\text{exp},t+1}(R_{t+1} - R_{ft})]$, across the different assets. This increase in pricing errors translates to a statistically significant increase in the HJ distance by 14.79% relative to the model with $a \neq 0_N$.¹² Second, we observe that the largest benefit in accounting for the compensation for asset-specific risk in reducing the pricing error is for the portfolios sorted by size and variance.¹³ Third, we find that the PCA-based model with two principal components is similar to our model with $a = 0_N$ and $K^{\text{mis}} = 2$, which is expected because the maximum-likelihood estimator corresponds to a weighted-PCA estimator (Bai and Li, 2012). Finally, consistent with Lettau and Pelger (2020) and Kozak, Nagel, and Santosh (2020), we notice that a naive strategy of including a larger number of principal components leads to inferior pricing performance of the corresponding model because of overfitting.

A common approach for evaluating asset pricing factor models is to overlay the average excess returns on the test assets with the model-implied risk premia. We follow this approach in Figure 3, which shows that, as expected, our model exhibits the perfect fit. This is a consequence of our main insight, namely to consider the elements of the vector a as

¹²For statistical inference, we run bootstrap.

¹³The corresponding results are available upon request.

Table 1: Relative HJ-distance of various models

This table reports the HJ distances of alternative candidate models and their corrected versions relative to the corrected version of the model with zero risk factors, $(\text{HJ}^Y/\text{HJ}^X - 1) \times 100\%$, where X indicates the corrected version of the candidate model with zero risk factors, highlighted in bold, and Y indicates any of the other considered models.

Class of models	Candidate Model	Relative HJ (%)
Traditional	CAPM	13.87
	C-CAPM	40.62
	FF3	13.92
PCA-based	PCA1	14.67
	PCA2	14.79
	PCA3	16.53
	PCA4	25.64
	PCA5	25.56
Fully Corrected	Zero risk factors + correction	0.00
	CAPM + correction	0.57
	C-CAPM + correction	0.00
	FF3 + correction	-1.15
Partially corrected	Zero risk factors + $K^{\text{mis}} = 2$ but $a = 0$	14.79

compensation for exposure to individual asset-specific shocks, as opposed to simply pricing errors.¹⁴

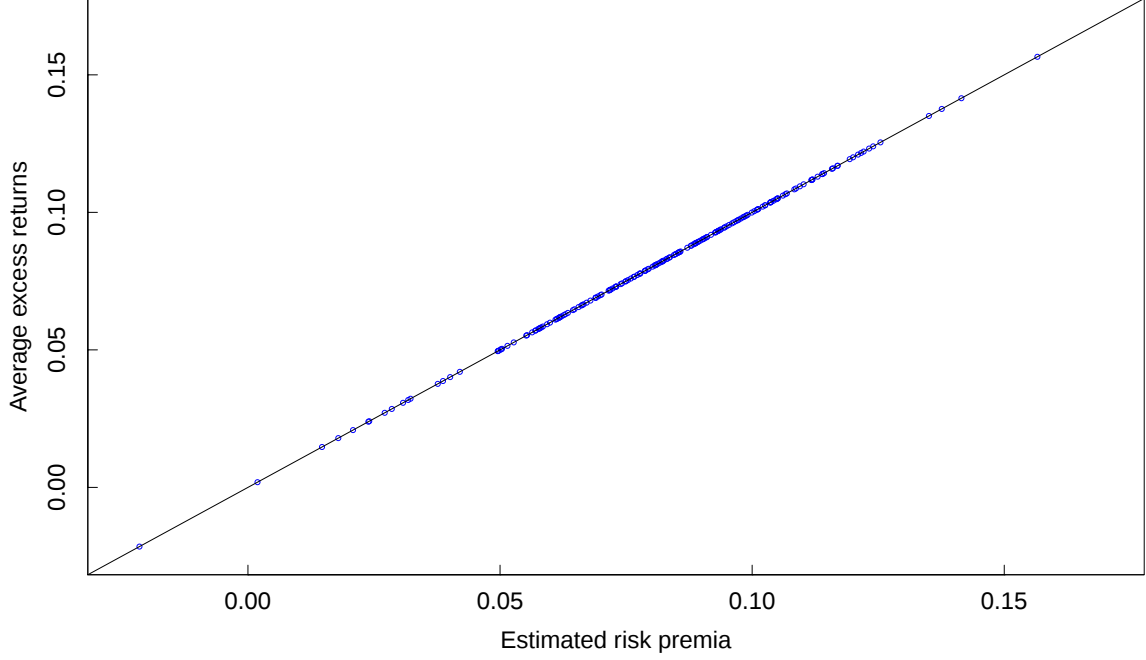
Next, we study the time-series properties of the estimated SDF $\hat{M}_{\text{exp},t+1}$ and its components $\hat{M}_{\text{exp},t+1}^a$ and $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$. Figure 4 shows that both $\hat{M}_{\text{exp},t+1}^a$ and $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$ exhibit sizable volatility during recessions and also during normal times. We see that different components of the SDF dominate variation in the SDF 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 beginning of 2000s, it is the aggregate measure of asset-specific risk that generates a spike in the volatility of the SDF. Thus both common and asset-specific shocks contribute meaningfully to explaining asset valuations.

We perform a formal variance decomposition of the SDF and report its results in Table 2. The standard deviation of $\log \hat{M}_{t+1}^a$ implies that an annual Sharpe ratio for the aggregate measure of asset-specific risk (that is, the premium for the assets' exposure per unit of volatility of the aggregate measure of asset-specific risk), is $\text{SR} = 0.59$. This result is

¹⁴Recall that the vector a is a ridge estimator of the asset-specific risk premia (see formula given in expression (B5) in Appendix B). For this exercise, we undo the shrinkage by multiplying each element of the vector a by the estimated value of the Lagrange multiplier κ plus 1. The estimated value of κ is 19.16.

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 resulting from the ridge estimator for a .



interesting for two reasons. First, the compensation for one unit of the aggregate measure of asset-specific risk is sizeable as it is in the same ballpark as the market Sharpe ratio. Second, given that the elements of the vector a are very small, it becomes immediately clear that what matters for pricing assets is not individual asset-specific shocks but a specific combination of these shocks that represent the component $\hat{M}_{\text{exp},t+1}^a$ of the admissible SDF.

Relatedly, 56% of variation of the SDF can be attributed to the aggregate measure of asset-specific risk with the rest due to systematic sources of risk. These results are consistent with [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 \(2019\)](#). The authors document that the portfolios of stock returns that hedge factor risk exposure exhibit high positive expected returns. These high expected returns could simply reflect compensation for the aggregate measure of asset-specific risk.

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

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

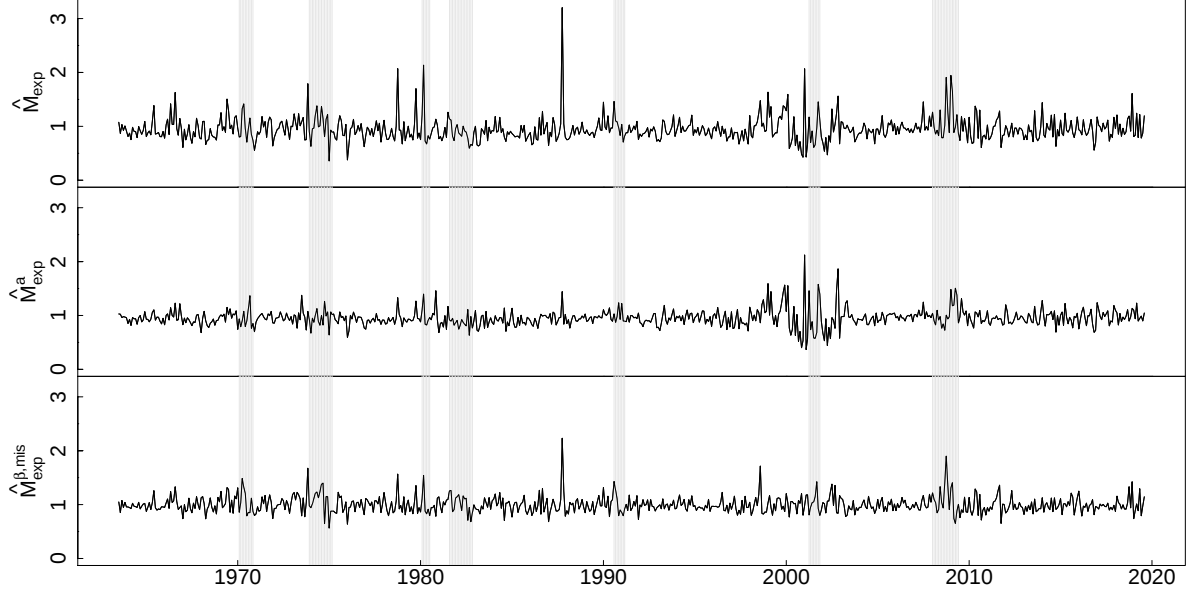


Table 2: Analysis of the SDF if candidate model has zero risk factors

This table reports the moments of the log SDF and its components. The last column gives the percentage of the variation of the SDF that is explained by each component.

	Mean	StdDev	Var	Skewness	Kurtosis	Variation (%)
$\log(\hat{M}_{\text{exp},t+1})$	-0.08	0.22	0.05	0.37	6.56	100.00
$\log(\hat{M}_{\text{exp},t+1}^a)$	-0.06	0.17	0.03	-0.36	8.02	56.21
$\log(\hat{M}_{\text{exp},t+1}^{\beta, \text{can}})$	-0.00	0.00	0.00	-0.63	3.70	0.05
$\log(\hat{M}_{\text{exp},t+1}^{\beta, \text{mis}})$	-0.01	0.15	0.02	0.53	5.33	43.74

Furthermore, an unreported but available upon request regression analysis indicates that $\hat{M}_{\text{exp},t+1}^a$ is a-cyclical: $\log(\hat{M}_{\text{exp},t+1}^a)$ does not significantly correlate with any business-cycle indicator. In contrast to $\log(\hat{M}_{\text{exp},t+1}^a)$, the systematic component $\log(\hat{M}_{\text{exp},t+1}^{\beta, \text{mis}})$ has a significant correlation with the NBER recession dummies.¹⁵ Given that formulae $\hat{M}_{\text{exp},t+1}^{\beta, \text{mis}}$ and $\hat{M}_{\text{exp},t+1}^a$ 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.

¹⁵In regression analysis, we use the log SDF because the log SDF is linear in common risk factors and asset-specific shocks.

Table 3: Strategies highly correlated with $\log(\hat{M}_{\text{exp},t+1}^a)$

This table reports trading strategies that are highly correlated with $\log(\hat{M}_{\text{exp},t+1}^a)$. The first column gives the name of the cluster to which the strategy belongs, using the classification scheme in [Jensen, Kelly, and Pedersen \(2021\)](#). 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 last column reports the R^2 of the univariate regressions of $\log(\hat{M}_{\text{exp},t+1}^a)$ on the return of each individual strategy. The clusters, and within each cluster the sources, are listed in alphabetical order.

Cluster name	Source	Variable name	R^2 (%)
Investment	Fama and French (2015)	cma	30.21
	Hou, Xue, and Zhang (2015)	r_ia	30.48
	Ortiz-Molina and Phillips (2014)	aliq_at	32.21
	Stambaugh and Yuan (2016)	mgmt	33.59
	Xing (2008)	invcap	39.26
Leverage	Fama and French (1992)	am	38.42
	Fama and French (1992)	bookleverage	39.58
	Palazzo (2012)	cash_at	36.66
	Penman Richardson and Tuna (2007)	netdebt_me	34.61
Low risk	Ang, Chen and Xing (2006)	betadown_252d	32.11
	Bradshaw, Richardson, Sloan (2006)	netequityfinance	37.50
	Bradshaw, Richardson, Sloan (2006)	xfin	39.68
	Frazzini and Pedersen (2014)	bab	46.65
	Pontiff and Woodgate (2008)	shareiss1y	37.48
Profitability	Frankel and Lee (1998)	analystvalue	38.00
	La Porta (1996)	fgr5yrlag	50.71
Unclassified	Daniel, Hirshleifer, and Sun (2019)	beh_fin	44.85
	Li (2011)	rdcap	33.24
	Ritter (1991)	indipo	36.42
Value	Barbee, Mukherji and Raines (1996)	sp	33.09
	Basu (1977)	ep	32.65
	Bhandari (1988)	leverage	43.21
	Boudoukh et al. (2007)	eqnpo_me	34.34
	Daniel and Titman (2006)	eqnpo_12m	30.31
	Dechow, Sloan and Soliman (2004)	equityduration	40.83
	Fama and French (1992)	hml	37.32
	Litzenberger and Ramaswamy (1979)	div12m_me	31.50

Having established the quantitative importance of the aggregate measure of asset-specific risk $\hat{M}_{\text{exp},t+1}^a$, we examine which trading strategies reflect exposures to this component of the SDF. As a first step, we run individual regressions of $\log(\hat{M}_{\text{exp},t+1}^a)$ on the excess returns of 457 strategies. We find 27 strategies with an R^2 larger than 30%; these are listed in Table 3. Examining these strategies closely, we find that there is large overlap across these strategies and that they fall into the following clusters, adopting the classification

of [Jensen, Kelly, and Pedersen \(2021\)](#): 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 \(2011\)](#)).

Next, we compute the risk premia associated with the exposures of these 457 trading strategies to the aggregate measure of asset-specific risk as the negative covariance of the return on the strategy and $M_{\text{exp},t+1}^a$:

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

As expected, we find that many of the strategies that are highly correlated with $\log(M_{\text{exp},t+1}^a)$ are associated with large risk premia. However, there are also some strategies that are not as highly correlated with $\log(M_{\text{exp},t+1}^a)$ but that still command sizable risk premia; for example, momentum strategies. Table 4 provides a list of strategies with the absolute value of risk premium exceeding 5% per annum. In the literature, the majority of these strategies are associated with frictions and behavioral biases.

Examining the composition of $\hat{M}_{\text{exp},t+1}^a$, we identify a substantial contribution of small stocks. Specifically, of the 34 basis assets with the highest contribution to the variation of $\hat{M}_{\text{exp},t+1}^a$, 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 aggregate measure of asset-specific risk is our estimation result, not an assumption hardwired into the corrected candidate model. Proposition 3 shows that the relative weight of a basis asset in $M_{\text{exp},t+1}^a$ 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 small stocks have higher volatility. Therefore, ex-ante it is not clear if small stocks feature prominently in the aggregate measure of asset-specific risk.

