Correcting Misspecified Stochastic Discount Factors*

Raman Uppal Paolo Zaffaroni Irina Zviadadze

December 16, 2019

Abstract

We show how to correct a misspecified stochastic discount factor (SDF) to obtain an admissible SDF, namely an SDF that prices a given set of assets correctly. We construct the admissible SDF in the context of the traditional Arbitrage Pricing Theory (APT), which we extend to capture misspecification from pervasive (systematic) risk in addition to idiosyncratic risk. If the number of assets is large, the admissible SDF recovers fully (1) the contribution of the missing pervasive factors without requiring one to identify the missing factors and (2) the effect of idiosyncratic risk, which can play a significant role in asset pricing. Our approach applies both to reduced-form and equilibrium models, either linear or nonlinear. In an empirical application, we use our methodology to correct the SDF based on the consumption CAPM model and find that Size and Profitability are the most significant missing risk factors, whereas Value and Intermediary Capital are not significant.

JEL classification: C52, C58, G11, G12.
Keywords: Asset pricing, factor models, representative-agent models, model misspecification.

*Uppal is affiliated with EDHEC Business School and CEPR; email: Raman.Uppal@edhec.edu. Zaffaroni is affiliated with Imperial College Business School; email: P.Zaffaroni@imperial.ac.uk. Irina Zviadadze is affiliated with the Stockholm School of Economics, HEC Paris, and CEPR; email: Irina.Zviadadze@hhs.se. We gratefully acknowledge comments from Torben Andersen, Harjoat Bhamra, Svetlana Bryzgalova, Ines Chaieb, Magnus Dahlquist, Victor DeMiguel, René Garcia, Lars Hansen, Ravi Jagannathan, Christian Julliard, Robert Korajczyk, Stefan Nagel, Valentina Raponi, Cesare Robotti, Simon Rottke, Olivier Scaillet, Alberto Teguia, Viktor Todorov, Fabio Trojani, Dacheng Xiu, and seminar participants at Imperial College Business School, Kellogg School of Management, Stockholm School of Economics, University of Chicago, University of Geneva, and the 6th SAFE Asset Pricing Workshop, 12th Annual SoFiE Conference, 16th Annual Conference in Financial Economics Research at Arison School of Business, European Finance Association Meetings, Financial Econometrics Conference at Toulouse School of Economics, and UBC Finance Summer Conference

1 Introduction

Hansen and Jagannathan (1991) provide the bound that any admissible stochastic discount factor (SDF) must satisfy and Hansen and Jagannathan (1997) characterize the distance between any given SDF and the set of admissible SDFs and provide the linear correction to make any misspecified SDF admissible. There is a large literature that builds on these papers: some of these papers show how to sharpen the bounds while others generalize the correction; see, for example, Ghosh, Julliard, and Taylor (2017) and Orlowski, Sali, and Trojani (2016). Our objective in this paper is to develop a new methodology to construct an admissible SDF from a given candidate SDF that depends on K *observed* risk factors, as commonly used in financial economics.

There are a number of challenges in going from the candidate factor SDF, called the beta SDF, to an SDF that prices the N assets correctly, especially when N is large. The first is the presence of model misspecification. For instance, the K risk factors may not span the entire space of factors that are priced. Other sources of misspecification include mismeasured factors, idiosyncratic pricing errors such as sentiment, and omitted nonlinear terms that should have been included according to an equilibrium asset-pricing model. In order to account for these sources of misspecification, we extend the traditional APT so that it captures not just small idiosyncratic pricing errors but also pricing errors that are large and pervasive, while still satisfying the no-arbitrage condition. Then, we use the extended APT to show how the beta SDF can be corrected to obtain an admissible SDF; we call the correction term the alpha SDF. The theory we provide ensures that the corrected admissible SDF is nonnegative.

The next challenge is that the alpha SDF is a function of idiosyncratic risk, which is not directly observable. We show how to construct a version of the corrected SDF using only observable quantities based on the notion of linear projections; we label this the *projection* SDF. To understand the economic forces driving the alpha SDF, we demonstrate that for large N, the alpha SDF recovers *fully* the contribution of the missing factors, mismeasured factors, and nonlinearities *without* requiring one to identify the missing factors. Indeed, the regression R^2 from projecting the alpha SDF on the space spanned by a set of candidate

missing factors converges to one as N increases, if the candidate factors span the true set of missing factors. This result holds regardless of the length of the time series, T.

The final challenge is to estimate the corrected SDF. The general nonparametric formulation of the admissible SDF by Hansen and Jagannathan (1991, eq. (5)) involves estimating the covariance matrix of returns, which depends on $N(N+1)/2$ quantities. The extended APT addresses precisely this problem. It captures the rich cross-sectional structure of asset payoffs, while ensuring that the second-moment payoff matrix is nonsingular even when $N > T$. Moreover, the extended APT accomplishes this with only a small number of parameters, *of the order of* N, once one imposes a suitable sparsity structure on the covariance matrix of return idiosyncratic innovations. This sparsity structure is not overly restrictive exactly because the cross-sectional dependence of payoffs is captured by the observed and latent factors accommodated by the extended APT.

We evaluate our methodology using simulations based on the Fama-French five factor model. The simulations demonstrate that our methodology is remarkably effective in correcting misspecification arising either from missing factors or idiosyncratic risk. In an empirical application, we use our methodology to correct the beta SDF based that is based on the consumption-CAPM. We then use the correction term to evaluate the contribution of factors that are missing in the consumption CAPM but are prominent in the empirical asset-pricing literature. We find that the Size and Profitability are missing risk factors, whereas Value and Intermediary Capital are not.

The rest of the paper is arranged as follows. We discuss the related literature in Section 2. In Section 3, we describe our modeling assumptions and the extended APT. In Section 4, we characterize the SDF implied by the extended APT and study its behavior as the number of assets increases. In Section 5, we provide two representations of the SDF—the first in terms of returns and the second in terms of a one-factor beta model. In Section 6, we describe how our approach applies also to nonlinear SDFs from equilibrium asset pricing models, such as Breeden (1979), Campbell and Cochrane (1999), and Bansal and Yaron (2004). In Section 7, we explain in greater detail how our approach compares to that in Hansen and Jagannathan (1997), Ghosh, Julliard, and Taylor (2017), and Kozak, Nagel,

and Santosh (2018). We illustrate our theoretical results using a simulation experiment in Section 9 and using empirical data in Section 10. We conclude in Section 11.

Technical results are collected in a series of appendices. Appendix A contains the proofs for all the theorems. A decomposition of the SDF return in terms of the returns on "alpha" and "beta" portfolios is given in Appendix B. Different forms of misspecification that our framework can capture are described in Appendix C. Basic notions about the SDF from the existing literature are described in Appendix D. Details of how to estimate the extended APT are given in Appendix E and our approach for computing p-values is described in Appendix F.

2 Related Literature

Arrow (1964) introduces the notion of state prices that paved the way for the concept of the SDF.[1] Ross (1978) shows that the absence of arbitrage opportunities implies the existence of a strictly positive SDF when there are only a finite number of states of the world; for a survey of work in this area, including the extension to the case where the number of states is infinite, see Delbaen and Schachermayer (2006). Chamberlain and Rothschild (1983) show that the law of one price, a concept weaker than no arbitrage, implies the existence of an SDF that is not necessarily positive when asset payoffs have finite variances. The existence of an SDF for an infinite sequence of assets, as considered in our work, does not follow directly from the law of one price but requires a form of asymptotic no arbitrage. We complement these papers by providing a closed-form expression for the SDF. We allow for correlation between latent and observed factors, which is not a problem when all factors are unobserved, as is assumed in these papers.

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. Snow (1991) shows how to bound higher moments of the pricing kernel. Stutzer (1995) shows how, using the

[1] For a comprehensive treatment of the SDF in the absence of misspecification, see the excellent textbooks by Cochrane (2005) and Back (2017).

Kullback-Leibler Information Criterion, one can construct an entropy bound for the risk-neutral probability measure that naturally imposes the nonnegativity constraint on the SDF. Ghosh, Julliard, and Taylor (2017) build on this approach to derive bounds on the SDF (and, as explained below, also show how to correct the SDF for misspecification). Bansal and Lehman (1997) and Alvarez and Jermann (2005) derive restrictions on the entropy bound to decompose the SDF into transitory and permanent components.

Instead of the least-square projections in Hansen and Jagannathan (1997), Almeida and Garcia (2012) consider minimum-discrepancy projections that take into account higher moments of asset returns. Backus, Chernov, and Zin (2014) exploit the term structure of entropy to measure the pricing ability of the SDF implied by models with recursive utility and habit. Liu (2015) generalizes the basic entropy bounds of Stutzer (1995) and Backus, Chernov, and Zin (2014) by developing bounds on the SDF based on a generalized entropy function and uses these bounds to estimate the distribution of rare events. Almeida and Garcia (2017) derive SDF bounds that generalize the variance (Hansen and Jagannathan, 1991), entropy (Stutzer, 1995; Bansal and Lehman, 1997; Backus, Chernov, and Zin, 2014), and higher-moment bounds (Snow, 1991) that allow one to distinguish models where dispersion comes mainly from skewness from those where it comes from kurtosis. Orlowski, Sali, and Trojani (2016) extend and unify the literature on bounds on SDFs by showing how variance, entropy, and Hellinger bounds can be obtained from the same minimization problem; an application of this theory to puzzles in international finance is presented in Sandulescu, Trojani, and Vedolin (2017). 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. As a by-product of our analysis, the corrected SDF, expressed as a projection on the set of payoffs, satisfies the variance bound exactly.

In contrast to the large literature on bounds on the SDF, there is less work in the area of correcting the SDF for misspecification error because of the intrinsic difficulty in finding a satisfactory solution. Hansen and Jagannathan (1997) explicitly recognize that, when using SDFs, there is the possibility of pricing errors, which may arise either because the model used is an approximation to the true model or because there is an error in measuring the relevant factors. They address the question: "*How large is the misspecification of the stochastic discount factor proxy* [their emphasis]?" In doing so, they 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 nonparametric correction using a Kullback-Leibler entropy-minimization approach. These papers provide a non-parametric specification of the correction. Although this ensures admissibility of the corrected SDF, it is challenging to estimate the correction accurately when the number of assets to be priced is large relative to the number of observations.[2] Our work contributes to this stream of the literature by using the extended APT to identify the required correction for the SDF that, on the one hand is sufficiently flexible to mitigate various sources of misspecification and, on the other hand, can be estimated accurately even when the number of assets is large; in fact, having a large number of assets allows us to reconstruct precisely the correction required to account for model misspecification because the larger the number of assets the better one can span the systematic variation in returns. Moreover, our approach allows us to tease out the source of misspecification and to understand why is the correction leading to an admissible SDF: we can distinguish between idiosyncratic and systematic pricing errors.

In contrast to the nonparametric approach adopted in these papers, Kozak, Nagel, and Santosh (2018) assume that the SDF is spanned by a set of pervasive latent factors, which they estimate using principal components. By assuming that all the factors are latent, they successfully mitigate the risk of misspecification from omitted pervasive factors, assuming that the correct number of factors is identified. Just like them, we allow for the possibility of latent pervasive factors; however, we allow also for the possibility that the SDF depends on a set of observed pervasive factors. More importantly, our approach allows for the possibility of sentiment or firm-specific characteristics to influence the SDF. Feng, Giglio, and Xiu (2019) provide one method for identifying the set of observed factors to include in the specification of the beta SDF, while latent factors are reflected in the alpha SDF.

There are a number of papers that show how one can exploit the duality between mean-variance optimal portfolios and SDFs. Chamberlain and Rothschild (1983) show that the mean-variance frontier is spanned by the projection of the SDF on the space of payoffs

[2]This is precisely the same problem that is encountered when estimating the portfolio weights in a mean-variance setting using the sample means and covariances of the individual asset returns; see, for instance, DeMiguel, Garlappi, and Uppal (2009).

("projection-SDF") together with the mean functional. This result has been extended by Hansen and Richard (1987), who also show that the projection-SDF has minimum second moment. This implies that the projection-SDF in terms of returns belongs to the lower (inefficient) branch of the Markowitz mean-variance frontier. We show that both the candidate (beta) SDF and the correction (alpha) SDF can be expressed as inefficient returns *not* belonging to the SDF frontier; however, the sum of the two is on the SDF frontier.[3]

Our analysis of the SDF in the presence of model misspecification is founded on the classical APT, which allows for idiosyncratic pricing errors. The classical APT of Ross (1976) is formalized by Chamberlain (1983), Chamberlain and Rothschild (1983), Huberman (1982), and Ingersoll (1984). Just as in Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals to be diagonal; that is, we allow for correlated error terms. All these models deal with a large but countable number of assets. Building on the work of Al-Najjar (1998), Gagliardini, Ossola, and Scaillet (2016) extend the APT to allow for an uncountable number of assets and also relax the boundedness assumption of the maximum eigenvalue of the residual covariance matrix. In particular, Gagliardini, Ossola, and Scaillet (2016) show that the APT bound-inequality leads to zero pricing error for each asset when there is a continuum of assets. This is partly a consequence of the fact that they consider the *unweighted* sum of the squared pricing errors, as in the traditional APT setting. In the same setting with a continuum of assets, Renault, van der Heijden, and Werker (2017) extend the APT to squared excess returns allowing one to price common factors in the idiosyncratic variance of returns. We extend the classical APT and show that it can allow not just for idiosyncratic pricing errors but also for pervasive (systematic) pricing errors.

3 The Extended APT: A No-Arbitrage Model with Misspecification

In contrast to the traditional APT, which allows only for idiosyncratic pricing errors, in this section we extend the APT so that it can capture also systematic pricing errors that would

[3]Moreover, we show that these returns can be expressed as the returns of the alpha and beta portfolios, which are the two inefficient portfolios that span the entire Markowitz efficient frontier.

arise if there are missing pervasive factors in the return-generating model.[4] Then, in the next section, we demonstrate how the extended APT model can be used to construct the term that is required to make a misspecified SDF admissible; that is, the SDF that prices all assets correctly.