Table 4: Strategies with high asset-specific risk premium

This table reports the strategies with high asset-specific risk premium. The first column gives the name of the cluster to which the strategy belongs, using the classification scheme in [Jensen, Kelly, and Pedersen \(2021\)](#). 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 last column reports the risk premium. The clusters, and within each cluster the sources, are listed in alphabetical order.

Cluster name	Source	Variable name	RP ^a (%)
Investment	Daniel, Hirshleifer, and Sun (2019)	beh_fin	5.33%
	Ritter (1991)	ageipo	5.83%
	Xing (2008)	invcap	6.06%
Leverage	Bhandari (1988)	leverage	5.54%
	Fama and French (1992)	am	5.16%
	Palazzo (2012)	cash	−5.98%
Low Risk	Ang, Chen, and Xing (2006)	betadown_252d	−5.03%
	Ang, Hodrick, Xing, and Zhang (2006)	ivol	6.14%
	Ang, Hodrick, Xing, and Zhang (2006)	rvol_21d	−5.25%
	Ali, Hwang, and Trombley (2003)	idiovolaht	5.88%
	Bali, Cakici, and Whitelaw (2010)	maxret	6.73%
	Bradshaw, Richardson, Sloan (2006)	xfin	6.54%
	Bradshaw, Richardson, Sloan (2006)	netequityfinance	5.24%
	Fama and MacBeth (1973)	beta	−6.05%
	Frazzini and Pedersen (2014)	betafp	−6.93%
	Jegadeesh and Titman (1993)	mom6m	5.51%
Momentum	Jegadeesh and Titman (1993)	mom12m	5.36%
	Chen, Novy-Marx, Zhang (2011)	rome	5.11%
Profitability	Diether, Malloy and Scherbina (2002)	forecastdispersion	5.05%
	Frankel and Lee (1998)	predictedfe	5.46%
	La Porta (1996)	fgr5yrlag	8.19%
	Cen, Wei, and Zhang (2006)	feps	6.15%
Unclassified	Cooper et al. (2008)	betaarb	6.19%
	Datar et al. (1998)	shvol	5.88%
	Easley, Hvidkjaer and O’Hara (2002)	probinformedtrading	6.20%
	Elgers, Lo and Pfeiffer (2001)	sfe	7.00%
Value	Basu (1977)	ep	5.33%
	Dechow, Sloan and Soliman (2004)	equityduration	5.11%

Our finding regarding the role of small stocks in the aggregate measure of asset-specific risk 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 $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$.¹⁶ In contrast, our central result is about the

¹⁶Recall that the candidate model has zero risk factors, so that $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$ reflects all systematic risk factors.

importance of idiosyncratic shocks to the returns of small companies that drive the acyclical component of the stochastic discount factor, $\hat{M}_{\text{exp},t+1}^a$.

We now turn our attention to $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$ the component of the SDF related to systematic risk factors. We find that the market factor of Sharpe (1964) exhibits the highest explanatory power for $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$ with $R^2 = 0.95$. 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 non-traded factors (shocks in VIX, intermediary capital (He, Kelly, and Manela, 2017), and dividend yield) that each individually explain more than 30% of variation in $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$. Because the market factor already explains a large proportion of the variation in $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$, 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 $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$.¹⁷ These results are in line with findings of Kozak, Nagel, and Santosh (2020) and Bryzgalova, Huang, and Julliard (2020) about the non-sparsity of the SDF in characteristics and the existence of several combinations of trading strategies that deliver similar cross-sectional fit.

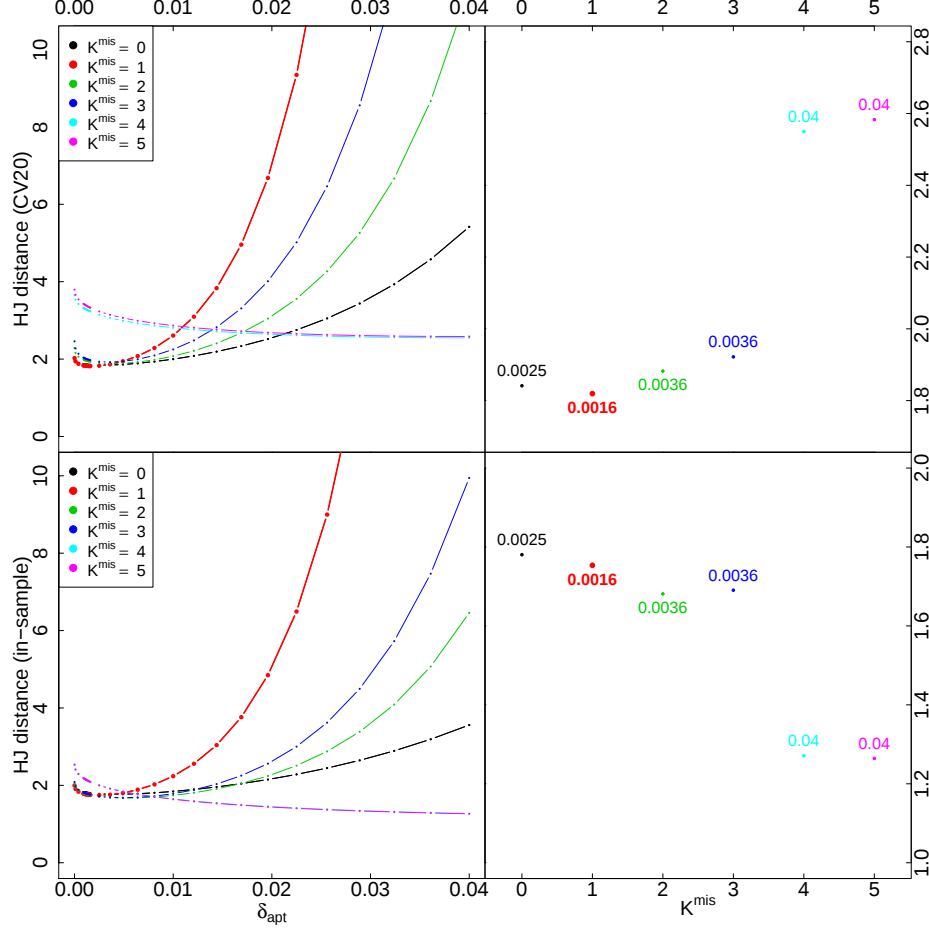
5.2 Popular Candidate Factor Models

To illustrate how our methodology brings new insights about factor models, we now consider three classic 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 $M_{\text{exp},t+1}^{\beta,\text{can}}$ implied by each of these three candidate factor models, we estimate the required correction terms $\hat{M}_{\text{exp},t+1}^a$ and $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$ and characterize their properties.

¹⁷These 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, Meulbroeck, and Sloan, 2001).

Figure 5: Correction of the market 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 top two panels show the estimation results based on cross validation, while the bottom two panels show the in-sample results. The panels on the left plot the HJ distance for a given choice of K^{mis} as one varies δ_{apt} (numbers inside the boxes). The panels on the right display the optimal values of the HJ distance for a given choice of K^{mis} .



5.2.1 The Market Model

Figure 5 shows that for the candidate model with the market as a sole factor, the estimation procedure selects $K^{\text{mis}} = 1$ and $\delta_{\text{apt}} = 0.0016$. 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. Furthermore, Table 5 shows that the admissible SDF obtained from correcting the market model is perfectly correlated with that obtained from correcting the candidate model with zero risk factors.

Table 5: Correlation matrix of corrected SDFs

This table reports the correlation matrix of admissible SDFs obtained from correcting different candidate models: the model with zero risk factors, CAPM, C-CAPM, and FF3.

	Zero risk factors	CAPM	C-CAPM	FF3
Zero risk factors	1.00			
CAPM	1.00	1.00		
C-CAPM	0.98	0.98	1.00	
FF3	0.99	0.99	0.98	1.00

Table 1 shows that the pricing errors under the candidate market model are substantially larger than those under the corrected market model, in which the market factor is augmented by one latent factor and the vector a of the asset-specific components in expected returns. Once the candidate model is corrected, the HJ distance drops by almost 14%, with the drop being statistically significant. 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 relative importance of asset-specific risk is evident from Figure 6 and Table 6. Figure 6 shows that $\hat{M}_{\text{exp},t+1}^a$ exhibits the higher volatility than $\hat{M}_{\text{exp},t+1}^{\beta,\text{can}}$ and $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$, and Table 6 confirms this finding by showing that $\log(\hat{M}_{\text{exp},t+1}^a)$ explains 58% of the variation in the admissible SDF. The high standard deviation of $\log(\hat{M}_{t+1}^a)$ translates into an annual Sharpe ratio for the aggregate measure of asset-specific risk $\text{SR} = 0.59$.

The remaining 42% of the variation in the log SDF is due to the combination of $\hat{M}_{\text{exp},t+1}^{\beta,\text{can}}$ and $\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}}$, with $32.26/42.25 \approx 76.36\%$ of this variation attributable to the market factor. Recall that when starting from the model with zero risk factors as a candidate model, we find that the market factor explains 95% of the systematic component of the admissible SDF. The quantitative difference in the role of the observable market factor is because the market factor is a scaled proxy for the genuine latent risk factor recovered when considering the case of the candidate model with no observable risk factors.¹⁸

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

¹⁸The ratio of the standard deviations of the systematic component of the SDF explained by the first systematic component in the case of the candidate model with zero risk factors to that explained by the market factor in the case of the CAPM candidate model is 1.17.

Figure 6: Time series of the SDF and its components when the market model is the candidate factor model

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

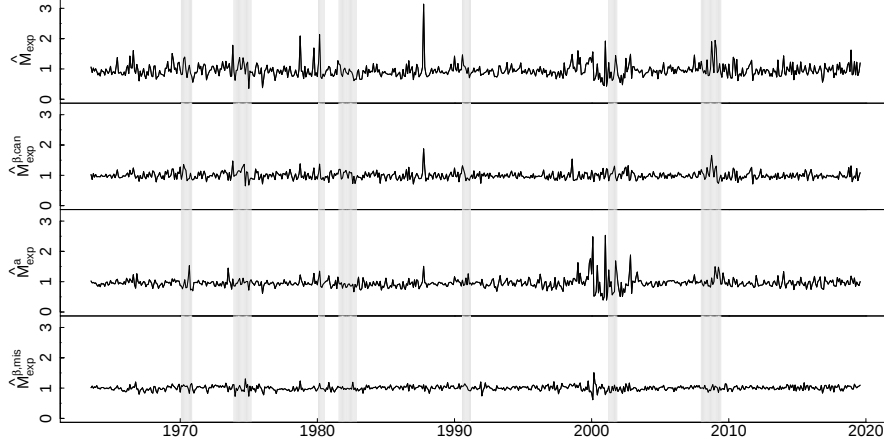


Table 6: Analysis of the SDF if candidate model is the market model

This table reports the moments of the log SDF and its components. The last column gives the percentage of the variation of the SDF that is explained by each component.

	Mean	StdDev	Var	Skewness	Kurtosis	Variation (%)
$\log(\hat{M}_{\text{exp},t+1})$	-0.07	0.21	0.05	0.36	6.57	100.00
$\log(\hat{M}_{\text{exp},t+1}^a)$	-0.06	0.17	0.03	0.16	10.42	57.75
$\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{can}})$	-0.01	0.12	0.01	0.46	4.92	32.26
$\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$	-0.00	0.08	0.01	-0.53	6.48	9.99

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

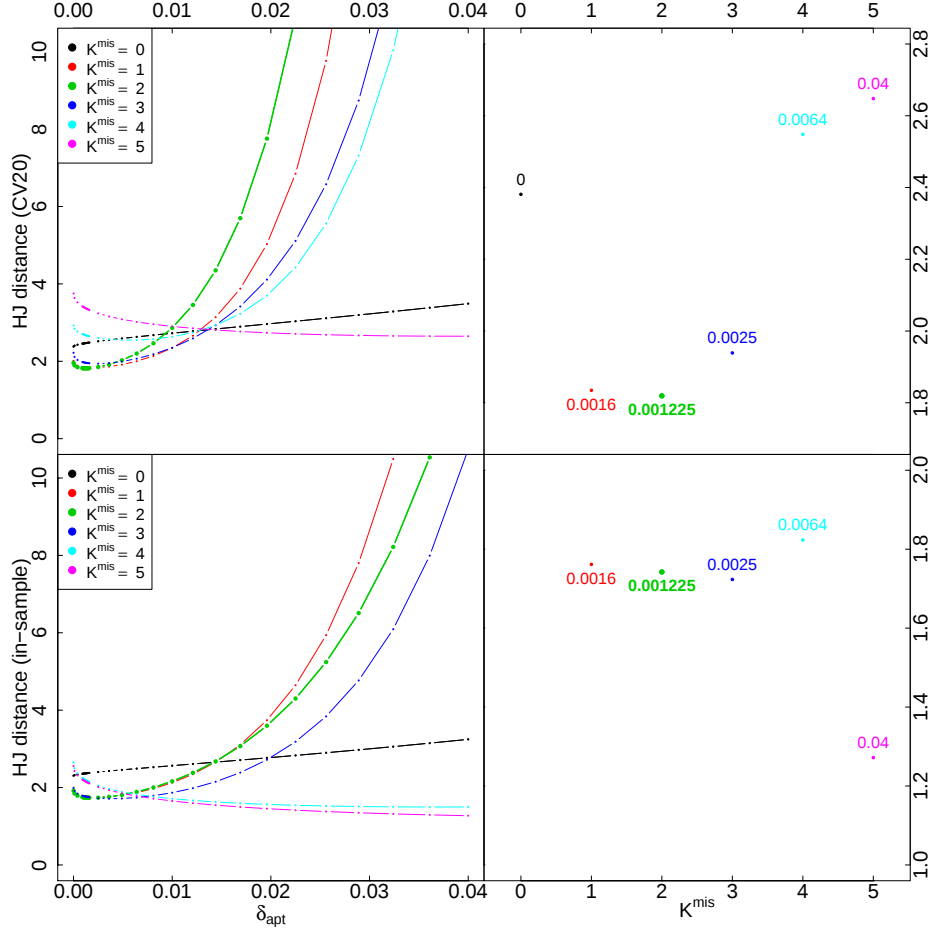
5.2.2 The Model with a Consumption-Mimicking Portfolio

We now consider the case in which the candidate factor model is the model with a consumption-mimicking portfolio. We follow the standard approach of Breeden, Gibbons, and Litzenberger (1989) to construct the consumption-mimicking portfolio.¹⁹ Figure 7 shows that if

¹⁹As outlined in Giglio and Xiu (2021), construction of factor mimicking portfolios can be sensitive to the choice of basis assets. However, we do not have an alternative. The three-stage procedure of Giglio and

Figure 7: Correction of the model with a consumption-mimicking portfolio 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 [Breedon, Gibbons, and Litzenberger \(1989\)](#). The top two panels show the estimation results based on cross validation, while the bottom two panels show the in-sample results. The panels on the left plot the HJ distance for a given choice of K^{mis} as one varies δ_{apt} (numbers inside the boxes). The panels on the right display the optimal values of the HJ distance for a given choice of K^{mis} .



one starts from a candidate model with the consumption-mimicking portfolio as a single factor, then the estimation procedure selects $K^{\text{mis}} = 2$ and $\delta_{\text{apt}} = 0.0012$. The consumption-mimicking portfolio does not highly correlate with either of the latent factors estimated when correcting the candidate factor model with zero risk factors – the correlations are 0.3 and 0 – and therefore, two additional factors are still required to capture the common variation in asset returns.

Xiu (2021), which is insensitive to the choice of test assets, assumes no role for asset-specific risk, while we document that the asset-specific risk plays a big role in the risk-return trade-off.

Table 7: Analysis of SDF if candidate model is consumption-mimicking portfolio

This table reports the moments of the log SDF and its components when the candidate model includes only the consumption-mimicking portfolio of [Breedon, Gibbons, and Litzenberger \(1989\)](#). The last column gives the percentage of variation of the SDF that is explained by each component.

	Mean	StdDev	Var	Skewness	Kurtosis	Variation (%)
$\log(\hat{M}_{\text{exp},t+1})$	-0.08	0.23	0.05	0.28	6.86	100.00
$\log(\hat{M}_{\text{exp},t+1}^a)$	-0.05	0.14	0.02	-0.36	7.79	44.87
$\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{can}})$	-0.03	0.10	0.01	0.18	9.07	34.11
$\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$	-0.00	0.12	0.02	0.30	4.66	21.02

The results reported in Table 1 imply that the pricing errors under the factor model based on C-CAPM are much larger than those under the corrected factor model, in which the consumption-mimicking portfolio factor is augmented by two latent factors and the vector of the asset-specific components in expected returns a . Correcting the C-CAPM for misspecification leads to a substantial and statistically significant drop in the HJ distance by 40.62%. The pricing errors in the candidate model are centred around 0.06, whereas the pricing errors in the corrected model are centred around zero.²⁰ This finding indicates that the candidate C-CAPM model is missing a level factor. Therefore, it is not surprising that the observable market factor is the factor with the highest explanatory power for $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$ and that $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$ exhibits cyclicalities.

Given the standard deviation of $\log(\hat{M}_{t+1}^a)$ reported in Table 7, we find that the annual Sharpe ratio for the aggregate measure of asset-specific risk is $\text{SR} = 0.48$. Furthermore, Table 7 shows that there is a higher proportion of missing systematic risk in the C-CAPM candidate model relative to the market candidate model: in terms of variation in the systematic components of the admissible SDF, $21.02/55.13 \approx 38\%$ versus $9.99/42.25 \approx 24\%$. This finding is another manifestation of the special role of the market factor in explaining common variation in asset returns and systematic variation in the admissible SDF. Overall, as Table 5 shows, the admissible SDF obtained when correcting the candidate model with the consumption-mimicking portfolio almost perfectly correlates with those obtained when correcting the candidate model with zero risk factors and the candidate model with the market factor.

²⁰To save space, these results are not reported, but they are available upon request.

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

Figure 8 shows that if we use the three-factor model of Fama and French (1993) as the candidate model, our methodology selects *zero* missing sources of systematic risk and optimal $\delta_{\text{apt}} = 0.0025$. 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 candidate model with zero risk factors 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. The contribution of each individual remaining factor is so small that it is indistinguishable from asset-specific risk.

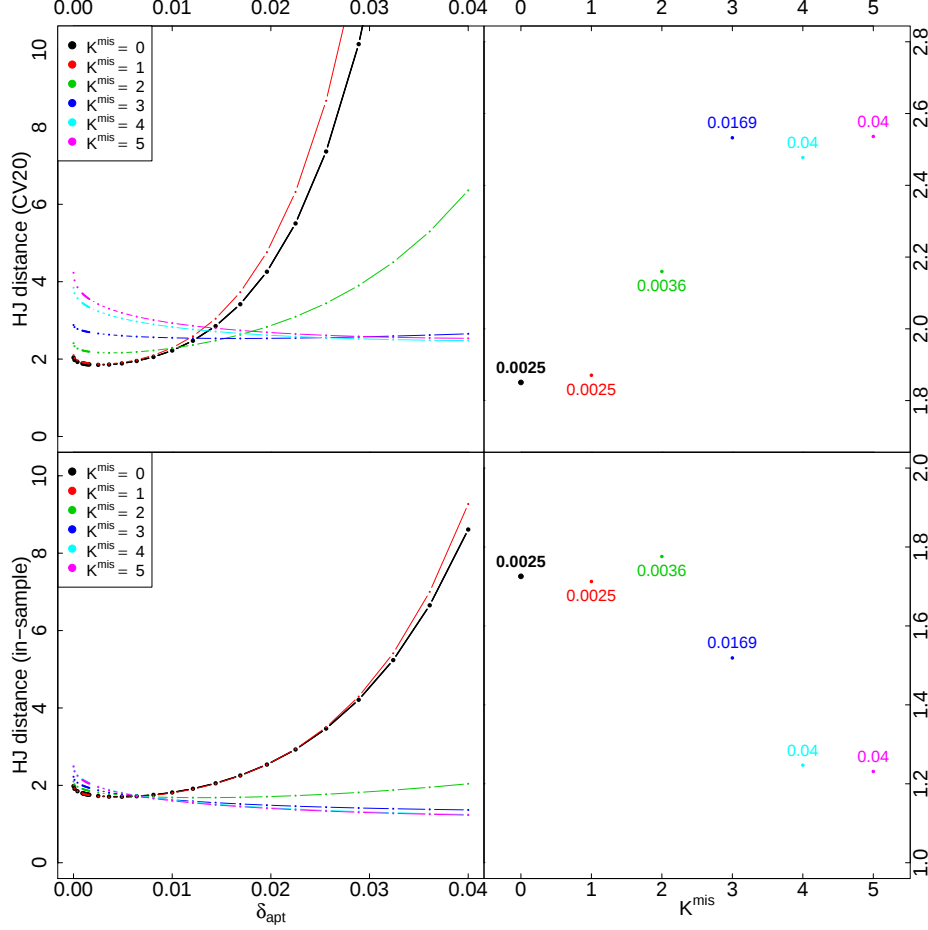
Differentiating between remaining common factors and asset-specific risk is especially challenging given that observable traded factors are noisy versions of the true risk factors that span $\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$ obtained in the case of the model with zero risk factors. As a result, when starting from FF3 as the candidate model, the estimated asset-specific shocks are mixed with weak latent factors. Moreover, the FF3 model includes the Value factor that correlates more strongly with the reference aggregate measure of asset-specific risk (the correlation is -0.61) than with the systematic component of the SDF (the correlation is 0.14). Thus, FF3 implicitly incorporates some asset-specific risk. Despite these challenges, the admissible SDF obtained after correcting the FF3 model is almost perfectly correlated with those obtained after correcting the other candidate factor models.

Table 8 shows that the estimated admissible SDF has higher volatility relative to the volatility of the SDFs obtained after correcting the other candidate models. This increase is due to the noise that observable factors implicitly introduce in the candidate factor model, consistent with Daniel, Mota, Rottke, and Santos (2020). The standard deviation of $\log(\hat{M}_{\text{exp},t+1}^a)$ implies that the Sharpe ratio for the asset-specific risk is 0.48 per annum. The lower value of this Sharpe ratio relative to those obtained after correcting alternative candidate models is because the volatility of returns adjusted for the three systematic factors of FF3 is higher than that of pure asset-specific risk. The pure asset-specific risk is estimated when starting from a model with zero risk factors.²¹ Keeping the total compensation for asset-specific risk fixed but increasing the volatility of asset-specific risk leads to a deterioration of the corresponding Sharpe ratio.

²¹The notion *pure* relates to the fact that the estimation of asset-specific risk is not conditional on assumed common factors in the cross-section of asset returns.

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 top two panels show the estimation results based on cross validation, while the bottom two panels show the in-sample results. The panels on the left plot 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} .



From the perspective of pricing, accounting for missing asset-specific risk improves the pricing performance of the model, especially with respect to Size-Momentum and Size-Variance portfolios. The HJ distance drops by a significant 14% once we account for compensation for asset-specific risk. 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.

Table 8: Analysis of the SDF if candidate model is FF3

This table reports the moments of the log SDF and its components when the candidate model is the three-factor model of [Fama and French \(1993\)](#). The last column gives the percentage of the variation of the SDF that is explained by each component.

	Mean	StdDev	Var	Skewness	Kurtosis	Variation (%)
$\log(\hat{M}_{\text{exp},t+1})$	-0.09	0.25	0.06	0.28	6.68	100.00
$\log(\hat{M}_{\text{exp},t+1}^a)$	-0.06	0.14	0.02	1.27	11.47	32.94
$\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{can}})$	-0.03	0.20	0.04	0.13	6.43	67.06
$\log(\hat{M}_{\text{exp},t+1}^{\beta,\text{mis}})$	0.00	0.00	0.00			0.00

5.2.4 Discussion

We could repeat our analysis for other candidate factor models. However, our main conclusion is not going to change – the aggregate measure of asset-specific risk accounts for the lion share of pricing of the cross-section of asset returns. If by chance a candidate factor model contains a factor that is correlated with the aggregate measure of asset-specific risk, then one may find that the role of missing asset-specific risk is biased down, as we saw in the case of the corrected FF3 model. Yet, including more and more factors into a candidate factor model would not lead to exact pricing of the cross-section. First, the inclusion of any observable risk factor into the candidate model necessarily adds noise into the admissible SDF, as we saw when correcting the C-CAPM and FF3 models. Second, and more importantly, the aggregate measure of asset-specific risk is a weak factor in the cross-section of asset returns, and therefore, the corresponding risk premia cannot be accurately estimated. In other words, one cannot recover the aggregate measure of asset-specific risk by including any number of tradable factors into the candidate factor model. We show this result explicitly in [Appendix A1](#). Our second main finding, that no candidate factor model can capture the aggregate measure of asset-specific risk, is consistent with the findings 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.

5.3 Out-of-Sample Analysis

To illustrate the robustness of our approach, we run out-of-sample analyses: we evaluate how the candidate and corrected models price two additional cross-section of stock returns. The first dataset, 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

Table 9: Cross-sectional Out-of-Sample Analysis

This table reports the relative improvement in the HJ distance when using the admissible SDF constructed by correcting a model with zero risk factors relative to the SDFs of the alternative original and corrected candidate factor models. We analyze the performance of these models on the set of 202 basis assets (column Benchmark) and on two additional datasets not used at the estimation. The improvement is tabulated in %.

Class of models	Model	Benchmark (%)	Dataset 1 (%)	Dataset2 (%)
Traditional models	CAPM	13.87	9.04	8.35
	C-CAPM	40.62	38.51	32.68
	FF3	13.92	10.80	10.06
PCA-based models	PCA1	14.67	9.84	9.07
	PCA2	14.79	10.03	9.24
	PCA3	16.53	14.57	13.00
	PCA4	25.64	24.42	21.56
	PCA5	25.56	24.34	21.51
Corrected SDFs	Fully-latent corrected	0.00	0.00	0.00
	CAPM corrected	0.81	0.47	0.47
	C-CAPM corrected	0.14	0.38	0.14
	FF3 corrected	-0.79	0.48	0.38

portfolios sorted by size and short-term reversal, and 49 industry portfolios.²² The second dataset 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.

Table 9 reports the relative *improvement* in the HJ distance when using the admissible SDF constructed by correcting a model with zero risk factors relative to the SDFs of the various traditional candidate models – both before and after they have been corrected for misspecification. This table shows that the admissible SDFs constructed by correcting the three candidate asset-pricing models that we have considered (CAPM, C-CAPM, FF3) have superior pricing performance compared to the candidate models that are based on just the observable factors or principal components.

6 Conclusion

A fundamental challenge in finance is to price 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

²²The 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.

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 asset-specific 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 asset-specific risk. Theoretically, we obtain this result through the lens of the SDF under the assumptions of the APT. We show that an aggregate measure of asset-specific risk, a component of the admissible SDF represented by a linear combination of asset-specific shocks, accounts for 56% of the variation in the admissible SDF.

What is missing in virtually all factor models is compensation for this asset-specific risk. We show that even though some conventional factors, for example, Value, have sizeable correlations with the aggregate measure of asset-specific risk, adding an arbitrary number of factors to a candidate factor model will not lead to an admissible SDF.

The methodology 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 methodology 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 asset-specific 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: Can one recover M_{t+1}^a using observable variables?

In this appendix, we show that when the asset-specific correction term M_{t+1}^a is correlated with some observed factors, then these factors are necessarily weak (Lettau and Pelger, 2020). Thus even if one could find the complete set of observed factors spanning M_{t+1}^a , one could not estimate accurately the corresponding risk premia.²³ As a result, the component M_{t+1}^a cannot be estimated the same way as $M_{t+1}^{\beta, \text{can}}$ and $M_{t+1}^{\beta, \text{mis}}$.

Proposition A1. *Under Assumptions 1 and 2, assume that there are no missing systematic risk factors, that is $K^{\text{mis}} = 0$, implying*

$$R_{t+1} - R_{ft} = a + \beta^{\text{can}} f_{t+1}^{\text{can}} + e_{t+1},$$

where

$$T^{-1} \sum_{t=1}^T (e_t - \bar{e})(e_t - \bar{e})' \rightarrow_p V_e,$$

$$T^{-1} \sum_{t=1}^T (f_t^{\text{can}} - \bar{f}^{\text{can}})(f_t^{\text{can}} - \bar{f}^{\text{can}})' \rightarrow_p V_f^{\text{can}} > 0,$$

as $T \rightarrow \infty$.

If, for an observed factor f_{t+1}^{idio} , $a'V_e^{-1}e_{t+1} = f_{t+1}^{\text{idio}} - \mathbb{E}(f_{t+1}^{\text{idio}})$, then f_{t+1}^{idio} must be a weak factor.