3.1 The Traditional APT

We now list the assumptions on which the APT is founded. We then describe how to relax these assumptions in order to obtain the extended APT. We study a market with an infinite number of assets.[5] Let the N-dimensional vector $\mathbf{R}_{N,t+1} = (R_{1,t+1}, R_{2,t+1}, \ldots, R_{N,t+1})'$ denote the vector of gross returns on the N risky assets with $\boldsymbol{\mu}_{N,t}$ its conditional mean at date t.

Let R_{ft} be the *gross* return on the risk-free asset. If a risk-free asset does not exist, then one needs to extend the set of payoffs with the unit payoff and use in place of the risk-free rate one of the following three returns: (1) the return on the minimum-variance portfolio; (2) the return on the zero-beta portfolio; (3) the constant-mimicking portfolio return.

The classical APT requires two main assumptions. The first assumption is of a linear factor structure for the returns $\mathbf{R}_{N,t+1}$. Let $\mathbf{f}_{t+1}$ be the $K \times 1$ *entire* vector of *observed* risk factors; $\mathbf{B}_{N,t} = (\boldsymbol{\beta}_{1,t}, \boldsymbol{\beta}_{2,t}, \ldots, \boldsymbol{\beta}_{N,t})'$ denotes an $N \times K$ full-rank matrix of factor loadings with ith row $\boldsymbol{\beta}_{i,t}'$; and, $\boldsymbol{\varepsilon}_{N,t+1} = (\varepsilon_{1,t+1}, \varepsilon_{2,t+1}, \ldots, \varepsilon_{N,t+1})'$ denotes an $N \times 1$ vector of idiosyncratic residuals. Our definition of the classical APT may appear to be different from the usual definition, which does not assume knowledge of the observed risk factors. However, the difference is only illusory because $\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})$ is still *unobserved* given that $\mathbb{E}_t(\mathbf{f}_{t+1})$ is, in general, unknown regardless of whether the $\mathbf{f}_t$ are observed or not. We prefer to formulate the classical APT in this way because we will formulate the *extended* APT below assuming that the K factors we observe are *not* the full set of risk factors.

[4] A statistical criterion to assess whether the error terms in a given model share at least one common (pervasive) factor is provided by Gagliardini, Ossola, and Scaillet (2017).

[5] Instead of considering a sequence of distinct economies, we consider a *fixed* infinite economy in which we study a sequence of nested subsets of assets. Therefore, in the Nth step, as a new asset is added to the first $N-1$ assets, the parameters of the first $N-1$ stay unchanged. These unchanging parameters can be interpreted as the parameters one would get in the limit as the number of assets becomes asymptotically large.

Assumption 3.1 (Linear factor model)**.** *We assume the N-dimensional vector $\mathbf{R}_{N,t}$ of gross asset returns can be characterized by:*

$$\mathbf{R}_{N,t+1} = \boldsymbol{\mu}_{N,t} + \mathbf{B}_{N,t}\big(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})\big) + \boldsymbol{\varepsilon}_{N,t+1},$$

where, at any time $t+1$, $\mathbf{f}_{t+1}$ has the $K \times K$ conditional covariance matrix $\boldsymbol{\Omega}_t$ and $\boldsymbol{\varepsilon}_{N,t+1}$ is distributed with zero conditional mean and $N \times N$ conditional covariance matrix $\boldsymbol{\Sigma}_{N,t}$, with $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Sigma}_{N,t}$ being positive definite. Moreover, $\boldsymbol{\varepsilon}_{N,t+1}$ and $\mathbf{f}_{t+1}$ are conditionally uncorrelated; that is, $\mathbb{E}_t(\boldsymbol{\varepsilon}_{N,t+1}\mathbf{f}'_{t+1}) = \mathbf{0}_{N\times K}$. Finally, assume that $K < N$.

The second assumption rules out arbitrage when the number of assets is large. Given an arbitrary portfolio strategy a with weights $\mathbf{w}^a_{N,t} = (w^a_{1,t}, w^a_{2,t}, \ldots, w^a_{N,t})'$ of N risky assets, and using $\mathbf{1}_N$ to denote an N-dimensional vector of ones, we define the associated portfolio gross return as $R^a_{t+1} = \mathbf{R}'_{N,t+1}\mathbf{w}^a_{N,t} + R_{ft}(1 - \mathbf{1}'_N\mathbf{w}^a_{N,t})$.

Assumption 3.2 (No arbitrage)**.** *There are no arbitrage opportunities for a sufficiently large number of assets; that is, there is no sequence of portfolios along some subsequence N' for which:*

$$\text{var}_t(\mathbf{R}'_{N,t+1}\mathbf{w}^a_{N',t}) \to 0 \quad \textit{as} \quad N' \to \infty \quad \textit{and} \quad (\boldsymbol{\mu}_{N',t} - R_{ft}\mathbf{1}_N)'\mathbf{w}^a_{N',t} \geq \delta > 0 \textit{ for all } N',$$

where δ denotes an arbitrary positive scalar.[6]

Note that the above definition of no arbitrage for a large number of assets does not rule out non-negative SDFs (see Chamberlain and Rothschild (1983) and Back (2017)), in contrast to the definition of no arbitrage in Hansen and Richard (1987).

Under Assumptions 3.1 and 3.2, the expected *excess* return can be written as:

$$\mathbb{E}_t(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N) = \boldsymbol{\mu}_{N,t} - R_{ft}\mathbf{1}_N = \boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t,$$

where the vector of pricing errors is $\boldsymbol{\alpha}_{N,t} = (\boldsymbol{\mu}_{N,t} - R_{ft}\mathbf{1}_N) - \mathbf{B}_{N,t}\boldsymbol{\lambda}_t$, and the vector of risk premia, $\boldsymbol{\lambda}_t$, is the limit of $\left(\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t}\right)^{-1}\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}(\boldsymbol{\mu}_{N,t} - R_{ft}\mathbf{1}_N)$ as $N \to \infty$; Ingersoll (1984) derives the precise condition for this limit to exist: Let $g_{iM}(\mathbf{A})$ denote the ith eigenvalue of a symmetric matrix $\mathbf{A}$ in decreasing order for $1 \leq i \leq M$. Then, if

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

$g_{1K}((\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t})^{-1}) \to 0$ as $N \to \infty$ and Assumptions 3.1 and 3.2 hold, then Ingersoll (1984, Theorem 3 and ftn. 10) shows that $\boldsymbol{\lambda}_t$ is unique and prices assets with bounded squared error. Obviously, if the observed factors are traded portfolio (gross) returns, then $\boldsymbol{\lambda}_t = \mathbb{E}_t(\mathbf{f}_{t+1}) - R_{ft}\mathbf{1}_K$.

More importantly, Ross (1976), Huberman (1982), Chamberlain and Rothschild (1983), and Ingersoll (1984) show that if $\boldsymbol{\Sigma}_{N,t}$ has bounded eigenvalues for large N, then the APT implies that the unique $\boldsymbol{\lambda}_t$ prices assets and the resulting pricing errors, $\boldsymbol{\alpha}_{N,t}$, satisfy the following bound:

$$\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t} \leq \delta_{\text{apt}} < \infty, \tag{1}$$

where δ_{apt} is some arbitrary positive scalar.

Several comments apply to our specification of the data-generating process for asset returns. First, the above arguments, such as the existence of the risk premia corresponding to the observed factors $\mathbf{f}_{t+1}$, require the associated loadings matrix $\mathbf{B}_{N,t}$ to be of full rank for any N and t. In turn, this means that we are not allowing any of the observed factors, to be spurious or almost-spurious, in the sense of Jagannathan and Wang (1998), Kan and Zhang (1999) and Kleibergen (2009).[7] Therefore, as customary in empirical asset pricing, for instance when estimating risk premia, one needs to use the various tests for spurious factors before implementing our method empirically. Second, all our theoretical results hold at every instant t; hence our specification of the APT as a *conditional* asset pricing model. When taking our model to the data, if time-variation is desired one needs to parametrize this; we explain in Section 8 how this can be done using state variables.

3.2 The Extended APT

With respect to the restriction in (1), there are two possible cases for $\boldsymbol{\Sigma}_{N,t}$ as $N \to \infty$. The existing APT literature has focused on studying the first case in which $\mathbf{f}_{t+1}$ indeed includes the entire set of risk factors, and thus, the pricing errors are idiosyncratic (implying that all the eigenvalues of $\boldsymbol{\Sigma}_{N,t}$ are bounded a.s.). We now extend the APT model in the existing literature to the second case in which $\mathbf{f}_{t+1}$ does *not* include *all* the risk factors, and therefore,

[7] More generally, we are not allowing any column of $\mathbf{B}_{N,t}$ to be a linear combination of the other columns.

the pricing errors are related to pervasive factors; that is, at least one of the eigenvalues of $\mathbf{\Sigma}_{N,t}$ is *unbounded*.

Theorem 3.1 (No-arbitrage constraint on $\boldsymbol{\alpha}_N$ with large pricing errors)**.** *Suppose that the vector of asset returns, $\mathbf{R}_{N,t+1}$, satisfies Assumptions 3.1 and 3.2. Suppose that for some finite $1 \leq p < N$ the following three conditions hold: (i) $\sup_N g_{pN}(\mathbf{\Sigma}_{N,t}) = \infty$; (ii) $\sup_N g_{p+1\,N}(\mathbf{\Sigma}_{N,t}) \leq \delta < \infty$; and, (iii) $\inf_N g_{NN}(\mathbf{\Sigma}_{N,t}) \geq \delta > 0$. Then, the APT restriction in* (1) *is satisfied by the pricing error $\boldsymbol{\alpha}_N$, represented as*

$$\boldsymbol{\alpha}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} + \mathbf{a}_{N,t}, \tag{2}$$

and the idiosyncratic covariance matrix given by:

$$\mathbf{\Sigma}_{N,t} = \mathbf{A}_{N,t}\mathbf{A}'_{N,t} + \mathbf{C}_{N,t}, \tag{3}$$

where $\mathbf{C}_{N,t}$ is a symmetric matrix with bounded eigenvalues, $\mathbf{A}_{N,t}$ is an $N \times p$ matrix whose jth column equals $g_{jN}^{\frac{1}{2}}(\mathbf{\Sigma}_{N,t})\,\mathbf{v}_{jN}(\mathbf{\Sigma}_{N,t})$, where $1 \leq j \leq p$, $\mathbf{v}_{jN}(\mathbf{\Sigma}_{N,t})$ is the eigenvector of $\mathbf{\Sigma}_{N,t}$ associated with the eigenvalue $g_{jN}(\mathbf{\Sigma}_{N,t})$, $\boldsymbol{\lambda}_{miss,t}$ is some $p \times 1$ vector, and $\mathbf{a}_{N,t}$ is some non-zero $N \times 1$ vector that satisfies $\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{a}_{N,t} \leq \delta < \infty$.

Theorem 3.1 shows that the common perception that the pricing error $\boldsymbol{\alpha}_N$ needs to be small in the APT is not accurate. In fact, it states that if the pricing errors are large so that the maximum eigenvalue of $\mathbf{\Sigma}_N$ is asymptotically unbounded, then the contribution of the pricing error to the portfolio return could also be large, but for this to satisfy the no-arbitrage condition, any portfolio earning this high return would *not* be well diversified and would be bearing idiosyncratic risk.[8]

Recall that, in general, latent factors can be determined only up to a rotation. Our formulation of $\boldsymbol{\alpha}_{N,t}$ and $\mathbf{\Sigma}_{N,t}$ in Theorem 3.1 implies a specific rotation where, in particular, we follow Chamberlain (1983), and assume that

$$cov_t(\mathbf{f}_{miss,t+1}) = \mathbf{I}_p. \tag{4}$$

[8]For additional details of the relation between well-diversified portfolios and deviation from exact pricing, see Chamberlain (1983, Corollary 1).

However, other rotations might be considered although this one provides a nice interpretation of the risk premia $\boldsymbol{\lambda}_{miss,t}$ as (multivariate) Sharpe ratios.[9]

Observe that the first term in (2), $\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}$, is associated with p latent or missing pervasive factors, indicated by $\mathbf{f}_{miss,t}$, in which $\mathbf{A}_{N,t}$ are the factor loadings and $\boldsymbol{\lambda}_{miss,t}$ are the risk premia for these missing factors. In particular, pervasiveness of the latent missing factors is implied by $\mathbf{A}'_{N,t}\mathbf{A}_{N,t} \to \infty$ as N increases; see Connor, Goldberg, and Korajczyk (2010) for further discussion. The second term in (2), $\mathbf{a}_{N,t}$, is the asset-specific part of the pricing error $\boldsymbol{\alpha}_{N,t}$; for instance, $\mathbf{a}_{N,t}$ could be interpreted as representing managerial skills or views of analysts about specific firms which, in contrast to $\mathbf{A}_{N,t}$, must satisfy $\mathbf{a}'_{N,t}\mathbf{a}_{N,t} \leq \delta < \infty$ as N increases because of no arbitrage.

Our result in (3) implies that the *purely* idiosyncratic covariance matrix has bounded eigenvalues for any N. Although this can be interpreted as a form of sparsity, for practical implementation of our methodology one needs to parametrize $\mathbf{C}_{N,t}$, in particular, imposing that it is a function of a number of parameters of the order of $O(N)$. This implies an even stronger form of sparsity than the one expressed by the bounded-eigenvalue condition. For instance, diagonal or even spherical $\mathbf{C}_{N,t}$ represent special, important, cases of possible parameterizations, as explained in our estimation section below. However, note that thanks to the factor structure specified for the observed and latent common risk factors, $\mathbf{f}_{t+1}$ and $\mathbf{f}_{miss,t+1}$ respectively, assuming such sparsity is reasonable because it does not limit the ability of the extended APT to capture cross-sectional dependence in asset returns.