Proof: Without loss of generality, given that the e_t is uncorrelated with f_t^{can} by Assumption 1, assume that f_t^{can} and f_t^{idio} are orthogonal in sample, that is $\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})f_t^{\text{can}} = 0_{K^{\text{can}}}$. Considering the time-series regression

$$R_{it} - R_{ft} = \beta_0 + \beta^{\text{can}} f_t^{\text{can}} + \beta^{\text{idio}} f_t^{\text{idio}} + u_t,$$

the OLS estimator of β^{idio} satisfies

$$\begin{aligned} \hat{\beta}_i^{\text{idio}} &= \frac{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})(R_{it} - R_{ft})}{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})^2} = \frac{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})(a_i + \beta_i^{\text{can}'} f_t^{\text{can}} + e_{it})}{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})^2} \\ &= \frac{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})a_i}{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})^2} + \frac{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})e_{it}}{\sum_{t=1}^T (f_t^{\text{idio}} - \bar{f}^{\text{idio}})^2} \end{aligned}$$

²³See Anatolyev and Mikusheva (2021) for the full asymptotic analysis of estimated risk premia associated with observable weak factors.

$$\begin{aligned}
&= \frac{\sum_{t=1}^T (a'V_e^{-1}(e_{t+1} - \bar{e}))a_i}{\sum_{t=1}^T (a'V_e^{-1}(e_{t+1} - \bar{e}))^2} + \frac{\sum_{t=1}^T (a'V_e^{-1}(e_{t+1} - \bar{e}))e_{it}}{\sum_{t=1}^T (a'V_e^{-1}(e_{t+1} - \bar{e}))^2} \\
&= 0 + \frac{\sum_{t=1}^T (a'V_e^{-1}(e_{t+1} - \bar{e}))e_{it}}{\sum_{t=1}^T (a'V_e^{-1}(e_{t+1} - \bar{e}))^2} \rightarrow_p \frac{a'V_e^{-1}V_e\iota_i}{a'V_e^{-1}V_eV_e^{-1}a} = \frac{a_i}{a'V_e^{-1}a} = \frac{a_i}{\delta_{apt}},
\end{aligned}$$

where we set ι_i equal to the i th row/column of I_N , and where we assumed, without loss of generality, that the APT constraint binds, i.e., $a'V_e^{-1}a = \delta_{apt}$. $\square$

Proposition A1 implies that f_t^{idio} is a weak factor because its loading satisfies $\beta^{\text{idio}'}\beta^{\text{idio}} < \delta < \infty$ for any N . This makes estimation of the corresponding risk premium problematic. In fact, its second-pass estimator satisfies

$$\hat{\lambda}^{\text{idio}} = 1 + \frac{O_p(N^{-1}T^{-\frac{1}{2}})}{O(N^{-1})}. \quad (\text{A1})$$

Therefore, $\hat{\lambda}^{\text{idio}}$ is meaningless especially when N is large. For a formal analysis see [Anatolyev and Mikusheva \(2021\)](#). To see how (A1) arises, it is enough to study the behaviour of the simple case of the two-pass estimator when $K^{\text{can}} = K^{\text{idio}} = 1$, satisfying

$$\begin{aligned}
\hat{\lambda} &= \begin{bmatrix} \hat{\lambda}^{\text{can}} \\ \hat{\lambda}^{\text{idio}} \end{bmatrix} = \begin{bmatrix} \beta^{\text{can}'}\beta^{\text{can}} & \beta^{\text{can}'}\beta^{\text{idio}} \\ \beta^{\text{idio}'}\beta^{\text{can}} & \beta^{\text{idio}'}\beta^{\text{idio}} \end{bmatrix}^{-1} \begin{bmatrix} \beta^{\text{can}'}(\bar{R} - \bar{R}_f) \\ \beta^{\text{idio}'}(\bar{R} - \bar{R}_f) \end{bmatrix} \\
&= \begin{bmatrix} \beta^{\text{can}'}\beta^{\text{can}} & \beta^{\text{can}'}a \\ a'\beta^{\text{can}} & a'a \end{bmatrix}^{-1} \begin{bmatrix} \beta^{\text{can}'}(\bar{R} - \bar{R}_f) \\ a'(\bar{R} - \bar{R}_f) \end{bmatrix} \\
&= \begin{bmatrix} \lambda^{\text{can}} + \bar{f}^{\text{can}} - \mathbb{E}f_t^{\text{can}} \\ 1 \end{bmatrix} + \begin{bmatrix} \beta^{\text{can}'}\beta^{\text{can}} & \beta^{\text{can}'}a \\ a'\beta^{\text{can}} & a'a \end{bmatrix}^{-1} \begin{bmatrix} \beta^{\text{can}'}\bar{e} \\ a'\bar{e} \end{bmatrix} \\
&= \begin{bmatrix} \lambda^{\text{can}} + \bar{f}^{\text{can}} - \mathbb{E}f_t^{\text{can}} \\ 1 \end{bmatrix} + \\
&\quad \frac{1}{\frac{a'a}{N}\frac{\beta^{\text{can}'}\beta^{\text{can}}}{N} - (\frac{\beta^{\text{can}'}a}{N})^2} \begin{bmatrix} \frac{a'a}{N} & -\frac{\beta^{\text{can}'}a}{N} \\ -\frac{\beta^{\text{can}'}a}{N} & \frac{\beta^{\text{can}'}\beta^{\text{can}}}{N} \end{bmatrix} \begin{bmatrix} \frac{\beta^{\text{can}'}\bar{e}}{N} \\ \frac{a'\bar{e}}{N} \end{bmatrix} \\
&= \begin{bmatrix} \lambda^{\text{can}} + \bar{f}^{\text{can}} - \mathbb{E}f_t^{\text{can}} \\ 1 \end{bmatrix} + \frac{1}{O(\frac{1}{N})} \begin{bmatrix} O(\frac{1}{N}) & O(\frac{1}{N^{\frac{1}{2}}}) \\ O(\frac{1}{N^{\frac{1}{2}}}) & O(1) \end{bmatrix} \begin{bmatrix} O_p(\frac{1}{(NT)^{\frac{1}{2}}}) \\ O_p(\frac{1}{NT^{\frac{1}{2}}}) \end{bmatrix} \\
&= \begin{bmatrix} \lambda^{\text{can}} + \bar{f}^{\text{can}} - \mathbb{E}f_t^{\text{can}} \\ 1 \end{bmatrix} + \begin{bmatrix} \frac{O_p(N^{-\frac{3}{2}}T^{-\frac{1}{2}})}{O(N^{-1})} \\ \frac{O_p(N^{-1}T^{-\frac{1}{2}})}{O(N^{-1})} \end{bmatrix}.
\end{aligned}$$

where we set $\delta_{apt} = 1$ for simplicity.

Whereas the two-pass estimator for the candidate risk factor is always consistent for $\lambda^{can} + \bar{f}^{can} - \mathbb{E}f_t^{can}$ when either N or T or both diverge, coinciding with λ^{can} when $T \rightarrow \infty$, the estimated risk premium $\hat{\lambda}^{idio}$ of f_{t+1}^{idio} has an undefined limit, especially when N becomes large. Therefore, the traditional two-pass regression approach does not permit to accurately estimate M_{t+1}^a . In contrast, the methodology described in this paper explains how to construct an accurate estimate of M_{t+1}^a , which does not even rely on the existence and identification of f_{t+1}^{idio} .

Theorem A1 extends to the multivariate case, that is when f_{t+1}^{idio} is a vector. It also extends to the case when f_t^{idio} spans the asset-specific risk *imperfectly*, that is for

$$(f_{t+1}^{idio} - E(f_{t+1}^{idio})) = \gamma a' V_e^{-1} e_{t+1} + \eta_{t+1}$$

where $E\eta_{t+1} = 0$, $corr(a' V_e^{-1} e_{t+1}, \eta_{t+1}) = 0$, $var(\eta_{t+1}) = \sigma_\eta^2 = var(f_{t+1}^{idio})(1 - \rho_{f^{idio}, M^a}^2)$. Then

$$\hat{\beta}_i^{idio} = \beta_i^{idio} + O_p(T^{-1/2}) = \frac{\gamma a_i}{\gamma^2 a' V_e^{-1} a + \sigma_\eta^2} + O_p(T^{-1/2}),$$

implying that f_{t+1}^{idio} continues to be a weak factor, given $\beta^{idio'} \beta^{idio} < \infty$, and estimation of its risk premia is still problematic, as discussed above.

B Appendix: Proofs

This appendix contains the proofs for all the propositions in the manuscript. We use the following notation: I_a denotes the identity matrix of dimension $a \times a$; $0_{a \times b}$ denotes a matrix of zeros with a rows and b columns and 0_a denotes a vector of zeros with a entries ; $A > 0$ means that the matrix A is positive definite; $\|\cdot\|$ denotes the Euclidean norm; $\mathbb{E}(\cdot)$ denotes the expectation operator; $a_N = O(b_N)$ and $a_N = o(b_N)$ with $b_N > 0$ it means that $|a_N|/b_N$ is bounded and $|a_N|/b_N \rightarrow 0$, respectively, as $N \rightarrow \infty$; $\xrightarrow{p}$ denotes convergence in probability; $a_N = O_p(b_N)$ and $a_N = o_p(b_N)$ with $b_N > 0$ it means that $|a_N|/b_N$ is bounded in probability and $|a_N|/b_N \xrightarrow{p} 0$, respectively, as $N \rightarrow \infty$.

B.1 Lemmas

We start by providing a set of lemmas needed to prove our results.

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

(i)

$$\mathbb{E}(e^{d'z}) = e^{d'\mu_z + \frac{1}{2}d'\Sigma_z d}.$$

(ii)

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

setting

$$\mu^* = (\mu_z + \Sigma_z d).$$

An alternative expression is

$$\mathbb{E}(ze^{d'z}) = (\mu_z + \Sigma_z d) e^{\frac{1}{2}d'\Sigma_z d + \mu_z' d}.$$

Proof: (i) is well-known. For (ii), denoting by n_z the dimensionality of the vector z ,

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

Then

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

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

implying

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

□

Lemma B.2. Let $V_R = \beta V_f \beta' + V_\varepsilon$ with $N \times N$ and $K \times K$ matrices $V_\varepsilon > 0$ and $V_f > 0$, and a full-column rank $N \times K$ matrix β satisfying $\beta' V_\varepsilon^{-1} \beta / N \rightarrow D > 0$ for some matrix D . Then:

$$\beta' V_R^{-1} \beta \rightarrow V_f^{-1}.$$

Proof: By the Sherman Morrison formula,

$$V_R^{-1} = V_\varepsilon^{-1} - V_\varepsilon^{-1} \beta (V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1} \beta' V_\varepsilon^{-1},$$

pre-multiplying by β' , and re-arranging terms, yields

$$\begin{aligned}
\beta' V_R^{-1} &= \beta' V_\varepsilon^{-1} - \beta' V_\varepsilon^{-1} \beta (V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1} \beta' V_\varepsilon^{-1} \\
&= (I_K - \beta' V_\varepsilon^{-1} \beta (V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1}) \beta' V_\varepsilon^{-1} \\
&= ((V_f^{-1} + \beta' V_\varepsilon^{-1} \beta) - \beta' V_\varepsilon^{-1} \beta) (V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1} \beta' V_\varepsilon^{-1} \\
&= V_f^{-1} (V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1} \beta' V_\varepsilon^{-1}.
\end{aligned}$$

Post-multiplying by β and taking the limit as $N \rightarrow \infty$ gives

$$\beta' V_R^{-1} \beta \rightarrow V_f^{-1},$$

because $(V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1} \beta' V_\varepsilon^{-1} \beta \rightarrow I_K$. □

Lemma B.3. Under the assumptions of Lemma B.2 and for a random vector e with mean zero and covariance V_ε :

$$\beta' V_R^{-1} e = O_p(N^{-\frac{1}{2}}).$$

Proof: Pre-multiplying by β' and post-multiplying by e one obtains:

$$\beta' V_R^{-1} e = V_f^{-1} (V_f^{-1} + \beta' V_\varepsilon^{-1} \beta)^{-1} \beta' V_\varepsilon^{-1} e.$$

The result follows noticing that $\beta' V_\varepsilon^{-1} e = O_p((\beta' V_\varepsilon^{-1} \beta)^{\frac{1}{2}})$ using the result $X = O_p((\mathbb{E}(X))^{\frac{1}{2}})$ for any random variable X with finite second moment. □

Lemma B.4. Under the assumptions of Lemma B.2 and letting $V_\varepsilon = \beta^{\text{mis}}\beta^{\text{mis}'} + V_e$ for a column full-rank $N \times K^{\text{mis}}$ matrix β^{mis} and a $N \times N$ matrix $V_e > 0$ such that $\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}/N \rightarrow E > 0$ for some matrix E , then:

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

Proof: Along the same lines of the proof to Lemma B.2

$$\beta^{\text{mis}'}V_\varepsilon^{-1}\beta = (I_{K^{\text{mis}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}\beta = \left(\frac{I_{K^{\text{mis}}}}{N} + \frac{\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}}{N}\right)^{-1}\frac{\beta^{\text{mis}'}V_e^{-1}\beta}{N} = O(1),$$

where, by the Schwartz inequality, $\|\beta^{\text{mis}'}V_e^{-1}\beta\| \leq \|\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\|^{\frac{1}{2}}\|\beta^{\text{mis}'}V_e^{-1}\beta\|^{\frac{1}{2}}$. $\square$

Lemma B.5. Under the assumptions of Lemma B.4:

$$\beta^{\text{mis}'}V_\varepsilon^{-1}\beta^{\text{mis}} \rightarrow I_{K^{\text{mis}}}.$$

Proof: This is a special case of Lemma B.2. $\square$

Lemma B.6. Under the assumptions of Lemma B.4 and for a random vector e with mean zero and covariance V_e : Then:

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

Proof: This is a special case of Lemma B.3. $\square$

Lemma B.7. Under the assumptions of Lemma B.4, setting $V_R = \beta V_f \beta' + V_\varepsilon$, $\alpha = a + \beta^{\text{mis}}\lambda_m$ for an $K^{\text{can}} \times 1$ vector of constants λ_m and $a'V_e^{-1}a = O(1)$, then:

$$\alpha'V_R^{-1}\beta = O(N^{-\frac{1}{2}}).$$

Proof: Given

$$\begin{aligned} \alpha'V_R^{-1}\beta &= a'V_R^{-1}\beta + \lambda_m'\beta^{\text{mis}'}V_R^{-1}\beta \\ &= a'(V_\varepsilon^{-1} - V_\varepsilon^{-1}\beta(V_f^{-1} + \beta'V_\varepsilon^{-1}\beta)^{-1}\beta'V_\varepsilon^{-1})\beta \\ &\quad + \lambda_m'\beta^{\text{mis}'}(V_\varepsilon^{-1} - V_\varepsilon^{-1}\beta(V_f^{-1} + \beta'V_\varepsilon^{-1}\beta)^{-1}\beta'V_\varepsilon^{-1})\beta \\ &= a'V_\varepsilon^{-1}\beta(V_f^{-1} + \beta'V_\varepsilon^{-1}\beta)^{-1}V_f^{-1} + \lambda_m'\beta^{\text{mis}'}V_\varepsilon^{-1}\beta(V_f^{-1} + \beta'V_\varepsilon^{-1}\beta)^{-1}V_f^{-1} \\ &= O(N^{-\frac{1}{2}}) + O(N^{-1}), \end{aligned}$$

by Lemma B.4, the bound $\|a'V_\varepsilon^{-1}\beta\| \leq \|a'V_\varepsilon^{-1}a\|^{\frac{1}{2}}\|\beta'V_\varepsilon^{-1}\beta\|^{\frac{1}{2}}$ and

$$|a'V_\varepsilon^{-1}a| = |a'(V_e^{-1} - V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1})a|$$

$$\begin{aligned}
&\leq |a'V_e^{-1}a| + |aV_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}a| \\
&\leq |a'V_e^{-1}a| + |aV_e^{-1}a|^{\frac{1}{2}} \|\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\|^{\frac{1}{2}} \|(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\| |aV_e^{-1}a|^{\frac{1}{2}} \|\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\|^{\frac{1}{2}} \\
&= |a'V_e^{-1}a| + |aV_e^{-1}a| \|\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\| \|(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\| = O(1),
\end{aligned}$$

together with $\|\beta'V_\varepsilon^{-1}\beta\|^{\frac{1}{2}} = O(N^{\frac{1}{2}})$. □

B.2 Proof of Proposition 1

The idea of the proof is to apply the APT mathematics to $R_{t+1} - \beta(f_{t+1} + \lambda - \mathbb{E}(f_{t+1})) = \alpha + \varepsilon_{t+1}$. By [Chamberlain and Rothschild \(1983, Theorem 4\)](#) the error covariance matrix has an *approximate* factor structure, and satisfies

$$V_\varepsilon = \beta^{\text{mis}}\beta^{\text{mis}'} + V_e,$$

where $V_e > 0$ with uniformly bounded eigenvalues, and by [Chamberlain and Rothschild \(1983, Corollary 2\)](#) 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 made by the K^{mis} dominant eigenvectors of V_ε (the eigenvectors associated with the largest K^{can} eigenvalues), each multiplied by the square-root of the corresponding eigenvalues. We set $a = \alpha - \beta^{\text{mis}}\lambda^{\text{mis}}$.

By the Sherman-Morrison-Woodbury decomposition,

$$V_\varepsilon^{-1} = V_e^{-1} - V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}.$$

Therefore, by substitution,

$$\begin{aligned}
\alpha'V_\varepsilon^{-1}\alpha &= \alpha'V_e^{-1}\alpha - \alpha'V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}\alpha \\
&= (\beta^{\text{mis}}\lambda^{\text{mis}} + a)'V_e^{-1}(\beta^{\text{mis}}\lambda^{\text{mis}} + a) \\
&\quad - (\beta^{\text{mis}}\lambda^{\text{mis}} + a)'V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}(\beta^{\text{mis}}\lambda^{\text{mis}} + a) \\
&= \lambda^{\text{mis}'}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\lambda^{\text{mis}} \\
&\quad - \lambda^{\text{mis}'}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\lambda^{\text{mis}} \\
&\quad + a'V_e^{-1}a - a'V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}a \\
&\quad + 2a'V_e^{-1}\beta^{\text{mis}}\lambda^{\text{mis}} - 2a'V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\lambda^{\text{mis}}.
\end{aligned}$$

We now show that $\alpha'V_\varepsilon^{-1}\alpha$ is bounded for any N . We study each of the terms on the right-hand side of the last equality sign, one by one. The first and second term satisfy

$$\lambda^{\text{mis}'}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\lambda^{\text{mis}} - \lambda^{\text{mis}'}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}(I_{K^{\text{can}}} + \beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}})^{-1}\beta^{\text{mis}'}V_e^{-1}\beta^{\text{mis}}\lambda^{\text{mis}}$$

$$\begin{aligned}
&= \lambda^{\text{mis}'} (I_N - \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1}) \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}} \lambda^{\text{mis}} \\
&= \lambda^{\text{mis}'} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}} \lambda^{\text{mis}} \leq \lambda^{\text{mis}'} \lambda^{\text{mis}},
\end{aligned}$$

because $I_{K^{\text{can}}} - (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}}$ is positive semidefinite. Next, for the third term, denoting by $g_{NN}(V_e)$ the smallest eigenvalue of V_e ,

$$a' V_e^{-1} a \leq a' a / g_{NN}(V_e) = O(1).$$

Now, the j th element of $a' V_e^{-1} \beta^{\text{mis}}$, obtained by considering the j th column of β^{mis} , for every $1 \leq j \leq K^{\text{mis}}$, satisfies

$$|a' V_e^{-1} \beta_j^{\text{mis}}| \leq (a' V_e^{-1} a)^{\frac{1}{2}} (\beta_j^{\text{mis}'} V_e^{-1} \beta_j^{\text{mis}})^{\frac{1}{2}} = O(N^{\frac{1}{2}}),$$

Moreover, the (i, j) th element, for every $1 \leq i, j \leq K^{\text{can}}$, of $(\beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})$ is equal to $\beta_j^{\text{mis}'} V_e^{-1} \beta_i^{\text{mis}}$. Therefore, $(I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1}$ decreases at rate $O(N^{-1})$. On the other hand, along the same lines, the elements of the vector $\beta^{\text{mis}'} V_e^{-1} a$ diverge at most at rate $O(N^{\frac{1}{2}})$. Collecting terms, the fourth term satisfies:

$$|a' V_e^{-1} \beta^{\text{mis}} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_e^{-1} a| = O(1).$$

Concerning the last two terms, it turns out that their difference converges to zero. In fact,

$$\begin{aligned}
&|2a' V_e^{-1} \beta^{\text{mis}} \lambda^{\text{mis}} - 2a' V_e^{-1} \beta^{\text{mis}} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}} \lambda^{\text{mis}}| \\
&= 2|a' V_e^{-1} \beta^{\text{mis}} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \lambda^{\text{mis}}| \\
&\leq (a' V_e^{-1} \beta^{\text{mis}} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_e^{-1} a)^{\frac{1}{2}} (\lambda^{\text{mis}'} (I_{K^{\text{can}}} + \beta^{\text{mis}'} V_e^{-1} \beta^{\text{mis}})^{-1} \lambda^{\text{mis}})^{\frac{1}{2}} \\
&= O(N^{-\frac{1}{2}}).
\end{aligned}$$

□

Observe that Proposition 1 assumes the presence of at least one omitted systematic risk factor. The case when there are no missing systematic factors, that is, $K^{\text{mis}} = 0$, in turn implied when V_e has all bounded eigenvalues, coincides with the classical APT.

B.3 Proof of Proposition 2

In general, the SDF can always be re-written as *linear* in payoffs or excess returns. Without loss of generality we assume that the candidate factors are traded and expressed as excess returns, or as difference of excess returns (such as highest decile portfolio minus lowest

decile portfolio) implying $\mathbb{E}f_{t+1}^{\text{can}} = \lambda^{\text{can}}$. In fact, if the factors are non traded, by standard arguments one replaces them with the corresponding (traded) mimicking portfolios.

Therefore, given that in the APT payoffs (excess returns) are linear in f_t^{can} , $\varepsilon_{N,t}$, and 1, then the SDF under the APT satisfies:

$$M_{t+1} = \mathbb{E}(M_{t+1}) + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1}, \quad (\text{B1})$$

for some given coefficient vector b , which is $K^{\text{can}} \times 1$, and coefficient vector c , which is $N \times 1$. We determine b and c below whereas $\mathbb{E}(M_{t+1}) = R_{ft}^{-1}$, and given that we assumed the existence of the risk-free asset, $R_{ft} = 1 + r_{ft}$, it must be that:

$$\begin{aligned} 0_{K^{\text{can}}} &= \mathbb{E}(M_{t+1}f_{t+1}^{\text{can}}), \\ 0_N &= \mathbb{E}(M_{t+1}(R_{t+1} - R_{ft}1_N)), \end{aligned}$$

leading to a total of $K^{\text{can}} + N$ constraints. Substituting M_{t+1} from (B1) one gets:

$$\begin{aligned} 0_{K^{\text{can}}} &= \mathbb{E} \left[\left(\mathbb{E}(M_{t+1}) + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1} \right) f_{t+1}^{\text{can}} \right] \\ &= \mathbb{E} \left[\left(R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1} \right) f_{t+1}^{\text{can}} \right] \\ &= R_{ft}^{-1} \mathbb{E}(f_{t+1}^{\text{can}}) + \mathbb{E} \left(f_{t+1}^{\text{can}} (f_{t+1}^{\text{can}} - \lambda^{\text{can}})' b \right) \\ &\quad + \mathbb{E} \left(f_{t+1}^{\text{can}} \varepsilon_{N,t+1}' \right) c \\ &= R_{ft}^{-1} \lambda^{\text{can}} + V_{f^{\text{can}}} b, \end{aligned}$$

implying that

$$b = -R_{ft}^{-1} V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}.$$

Next,

$$\begin{aligned} 0_N &= \mathbb{E} \left[\left(R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1} \right) (R_{t+1} - R_{ft}1_N) \right] \\ &= \mathbb{E} \left[\left(R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1} \right) \times \right. \\ &\quad \left. \left(\alpha + \beta^{\text{can}} \lambda^{\text{can}} + \beta^{\text{can}} (f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + \varepsilon_{t+1} \right) \right] \\ &= R_{ft}^{-1} (\alpha + \beta^{\text{can}} \lambda^{\text{can}}) + \beta^{\text{can}} V_{f^{\text{can}}} b + V_{\varepsilon} c \\ &= R_{ft}^{-1} (\alpha + \beta^{\text{can}} \lambda^{\text{can}}) - R_{ft}^{-1} \beta^{\text{can}} V_{f^{\text{can}}} V_{f^{\text{can}}}^{-1} \lambda^{\text{can}} + V_{\varepsilon} c \\ &= R_{ft}^{-1} (\alpha + \beta^{\text{can}} \lambda^{\text{can}}) - R_{ft}^{-1} \beta^{\text{can}} \lambda^{\text{can}} + V_{\varepsilon} c \\ &= R_{ft}^{-1} \alpha + V_{\varepsilon} c, \end{aligned}$$

implying that

$$c = -R_{ft}^{-1} V_{\varepsilon}^{-1} \alpha.$$

□

B.4 Proof of Proposition 3

Assuming, without loss of generality, that the candidate model has only tradable factors represented by either factor returns in excess of the risk-free rate (for example, market factor) or long-minus-short strategies, we define the non-negative SDF to be:

$$M_{\text{exp},t+1} = \exp(\mu_m^+ + (b^+)'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + (c^+)' \varepsilon_{t+1}),$$

which implies that to identify $M_{\text{exp},t+1}$, we need to find: μ_m^+ , b^+ , and c^+ .

Imposing the following $1 + K^{\text{can}} + N$ constraints,

$$R_{ft}^{-1} = \mathbb{E}(M_{\text{exp},t+1}),$$

$$0_{K^{\text{can}}} = \mathbb{E}(M_{\text{exp},t+1} f_{t+1}^{\text{can}}),$$

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

allows one to identify $M_{\text{exp},t+1}$, as we show below. Starting with the first restriction, using Lemma B.1 below, we get:

$$\begin{aligned} R_{ft}^{-1} &= \mathbb{E}(M_{\text{exp},t+1}) = \mathbb{E}(\exp[\mu_m^+ + (b^+)'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c^{+'} \varepsilon_{t+1}]) \\ &= \exp[\mu_m^+] \exp\left[\left(\frac{1}{2} b^{+'} V_{f^{\text{can}}} b^+ + \frac{1}{2} c^{+'} \Sigma_{N,t} c^+\right)\right] \end{aligned}$$

implying

$$\exp[\mu_m^+] = R_{ft}^{-1} \exp\left[-\left(\frac{1}{2} b^{+'} V_f b^+ + \frac{1}{2} c^{+'} V_\varepsilon c^+\right)\right].$$

Next, considering the K^{can} restrictions and using Lemma B.1 again, we obtain:

$$\begin{aligned} 0_{K^{\text{can}}} &= \mathbb{E}(M_{\text{exp},t+1} f_{t+1}^{\text{can}}) \\ &= \mathbb{E}(M_{\text{exp},t+1} (f_{t+1}^{\text{can}} - \lambda^{\text{can}})) + \lambda^{\text{can}} \mathbb{E}(M_{\text{exp},t+1}) \\ &= \lambda^{\text{can}} \mathbb{E}(M_{\text{exp},t+1}) + e^{\mu_m^+} \mathbb{E}(e^{c^{+'} \varepsilon_{t+1}}) \mathbb{E}(e^{b^{+'} (f_{t+1}^{\text{can}} - \lambda^{\text{can}})}) (f_{t+1}^{\text{can}} - \lambda^{\text{can}}) \\ &= \lambda^{\text{can}} R_{ft}^{-1} + e^{(\mu_m^+ + \frac{1}{2} c^{+'} V_\varepsilon c^+ + \frac{1}{2} b^{+'} V_{f^{\text{can}}} b^+)} V_{f^{\text{can}}} b^+ \\ &= \lambda^{\text{can}} R_{ft}^{-1} + R_{ft}^{-1} V_{f^{\text{can}}} b^+ \end{aligned}$$

yielding

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

Finally, imposing the N restrictions and using Lemma B.1 again, we get:

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

$$\begin{aligned}
&= \mathbb{E}(M_{\text{exp},t+1}(\alpha + \beta^{\text{can}}\lambda^{\text{can}} + \beta^{\text{can}}(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + \varepsilon_{t+1})) \\
&= (\alpha + \beta^{\text{can}}\lambda^{\text{can}})\mathbb{E}(M_{\text{exp},t+1}) + \mathbb{E}(M_{\text{exp},t+1}\beta^{\text{can}}(f_{t+1}^{\text{can}} - \lambda^{\text{can}})) + \mathbb{E}(M_{\text{exp},t+1}\varepsilon_{t+1}) \\
&= (\alpha + \beta^{\text{can}}\lambda^{\text{can}})\mathbb{E}(M_{\text{exp},t+1}) - \beta^{\text{can}}\lambda^{\text{can}}\mathbb{E}(M_{\text{exp},t+1}) + R_{ft}^{-1}V_{\varepsilon}c^+,
\end{aligned}$$

implying that

$$c^+ = -V_{\varepsilon}^{-1}\alpha,$$

where we used $\mathbb{E}(M_{\text{exp},t+1}\varepsilon_{t+1}) = e^{(\mu_m^+ + \frac{1}{2}b^+V_{f^{\text{can}}}b^+)}\mathbb{E}(e^{c^+\varepsilon_{t+1}}\varepsilon_{t+1}) = R_{ft}^{-1}V_{\varepsilon}c^+$. Putting terms together

$$M_{\text{exp},t+1} = R_{ft}^{-1}e^{-[\lambda^{\text{can}}V_{f^{\text{can}}}^{-1}(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + \frac{1}{2}\lambda^{\text{can}}V_{f^{\text{can}}}^{-1}\lambda^{\text{can}}]}e^{-[\alpha'V_{\varepsilon}^{-1}\varepsilon_{t+1} + \frac{1}{2}\alpha'V_{\varepsilon}^{-1}\alpha]}.$$

□

B.5 Proof of Proposition 4

The result follows from Lemmas B.2 and B.3.