[9] To better understand the implications of such a rotation, consider the factor structure in the (standardized) latent factors:

$$\boldsymbol{\alpha}_{N,t} + \mathbf{A}_{N,t}(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})) + \boldsymbol{\eta}_{t+1},$$

where $\boldsymbol{\eta}_{t+1}$ has zero conditional mean, conditional covariance $\mathbf{C}_{N,t}$, and is conditionally uncorrelated with $\mathbf{f}_{miss,t+1}$. Then, such factor structure has mean $\boldsymbol{\alpha}_{N,t}$ and covariance $\boldsymbol{\Sigma}_{N,t}$, which for any symmetric non-singular matrix $\boldsymbol{\Omega}_{miss,t}$, can be expressed as:

$$\begin{aligned}
&\boldsymbol{\alpha}_{N,t} = \mathbf{a}_{N,t} + \mathbf{A}_{N,t}\boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}\boldsymbol{\Omega}_{miss,t}^{\frac{1}{2}}\boldsymbol{\lambda}_{miss,t} = \mathbf{a}_{N,t} + \mathbf{A}^{\dagger}_{N,t}\boldsymbol{\lambda}^{\dagger}_{miss,t}, \quad \text{and} \\
&\boldsymbol{\Sigma}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}\boldsymbol{\Omega}_{miss,t}\boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}\mathbf{A}'_{N,t} + \mathbf{C}_{N,t} = \mathbf{A}^{\dagger}_{N,t}\boldsymbol{\Omega}_{miss,t}\mathbf{A}^{\dagger\prime}_{N,t} + \mathbf{C}_{N,t},
\end{aligned}$$

where $\mathbf{A}^{\dagger}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}$. In other words, we are assuming that the latent factors in Theorem 3.1 are a rotation, by means of the matrix $\boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}$, of some *un-normalized* missing factors, say $\mathbf{f}^{\dagger}_{miss,t+1}$ with risk premia $\boldsymbol{\lambda}^{\dagger}_{miss,t} = \boldsymbol{\Omega}_{miss,t}^{\frac{1}{2}}\boldsymbol{\lambda}_{miss,t}$, covariance $\boldsymbol{\Omega}_{miss,t}$, and loadings $\mathbf{A}^{\dagger}_{N,t}$. Therefore, $\boldsymbol{\lambda}_{miss,t} = \boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}\boldsymbol{\lambda}^{\dagger}_{miss,t}$; that is the risk premia $\boldsymbol{\lambda}_{miss,t}$ have the interpretation of Sharpe ratios. The same identification assumption (4) is obtained when considering any rotation by means of $\mathbf{H}_t\boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}}$, for any *orthogonal* matrix $\mathbf{H}_t$. Therefore, each of the elements of $\mathbf{f}_{miss,t+1}$ is a given linear combination of *all* the elements of $\mathbf{f}^{\dagger}_{miss,t+1}$.

4 The SDF with Model Misspecification

In this section, we report our main results about the SDF in the presence of model misspecification.[10] We consider a situation where one wishes to price a given set of N assets using a candidate SDF that is linear in a set of K *observed* risk factors; the nonlinear case is treated in Section 6.

4.1 The SDF under the Extended APT

We now derive the closed-form expression for the SDF when the extended APT is used to capture model misspecification. These results complement Chamberlain (1983), who shows existence and continuity of the "cost functional" (i.e. the SDF) under the classical APT, without providing closed-form expressions.

The salient aspect of the extended APT is that the implied SDF depends on both common and idiosyncratic risk premia and risks (unless $\boldsymbol{\alpha}_{N,t} = \mathbf{0}_N$). In turn, this reflects the possibility that *pure* idiosyncratic risk, arising when the elements of $\boldsymbol{\varepsilon}_{N,t+1}$ are only mildly cross-sectionally dependent, affects asset prices. This occurs when the idiosyncratic component of expected returns, $\mathbf{a}_{N,t}$ are not all zero. For this case, Chamberlain (1983, Section 3) shows that the properties of the mean-variance frontier are still valid, although the frontier will not be well-diversified. Raponi, Uppal, and Zaffaroni (2019) formalize fund separation, namely that the mean-variance frontier is now spanned by two *inefficient* portfolios, with special properties: one portfolio contains only common risk and the other only idiosyncratic risk. However, the idiosyncratic component $\boldsymbol{\varepsilon}_{N,t+1}$ can also mask *unspanned* common risk, due to omitted pervasive factors. In this case, the elements of $\boldsymbol{\varepsilon}_{N,t+1}$ will be strongly cross-sectionally correlated because the missing pervasive factors induce a factor structure. Although we start our analysis with the simplifying assumption of conditional independence between the common observed factors $\mathbf{f}_{t+1}$ and the idiosyncratic shock $\boldsymbol{\varepsilon}_{N,t+1}$, the possibility that $\boldsymbol{\varepsilon}_{N,t+1}$ contains missing pervasive factors requires us to generalize to the case when the two components are correlated.

[10] For background information, we also described existing results about the SDF in the *absence* of model misspecification in Appendix D, where we follow Hansen and Richard (1987) and often refer to Chamberlain and Rothschild (1983); for textbook treatment, see Cochrane (2005) and Back (2017).

Theorem 4.1 (SDF in closed form and its linear projection)**.** *Under Assumptions 3.1 and 3.2 of the APT, for a given $\mu_{m,t}$, there exists an admissible SDF of the form*

$$m_{t+1} = \mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1},$$

with

$$\begin{aligned} \mathbf{b}_t &= -\mu_{m,t}\,\boldsymbol{\Omega}_t^{-1}\,\boldsymbol{\lambda}_t, \\ \mathbf{c}_{N,t} &= -\mu_{m,t}\,\boldsymbol{\Sigma}_{N,t}^{-1}\,\boldsymbol{\alpha}_{N,t}. \end{aligned} \tag{5}$$

and in terms of a linear projection on the set of payoffs $(1, \mathbf{R}_{N,t+1}^e)$, the projection SDF is

$$\begin{aligned} m_{t+1}^* &= \text{proj}(m_{t+1}|(1, \mathbf{R}_{N,t+1}^e)) \\ &= \mu_{m,t} + (\mathbf{b}_t'\boldsymbol{\Omega}_t\mathbf{B}_{N,t}' + \mathbf{c}_{N,t}'\boldsymbol{\Sigma}_{N,t})\mathbf{V}_{N,t}^{-1}\big(\mathbf{R}_{N,t+1}^e - \mathbb{E}_t(\mathbf{R}_{N,t+1}^e)\big), \end{aligned}$$

where $\mathbf{V}_{N,t}$ is the (conditional) covariance matrix of excess returns.

Remark 4.1.1. Regarding m_{t+1}^*, note that $\mathbb{E}_t(\mathbf{R}_{N,t+1}^e) = \boldsymbol{\mu}_{N,t} - R_{ft}\mathbf{1}_N = \boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t$, implying that one *always* recovers the true expected excess return by combining the alpha and beta components of returns. In other words, the extended APT mitigates completely model misspecification.

Remark 4.1.2. Notice that, although they have the same pricing implications, $\mathbb{E}(m_{t+1}^*)^2 \leq \mathbb{E}(m_{t+1})^2$ because m_{t+1}^* is the unique minimum-variance SDF for a given mean $\mu_{m,t}$.

Remark 4.1.3. Chamberlain (1983, Cor. 1 (i)) shows that exact pricing (i.e. when expected excess returns are an exact linear combination of the betas) holds if and only if the SDF (i.e. the cost functional) is well diversified. This is confirmed by our result in (5), from which we see that $\mathbf{c}_{N,t} = \mathbf{0}_N$ if and only if the pricing errors $\boldsymbol{\alpha}_{N,t} = \mathbf{0}_N$. This implies that the SDF m_{t+1} is not a function of idiosyncratic risk $\boldsymbol{\varepsilon}_{N,t+1}$ if and only if exact pricing holds.[11]

Remark 4.1.4. The result in the theorem above holds for both traded and nontraded observed factors, $\mathbf{f}_t$. In both cases, these factors can be constructed as the limit of portfolios that have zero idiosyncratic risk. Indeed, for the case of nontraded factors, one just replaces them with the corresponding mimicking portfolios that are asymptotically valid when N is large, which is exactly the setting for which the APT is designed.

[11] The term $\mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}$ has zero mean and (conditional) variance $\mathbf{c}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\mathbf{c}_{N,t}$. The latter is zero, implying $\mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}$ is zero a.s., if and only if $\mathbf{c}_{N,t} = \mathbf{0}_N$, which by non-singularity of $\boldsymbol{\Sigma}_{N,t}$, is equivalent to $\boldsymbol{\alpha}_{N,t} = \mathbf{0}_N$.

We now show how the SDF under the APT is related to the SDF under a traditional factor model, as discussed above.

Theorem 4.2 (Decomposition of SDF and its linear projection)**.** *Under Assumptions 3.1 and 3.2, for any $\mu_{m,t}$, the admissible SDF can be decomposed as*

$$m_{t+1} = m^{\alpha}_{t+1} + m^{\beta}_{t+1},$$

where

$$\begin{aligned} m^{\beta}_{t+1} &= \mu_{m,t} - \mu_{m,t}(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}))'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t, \\ m^{\alpha}_{t+1} &= -\mu_{m,t}\boldsymbol{\varepsilon}'_{N,t+1}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}, \end{aligned} \tag{6}$$

such that $cov_t(m^{\alpha}_{t+1}, m^{\beta}_{t+1}) = 0$.

The projection SDF can also be decomposed in terms of linear projection on the set of payoffs $(1, \mathbf{R}^e_{N,t+1})$:

$$m^{*}_{t+1} = m^{\alpha *}_{t+1} + m^{\beta *}_{t+1},$$

where

$$\begin{aligned} m^{\beta *}_{t+1} &= \mu_{m,t} - \mu_{m,t}\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big), \quad \textit{and} \\ m^{\alpha *}_{t+1} &= -\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big). \end{aligned}$$

Remark 4.2.1. The implications of the above theorem for the pricing of asset returns are:

$$\begin{aligned} \mathbb{E}_t\left(m^{\beta}_{t+1}\begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix}\right) &= \mu_{m,t}\begin{bmatrix} 1 \\ \boldsymbol{\alpha}_{N,t} \end{bmatrix} \\ \mathbb{E}_t\left(m^{\alpha}_{t+1}\begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix}\right) &= \mu_{m,t}\begin{bmatrix} 0 \\ -\boldsymbol{\alpha}_{N,t} \end{bmatrix}, \end{aligned}$$

and, because $m_{t+1} = m^{\alpha}_{t+1} + m^{\beta}_{t+1}$, we have:

$$\mathbb{E}_t\left(m_{t+1}\begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix}\right) = \mu_{m,t}\begin{bmatrix} 1 \\ \mathbf{0}_N \end{bmatrix}.$$

However, notice that although m^{β}_{t+1} is misspecified for the excess returns $\mathbf{R}^e_{N,t+1}$, it prices the observed factors correctly, because $\mathbb{E}_t(m^{\beta}_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)) = \mathbf{0}_K$.

Remark 4.2.2. In terms of the first two moments of the decomposition of the SDF, we have that

$$\mathbb{E}_t(m^{\beta}_{t+1}) = \mu_{m,t}, \qquad \mathbb{E}_t\big((m^{\beta}_{t+1})^2\big) = \mu^2_{m,t}(1 + \boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t),$$
$$\mathbb{E}_t(m^{\alpha}_{t+1}) = 0, \qquad \mathbb{E}_t\big((m^{\alpha}_{t+1})^2\big) = \mu^2_{m,t}\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}.$$

Remark 4.2.3. The term m^{β}_{t+1} in (6) is exactly the "classic" factor SDF used in the existing asset-pricing literature.

Remark 4.2.4. The term m^{α}_{t+1} is *not* an admissible SDF; for instance, $\mathbb{E}_t(m^{\alpha}_{t+1}) = 0$ for any $\mu_{m,t}$.

4.2 Nonnegativity of the SDF

The SDF characterized in the previous section may not always be nonnegative. In this section, we show how one can identify the SDF implied by the extended APT so that the SDF is always nonnegative. There are at least two approaches for addressing this problem. The first approach is to specify that the SDF is an *exponential* function of the payoffs (see Ghosh, Julliard, and Taylor (2017) and Gourieroux and Monfort (2007)), which then leads to an SDF that is nonnegative by construction; this approach is illustrated below. The second approach, not pursued here, is that of Hansen and Jagannathan (1997, Eq. (24)), who show that it is convenient to express the SDF corrected for model misspecification as the payoff to an option, which is always nonnegative. Both approaches require an assumption regarding the distribution of payoffs.[12] For simplicity, we assume throughout that returns are conditionally normal but one can consider other distributions, allowing for asymmetry and fat tails.[13]

Theorem 4.3 (Nonnegative SDF in closed form)**.** *Under Assumptions 3.1 and 3.2 of the APT and assuming that returns are conditionally normally distributed, there exists an admissible SDF m^{+}_{t+1}, with the given mean $\mu_{m,t}$, of the form*

$$m^{+}_{t+1} = \exp\Big[\mu^{+}_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\Big],$$

[12] If the distribution of payoffs is absolutely continuous but time is discrete, then markets will be incomplete and the admissible SDF is not uniquely defined.

[13] Examples are the generalized-elliptical distribution, the generalized-error distribution (Box and Tiao, 1973), and the variance-gamma distribution (Madan and Seneta, 1990).

with

$$\mu^{+}_{m,t} = \ln \mu_{m,t} - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\,\boldsymbol{\lambda}_t - \frac{1}{2}\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\,\boldsymbol{\alpha}_{N,t},$$

$$\mathbf{b}_t = -\boldsymbol{\Omega}_t^{-1}\,\boldsymbol{\lambda}_t,$$

$$\mathbf{c}_{N,t} = -\boldsymbol{\Sigma}_{N,t}^{-1}\,\boldsymbol{\alpha}_{N,t}.$$

When the risk-free asset is available

$$\mu^{+}_{m,t} = -\ln R_{ft} - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\,\boldsymbol{\lambda}_t - \frac{1}{2}\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\,\boldsymbol{\alpha}_{N,t}.$$

Theorem 4.4 (Decomposition of the nonnegative SDF)**.** *Under Assumptions 3.1 and 3.2 of the APT and that returns are conditionally normally distributed, the admissible nonnegative SDF can be decomposed as*

$$m^{+}_{t+1} = m^{\alpha+}_{t+1}\, m^{\beta+}_{t+1},$$

where

$$m^{\beta+}_{t+1} = \mu_{m,t} \exp\Big[-\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t\Big],$$

$$m^{\alpha+}_{t+1} = \exp\Big[-\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{N,t+1} - \frac{1}{2}\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t}\Big],$$

such that $cov_t(m^{\alpha+}_{t+1}, m^{\beta+}_{t+1}) = 0$.