□

B.6 Proof of Proposition 5

Consider

$$\begin{aligned}
\hat{M}_{\text{exp},t+1}^{\alpha} &= \exp \left[-\alpha'V_R^{-1}(R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2}\alpha'V_R^{-1}\alpha \right] \\
&= \exp \left[-\alpha'(V_{\varepsilon}^{-1} - V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1})(R_{t+1} - \mathbb{E}[R_{t+1}]) - \frac{1}{2}\alpha'V_R^{-1}\alpha \right],
\end{aligned}$$

where the first term of the exponent can be written as the sum of three components:

$$\begin{aligned}
&\alpha'(V_{\varepsilon}^{-1} - V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1})(R_{t+1} - \mathbb{E}[R_{t+1}]) \\
&= \alpha'V_{\varepsilon}^{-1}\varepsilon_{t+1} - \alpha'V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1}\varepsilon_{t+1} \\
&+ \alpha'(V_{\varepsilon}^{-1} - V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1})\beta^{\text{can}}(f_{t+1}^{\text{can}} - \lambda^{\text{can}}).
\end{aligned}$$

As $N \rightarrow \infty$, the second and third components vanish and only the first component remains.

In fact, given $\varepsilon_{t+1} = \beta^{\text{mis}}(f_{t+1}^{\text{mis}} - \lambda^{\text{mis}}) + e_{t+1}$ and $\alpha = \beta^{\text{mis}}\lambda^{\text{mis}} + a$, one obtains

$$\begin{aligned}
&\alpha'V_{\varepsilon}^{-1}\varepsilon_{t+1} - \alpha'V_{\varepsilon}^{-1}\varepsilon_{t+1} \xrightarrow{p} 0, \text{ and} \\
&\alpha'V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1}\varepsilon_{t+1} = \\
&\alpha'V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{mis}}(f_{t+1}^{\text{mis}} - \lambda^{\text{mis}}) \\
&+ \alpha'V_{\varepsilon}^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}}V_{\varepsilon}^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}}V_{\varepsilon}^{-1}e_{t+1}
\end{aligned}$$

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

$$\alpha'(V_\varepsilon^{-1} - V_\varepsilon^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}'}V_\varepsilon^{-1}\beta^{\text{can}})^{-1}\beta^{\text{can}'}V_\varepsilon^{-1})\beta^{\text{can}}(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) =$$

$$\alpha'V_\varepsilon^{-1}\beta^{\text{can}}(V_{f^{\text{can}}}^{-1} + \beta^{\text{can}'}V_\varepsilon^{-1}\beta^{\text{can}})^{-1}V_{f^{\text{can}}}^{-1}(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) = o_p(N^{-\frac{1}{2}}),$$

making use of $\beta^{\text{can}'}V_\varepsilon^{-1}e_{t+1} = O_p((\beta^{\text{can}'}V_\varepsilon^{-1}V_eV_\varepsilon^{-1}\beta^{\text{can}})^{\frac{1}{2}}) = O_p(N^{\frac{1}{2}})$ and $\alpha'V_\varepsilon^{-1}\beta^{\text{can}} = \alpha'V_\varepsilon^{-1}\beta^{\text{can}} + \lambda^{\text{mis}'}\beta^{\text{mis}'}V_\varepsilon^{-1}\beta^{\text{can}} = o(N^{\frac{1}{2}}) + o(N^{\frac{1}{2}})$, recalling $V_e = \beta^{\text{can}}\beta^{\text{can}'} + V_e$. $\square$

Proof of Proposition 6

By Proposition 5, $\hat{M}_{t+1}^{\beta, \text{mis}} \xrightarrow{p} -R_{ft}^{-1}\lambda^{\text{mis}'}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}))$, setting for simplicity $M_{1_T} = I_T - 1_T 1_T'/T$, one obtains, given $M_{1_T} 1_T = O_{T \times T}$,

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

Thus, regarding the limit of R_g^2 , its numerator simplifies to

$$\begin{aligned} \gamma_1'(G' M_{1_T} G) \gamma_1 &= (R_{ft})^{-2} \lambda^{\text{mis}'} Q^{-1} Q (F^{\text{mis}'} M_{1_T} F^{\text{mis}}) Q' (Q')^{-1} \lambda^{\text{mis}} \\ &= (R_{ft})^{-2} \lambda^{\text{mis}'} (F^{\text{mis}'} M_{1_T} F^{\text{mis}}) \lambda^{\text{mis}}, \end{aligned}$$

and its denominator becomes

$$\begin{aligned} &= (R_{ft})^{-2} \lambda^{\text{mis}'} (F^{\text{mis}} - 1_T \mathbb{E}(f_{t+1}^{\text{mis}}))' M_{1_T} (F^{\text{mis}} - 1_T \mathbb{E}(f_{t+1}^{\text{mis}})) \lambda^{\text{mis}} \\ &= (R_{ft})^{-2} \lambda^{\text{mis}'} (F^{\text{mis}'} M_{1_T} F^{\text{mis}}) \lambda^{\text{mis}}, \end{aligned}$$

and thus identical to the numerator of the limit of R_g^2 .

The case of orthogonal G and F^{mis} is straightforward and so we omit details. $\square$

B.7 Proposition B1

Proposition B1 (Parameter estimates of APT). *Suppose that the vector of asset returns R_{t+1} satisfies the data-generating process given in equations (1) and (2). Assume that the number of missing factors in the candidate model, K^{mis} , is known, that the sample covariance matrix of candidate factors $\hat{V}_{f^{\text{can}}}$ is nonsingular, with $\hat{V}_{f^{\text{can}}} = \hat{M}_{f^{\text{can}}} - \bar{f}^{\text{can}} \bar{f}^{\text{can}'}$, setting $\hat{M}_{f^{\text{can}}} = T^{-1} \sum_{t=1}^T f_t^{\text{can}} f_t^{\text{can}'}$, $\bar{f}^{\text{can}} = T^{-1} \sum_{t=1}^T f_t^{\text{can}}$, and setting $\hat{M}_{R f^{\text{can}}} = \frac{1}{T} \sum_{t=1}^T (R_t - R_{ft} 1_N) f_t^{\text{can}'}$, $\bar{R} = \frac{1}{T} \sum_{t=1}^T R_t$, and $\bar{R}_f = \frac{1}{T} \sum_{t=1}^T R_{ft}$.*

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier $\hat{\kappa}$ is greater than zero, then

$$\text{vec}(\hat{\beta}^{\text{can}}) = ((\hat{M}_{f^{\text{can}}} \otimes I_N) - (\bar{f}^{\text{can}} \bar{f}^{\text{can}'} \otimes G))^{-1} \text{vec}(\hat{M}_{Rf^{\text{can}}} - G(\bar{R} - R_f 1_N) \bar{f}^{\text{can}'}), \quad (\text{B2})$$

$$\hat{\lambda}^{\text{mis}} = (\hat{\beta}^{\text{mis}'} \hat{V}_\varepsilon^{-1} \hat{\beta}^{\text{mis}})^{-1} \hat{\beta}^{\text{mis}'} \hat{V}_\varepsilon^{-1} (\bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}} \bar{f}^{\text{can}}), \quad (\text{B3})$$

$$\hat{a} = \frac{1}{\hat{\kappa} + 1} (\bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}} \bar{f}^{\text{can}} - \hat{\beta}^{\text{mis}} \hat{\lambda}^{\text{mis}}). \quad (\text{B4})$$

in which

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

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

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

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

$$\hat{\alpha} = \bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}} \bar{f}^{\text{can}}, \quad (\text{B5})$$

and the expression for $\text{vec}(\hat{\beta}^{\text{can}})$ can be obtained by setting $\hat{\kappa} = 0$ in (B2). The expressions for $\hat{\lambda}^{\text{can}}$ and $V_{f^{\text{can}}}$ are unchanged, and, as before, the expressions for the estimators of $\hat{\beta}^{\text{mis}}$ and $\hat{V}_e$ do not admit a closed-form solution.

Proof of Proposition B1

Defining by $\tilde{\theta}$ the constrained maximum likelihood estimator corresponding to $\tilde{\kappa} = 0$, this is infeasible whenever $\tilde{a}' \tilde{V}_\varepsilon^{-1} \tilde{a} > \delta_{\text{apt}}$. Similarly, the case $\hat{\kappa} > 0$ is infeasible whenever, for every $\hat{\kappa} > 0$,

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

because

$$(1 + \hat{\kappa})^2 = \frac{(\bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}} \bar{f}^{\text{can}} - \hat{\beta}^{\text{mis}} \hat{\lambda}^{\text{mis}})' \hat{V}_\varepsilon^{-1} (\bar{R} - \bar{R}_f 1_N - \hat{\beta}^{\text{can}} \bar{f}^{\text{can}} - \hat{\beta}^{\text{mis}} \hat{\lambda}^{\text{mis}})}{\delta_{\text{apt}}}.$$

When both cases are feasible, the optimal value for the Karush-Kuhn-Tucker multiplier $\tilde{\kappa}$ will be either 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 $\hat{\kappa} > 0$ then $\hat{a}'\hat{V}_\varepsilon^{-1}\hat{a} = \delta_{\text{apt}}$ by construction.

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

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

where recall that $V_\varepsilon = \beta^{\text{mis}} \beta^{\text{mis}'} + V_e$. It is straightforward to see that, because of the APT restriction, λ^{mis} and a can now be identified separately, as long as $\hat{\kappa} > 0$. In fact, the above system of linear equations can be solved because the matrix premultiplying $\hat{\lambda}^{\text{mis}}$ and $\hat{a}$ is non-singular for every $\hat{\kappa} > 0$, leading to the closed-form solution (B3) and (B4).

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

$$-\frac{1}{T} \sum_{t=1}^T \left(R_t - R_{ft} 1_N - \beta^{\text{mis}} \lambda^{\text{mis}} - a - \hat{\beta}^{\text{can}} f_t^{\text{can}} \right)' V_\varepsilon^{-1} \left(-\frac{\partial \beta^{\text{can}}}{\partial B_{2ab}} f_t \right) = 0,$$

which can be re-arranged as

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

Inserting (B3) and (B4), and G , and re-arranging terms yields

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

which can be rewritten more succinctly as

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

with $g_t = \left(R_t - R_{ft} 1_N - G(\bar{R} - \bar{R}_f 1_N) - \hat{\beta}^{\text{can}} f_t^{\text{can}} + G \hat{\beta}^{\text{can}} \bar{f}^{\text{can}} \right)$. Taking the vec operator and solving for $\hat{\beta}^{\text{can}}$ gives the desired expression in (B2).

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

$$\begin{aligned} & \left((\hat{M}_{f^{\text{can}}} \otimes I_N) - (\bar{f}^{\text{can}} \bar{f}^{\text{can}'} \otimes G) \right) \\ &= \left(((\hat{M}_{f^{\text{can}}} - \bar{f}^{\text{can}} \bar{f}^{\text{can}'}) \otimes I_N) + (\bar{f}^{\text{can}} \bar{f}^{\text{can}'} \otimes (I_N - G)) \right). \end{aligned}$$

The first matrix on the right hand side is non-singular by assumption. One then just needs to show that the second matrix is positive semi-definitive. This follows because $(I_N - G)$ is positive semi-definite.

$$\begin{aligned}
I_N - G &= I_N - \frac{1}{(\hat{\kappa} + 1)} I_N - \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}}\right) \beta^{\text{mis}} (\beta^{\text{mis}'} V_\varepsilon^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_\varepsilon^{-1} \\
&= \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}}\right) (I_N - \beta^{\text{mis}} (\beta^{\text{mis}'} V_\varepsilon^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_\varepsilon^{-1}) \\
&= \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}}\right) V_\varepsilon (V_\varepsilon^{-1} - V_\varepsilon^{-1} \beta^{\text{mis}} (\beta^{\text{mis}'} V_\varepsilon^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} V_\varepsilon^{-1}) \\
&= \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}}\right) V_\varepsilon (V_\varepsilon^{-1})^{\frac{1}{2}} (I_N - (V_\varepsilon^{-1})^{\frac{1}{2}} \beta^{\text{mis}} (\beta^{\text{mis}'} V_\varepsilon^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} (V_\varepsilon^{-1})^{\frac{1}{2}}) (V_\varepsilon^{-1})^{\frac{1}{2}}.
\end{aligned}$$

The right-hand side is the product of positive-definite matrices, namely V_ε and $(V_\varepsilon^{-1})^{\frac{1}{2}}$, and of the matrix $I_N - (V_\varepsilon^{-1})^{\frac{1}{2}} \beta^{\text{mis}} (\beta^{\text{mis}'} V_\varepsilon^{-1} \beta^{\text{mis}})^{-1} \beta^{\text{mis}'} (V_\varepsilon^{-1})^{\frac{1}{2}}$, which is the projection matrix orthogonal to $(V_\varepsilon^{-1})^{\frac{1}{2}} \beta^{\text{mis}}$, and therefore, positive semidefinite itself.

Therefore, plugging $\hat{\beta}^{\text{can}}$ into $\hat{\lambda}^{\text{mis}}$ and $\hat{a}$, one obtains that

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

Substituting them, together with $\hat{\beta}^{\text{can}} = \hat{\beta}^{\text{can}}(\beta^{\text{mis}}, V_e)$, into $L(\theta) - \kappa(a' V_\varepsilon^{-1} a - \delta_{\text{apt}})$, gives the concentrated likelihood function, which is a function of only β^{mis} and V_e which will be maximized numerically, providing $\hat{\beta}^{\text{mis}}$ and $\hat{V}_e$.

Suppose now that $\hat{\kappa} = 0$ holds, and recall that in this case the MLC is indicated by $\hat{\theta}$.

One can clearly obtain a unique solution to the linear system of equations $(\beta^{\text{mis}}, I_N) \begin{pmatrix} \hat{\lambda}^{\text{mis}} \\ \hat{a} \end{pmatrix} = \beta^{\text{mis}} \hat{\lambda}^{\text{mis}} + \hat{a}$. However, to solve for $\hat{\lambda}^{\text{mis}}$ and $\hat{a}$ separately, one needs to invert the matrix

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

which is not possible because it is of dimension $(N + K^{\text{mis}}) \times (N + K^{\text{mis}})$ but of rank N , as the left-hand side shows that it is obtained from the product of two matrices of dimension $(N + K^{\text{mis}}) \times N$. All the other parameters are identified separately, and their expressions follow from differentiating $L(\theta)$ and solving the resulting first-order conditions. For instance, the formula for $\hat{\beta}^{\text{can}}$ follows by setting $G = I_N$ into (B2). $\square$

C Online Appendix with Additional Results

This appendix contains additional results and generalizations of some of the results reported in the main text.

C.1 The SDF with Non-Orthogonal Components

All the previous results were derived under the assumption that the candidate risk factors f_t^{can} and the unobserved idiosyncratic shock ε_{t+1} are (conditionally) orthogonal, as formalized in Assumption 1. However, one can envisage situations where orthogonality does not necessarily hold, the best example being when there are missing systematic risk factors that are *hidden* in the idiosyncratic shock and are correlated with the observed risk factors.²⁴

In this case, note that an *observationally equivalent* representation of the SDF M_{t+1} exists such that the observed risk factors f_{t+1}^{can} and the unobserved idiosyncratic shock ε_{t+1} are orthogonal. In particular, recalling that $\varepsilon_{t+1} = e_{t+1} + \beta^{\text{mis}}(f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}))$, as before, with e_{t+1} and f_{t+1}^{mis} being mutually uncorrelated,

$$M_{t+1} = R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + c'\varepsilon_{t+1}, \quad (\text{C1})$$

$$= R_{ft}^{-1} + \tilde{b}'(f_{t+1}^{\text{can}} - E_t(f_{t+1}^{\text{can}})) + c'\tilde{\varepsilon}_{t+1}, \quad (\text{C2})$$

setting the $K^{\text{can}} \times K^{\text{mis}}$ matrix of covariances $Q = \text{cov}(f_{t+1}^{\text{can}}, f_{t+1}^{\text{mis}'})$ with

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

$$\tilde{\varepsilon}_{t+1} = e_{t+1} + \beta^{\text{mis}}(\tilde{f}_{t+1}^{\text{mis}} - \mathbb{E}(\tilde{f}_{t+1}^{\text{mis}})), \quad \text{where}$$

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

Notice that by construction $\text{cov}(f_{t+1}^{\text{can}}, \tilde{f}_{t+1}^{\text{mis}'}) = 0_{K^{\text{can}} \times K^{\text{mis}}}$, 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 $f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})$.

Although the two representations (C1) and (C2) are observationally equivalent, the one based on correlated components, that is (C1), has the advantage of ensuring a cleaner interpretation of the parameters, such as the ones for loadings and risk premia. For instance, the loadings associated with f_{t+1}^{can} in representation (C2) differ from the (true) loadings of f_{t+1}^{can} in representation (C1), a consequence of the omitted-variable bias. This can be seen

²⁴We report only the results valid for finite N . The large N analysis follow along the steps outlined in Section 2.3.2.

by comparing the extended APT in the orthogonal and non-orthogonal representations:

$$\begin{aligned} R_{t+1} - R_{ft}1_N &= a + (\beta^{\text{mis}}, \beta^{\text{can}}) \begin{pmatrix} \lambda^{\text{mis}} \\ \lambda^{\text{can}} \end{pmatrix} + (\beta^{\text{mis}}, \beta^{\text{can}}) \begin{pmatrix} f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}) \\ f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}) \end{pmatrix} + e_{t+1}, \\ &= a + (\tilde{\beta}^{\text{mis}}, \tilde{\beta}^{\text{can}}) \begin{pmatrix} \tilde{\lambda}^{\text{mis}} \\ \lambda^{\text{can}} \end{pmatrix} + (\tilde{\beta}^{\text{mis}}, \tilde{\beta}^{\text{can}}) \begin{pmatrix} \tilde{f}_{t+1}^{\text{mis}} - \mathbb{E}(\tilde{f}_{t+1}^{\text{mis}}) \\ f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}) \end{pmatrix} + e_{t+1}, \end{aligned}$$

where

$$\begin{aligned} \tilde{\beta}^{\text{mis}} &= \beta^{\text{mis}}(I_{K^{\text{mis}}} - Q'V_{f^{\text{can}}}^{-1}Q)^{\frac{1}{2}}, \\ \tilde{\beta}^{\text{can}} &= \beta^{\text{can}} + \beta^{\text{mis}}Q'V_{f^{\text{can}}}^{-1}, \\ \tilde{\lambda}^{\text{mis}} &= (I_{K^{\text{mis}}} - Q'V_{f^{\text{can}}}^{-1}Q)^{-\frac{1}{2}}(\lambda^{\text{mis}} - Q'V_{f^{\text{can}}}^{-1}\lambda^{\text{can}}), \end{aligned}$$

and $\tilde{f}_{t+1}^{\text{mis}}$ has (conditionally) unit covariance matrix and is uncorrelated with f_{t+1}^{can} .²⁵ Notice that $(\beta^{\text{mis}}, \beta^{\text{can}}) \begin{pmatrix} \lambda^{\text{mis}} \\ \lambda^{\text{can}} \end{pmatrix} = (\tilde{\beta}^{\text{mis}}, \tilde{\beta}^{\text{can}}) \begin{pmatrix} \tilde{\lambda}^{\text{mis}} \\ \lambda^{\text{can}} \end{pmatrix}$ as the possibility of a non-zero Q does not affect expected excess returns $\mathbb{E}(R_{t+1} - R_{ft}1_N)$.

We now show how all our results can be generalized to allow for the case of correlated observed and missing factors. In particular, we need to generalize Assumption 1 to:

Assumption C1 (Linear factor model: correlated case). *Assumption 1 holds with*

$$\mathbb{E}(f_{t+1}^{\text{can}}\varepsilon'_{t+1}) = P,$$

for some non-zero $K^{\text{can}} \times N$ matrix P such that perfect (conditional) correlation between f_{t+1}^{can} and ε_{t+1} is ruled out:

$$I_N - (V_\varepsilon^{-1})^{\frac{1}{2}}P'V_{f^{\text{can}}}^{-1}P(V_\varepsilon^{-1})^{\frac{1}{2}} > 0.$$

When $V_\varepsilon = \beta^{\text{mis}}\beta^{\text{mis}'} + V_e$, then

$$P = Q\beta^{\text{mis}}, \tag{C3}$$

but we keep the more general notation in terms of P in order to provide the SDF formulae. However, when constructing estimators for the model's parameters, we will impose (C3).

²⁵In particular, $\tilde{f}_{t+1}^{\text{mis}}$ is given by:

$$\tilde{f}_{t+1}^{\text{mis}} = (I_{K^{\text{mis}}} - Q'V_{f^{\text{can}}}^{-1}Q)^{-\frac{1}{2}}(I_{K^{\text{mis}}} - Q'V_{f^{\text{can}}}^{-1}) \begin{pmatrix} f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}}) \\ f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}}) \end{pmatrix}.$$

Although the expression for expected excess returns is unchanged, the variance for excess returns becomes:

$$\text{cov}(R_{t+1} - R_{ft}1_N) = V_R = \beta^{\text{can}} V_f \beta^{\text{can}'} + V_\varepsilon + P' \beta^{\text{can}'} + \beta^{\text{can}} P.$$

We show how the expressions for the linear and exponential SDF change in the absence of orthogonality.

Proposition C1 (SDF: Correlated case). *Under Assumptions C1 and 2 of the APT, there exists an admissible SDF of the form*

$$\begin{aligned} M_{t+1} &= R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + c'\varepsilon_{t+1}, \quad \text{with} \\ b &= -R_{ft}^{-1} \left(V_{f^{\text{can}}}^{-1} \lambda^{\text{can}} - V_{f^{\text{can}}}^{-1} P H^{-1} (\alpha - P' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}) \right), \\ c &= -R_{ft}^{-1} \left(H^{-1} (\alpha - P' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}) \right), \quad \text{where} \\ H &= V_\varepsilon - P' V_{f^{\text{can}}}^{-1} P. \end{aligned}$$

When expressed in terms of a linear projection on the set of payoffs $(1, R_{t+1} - R_{ft}1_N)$, the SDF is

$$\begin{aligned} \hat{M}_{t+1} &= R_{ft}^{-1} + (b'[V_{f^{\text{can}}} \beta^{\text{can}'} + P] + c'[V_\varepsilon + P' \beta^{\text{can}'}]) V_R^{-1} (R_{t+1} - \mathbb{E}(R_{t+1})) \\ &= R_{ft}^{-1} - (\alpha + \beta^{\text{can}} \lambda^{\text{can}})' V_R^{-1} (R_{t+1} - \mathbb{E}(R_{t+1})). \end{aligned}$$

Proof of Proposition C1

We start with the conjecture that the SDF is still linear in the observed factors f_{t+1}^{can} and idiosyncratic risk ε_{t+1} , although now these can be cross-correlated. Stating the $K^{\text{can}} + N$ pricing equations:

$$\begin{aligned} 0_{K^{\text{can}}} &= \mathbb{E}(M_{t+1} f_{t+1}^{\text{can}}) \\ 0_N &= \mathbb{E}(M_{t+1} (R_{t+1} - R_{ft}1_N)), \end{aligned}$$

yields,

$$\begin{aligned} 0_{N+K^{\text{can}}} &= \begin{pmatrix} \mathbb{E} \left[\left(R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1} \right) f_{t+1}^{\text{can}} \right] \\ \mathbb{E} \left[\left(R_{ft}^{-1} + b'(f_{t+1}^{\text{can}} - \lambda^{\text{can}}) + c'\varepsilon_{t+1} \right) (R_{t+1} - R_{ft}1_N) \right] \end{pmatrix} \\ &= R_{ft}^{-1} \begin{pmatrix} \lambda^{\text{can}} \\ \beta^{\text{can}} \lambda^{\text{can}} + \alpha \end{pmatrix} + \begin{pmatrix} V_{f^{\text{can}}} b + P c \\ (\beta^{\text{can}} V_{f^{\text{can}}} + P') b + (V_\varepsilon + \beta^{\text{can}} P) c \end{pmatrix} \end{aligned}$$

$$= R_{ft}^{-1} \begin{pmatrix} \lambda^{\text{can}} \\ \beta^{\text{can}} \lambda^{\text{can}} + \alpha^{\text{can}} \end{pmatrix} + \begin{pmatrix} V_{f^{\text{can}}} & P \\ (\beta^{\text{can}} V_{f^{\text{can}}} + P') & (V_\varepsilon + \beta^{\text{can}} P) \end{pmatrix} \begin{pmatrix} b \\ c \end{pmatrix}.$$

Using the blockwise formula for the inverse of a matrix, in view of the lack of perfect correlation between the f_{t+1}^{can} and the ε_{t+1} , one obtains the solution:

$$\begin{aligned} \begin{pmatrix} b \\ c \end{pmatrix} &= -R_{ft}^{-1} \begin{pmatrix} V_{f^{\text{can}}} & P \\ (\beta^{\text{can}} V_{f^{\text{can}}} + P') & (V_\varepsilon + \beta^{\text{can}} V_{f^{\text{can}}}) \end{pmatrix}^{-1} \begin{pmatrix} \lambda^{\text{can}} \\ \beta^{\text{can}} \lambda^{\text{can}} + \alpha \end{pmatrix} \\ &= -R_{ft}^{-1} \begin{pmatrix} V_{f^{\text{can}}}^{-1} + V_{f^{\text{can}}}^{-1} P H^{-1} (\beta^{\text{can}} + P' V_{f^{\text{can}}}^{-1}) & -V_{f^{\text{can}}}^{-1} P H^{-1} \\ -H^{-1} (\beta^{\text{can}} + P' V_{f^{\text{can}}}^{-1}) & H^{-1} \end{pmatrix} \begin{pmatrix} \lambda^{\text{can}} \\ \beta^{\text{can}} \lambda^{\text{can}} + \alpha \end{pmatrix} \\ &= -R_{ft}^{-1} \begin{pmatrix} V_{f^{\text{can}}}^{-1} \lambda^{\text{can}} - V_{f^{\text{can}}}^{-1} P H^{-1} (\alpha - P' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}) \\ H^{-1} (\alpha - P' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}) \end{pmatrix}. \end{aligned}$$

We now establish the result for the projection SDF $\hat{M}_{t+1}$. By construction, setting $X_{t+1} = (1, R'_{t+1} - R_{ft} 1'_N)'$ and for simplicity using $\mu = \mathbb{E}(R_{t+1} - R_{ft} 1_N)$,

$$\begin{aligned} \hat{M}_{t+1} &= \mathbb{E}(M_{t+1} X'_{t+1}) (\mathbb{E}(X_{t+1} X'_{t+1}))^{-1} X_{t+1} \\ &= (R_{ft}^{-1}, R_{ft}^{-1} \mu' + b' [V_{f^{\text{can}}} \beta^{\text{can}'} + P] + c' [V_\varepsilon + P' \beta^{\text{can}'}]) \begin{pmatrix} 1 + \mu' V_R^{-1} \mu & -\mu' V_R^{-1} \\ -V_R^{-1} \mu & V_R^{-1} \end{pmatrix} X_{t+1} \\ &= (R_{ft}^{-1} - (b' [V_{f^{\text{can}}} \beta^{\text{can}'} + P] + c' [V_\varepsilon + P' \beta^{\text{can}'}]) V_R^{-1} \mu, (b' [V_{f^{\text{can}}} \beta^{\text{can}'} + P] + c' [V_\varepsilon + P' \beta^{\text{can}'}]) V_R^{-1} X_{t+1} \\ &= R_{ft}^{-1} + (b' [V_{f^{\text{can}}} \beta^{\text{can}'} + P] + c' [V_\varepsilon + P' \beta^{\text{can}'}]) V_R^{-1} (R_{t+1} - R_{ft} 1_N - \mu), \end{aligned}$$

where we apply the block formula for the inverse of a square matrix to $\mathbb{E}(X_{t+1} X'_{t+1})$, which exists because of our assumption of less-than-perfect correlation between observed factors and idiosyncratic shocks. Finally, by means of algebraic manipulations,

$$b' [V_{f^{\text{can}}} \beta^{\text{can}'} + P] + c' [V_\varepsilon + P' \beta^{\text{can}'}] = \lambda^{\text{can}'} \beta^{\text{can}'} + \alpha',$$

which gives the desired result. $\square$

Note that although the expressions for the coefficients in the SDF, namely b and c , differ from the case $P = 0_{K^{\text{can}} \times N}$, one still obtains the decomposition into the alpha- and beta-SDFs:

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

where

$$M_{t+1}^{\beta, \text{can}} = R_{ft}^{-1} - R_{ft}^{-1} b' (f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) \quad \text{and} \quad M_{t+1}^\alpha = -R_{ft}^{-1} \varepsilon'_{t+1} c,$$

where b and c are defined in Proposition C1. In terms of the pricing of asset returns:

$$\mathbb{E} \left(M_{t+1}^{\beta, \text{can}} \begin{bmatrix} 1 \\ R_{t+1} - R_{ft} 1_N \end{bmatrix} \right) = R_{ft}^{-1} \begin{bmatrix} 1 \\ (V_\varepsilon + \beta^{\text{can}} P) c \end{bmatrix}$$

$$\mathbb{E} \left(M_{t+1}^\alpha \begin{bmatrix} 1 \\ R_{t+1} - R_{ft} 1_N \end{bmatrix} \right) = R_{ft}^{-1} \begin{bmatrix} 0 \\ -(V_\varepsilon + \beta^{\text{can}} P)c \end{bmatrix}.$$

Notice that now $\text{cov}(M_{t+1}^\alpha, M_{t+1}^{\beta, \text{can}}) = R_{ft}^2 b' P c \neq 0$. Despite this, as for the earlier orthogonal case, the misspecified M_{t+1}^{can} prices the observed factors correctly, that is $\mathbb{E}(M_{t+1}^{\beta, \text{can}} f_{t+1}^{\text{can}}) = 0_{K^{\text{can}}}$.