Remark 4.4.1. The implications of the above theorem for the pricing of asset returns are:

$$\mathbb{E}_t\left(m^{\beta+}_{t+1} \begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix} \right) = \mu_{m,t} \begin{bmatrix} 1 \\ \boldsymbol{\alpha}_{N,t} \end{bmatrix}$$

$$\mathbb{E}_t\left(m^{\alpha+}_{t+1} \begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix} \right) = \begin{bmatrix} 1 \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t \end{bmatrix}.$$

Given that $m^{+}_{t+1} = m^{\alpha+}_{t+1} m^{\beta+}_{t+1}$, we have:

$$\begin{aligned}
\mathbb{E}_t\left(m^{+}_{t+1} \begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix} \right) &= cov_t(m^{\alpha+}_{t+1} \begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix}, m^{\beta+}_{t+1}) + \mathbb{E}_t(m^{\alpha+}_{t+1} \begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1} \end{bmatrix})\mathbb{E}_t(m^{\beta+}_{t+1}) \\
&= cov_t(m^{\alpha+}_{t+1} \begin{bmatrix} 1 \\ \mathbf{B}_{N,t}(\mathbf{f}_t - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) \end{bmatrix}, m^{\beta+}_{t+1}) + \mu_{m,t} \begin{bmatrix} 1 \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t \end{bmatrix} \\
&= \mathbb{E}_t(m^{\alpha+}_{t+1}) cov_t(\begin{bmatrix} 1 \\ \mathbf{B}_{N,t}(\mathbf{f}_t - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) \end{bmatrix}, m^{\beta+}_{t+1}) + \mu_{m,t} \begin{bmatrix} 1 \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t \end{bmatrix}
\end{aligned}$$

$$= \mathbb{E}_t(m_{t+1}^{\alpha+})\mathbb{E}_t\Big(\begin{bmatrix} 0 \\ \mathbf{B}_{N,t}(\mathbf{f}_t - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) \end{bmatrix} m_{t+1}^{\beta+}\Big) + \mu_{m,t}\begin{bmatrix} 1 \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t \end{bmatrix}$$

$$= \mu_{m,t}\begin{bmatrix} 1 \\ \mathbf{0}_N \end{bmatrix},$$

because $\mathbb{E}_t(\mathbf{f}_t - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t)m_{t+1}^{\beta+}) = -\mu_{m,t}\boldsymbol{\lambda}_t$ and $\mathbb{E}_t(m_{t+1}^{\alpha+}) = 1$.

Notice that, just like in the case where the SDF is a linear function of the factors, $m_{t+1}^{\beta+}$ prices correctly $\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K$ (and the risk-free asset), because $\mathbb{E}_t(m_{t+1}^{\beta+}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)) = \mathbf{0}_K$, although it is misspecified for the returns $\mathbf{R}_{N,t+1}^e$, unless $\boldsymbol{\alpha}_{N,t} = \mathbf{0}_N$. Moreover, the pricing errors $\boldsymbol{\alpha}_{N,t}$ still have the interpretation of *prices* for the idiosyncratic shocks because $\mathbb{E}_t(m_{t+1}^{\alpha+}\boldsymbol{\varepsilon}_{N,t+1}) = -\boldsymbol{\alpha}_{N,t}$.

Remark 4.4.2. In terms of the first two moments of the decomposition of the SDF, we have that

$$\mathbb{E}_t(m_{t+1}^{\beta+}) = \mu_{m,t}, \qquad \mathbb{E}_t\big((m_{t+1}^{\beta+})^2\big) = \mu_{m,t}^2 e^{\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t},$$

$$\mathbb{E}_t(m_{t+1}^{\alpha+}) = 1, \qquad \mathbb{E}_t\big((m_{t+1}^{\alpha+})^2\big) = e^{\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t}}.$$

4.3 Projection of Components of Nonnegative SDF under Extended APT

Just like in the case where the SDF is a linear function of the factors, m_{t+1}^{+} is a function of the unobservable quantity $\boldsymbol{\varepsilon}_{N,t+1}$, and hence, cannot be implemented. Thus, we develop a projection version for it that is feasible.

Corollary 4.4.1 (Representation and decomposition of the nonnegative SDF in terms of *nonlinear* projection)**.** *The nonnegative SDF m_{t+1}^{+} can be represented in terms of the exponential function of the linear projections on the set of payoffs $(1, \mathbf{R}_{N,t+1}^e)$, which can be decomposed as:*

$$m_{t+1}^{*+} = m_{t+1}^{\alpha*+}\, m_{t+1}^{\beta*+},$$

where

$$m_{t+1}^{\beta*+} = \mu_{m,t}\exp\Big[-\boldsymbol{\lambda}_t'\mathbf{B}_{N,t}'\mathbf{V}_{N,t}^{-1}(\mathbf{R}_{N,t+1}^e - \mathbb{E}_t(\mathbf{R}_{N,t+1}^e) - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t\Big], \quad \textit{and}$$

$$m_{t+1}^{\alpha*+} = \exp\Big[-\boldsymbol{\alpha}_{N,t}'\mathbf{V}_{N,t}^{-1}\big(\mathbf{R}_{N,t+1}^e - \mathbb{E}_t(\mathbf{R}_{N,t+1}^e)\big) - \frac{1}{2}\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t}\Big].$$

Remark 4.4.3. It follows that:

$$\mathbb{E}_t(m_{t+1}^{*+}\,\mathbf{R}_{N,t+1}^e) = \mathbf{0}_N \text{ and } \mathbb{E}_t(m_{t+1}^{*+}) \to \mu_{m,t} \text{ as } N \to \infty.$$

Therefore, the *feasible* m_{t+1}^{*+} prices correctly the risky assets, and it prices correctly the unit payoff asymptotically, whilst maintaining nonnegativity.[14]

Remark 4.4.4. Note that the *linear* projection versions of both m_{t+1} and m_{t+1}^{+}, in view of their admissibility property, are identical and equal to m_{t+1}^{*}, which is given in Theorem 4.1. However, because m_{t+1}^{*} has the form of a portfolio return, there is no guarantee that it is nonnegative, unlike m_{t+1}^{*+}.

Remark 4.4.5. Close analogies exist between m_{t+1}^{*+} and the nonnegative nonparametric SDF of Ghosh, Julliard, and Taylor (2017); in particular, between our correction factor $m_{t+1}^{\alpha*+}$ and their equation (8), because the latter can be expressed as an exponential function of the payoffs $(1, \mathbf{R}_{N,t+1}^e)$, net of a constant. These analogies are discussed in greater detail in Section 7.

4.4 Characterizing Components of the SDF for Large N

In this section, we study the properties of the SDFs when the number of risky assets is large. In particular, we characterize the behavior of the components of the linear projection SDFs, $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\beta*}$, as the number of assets, N, increases to infinity. As an important by product, these results allow us to directly derive the behavior of the nonlinear projection SDFs, $m_{t+1}^{\alpha*+}$ and $m_{t+1}^{\beta*+}$, without additional arguments. These results are practically relevant but also economically important, because studying the behavior of the correction terms (the $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\alpha*+}$) for large N sheds light on the reasons why the alpha SDFs are able to successfully mitigate misspecification in the candidate beta SDFs.

Recalling the expression for the linear SDF given above,

$$m_{t+1} = \mu_{m,t} + \mathbf{b}_t'\big(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})\big) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1},$$

[14] One can construct also another, asymptotically equivalent, nonnegative SDF that prices the unit payoff correctly *for any* N by setting $m_{t+1}^{**+} = m_{t+1}^{\alpha**+} m_{t+1}^{\beta**+}$ with $m_{t+1}^{\beta**+} = \mu_{m,t} \exp\Big[- \boldsymbol{\lambda}_t' \mathbf{B}_{N,t}' \mathbf{V}_{N,t}^{-1} (\mathbf{R}_{N,t+1}^e - \mathbb{E}_t(\mathbf{R}_{N,t+1}^e)) - \frac{1}{2}\mathbb{E}_t'(\mathbf{R}_{N,t+1}^e)' \mathbf{V}_{N,t}^{-1} \mathbb{E}_t(\mathbf{R}_{N,t+1}^e)\Big]$, and $m_{t+1}^{\alpha**+} = \exp\Big[- \boldsymbol{\alpha}_{N,t}' \mathbf{V}_{N,t}^{-1} \big(\mathbf{R}_{N,t+1}^e - \mathbb{E}_t(\mathbf{R}_{N,t+1}^e)\big)\Big]$.

one needs to ensure that m_{t+1} is well defined, i.e. is not diverging when N is arbitrarily large. Noticing that the last term of the SDF is the only part that depends on N, a sufficient condition for this is that $\text{var}_t(\mathbf{c}'_{N,t}\boldsymbol{\varepsilon}_{N,t+1}) < \infty$ almost surely. It turns out that this is implied by the no-arbitrage condition that underlies the APT; in fact,

$$\text{var}_t(m^{\alpha}_{t+1}) = \mathbf{c}'_{N,t}\boldsymbol{\Sigma}_{N,t}\,\mathbf{c}_{N,t} = \mu^2_{m,t}\,\boldsymbol{\alpha}'_{N,t}\,\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\Sigma}_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t} < \mu^2_{m,t}\delta_{\text{apt}},$$

where δ_{apt} is defined in (1). Similarly, with respect to our nonnegative-SDF formulation,

$$\mathbb{E}_t\big((m^{\alpha+}_{t+1})^2\big) = e^{\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}} < e^{\delta_{\text{apt}}}.$$

Having ensured that our formulations of the SDF are, by non-arbitrage, well-defined for arbitrarily large N, we start by studying the behavior or $m^{\beta*}_{t+1}$, which is the SDF implied by a factor structure in the absence of any type of pricing error.

Theorem 4.5 (Properties of $m^{\beta*}_{t+1}$ and $m^{\beta*+}_{t+1}$ for large N)**.** *Under Assumptions 3.1, 3.2, and* $N^{-1}\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t} \to_p \mathbf{D}_t > 0$*, then as* $N \to \infty$*,*

$$m^{\beta*}_{t+1} \to_p m^{\beta}_{t+1} = \mu_{m,t} - \mu_{m,t}\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})),$$

$$m^{\beta*+}_{t+1} \to_p m^{\beta+}_{t+1} = \mu_{m,t}\exp\Big[-\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) - \frac{1}{2}\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t\Big].$$

Thus, the projection versions $m^{\beta*}_{t+1}$ and $m^{\beta*+}_{t+1}$ recover, respectively, m^{β}_{t+1} and $m^{\beta+}_{t+1}$ exactly with respect to the set of observed factors as $N \to \infty$. Of course, the SDF components m^{β}_{t+1} and $m^{\beta+}_{t+1}$ will still be potentially misspecified, in general, unless $\boldsymbol{\alpha}_{N,t} = \mathbf{0}_N$.

Next, we look at the component of the admissible SDF associated with misspecification, $m^{\alpha*}_{t+1}$ and $m^{\alpha*+}_{t+1}$. To simplify the analysis, we first consider the setting where this component depends only on firm-specific attributes, such as characteristics or sentiment; that is, $\boldsymbol{\alpha}_{N,t} = \mathbf{a}_{N,t}$. We then, subsequently, study the case when misspecification is solely due to missing pervasive factors. In practice, both effects are likely to be relevant when correcting a misspecified SDF. It is important to allow for firm-specific attributes because these cannot be captured by common (systematic) factors that are missing from the model. Our theory indicates that the pricing effect of such firm-specific attributes is first-order important and, indeed, our empirical work will show that these have a substantial affect on asset prices.

Theorem 4.6 (Properties of $m^{\alpha*}_{t+1}$ and $m^{\alpha*+}_{t+1}$ for large N when $\boldsymbol{\alpha}_{N,t} = \mathbf{a}_{N,t}$)**.** *Under Assumptions 3.1, 3.2, we have that* $N^{-1}\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t} \to_p \mathbf{D}_t > 0$ *and* $N^{-\frac{1}{2}}\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t} \to_p$

0, *then as* $N \to \infty$,

$$m_{t+1}^{\alpha*} - m_{t+1}^{\alpha} \to_p 0, \tag{7}$$

$$m_{t+1}^{\alpha*+} - m_{t+1}^{\alpha+} \to_p 0. \tag{8}$$

Remark 4.6.1. Equations (7) and (8) suggest that $\boldsymbol{\alpha}_{N,t}$, the idiosyncratic pricing error associated with firm-specific attributes, has a non-negligible contribution to the prices of assets, even though the no-arbitrage condition of the APT requires these pricing errors to be small, as specified in (1).

Remark 4.6.2. The additional assumption $N^{-\frac{1}{2}}\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t} \to_p 0$ is not strictly necessary for our result but it simplifies our exposition, namely that the projected and non-projected version of the alpha SDF, although both provide the required correction, have exactly the same limiting behavior. Moreover, such additional condition does not imply that $\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t}$ must equal zero at the limit, that is it allows some assets to exhibit a potentially very large pricing error.

Note that, unlike for the case of the beta SDF, Theorem 4.6 does not specify the exact form of the limit of $m_{t+1}^{\alpha*}$ and f $m_{t+1}^{\alpha*+}$. However, this can be derived as follows. Further to the regularity conditions made, now assume that

$$\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{t+1} \to_d N(0, \delta_t), \text{ where } \boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t} \to_p \delta_t.$$

Then, $m_{t+1}^{\alpha*}$ and m_{t+1}^{α} share the same limit (in distribution), namely

$$m_{t+1}^{\alpha} \to_d \mu_{m,t}\delta_t^{\frac{1}{2}}\eta_{t+1},$$

where $\eta_{t+1} \sim N(0,1)$. Therefore, even though the APT condition in (1) implies that the large majority of the pricing error $\mathbf{a}_{N,t}$ vanishes asymptotically in N, both m_{t+1}^{α} and $m_{t+1}^{\alpha*}$ do not vanish asymptotically. In other words, m_{t+1}^{α} and $m_{t+1}^{\alpha*}$ have a first-order effect on asset prices that does not dissipate even for large N. The same applies to the nonnegative SDFs, namely $m_{t+1}^{\alpha+*}$ and $m_{t+1}^{\alpha+}$ share the same limit (distribution), namely

$$m_{t+1}^{\alpha+} \to_d \exp\Big[-\delta_t^{\frac{1}{2}}\eta_{t+1} - \frac{1}{2}\delta_t\Big].$$

When N is very large, it appears that the nonnegative formulation is more useful as the limit alpha SDF will exhibit a Gaussian distribution (around zero), and thus it will take

negative values with probability of 50%. In turn, this could lead to negative values of the limiting corrected SDF, although with a smaller probability, than 50%, as the corrected SDF is centered around $\mu_{m,t}$.

Next, in contrast to the case above, we consider the case where the misspecification in the model is related only to pervasive factors; that is, $\boldsymbol{\alpha}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}$ with $\mathbf{a}_{N,t} = \mathbf{0}_N$.

Theorem 4.7 (Properties of $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\alpha*+}$ for large N when $\boldsymbol{\alpha}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}$)**.** *Under Assumptions 3.1, 3.2,* $N^{-1}\mathbf{B}_{N,t}'\boldsymbol{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t} \to_p \mathbf{D}_t > 0$, $N^{-1}\mathbf{B}_{N,t}'\mathbf{C}_{N,t}^{-1}\mathbf{B}_{N,t} \to_p \mathbf{F}_t > 0$, $N^{-1}\mathbf{A}_{N,t}'\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t} \to_p \mathbf{E}_t > 0$, *and* $\mathbf{A}_{N,t}$ *and* $\mathbf{B}_{N,t}$ *are not asymptotically collinear,*[15] *then as* $N \to \infty$,

$$m_{t+1}^{\alpha*} - m_{t+1}^{\alpha} \to_p 0 \qquad \textit{with} \quad m_{t+1}^{\alpha} \to_p -\mu_{m,t}\boldsymbol{\lambda}_{miss,t}'(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})),$$

$$m_{t+1}^{\alpha*+} - m_{t+1}^{\alpha+} \to_p 0 \qquad \textit{with} \quad m_{t+1}^{\alpha+} \to_p \exp\Big[-\boldsymbol{\lambda}_{miss,t}'(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})) - \frac{1}{2}\boldsymbol{\lambda}_{miss,t}'\boldsymbol{\lambda}_{miss,t}\Big],$$

where $\mathbf{f}_{miss,t+1}$ *is the* $p \times 1$ *vector of missing factors and* $\boldsymbol{\lambda}_{miss,t}$ *are the associated risk premia.*

Remark 4.7.1. The limit of $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\alpha*+}$ will have precisely the same mathematical form as an SDF corresponding to a factor model (i.e. just like $m_{t+1}^{\beta*}$ and $m_{t+1}^{\beta*}$), but with respect to the set of missing factors $\mathbf{f}_{miss,t+1}$. That is, in light of the identification condition that $\text{var}_t(\mathbf{f}_{miss,t+1}) = \mathbf{I}_p$, the above result becomes

$$m_{t+1}^{\alpha*} \to_p -\mu_{m,t}\boldsymbol{\lambda}_{miss,t}'\big(\text{var}_t(\mathbf{f}_{miss,t+1})\big)^{-1}\big(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})\big).$$

Therefore, the corrected SDF m_{t+1}^*, satisfies

$$m_{t+1}^* = m_{t+1}^{\beta*} + m_{t+1}^{\alpha*} \to_p \mu_{m,t} - \mu_{m,t}(\mathbf{F}_{t+1} - \mathbb{E}_t(\mathbf{F}_{t+1}))'(\text{var}_t(\mathbf{F}_{t+1}))^{-1}\boldsymbol{\Lambda}_t = m_{t+1},$$

where we define: $\mathbf{F}_{t+1} = (\mathbf{f}_{t+1}', \mathbf{f}_{miss,t}')'$ and $\boldsymbol{\Lambda}_t = (\boldsymbol{\lambda}_t', \boldsymbol{\lambda}_{miss,t}')'$. Notice that $\text{var}_t(\mathbf{F}_{t+1})$ is block-diagonal, as we assumed uncorrelatedness between the observed and missing factors; below, we will show how to modify our results when this assumption is relaxed.

[15] By *asymptotic collinearity* of the generic matrices $\mathbf{C}_N, \mathbf{D}_N$ we mean that, as $N \to \infty$, either $\mathbf{C}_N'\mathbf{M}_{D_N}\mathbf{C}_N \to 0$ or $\mathbf{D}_N'\mathbf{M}_{C_N}\mathbf{D}_N \to 0$ or both, depending on whether the number of unobserved factors $p \leq K$, $p \geq K$ or $p = K$, where $\mathbf{M}_{C_N} = \mathbf{I}_N - \mathbf{C}_N(\mathbf{C}_N'\mathbf{C}_N)^{-1}\mathbf{C}_N'$ is the matrix that spans the space orthogonal to any full-column-rank matrix $\mathbf{C}_N$, where $\mathbf{C}_N$ is a $N \times p$ matrix and $\mathbf{D}_N$ is a $N \times K$ matrix. When $p \leq K$, a sufficient condition for asymptotic collinearity is $\mathbf{C}_N = \mathbf{D}_N\delta + \mathbf{G}_N$, for some constant $K \times p$ matrix δ and some residual matrix $\mathbf{G}_N$ satisfying $\mathbf{G}_N'\mathbf{G}_N \to 0$. When $\mathbf{G}_N$ is a matrix of zeroes, then $\mathbf{C}_N$ and $\mathbf{D}_N$ are perfectly collinear.

Similarly, for the nonnegative corrected SDF

$$m_{t+1}^{*+} = m_{t+1}^{\beta*+} m_{t+1}^{\alpha*+} \to_p \mu_{m,t} \exp\left[-(\mathbf{F}_{t+1} - \mathbb{E}_t(\mathbf{F}_{t+1}))'(\text{var}_t(\mathbf{F}_{t+1}))^{-1}\boldsymbol{\Lambda}_t - \frac{1}{2}\boldsymbol{\Lambda}_t'(\text{var}_t(\mathbf{F}_{t+1}))^{-1}\boldsymbol{\Lambda}_t\right] = m_{t+1}^{+}.$$

It is important to note that preliminary estimation of the missing factors, $\mathbf{f}_{t,miss}$ is not required for deriving the corrected, admissible, SDF, reducing the impact of sampling variability of the estimated SDF.

Remark 4.7.2. Notice that the m_{t+1}^{α} is rotation-free, meaning that it is independent of the rotation assumed for the latent factors. In other words, we are recovering the SDF associated with the *true* missing factors, $\mathbf{f}_{miss,t+1}^{\dagger}$. In fact,

$$-\mu_{m,t}\boldsymbol{\lambda}_{miss,t}'(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})) = -\mu_{m,t}\boldsymbol{\lambda}_{miss,t}'\boldsymbol{\Omega}_{miss,t}^{\frac{1}{2}}\boldsymbol{\Omega}_{miss,t}^{-1}\boldsymbol{\Omega}_{miss,t}^{\frac{1}{2}}(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}))$$

$$= -\mu_{m,t}\boldsymbol{\lambda}_{miss,t}^{\dagger\prime}\boldsymbol{\Omega}_{miss,t}^{-1}(\mathbf{f}_{miss,t+1}^{\dagger} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}^{\dagger})),$$

and

$$\boldsymbol{\lambda}_{miss,t}'\boldsymbol{\lambda}_{miss,t} = \boldsymbol{\lambda}_{miss,t}^{\dagger\prime}\boldsymbol{\Omega}_{miss,t}^{-1}\boldsymbol{\lambda}_{miss,t}^{\dagger},$$

recalling that $\boldsymbol{\lambda}_{miss,t}^{\dagger}$ and $\boldsymbol{\Omega}_{miss,t}$ define the risk premia and covariance matrix associated with the $\mathbf{f}_{miss,t+1}^{\dagger}$.

Recall that, as pointed out above, $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\beta*}$, the projections of m_{t+1}^{α} and m_{t+1}^{β}, are not orthogonal for any finite N. However, as $N \to \infty$, $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\beta*}$ becomes (conditionally) orthogonal; that is, $cov_t(m_{t+1}^{\alpha*}, m_{t+1}^{\beta*}) \to_p 0$, implying that:

$$\mathbb{E}_t\left(m_{t+1}^{\beta*}m_{t+1}^{*}\right) = \mathbb{E}_t\left((m_{t+1}^{\beta*})^2\right) + O_p(N^{-\frac{1}{2}}) = \mathbb{E}_t\left((m_{t+1}^{\beta})\right)^2 + O_p(N^{-\frac{1}{2}}).$$

4.5 Detecting the Missing Factors

Our result above shows that, for the case of missing pervasive factors, $m_{t+1}^{\alpha*}$ converges to a linear function of the missing factors themselves. Note that this result does not require identification of the factors because the spanning arises automatically as N becomes large. The same argument applies to the logarithm of $m_{t+1}^{\alpha+*}$. Although not necessary for the sake of pricing, economically it is interesting to detect such missing factors. It turns out that, in our framework, a simple regression approach achieves this goal.

In particular, the R^2 of the regression of m_t^α on an intercept and the missing factors $\mathbf{f}_{miss,t}$ is defined as:

$$R^2_{miss} = \frac{\hat{\boldsymbol{\gamma}}'_{miss}\mathbf{F}'_{miss}\mathbf{M}_{1_T}\mathbf{F}_{miss}\hat{\boldsymbol{\gamma}}_{miss}}{\mathbf{m}^{\alpha\prime}\mathbf{M}_{1_T}\mathbf{m}^\alpha},$$

in which $\mathbf{F}_{miss} = (\mathbf{f}_{miss,1}\cdots\mathbf{f}_{miss,T})'$, $\mathbf{m}^\alpha = (m_1^\alpha\cdots m_T^\alpha)'$ and

$$\hat{\boldsymbol{\gamma}}_{miss} = (\mathbf{F}'_{miss}\mathbf{M}_{1_T}\mathbf{F}_{miss})^{-1}\mathbf{F}'_{miss}\mathbf{M}_{1_T}\mathbf{m}^\alpha = (\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss})^{-1}\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{m}}^\alpha.$$

Defining $\tilde{\mathbf{A}} = \mathbf{M}_{1_T}\mathbf{A}$ for any $T\times a$ matrix $\mathbf{A}$ and the projecting matrix $\mathbf{P}_\mathbf{A} = \mathbf{I}_a - \mathbf{M}_\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, for any full-column rank matrix $\mathbf{A}$ of rank a we get the following result.

Theorem 4.8 (Detecting the missing factors)**.** *Under Assumptions 3.1 and 3.2, as $N\to\infty$,*

(i) If $\boldsymbol{\alpha}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}$. If $\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t} \to_p \mathbf{D}_t > 0$, $\mathbf{B}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{B}_{N,t} \to_p \mathbf{F}_t > 0$, $\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t} \to_p \mathbf{E}_t > 0$ and $\mathbf{A}_{N,t}$ and $\mathbf{B}_{N,t}$ are not asymptotically collinear, then as $N\to\infty$:

$$\hat{\gamma}_{miss} \to_p \gamma_A = -(\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss})^{-1}(\sum_{t=1}^{T}\tilde{\mathbf{f}}_{miss,t}\xi_{At})$$

$$\text{and} \qquad R^2_{miss} \to_p \frac{\gamma'_A\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\gamma_A}{\boldsymbol{\xi}'_A\mathbf{M}_{1_T}\boldsymbol{\xi}_A} = \frac{\boldsymbol{\xi}'_A\mathbf{P}_{\tilde{F}_{miss}}\boldsymbol{\xi}_A}{\boldsymbol{\xi}'_A\mathbf{M}_{1_T}\boldsymbol{\xi}_A} = \frac{\tilde{\boldsymbol{\xi}}'_A\mathbf{P}_{\tilde{F}_{miss}}\tilde{\boldsymbol{\xi}}_A}{\tilde{\boldsymbol{\xi}}'_A\tilde{\boldsymbol{\xi}}_A},$$

in which $\boldsymbol{\xi}_A = (\xi_{A1},\cdots,\xi_{AT})'$ with $\boldsymbol{\xi}_{At} = \mu_{m,t-1}(\mathbf{f}_{miss,t} - \mathbb{E}_{t-1}(\mathbf{f}_{miss,t}))'\boldsymbol{\lambda}_{miss,t-1}$.

Moreover, when $\mu_{m,t} = \mu_m, \boldsymbol{\lambda}_{miss,t} = \boldsymbol{\lambda}_{miss}$ and $\mathbb{E}_{t-1}(\mathbf{f}_{miss,t}) = \mathbb{E}(\mathbf{f}_{miss,t})$ then:

$$\hat{\gamma}_m \to_p \gamma_A = -\mu_m\boldsymbol{\lambda}_{miss}, \quad \text{and}$$
$$R^2_{miss} \to_p 1.$$

(ii) If $\boldsymbol{\alpha}_{N,t} = \mathbf{a}_{N,t}$. If the conditions of Remark 4.6.2 hold, then as $N\to\infty$:

$$\hat{\gamma}_{miss} \to_d \gamma_a = -(\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss})^{-1}(\tilde{\mathbf{F}}'_{miss}\boldsymbol{\xi}_a), \quad \text{and}$$

$$R^2_{miss} \to_d \frac{\gamma'_a\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\gamma_a}{\boldsymbol{\xi}'_a\mathbf{M}_{1_T}\boldsymbol{\xi}_a} = \frac{\boldsymbol{\xi}'_a\mathbf{P}_{\tilde{F}_{miss}}\boldsymbol{\xi}_a}{\boldsymbol{\xi}'_a\mathbf{M}_{1_T}\boldsymbol{\xi}_a} = \frac{\tilde{\boldsymbol{\xi}}'_a\mathbf{P}_{\tilde{F}_{miss}}\tilde{\boldsymbol{\xi}}_a}{\tilde{\boldsymbol{\xi}}'_a\tilde{\boldsymbol{\xi}}_a},$$

setting $\boldsymbol{\xi}_a = (\xi_{a1},\cdots,\xi_{aT})'$ with $\xi_{at} \sim N(0,\mu^2_{m,t-1}(\delta_{t-1} - h_{t-1}\mathbf{H}^{-1}_{t-1}h_{t-1}))$ mutually independent. Moreover, when $\mu_{m,t} = \mu_m$, the same result holds except that the limit R^2 will now be functionally independent of μ_m.