Likewise, one obtains the decomposition in terms of linear projections as:

$$\begin{aligned} \hat{M}_{t+1} &= \hat{M}_{t+1}^\alpha + \hat{M}_{t+1}^{\beta, \text{can}}, \quad \text{with} \\ \hat{M}_{t+1}^{\beta, \text{can}} &= R_{ft}^{-1} + b'[V_{f^{\text{can}}} \beta^{\text{can}'} + P]V_R^{-1}(R_{t+1} - \mathbb{E}(R_{t+1})), \\ \hat{M}_{t+1}^\alpha &= c'[V_\varepsilon + P' \beta^{\text{can}'}]V_R^{-1}(R_{t+1} - \mathbb{E}(R_{t+1})), \end{aligned}$$

where it can be shown that the previous large- N results extend also to the non-orthogonal case.²⁶

C.2 The Nonnegative SDF with Non-Orthogonal Components

Given the strong analogies between the specifications of the linear and nonnegative SDF cases, we introduce the nonnegative SDF for the case of correlated components, and its corresponding decomposition in terms of (nonlinear) projections, without a formal proof.

Proposition C2 (Nonnegative SDF: Correlated case). *Under Assumptions C1 and 2 of the APT and that returns are conditionally Gaussian, there exists an admissible SDF $M_{\text{exp}, t+1}$ of the form*

$$\begin{aligned} M_{\text{exp}, t+1} &= \exp \left[\mu_m^+ + b^{+'}(f_{t+1}^{\text{can}} - \mathbb{E}(f_{t+1}^{\text{can}})) + c^{+'}\varepsilon_{t+1} \right], \quad \text{with} \\ \mu_m^+ &= \ln(R_{ft}^{-1}) - \frac{1}{2}(b^{+'}, c^{+'}) \begin{pmatrix} V_{f^{\text{can}}} & P \\ P' & V_\varepsilon \end{pmatrix} \begin{pmatrix} b^+ \\ c^+ \end{pmatrix}, \\ b^+ &= - \left(V_{f^{\text{can}}}^{-1} \lambda^{\text{can}} - V_{f^{\text{can}}}^{-1} P H^{-1} (\alpha - P' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}) \right), \\ c^+ &= - \left(H^{-1} (\alpha - P' V_{f^{\text{can}}}^{-1} \lambda^{\text{can}}) \right), \end{aligned}$$

recalling $H = V_\varepsilon - P' V_{f^{\text{can}}}^{-1} P$.

Proof of Proposition C2

We omit the proof because it is similar to the proof for Proposition 3, but using Proposition C1. □

²⁶Details are available upon request.

Note that the relevant decomposition of the nonnegative SDF in terms of (nonlinear) projections is given by:

$$\hat{M}_{\text{exp},t+1} = \hat{M}_{\text{exp},t+1}^{\alpha} \hat{M}_{\text{exp},t+1}^{\beta,\text{can}}, \quad \text{where}$$

$$\hat{M}_{\text{exp},t+1}^{\beta,\text{can}} = R_{ft}^{-1} \exp \left[b^{+'} (V_{f^{\text{can}}} \beta^{\text{can}'} + P) V_R^{-1} (R_{t+1} - \mathbb{E}(R_{t+1}) - \frac{1}{2} b^{+'} V_{f^{\text{can}}} b^{+}) \right], \quad \text{and}$$

$$\hat{M}_{\text{exp},t+1}^{\alpha*+} = \exp \left[c^{+'} (V_{\varepsilon} + P' \beta^{\text{can}'} V_R^{-1} (R_{t+1} - \mathbb{E}(R_{t+1})) - \frac{1}{2} (b^{+'}, c^{+'}) \begin{pmatrix} 0_{K^{\text{can}} \times K^{\text{can}}} & P \\ P' & V_{\varepsilon} \end{pmatrix} \begin{pmatrix} b^{+} \\ c^{+} \end{pmatrix} \right],$$

where the previous large- N results extend also to this non-negative non-orthogonal case.²⁷

C.3 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

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

$$\alpha = a + \beta^{\text{mis}} \lambda^{\text{mis}}, \quad \varepsilon_{t+1} = \beta^{\text{mis}} (f_{t+1}^{\text{mis}} - \mathbb{E}(f_{t+1}^{\text{mis}})) + e_{t+1},$$

where we set $\beta^{\text{can}} = (\beta_1^{\text{can}}, \beta_2^{\text{can}})$, $V_f = \text{var}(f_{t+1})$, $f_{t+1} = (f_{1t+1}', f_{2t+1}^e)'$, 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, where $K^{\text{can}} = K_1^{\text{can}} + K_2^{\text{can}}$. We assume that the missing factors are uncorrelated with the observed factors.²⁸ Given that f_{2t}^{can} are excess returns, or difference of, on tradable assets, their risk premia satisfy $\lambda_2^{\text{can}} = E(\mathbf{f}_{2t}^{\text{can}})$ 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:

$$L(\tilde{\theta}) = -\frac{1}{2} \log(\det(\beta^{\text{mis}} \tilde{\beta}^{\text{mis}'} + \tilde{V}_e)) \tag{C4}$$

$$- \frac{1}{2T} \sum_{t=1}^T \left(R_t - R_{ft} 1_N - \beta^{\text{mis}} \tilde{\lambda}^{\text{mis}} - \tilde{a} - \tilde{\beta}_1^{\text{can}} (\tilde{\lambda}_1^{\text{can}} + f_{1t}^{\text{can}} - \tilde{\mathbb{E}}(f_{1t}^{\text{can}})) - \tilde{\beta}_2^{\text{can}} f_{2t}^{\text{can}} \right)'$$

$$\times (\beta^{\text{mis}} \tilde{\beta}^{\text{mis}'} + \tilde{V}_e)^{-1} \left(R_t - R_{ft} 1_N - \beta^{\text{mis}} \tilde{\lambda}^{\text{mis}} - \tilde{a} - \tilde{\beta}_1^{\text{can}} (\tilde{\lambda}_1^{\text{can}} + f_{1t}^{\text{can}} - \tilde{\mathbb{E}}(f_{1t}^{\text{can}})) - \tilde{\beta}_2^{\text{can}} f_{2t}^{\text{can}} \right)$$

²⁷Details are available upon request.

²⁸The estimator can be extended to the case of correlated observed and omitted risk factors; details are available upon request.

$$-\frac{1}{2} \log(\det(\tilde{V}_f)) - \frac{1}{2T} \sum_{t=1}^T (f_t - \tilde{\mathbb{E}}(f_t))' \tilde{V}_f^{-1} (f_t - \tilde{\mathbb{E}}(f_t)).$$

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 C3 (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}'} is nonsingular, where $\hat{M}_{f_2^{\text{can}}} = T^{-1} \sum_{t=1}^T f_{2t}^{\text{can}} f_{2t}^{\text{can}'}$ and $\bar{f}_2^{\text{can}} = T^{-1} \sum_{t=1}^T f_{2t}^{\text{can}}$. Then$*

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

where $L(\tilde{\theta})$ is defined in (C4), and $\hat{\theta} = (\hat{a}', \lambda^{\text{mis}'}', \hat{\lambda}_1^{\text{can}'}, \hat{\mathbb{E}}(f_{1t}^{\text{can}})', \hat{\mathbb{E}}(f_{2t}^{\text{can}})', \text{vec}(\beta^{\text{mis}})', \text{vec}(\hat{\beta}^{\text{can}})', \text{vech}(\hat{V}_e)', \text{vech}(\hat{V}_f)')$.

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

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

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

$$\text{vec}(\hat{B}_2^{\text{can}}) = \left((\hat{M}_{f_2^{\text{can}}} \otimes I_N) - (\bar{f}_2^{\text{can}} \bar{f}_2^{\text{can}'} \otimes \hat{G}) \right)^{-1} \text{vec} \left(\hat{M}_{h f_2^{\text{can}}} - \bar{G} \bar{h} \bar{f}_2^{\text{can}'} \right), \quad (\text{C5})$$

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

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

where $\hat{V}_\varepsilon = \beta^{\text{mis}} \beta^{\text{mis}'} + \hat{V}_e$, $\hat{M}_{h f_2^{\text{can}}} = T^{-1} \sum_{t=1}^T h_t f_{2t}^{\text{can}'}$, $\bar{h} = T^{-1} \sum_{t=1}^T h_t$ 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 $\hat{D} = (\beta^{\text{mis}}, \hat{B}_1^{\text{can}})$ and $\hat{V}_e$ do not admit a closed-form solution 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

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

and the expression for $\text{vec}(\hat{\beta}_2^{\text{can}})$ can be obtained by setting $\hat{\kappa} = 0$ in the terms that appear in (C5). 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,

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

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

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

where recall that $V_\varepsilon = \beta^{\text{mis}}\beta^{\text{mis}'} + V_e$, 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). It is straightforward to see that, because of the APT restriction, λ and a can now 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 non-singular for every $\hat{\kappa} > 0$, leading to the closed-form solution:

$$\hat{\lambda} = (D'V_\varepsilon^{-1}D)^{-1}D'V_\varepsilon^{-1}(\bar{R} - \bar{R}_f 1_N - \beta_2^{\text{can}} \bar{f}_2^{\text{can}}), \quad (\text{C6})$$

$$\hat{a} = \frac{1}{\hat{\kappa} + 1}(\bar{R} - \bar{R}_f 1_N - \beta_2^{\text{can}} \bar{f}_2^{\text{can}} - D\hat{\lambda}). \quad (\text{C7})$$

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 $1 \leq a \leq N$, $1 \leq b \leq K_2^{\text{can}}$, one obtains

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

Now, inserting (C6) and (C7) into the above expression, setting

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

and re-arranging terms yields, setting $\hat{M}_{f_1^{\text{can}} f_2^{\text{can}}} = 1/T \sum_{t=1}^T f_{1t}^{\text{can}} f_{2t}^{\text{can}'}$,

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

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 $g_t = (h_t - G\bar{h} - \hat{\beta}_2^{\text{can}} f_{2t}^{\text{can}} + G \hat{\beta}_2^{\text{can}} \bar{f}_2^{\text{can}})$. Taking the vec operator and solving for $\hat{\beta}_2^{\text{can}}$ gives the desired expression in (C5).

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

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

The first matrix on the right hand side is non-singular, given the assumptions made. One then just needs to show that the second matrix is positive semi-definitive, which follows from the proof of Theorem B1.

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 either $\hat{\kappa} = 0$ 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. All the other parameters are identified separately 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 (C5). $\square$

References

- ALMEIDA, C., AND R. GARCIA (2012): “Assessing Misspecified Asset Pricing Models With Empirical Likelihood Estimators,” *Journal of Econometrics*, 170, 519–537.
- ANATOLYEV, S., AND A. MIKUSHEVA (2021): “Factor models with many assets: Strong factors, weak factors, and the two-pass procedure,” *Journal of Econometrics*.
- 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 K. LI (2012): “Statistical analysis of factor models of high dimension,” *Annals of Statistics*, 40(1), 436–465.
- BAI, J., AND S. NG (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70, 191–221.
- 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.
- 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*.
- CHAIIEB, 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. ROTHSCILD (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. (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., L. MOTA, S. ROTTKE, AND T. SANTOS (2020): “The cross-section of risk and returns,” *Review of Financial Studies*, 33(5), 1927–1979.
- 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.
- DOUGLAS, G. W. (1969): “Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency,” in *Yale Economic Essays, Vol. 9*, vol. Vol. 9, pp. 3–45. Yale University.
- 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.
- 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 (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.

- GOURIEROUX, C., AND A. MONFORT (2007): “Econometric specification of stochastic discount factor models,” *Journal of Econometrics*, 136(2), 509–530.
- GOYAL, A., AND P. SANTA-CLARA (2003): “Idiosyncratic risk matters!,” *Journal of Finance*, 58(3), 975–1007.
- 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., B. KELLY, H. LUSTIG, AND S. VAN NIEUWERBURGH (2016): “The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications,” *Journal of Financial Economics*, 119(2), 249–283.
- 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., 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.
- KAN, R., AND C. ZHANG (1999): “Two-pass tests of asset pricing models with useless factors,” *Journal of Finance*, 54(1), 203–235.
- KLEIBERGEN, F. (2009): “Tests of risk premia in linear factor models,” *Journal of Econometrics*, 149(2), 149–173.
- 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.
- LEHMANN, B. N. (1990): “Residual Risk Revisited,” *Journal of Econometrics*, 45(1-2), 71–97.
- 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.
- LEVY, H. (1978): “Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio,” *American Economic Review*, 68(4), 643–658.
- LI, D. (2011): “Financial Constraints, R& D Investment, and Stock Returns,” *Review of Financial Studies*, 24(9), 2974–3007.
- 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.
- MALKIEL, B. G., AND Y. XU (2006): “Idiosyncratic Risk and Security Returns,” Working paper, Yale University.
- MCCRACKEN, M. W., AND S. NG (2015): “FRED-MD: A Monthly Database for Macroeconomic Research,” Working Paper.
- MEHRA, R., S. WAHAL, AND D. XIE (2021): “Is Idiosyncratic Risk Conditionally Priced?,” *Quantitative Economics*, 12(2), 625–646.
- 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.
- RAPONI, V., R. UPPAL, AND P. ZAFFARONI (2021): “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.
- 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.