Remark 4.8.1. When the nonnegative SDFs are considered, the same results above apply replacing m_t^{α} with $\log(m_t^{\alpha+})$.

Remark 4.8.2. In the above theorem, the result of part (i) holds for *every* finite T, as long as $T > K$. Regarding part (ii), under the following three mild conditions $T^{-1}\tilde{\boldsymbol{\xi}}_a'\tilde{\boldsymbol{\xi}}_a \to_p \boldsymbol{\Sigma}_{\xi a} > 0, \tilde{\mathbf{F}}'_{miss}\tilde{\boldsymbol{\xi}}_a = O_p(T^{\frac{1}{2}})$, and, $T^{-1}\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss} \to_p \boldsymbol{\Sigma}_{F_{miss}} > 0$:

$$R^2_{miss} \to_p 0 \text{ as } N, T \to \infty.$$

This implies that we can detect whether the pricing errors are driven only by $\mathbf{a}_N$ as opposed to missing factors.

Remark 4.8.3. If one erroneously considers a set of observed factors $\mathbf{F}_{wrong}$, none of which is spanning the α-SDF, assuming for simplicity that $\mathbf{a}_N = \mathbf{0}_N$ and assuming that $\mu_{m,t} = \mu_m, \boldsymbol{\lambda}_{miss,t} = \boldsymbol{\lambda}_{miss}, \mathbb{E}_{t-1}(\mathbf{f}_{miss,t}) = \mathbb{E}(\mathbf{f}_{miss,t})$, then as $N \to \infty$,

$$R^2_{miss} \to_p \frac{\boldsymbol{\lambda}'_{miss}\tilde{\mathbf{F}}'_{miss}\mathbf{P}_{\tilde{\mathbf{F}}_{wrong}}\tilde{\mathbf{F}}_{miss}\boldsymbol{\lambda}_{miss},}{\boldsymbol{\lambda}'_{miss}\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\boldsymbol{\lambda}_{miss}}. \tag{9}$$

Moreover, as $T \to \infty$, if the correct and the wrong missing factors, $\mathbf{F}_{miss}$ and $\mathbf{F}_{wrong}$, respectively, are uncorrelated (in population), the right-hand side of (9) converges to zero. This is contrast to part (ii) of the theorem. The right-hand side of (9) is identically equal to zero if $\mathbf{F}_{miss}$ and $\mathbf{F}_{wrong}$factors are orthogonal for a given T.

4.6 The SDF under the Extended APT: Non-Orthogonal Components

All the previous results were derived under the assumption that the observed risk factors $\mathbf{f}_t$ and the unobserved idiosyncratic shock $\boldsymbol{\varepsilon}_{t+1}$ are (conditionally) orthogonal, as formalized in Assumption 3.1. However, one can envisage situations where orthogonality does not necessarily hold, the best example being when there are missing pervasive factors that are *hidden* in the idiosyncratic shock and are correlated with the observed risk factors.

In this case, note that an *observationally equivalent* representation of the SDF m_{t+1} exists such that the observed risk factors $\mathbf{f}_t$ and the unobserved idiosyncratic shock $\boldsymbol{\varepsilon}_{N,t+1}$ are orthogonal. In particular, setting $\boldsymbol{\varepsilon}_{N,t+1} = \boldsymbol{\eta}_{N,t} + \mathbf{A}_{N,t}(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}))$, as from Section 3, with $\boldsymbol{\eta}_{N,t}$ and $\mathbf{f}_{miss,t+1}$ being mutually uncorrelated,

$$m_{t+1} = \mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - E_t(\mathbf{f}_{t+1})) + \mathbf{c}'_{N,t}\boldsymbol{\epsilon}_{N,t+1}, \tag{10}$$

$$= \mu_{m,t} + \tilde{\mathbf{b}}_t'(\mathbf{f}_{t+1} - E_t(\mathbf{f}_{t+1})) + \mathbf{c}_{N,t}'\tilde{\boldsymbol{\epsilon}}_{N,t+1}, \tag{11}$$

setting the $K \times p$ matrix of covariances $\mathbf{Q}_t = cov_t(\mathbf{f}_{t+1}, \mathbf{f}_{miss,t+1}')$ with

$$\tilde{\mathbf{b}}_t = \mathbf{b}_t + \boldsymbol{\Omega}_t^{-1}\mathbf{Q}_t\mathbf{A}_{N,t}'\mathbf{c}_{N,t},$$

$$\tilde{\boldsymbol{\epsilon}}_{N,t+1} = \boldsymbol{\eta}_{N,t+1} + \mathbf{A}_{N,t}\big(\tilde{\mathbf{f}}_{miss,t+1} - \mathbb{E}_t(\tilde{\mathbf{f}}_{miss,t+1})\big), \quad \text{where}$$

$$\tilde{\mathbf{f}}_{miss,t+1} = (\mathbf{I}_p, -\mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}) \begin{pmatrix} \mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}) \\ \mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}) \end{pmatrix}.$$

Notice that $cov_t(\mathbf{f}_{t+1}, \tilde{\mathbf{f}}_{miss,t+1}') = \mathbf{0}_{K\times p}$ by construction, because $\tilde{\mathbf{f}}_{miss,t+1}$ represent the linear-projection residual from projecting $\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})$ on $\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})$.

Although the two representations (10) and (11) are observationally equivalent, the one based on correlated components, that is (10), has the advantage of ensuring a clearer interpretation of the parameters, such as the ones for loadings and risk premia. For instance, the loadings associated with $\mathbf{f}_{t+1}$ in representation (11) differ from the (true) loadings of $\mathbf{f}_{t+1}$ in representation (10), a consequence of the omitted-variable bias. This can be immediately seen by comparing the extended APT in the orthogonal and non-orthogonal representations:

$$\mathbf{R}_{N,t+1}^e = \mathbf{a}_{N,t} + (\mathbf{A}_{N,t}, \mathbf{B}_{N,t}) \begin{pmatrix} \boldsymbol{\lambda}_{miss,t} \\ \boldsymbol{\lambda}_t \end{pmatrix} + (\mathbf{A}_{N,t}, \mathbf{B}_{N,t}) \begin{pmatrix} \mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}) \\ \mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}) \end{pmatrix} + \boldsymbol{\eta}_{t+1},$$

$$= \mathbf{a}_{N,t} + (\tilde{\mathbf{A}}_{N,t}, \tilde{\mathbf{B}}_{N,t}) \begin{pmatrix} \tilde{\boldsymbol{\lambda}}_{miss,t} \\ \boldsymbol{\lambda}_t \end{pmatrix} + (\tilde{\mathbf{A}}_{N,t}, \tilde{\mathbf{B}}_{N,t}) \begin{pmatrix} \tilde{\mathbf{f}}_{miss,t+1} - \mathbb{E}_t(\tilde{\mathbf{f}}_{miss,t+1}) \\ \mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}) \end{pmatrix} + \boldsymbol{\eta}_{t+1},$$

where

$$\tilde{\mathbf{A}}_{N,t} = \mathbf{A}_{N,t}(\mathbf{I}_p - \mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{Q}_t)^{\frac{1}{2}}, \tag{12}$$

$$\tilde{\mathbf{B}}_{N,t} = \mathbf{B}_{N,t} + \mathbf{A}_{N,t}\mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}, \tag{13}$$

$$\tilde{\boldsymbol{\lambda}}_{miss,t} = (\mathbf{I}_p - \mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{Q}_t)^{-\frac{1}{2}}(\boldsymbol{\lambda}_{miss,t} - \mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t), \tag{14}$$

and $\tilde{\mathbf{f}}_{miss,t+1}$ have (conditionally) unit covariance matrix and are uncorrelated with the $\mathbf{f}_{t+1}$.[16]

[16] In particular, the $\tilde{\mathbf{f}}_{miss,t+1}$ are given by:

$$\tilde{\mathbf{f}}_{miss,t+1} = (\mathbf{I}_p - \mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}\mathbf{Q}_t)^{-\frac{1}{2}}(\mathbf{I}_p, -\mathbf{Q}_t'\boldsymbol{\Omega}_t^{-1}) \begin{pmatrix} \mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}) \\ \mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}) \end{pmatrix}.$$

It follows that, unless one makes use of the mappings between the parameters of the orthogonal and non-orthogonal representations, namely (12)-(13)-(14), the orthogonal representation parameters confound completely the original parameters, in particular in terms of loadings and risk premia of the missing factors, without any additional computational and efficiency gain, because the total number of parameters remains unchanged. On the other hand, if one makes use of the mappings to back out one set from the other, then it is unclear whether keeping the orthogonal representation brings any advantages to the analysis. Based on this consideration, this section is written in terms of the non-orthogonal representation of the extended APT.

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 3.1 to:

Assumption 4.1 (Linear factor model: correlated case)**.** *Assumption 3.1 holds with*

$$\mathbb{E}_t(\mathbf{f}_{t+1}\boldsymbol{\varepsilon}'_{t+1}) = \mathbf{P}_{N,t},$$

for some non-zero $K \times N$ matrix $\mathbf{P}_{N,t}$ such that perfect (conditional) correlation between the $\mathbf{f}_{t+1}$ and the $\boldsymbol{\varepsilon}_{t+1}$ is ruled out:

$$\mathbf{I}_N - \boldsymbol{\Sigma}_{N,t}^{-\frac{1}{2}}\mathbf{P}'_{N,t}\boldsymbol{\Omega}_t^{-1}\mathbf{P}_{N,t}\boldsymbol{\Sigma}_{N,t}^{-\frac{1}{2}} > 0.$$

Although the expression for expected excess returns is unchanged, the (conditional) second moment for excess returns becomes:

$$cov_t(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N) = \mathbf{V}_{N,t} = \mathbf{B}_{N,t}\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \boldsymbol{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t} + \mathbf{B}_{N,t}\mathbf{P}_{N,t}.$$

We first show how the expression for the linear and exponential SDF changes, as a result of the lack of orthogonality. Then, we analyze their large-N behavior.

Theorem 4.9 (SDF in closed form for the correlated case)**.** *Under Assumptions 4.1 and 3.2 of the APT, for a given $\mu_{m,t}$, there exists an admissible SDF of the form*

$$m_{t+1} = \mu_{m,t} + \mathbf{b}'_t(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) + \mathbf{c}'_{N,t}\boldsymbol{\varepsilon}_{N,t+1},$$

with

$$\mathbf{b}_t = -\mu_{m,t}\left(\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\Omega}_t^{-1}\mathbf{P}_{N,t}\mathbf{H}_{N,t}^{-1}(\boldsymbol{\alpha}_{N,t} - \mathbf{P}'_{N,t}\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t)\right),$$

$$\mathbf{c}_{N,t} = -\mu_{m,t}\left(\mathbf{H}_{N,t}^{-1}(\boldsymbol{\alpha}_{N,t} - \mathbf{P}'_{N,t}\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t)\right),$$

setting

$$\mathbf{H}_{N,t} = \boldsymbol{\Sigma}_{N,t} - \mathbf{P}'_{N,t}\boldsymbol{\Omega}_t^{-1}\mathbf{P}_{N,t}.$$

When expressed in terms of a linear projection on the set of payoffs $(1, \mathbf{R}^e_{N,t+1})$*, the SDF is*

$$\begin{aligned} m^*_{t+1} &= \text{proj}(m_{t+1}|(1, \mathbf{R}^e_{N,t+1})) \\ &= \mu_{m,t} + (\mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}] + \mathbf{c}'_{N,t}[\boldsymbol{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t}])\mathbf{V}^{-1}_{N,t}\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big) \\ &= \mu_{m,t} - (\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)'\mathbf{V}^{-1}_{N,t}\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1}), \end{aligned}$$

where $\mathbf{V}_{N,t}$ *is the (conditional) covariance matrix of excess returns.*

Although the expressions for the coefficients in the SDF, namely $\mathbf{b}_t$ and $\mathbf{c}_{N,t}$ differ from before, one obtains the decomposition into the alpha and beta SDF:

$$m_{t+1} = m^{\alpha}_{t+1} + m^{\beta}_{t+1},$$

where

$$m^{\beta}_{t+1} = \mu_{m,t} - \mu_{m,t}\mathbf{b}'_t(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) \text{ and } m^{\alpha}_{t+1} = -\mu_{m,t}\boldsymbol{\varepsilon}'_{N,t+1}\mathbf{c}_{N,t},$$

where $\mathbf{b}_t$ and $\mathbf{c}_{N,t}$ are defined in Theorem 4.9. In terms of the pricing of asset returns:

$$\mathbb{E}_t\left(m^{\beta}_{t+1}\begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1}\end{bmatrix}\right) = \mu_{m,t}\begin{bmatrix} 1 \\ (\boldsymbol{\Sigma}_{N,t} + \mathbf{B}_{N,t}\mathbf{P}_{N,t})\mathbf{c}_{N,t}\end{bmatrix}$$

$$\mathbb{E}_t\left(m^{\alpha}_{t+1}\begin{bmatrix} 1 \\ \mathbf{R}^e_{N,t+1}\end{bmatrix}\right) = \mu_{m,t}\begin{bmatrix} 0 \\ -(\boldsymbol{\Sigma}_{N,t} + \mathbf{B}_{N,t}\mathbf{P}_{N,t})\mathbf{c}_{N,t}\end{bmatrix}.$$

Notice that now $cov_t(m^{\alpha}_{t+1}, m^{\beta}_{t+1}) = \mu^2_{m,t}\mathbf{b}'_t\mathbf{P}_{N,t}\mathbf{c}_{N,t} \neq 0$. Despite this, as for the previous orthogonal case, the misspecified m^{β}_{t+1} prices the observed factors correctly, that is $\mathbb{E}_t(m^{\beta}_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)) = \mathbf{0}_K$.

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

$$m^*_{t+1} = m^{\alpha *}_{t+1} + m^{\beta *}_{t+1},$$

with

$$m^{\beta *}_{t+1} = \mu_{m,t} + \mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}]\mathbf{V}^{-1}_{N,t}\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big),$$

$$m_{t+1}^{\alpha*} = \mathbf{c}'_{N,t}[\mathbf{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t}])\mathbf{V}_{N,t}^{-1}\big(\mathbf{R}_{N,t+1}^{e} - \mathbb{E}_t(\mathbf{R}_{N,t+1}^{e})\big).$$

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

Theorem 4.10 (Nonnegative SDF in closed form for the correlated case)**.** *Under Assumptions 4.1 and 3.2 of the APT and that returns are conditionally normally distributed, there exists an admissible SDF m_{t+1}^{+}, with the given mean $\mu_{m,t}$, of the form*

$$m_{t+1}^{+} = \exp\Big[\mu_{m,t}^{+} + \mathbf{b}_t^{+\prime}(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})) + \mathbf{c}_{N,t}^{+\prime}\boldsymbol{\varepsilon}_{N,t+1}\Big],$$

with

$$\begin{aligned}
\mu_{m,t}^{+} &= \ln\mu_{m,t} - \frac{1}{2}(\mathbf{b}_t^{+\prime}, \mathbf{c}_{N,t}^{+\prime})\begin{pmatrix} \mathbf{\Omega}_t & \mathbf{P}_{N,t} \\ \mathbf{P}'_{N,t} & \mathbf{\Sigma}_{N,t} \end{pmatrix}\begin{pmatrix} \mathbf{b}_t^{+} \\ \mathbf{c}_{N,t}^{+} \end{pmatrix}, \\
\mathbf{b}_t^{+} &= -\Big(\mathbf{\Omega}_t^{-1}\boldsymbol{\lambda}_t - \mathbf{\Omega}_t^{-1}\mathbf{P}_{N,t}\mathbf{H}_{N,t}^{-1}(\boldsymbol{\alpha}_{N,t} - \mathbf{P}'_{N,t}\mathbf{\Omega}_t^{-1}\boldsymbol{\lambda}_t)\Big), \\
\mathbf{c}_{N,t}^{+} &= -\Big(\mathbf{H}_{N,t}^{-1}(\boldsymbol{\alpha}_{N,t} - \mathbf{P}'_{N,t}\mathbf{\Omega}_t^{-1}\boldsymbol{\lambda}_t)\Big),
\end{aligned}$$

recalling $\mathbf{H}_{N,t} = \mathbf{\Sigma}_{N,t} - \mathbf{P}'_{N,t}\mathbf{\Omega}_t^{-1}\mathbf{P}_{N,t}$. When the risk-free asset is available one replaces $\ln\mu_{m,t}$ with $-\ln R_{ft}$ into $\mu_{m,t}^{+}$.

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

$$m_{t+1}^{*+} = m_{t+1}^{\alpha*+}\, m_{t+1}^{\beta*+},$$

where

$$m_{t+1}^{\beta*+} = \mu_{m,t}\exp\Big[\mathbf{b}_t^{+\prime}(\mathbf{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t})\mathbf{V}_{N,t}^{-1}(\mathbf{R}_{N,t+1}^{e} - \mathbb{E}_t(\mathbf{R}_{N,t+1}^{e}) - \frac{1}{2}\mathbf{b}_t^{+\prime}\mathbf{\Omega}_t\mathbf{b}_t^{+}\Big],$$

and

$$m_{t+1}^{\alpha*+} = \exp\Big[\mathbf{c}_{N,t}^{+\prime}(\mathbf{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t})\mathbf{V}_{N,t}^{-1}\big(\mathbf{R}_{N,t+1}^{e} - \mathbb{E}_t(\mathbf{R}_{N,t+1}^{e})\big) - \frac{1}{2}(\mathbf{b}_t^{+\prime}, \mathbf{c}_{N,t}^{+\prime})\begin{pmatrix} \mathbf{0}_{K\times K} & \mathbf{P}_{N,t} \\ \mathbf{P}'_{N,t} & \mathbf{\Sigma}_{N,t} \end{pmatrix}\begin{pmatrix} \mathbf{b}_t^{+} \\ \mathbf{c}_{N,t}^{+} \end{pmatrix}\Big].$$

We complete this section by showing how the previous large-N results extend to the non-orthogonal case. For simplicity we focus on the case when the pricing errors are only driven

by missing pervasive factors, that is $\boldsymbol{\alpha}_{N,t} = \mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}$. This implies

$$\mathbf{P}_{N,t} = cov_t(\mathbf{f}_{t+1}, \mathbf{f}'_{miss,t+1})\mathbf{A}'_{N,t} = \mathbf{Q}_t\mathbf{A}'_{N,t},$$

setting the $K \times p$ matrix of covariances $\mathbf{Q}_t = cov_t(\mathbf{f}_{t+1}, \mathbf{f}'_{miss,t+1})$.

For the linear SDFs, under Assumptions 3.1, 3.2, and $N^{-1}(\mathbf{A}_{N,t}, \mathbf{B}_{N,t})'\mathbf{C}_{N,t}^{-1}(\mathbf{A}_{N,t}, \mathbf{B}_{N,t}) \to_p \mathbf{D}_t > 0$, as $N \to \infty$,

$$\begin{aligned}
m^*_{t+1} &= m^{\alpha *}_{t+1} + m^{\beta *}_{t+1} \\
&= \mu_{m,t} - \mu_{m,t}(\boldsymbol{\lambda}'_{miss,t}, \boldsymbol{\lambda}'_t)(\mathbf{A}_{N,t}, \mathbf{B}_{N,t})'\mathbf{V}_{N,t}^{-1}\left((\mathbf{A}_{N,t}, \mathbf{B}_{N,t})\begin{pmatrix} \mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}) \\ \mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}) \end{pmatrix}\right) \\
&\to_p \mu_{m,t} - \mu_{m,t}(\boldsymbol{\lambda}'_{miss,t}, \boldsymbol{\lambda}'_t)\begin{pmatrix} \mathbf{I}_p & \mathbf{Q}'_t \\ \mathbf{Q}_t & \boldsymbol{\Omega}_t \end{pmatrix}\begin{pmatrix} \mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1}) \\ \mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}) \end{pmatrix} \\
&= \mu_{m,t} - \mu_{m,t}\boldsymbol{\Lambda}'_t(\mathrm{var}_t(\mathbf{F}_{t+1}))^{-1}(\mathbf{F}_{t+1} - \mathbb{E}_t(\mathbf{F}_{t+1})) = m_{t+1},
\end{aligned}$$

recalling that $\mathbf{F}_{t+1} = (\mathbf{f}'_{miss,t}, \mathbf{f}'_{t+1})'$ and $\boldsymbol{\Lambda}_t = (\boldsymbol{\lambda}'_{miss,t}, \boldsymbol{\lambda}'_t)'$, where now $\mathrm{var}_t(\mathbf{F}_{t+1}) = \begin{pmatrix} \mathbf{I}_p & \mathbf{Q}'_t \\ \mathbf{Q}_t & \boldsymbol{\Omega}_t \end{pmatrix}$ is not block-diagonal any longer.[17]

By the same arguments, regarding the (nonlinear) projection of the nonnegative SDF,

$$m^{+*}_{t+1} = m^{\alpha +*}_{t+1}m^{\beta +*}_{t+1} \to_p \mu_{m,t} exp\left[-\boldsymbol{\Lambda}'_t(\mathrm{var}_t(\mathbf{F}_{t+1}))^{-1}(\mathbf{F}_{t+1} - \mathbb{E}_t(\mathbf{F}_{t+1})) - \frac{1}{2}\boldsymbol{\Lambda}'_t(\mathrm{var}_t(\mathbf{F}_{t+1}))^{-1}\boldsymbol{\Lambda}_t\right].$$

Therefore, the correction carried out by the (feasible) $m^{\alpha *}_{t+1}$ and $m^{\alpha *+}_{t+1}$ delivers, for large N, precisely the *infeasible* SDFs, respectively in the linear and exponential form, that correspond to the case of $K + p$ common risk factors $(\mathbf{f}'_{miss,t}, \mathbf{f}'_{t+1})'$. As for the orthogonal case, we are selecting a specific rotation of the true factors $(\mathbf{f}^{\dagger\prime}_{miss,t}, \mathbf{f}'_{t+1})'$ such that

$$\begin{pmatrix} \mathbf{f}_{miss,t} \\ \mathbf{f}_{t+1} \end{pmatrix} = \mathbf{H}_t\begin{pmatrix} \boldsymbol{\Omega}_{miss,t}^{-\frac{1}{2}} & \mathbf{0}_{p\times K} \\ \mathbf{0}_{K\times p} & \mathbf{I}_K \end{pmatrix}\begin{pmatrix} \mathbf{f}^{\dagger}_{miss,t} \\ \mathbf{f}_{t+1} \end{pmatrix},$$

for any $p + K \times p + K$ orthogonal matrix $\mathbf{H}_t$.

[17] By an extension of Lemma 1, under our assumptions,

$$(\mathbf{A}_{N,t}, \mathbf{B}_{N,t})'\mathbf{V}_{N,t}^{-1}(\mathbf{A}_{N,t}, \mathbf{B}_{N,t}) \to_p \begin{pmatrix} \mathbf{I}_p & \mathbf{Q}'_t \\ \mathbf{Q}_t & \boldsymbol{\Omega}_t \end{pmatrix}^{-1},$$

recalling that $\mathbf{V}_{N,t} = (\mathbf{A}_{N,t}, \mathbf{B}_{N,t})\begin{pmatrix} \mathbf{I}_p & \mathbf{Q}'_t \\ \mathbf{Q}_t & \boldsymbol{\Omega}_t \end{pmatrix}(\mathbf{A}_{N,t}, \mathbf{B}_{N,t})' + \mathbf{C}_{N,t}$.

5 Representations of the SDF

In this section, we establish two important representations of the SDF under the extended APT. The first one is in terms of the returns of mean-variance efficient portfolios. The second is the one-factor beta representation.

5.1 The SDF Frontier in Terms of Returns

This section describe the interpretation of the SDF as returns. In particular, we extend the well-know duality between admissible SDF and efficient portfolio to a duality between *misspecified* SDFs and *inefficient* portfolios, yet with very special properties. In particular, once we map the SDF components $m_{t+1}^{\alpha*}$ and $m_{t+1}^{\beta*}$ into the returns' space, we show that they correspond to the (excess) returns of two portfolios, denominated as the alpha and beta portfolios, respectively, (these portfolios are described in detail in Raponi, Uppal, and Zaffaroni (2019)). These two portfolios turn out to be inefficient; that is, they lie *inside* the efficient frontier, as opposed of being on the efficient frontier. However, they satisfy two-fund separation: they span the efficient frontier, in particular the lower branch of the efficient frontier, which corresponds to the SDF frontier; that is, the set of admissible minimum-variance SDFs.

We first need some definitions. Under Assumption D.3 from Hansen and Richard (1987, Assumption 2.4), the return

$$R_{t+1}^* = \frac{m_{t+1}^*}{\mathbb{E}_t\big((m_{t+1}^*)^2\big)}$$

is well defined, implying that the space of returns $\boldsymbol{\mathcal{R}} = \{x_{t+1} \in \boldsymbol{\mathcal{X}} : p(x_{t+1}) = 1\}$ is not empty. We define R_{t+1}^* the *pricing functional*, namely the SDF mapped into the returns space. Hansen and Richard (1987, Lemma 3.1) show that the return R_{t+1}^* is the minimum conditional second moment return: $\mathbb{E}_t(R_{t+1}^*)^2 \leq \mathbb{E}_t(R_{t+1})^2$ for every $R_{t+1} \in \boldsymbol{\mathcal{R}}$. Moreover, R_{t+1}^* *prices* all excess returns: $\mathbb{E}_t(R_{t+1}^* R_{n,t+1}^e) = 0$ for every $R_{n,t+1}^e \in \boldsymbol{\mathcal{R}}^e = \{x_{t+1} \in \boldsymbol{\mathcal{X}} : p(x_{t+1}) = 0\}$).

We first show how the return pricing functional R^*_{t+1} is related to $m^{\alpha*}_{t+1}$ and $m^{\beta*}_{t+1}$. In particular, define:

$$R^{\alpha*}_{t+1} = \frac{m^{\alpha*}_{t+1}}{\mathbb{E}_t\left(m^{\alpha*}_{t+1}m^*_{t+1}\right)}, \qquad\qquad R^{\beta*}_{t+1} = \frac{m^{\beta*}_{t+1}}{\mathbb{E}_t\left(m^{\beta*}_{t+1}m^*_{t+1}\right)}.$$

Then, under Assumptions 3.1 and 3.2, the pricing functional R^*_{t+1} has the following decomposition in terms of returns:

$$R^*_{t+1} = \frac{m^*_{t+1}}{\mathbb{E}_t\left(m^*_{t+1}\right)^2} = \kappa^{\alpha*}_t R^{\alpha*}_{t+1} + \left(1-\kappa^{\alpha*}_t\right) R^{\beta*}_{t+1} \quad \text{with} \tag{15}$$

$$\kappa^{\alpha*}_t = \frac{\mathbb{E}_t\left(m^{\alpha*}_{t+1}m^*_{t+1}\right)}{\mathbb{E}_t\left(m^*_{t+1}\right)^2}.$$

Unlike m^{α}_{t+1} and m^{β}_{t+1}, the projections $m^{\alpha*}_{t+1}$ and $m^{\beta*}_{t+1}$, and thus $R^{\alpha*}_{t+1}$ and $R^{\beta*}_{t+1}$, are *not* orthogonal for any finite N. However, as $N\to\infty$, $R^{\alpha*}_{t+1}$ and $R^{\beta*}_{t+1}$ become (conditionally) orthogonal, namely: In fact,

$$cov_t(R^{\alpha*}_{t+1}, R^{\beta*}_{t+1}) = \frac{cov_t(m^{\alpha*}_{t+1}, m^{\beta*}_{t+1})}{\mathbb{E}_t(m^*_{t+1}m^{\alpha*}_{t+1})\mathbb{E}_t(m^*_{t+1}m^{\beta*}_{t+1})} = \frac{\mu^2_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t}{\mathbb{E}_t(m^*_{t+1}m^{\alpha*}_{t+1})\mathbb{E}_t(m^*_{t+1}m^{\beta*}_{t+1})} \to_p 0.$$

This implies that $0 \le \kappa^{\alpha*}_t \le 1$ as $N\to\infty$, although these bounds on $\kappa^{\alpha*}_t$ follows, for any N, whenever $-\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t \le \boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}$ and $-\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t \le 1 + \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t$.

In order to understand the relative positions of R^*_{t+1} and its components, in the returns' space, it is useful to study their first and second moments for large-N. In particular,

$$\mathbb{E}_t(R^*_{t+1}) = \frac{1}{\mu_{m,t}(1+\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}+\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t)} + O_p(N^{-\frac{1}{2}}),$$

where $\mathbb{E}_t\big(R^{\alpha*}_{t+1}\big) = 0$ and $\mathbb{E}_t\big(R^{\beta*}_{t+1}\big) = \frac{1}{\mu_{m,t}\left(1+\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t\right)} + O_p(N^{-\frac{1}{2}})$, and, in terms of (conditional) variances,

$$\mathrm{var}_t(R^*_{t+1}) = \frac{1}{\mu^2_{m,t}}\frac{(\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}+\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t)}{(1+\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}+\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t)^2} + O_p(N^{-\frac{1}{2}}),$$

where $\mathrm{var}_t(R^{\alpha*}_{t+1}) = \frac{1}{\mu^2_{m,t}}\frac{1}{(\boldsymbol{\alpha}_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t})} + O_p(N^{-\frac{1}{2}})$ and $\mathrm{var}_t(R^{\beta*}_{t+1}) = \frac{1}{\mu^2_{m,t}}\frac{\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t}{(1+\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t)^2} + O_p(N^{-\frac{1}{2}})$. Therefore, under our assumption that a risk-free asset is traded,

$$\mathbb{E}_t(R^*_{t+1}) < R_{f,t} = \frac{1}{\mu_{m,t}},$$

then it follows that R^*_{t+1} is on the *lower* part of the mean-variance frontier, because it earns an expected return *smaller* than R_{ft} but with a non-zero variance. It is exactly on the frontier because, as indicated above, it is has the minimum second moment. Under the same conditions for $0 \leq \kappa^{\alpha *} \leq 1$, always valid for large N, one obtains

$$0 = \mathbb{E}_t\big(R^{\alpha *}_{t+1}\big) < \mathbb{E}_t\big(R^*_{t+1}\big) < \mathbb{E}_t\big(R^{\beta *}_{t+1}\big) < R_{ft} = \frac{1}{\mu_{m,t}}.$$

A more formal way to understand the role of $R^{\alpha *}_{t+1}$ and $R^{\beta *}_{t+1}$ as returns in the mean-standard deviation space is presented in the next theorem, where we decompose the R^*_{t+1} in excess of the risk-free rate, in terms of the returns of two *special* inefficient portfolios, called the alpha and beta portfolios, with corresponding weights indicated by $\mathbf{w}^{\alpha}_N$ and $\mathbf{w}^{\beta}_N$ portfolios (see Raponi, Uppal, and Zaffaroni (2019) for details).

Theorem 5.1 (Decomposition of pricing functional R^*_{t+1} in terms of return on $\mathbf{w}^{\alpha}_N$ and $\mathbf{w}^{\beta}_N$ portfolios)**.** *Under Assumptions 3.1 and 3.2, for any $\mu_{m,t}$, R^*_{t+1} satisfies:*

$$R^*_{t+1} - R_{ft} = \phi^{\alpha}_t \mathbf{w}^{\alpha\prime}_{0,t} \mathbf{R}^e_{N,t+1} + \phi^{\beta}_t \mathbf{w}^{\beta\prime}_{0,t} \mathbf{R}^e_{N,t+1},$$

for some coefficients ϕ^{α}_t and ϕ^{β}_t (see Eq.? in the proof) satisfying $\phi^{\alpha}_t + \phi^{\beta}_t \to_p 1$ as N diverges, and $\mathbf{w}^{\alpha}_{\mu^*,t}$ and $\mathbf{w}^{\beta}_{\mu^*,t}$ are the α-portfolio and the β-portfolio, for given target mean μ^*, defined respectively by:

$$\mathbf{w}^{\alpha}_{\mu^*,t} = \frac{(\mu^* - R_{ft})}{\boldsymbol{\alpha}'_{N,t} \mathbf{V}^{-1}_{N,t} \boldsymbol{\alpha}_{N,t}} \mathbf{V}^{-1}_{N,t} \boldsymbol{\alpha}_{N,t}, \qquad \mathbf{w}^{\beta}_{\mu^*,t} = \frac{(\mu^* - R_{ft})}{\boldsymbol{\lambda}'_t \mathbf{B}'_{N,t} \mathbf{V}^{-1}_{N,t} \mathbf{B}_{N,t} \boldsymbol{\lambda}_t} \mathbf{V}^{-1}_{N,t} \mathbf{B}_{N,t} \boldsymbol{\lambda}_t.$$

Remark 5.1.1. In Raponi, Uppal, and Zaffaroni (2019), it is shown that the $\mathbf{w}^{\alpha}_N$ and $\mathbf{w}^{\beta}_N$ portfolios are both inefficient (that is, belong to the interior of the mean-variance frontier) but still satisfy the two-fund separation theorem (that is, span the efficient frontier). In fact, $\mathbf{w}^{\alpha}_N$ and $\mathbf{w}^{\beta}_N$ have very special properties: $\mathbf{w}^{\alpha}_N$ is the minimum-variance orthogonal portfolio to $\mathbf{w}^{\beta}_N$ and, vice-versa, $\mathbf{w}^{\beta}_N$ is the minimum-variance orthogonal portfolio to $\mathbf{w}^{\alpha}_N$. Moreover, although inefficient, both portfolios satisfy the *mean-variance* property by which their expected return is equal to the variance or, alternatively, their Sharpe ratios coincide with the portfolios' standard deviations. Therefore, all these properties are inherited by $R^{\alpha *}_{t+1}$ and $R^{\beta *}_{t+1}$ so, in particular, they are positioned in the interior of the SDF frontier, with $R^{\alpha *}_{t+1}$ on the horizontal axis.

5.2 Beta Representation of the SDF

Recall from the existing literature that the SDF can be used to obtain a single-beta representation for returns.[18] At the same time, when the SDF is linear in a set of K factors, expected returns follow a K-factor beta representation. In this section, we combine these two results to show that under the extended APT one obtains a $(K+1)$-factor representation of expected returns.

From the existing literature, if an admissible SDF satisfies a K-factor structure,

$$m_{t+1} = \mathbb{E}_t(m_{t+1}) + \mathbf{b}_t'\big(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})\big), \tag{16}$$

for some observed random vector $\mathbf{f}_t$ such that $\mathbb{E}_t(m_{t+1}) > 0$, then the K-factor beta representation follows:

$$\mathbb{E}_t(R^e_{n,t+1}) = \boldsymbol{\beta}_{n,t}'\boldsymbol{\lambda}_t, \tag{17}$$

where $\boldsymbol{\beta}_{n,t} = \boldsymbol{\Omega}_t^{-1}\,\mathrm{cov}_t(\mathbf{f}_{t+1}, R_{n,t+1})$. The converse is also true; namely, if (17) is true, then (16) is true.[19]

Regarding the single-factor representation in the existing literature, we need the following definition.

Definition 5.1 (Reference return; Hansen and Richard (1987, Equation (3.28))). *A return $R_{beta,t+1}$ is a reference return for a single-beta representation conditional on information at date t if prob$(\mathrm{var}_t(R_{beta,t+1}) = 0) = 0$ and*

$$\mathbb{E}_t(R_{n,t+1}) - a_t = \frac{cov_t(R_{n,t+1}), R_{beta,t+1}}{\mathrm{var}_t(R_{beta,t+1})}\left[\mathbb{E}_t(R_{beta,t+1}) - a_t\right] \quad \forall \quad R_{n,t+1} \in \boldsymbol{\mathcal{R}}.$$

If there is a unit payoff, then $a_t = R_{ft}$; otherwise, a_t is arbitrary.

Then, Hansen and Richard (1987, Lemma 3.5) show that, under Assumptions D.1, D.2, D.3, and the absence of risk neutrality and arbitrage opportunities, then $R_{beta,t+1}$ is a

[18] These well-known results have been pioneered by Ross (1978), Dybvig and Ingersoll, Jr. (1982), Cochrane (1996), and Lettau and Ludvigson (2001); for textbook treatment, see Cochrane (2005) and Back (2017).

[19] Notice that the factor-structure representation of the SDF in (16) does not rule out the possibility of a negative SDF, that is, $m_{t+1} < 0$. However, we know from Hansen and Richard (1987, Lemma 2.3) that in this case arbitrage opportunities are not ruled out. For a discussion of the conditions under which $m_{t+1} > 0$, see Back (2017). In particular, if markets are complete, then to rule out arbitrage one needs to restrict the support of the distribution of $\mathbf{f}_{t+1}$.

reference return for a conditional single-beta representation, if and only if,

$$R_{beta,t+1} = R^*_{t+1} + w_t^{beta}\, R^{e*}_{t+1},$$

where w_t^{beta} differs from $\frac{\mathbb{E}_t(R^*_{t+1})}{1-\mathbb{E}_t(R^{e*}_{t+1})}$ with probability one.

We now show how the admissible SDF implied by the extended APT leads to a $(K+1)$-beta representation for returns, where K refers to the number of observed factors corresponding to the beta SDF, m^{β}_{t+1}.

Theorem 5.2 ($(K+1)$ factor structure for expected returns)**.** *Under Assumptions 3.1 and 3.2, for any $\mu_{m,t} \neq 0$, every $R_{n,t+1}$ and $R^e_{n,t+1}$satisfy a $(K+1)$-factor structure:*

$$\mathbb{E}_t(R^e_{n,t+1}) = \beta^{\alpha}_{n,t}\lambda^{\alpha}_t + \boldsymbol{\beta}'_{n,t}\boldsymbol{\lambda}_t,$$

where $\boldsymbol{\beta}_{n,t} = \boldsymbol{\Omega}_t^{-1}\,\mathrm{cov}_t(\mathbf{f}_{t+1}, R_{n,t+1})$, $\beta^{\alpha}_{n,t} = \frac{\mathrm{cov}_t(m^{\alpha}_{t+1}, R_{n,t+1})}{var_t(m^{\alpha}_{t+1})}$, and $\lambda^{\alpha}_t = -R_{ft}\mathrm{var}_t(m^{\alpha}_{t+1})$.

Remark 5.2.1. If the risk-free rate is constant, then $\mathbb{E}_t(m^{\beta}_{t+1})$ is constant, implying that $\mathbf{b}'_t\mathbf{f}_{t+1}$ is serially uncorrelated.

Our result implies that, starting from a given beta SDF that has a K-factor representation, the admissible SDF has a $K+1$ factor structure, even when one does not know the exact number of missing factors. This is in contrast to the principal-components approach, where the admissible SDF is written as a function of the estimated factors. Identifying the precise number of factors that are needed for an admissible SDF could be problematic, and the consequences of both under- and over-estimating the number of true factors can lead to severe problems when pricing assets. In the case where the number of factors is underestimated, we face the typical problem arising from missing variables in regression analysis. In the case where the number of factors is overestimated, we face the problem of spurious factors, namely highly significant risk premia when the true risk premia is zero for the spurious factor.

6 The SDF for Equilibrium Asset-Pricing Models

Our methodology has been developed within the set of linear factor SDF. In this section, we explain how our methodology applies also to nonlinear SDFs, which arise in the context of equilibrium asset-pricing models.

Suppose that we are interested in a class of equilibrium SDFs generically specified as:

$$m_{t+1}^{eq} = m(\mathbf{f}_{t+1}),$$

for some function $m(\cdot)$, an S-dimensional vector of state-variables $\mathbf{f}_{t+1}$ (for example, consumption growth c_{t+1}) that are Gaussian conditional on past information. Note that above we have written m_{t+1} to indicate that for pricing purposes, the SDF is an argument of a one-step-ahead conditional expectation, the specification of any aspect of the SDF that is a function of information at dates prior to $t+1$ add no additional difficulty to our analysis.

For instance, the consumption-CAPM model of Breeden (1979) can be expressed as

$$m_{t+1}^{eq} = \zeta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma},$$

where ζ is the subjective rate of time preference, γ is the relative risk aversion, and C_t is aggregate consumption at date t, implying that the single state variable $f_t = \ln(C_t/C_{t-1})$. Similarly, the model of Campbell and Cochrane (1999) with external habit can be expressed as

$$m_{t+1}^{eq} = \zeta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\frac{S_{t+1}}{S_t} \right)^{-\gamma},$$

where $S_t = \frac{C_t - X_t}{C_t}$ denotes the surplus consumption ratio and X_t represents the level of external habit, where now the state-variable vector is $\mathbf{f_t} = \big(\ln(C_t/C_{t-1}), \ln(S_t/S_{t-1}) \big)'$. The model of Bansal and Yaron (2004) with long-run risk can be written as

$$m_{t+1}^{eq} = \zeta^{\theta} \left(\frac{C_{t+1}}{C_t} \right)^{-\theta/\rho} R_{c,t+1}^{\theta-1},$$

where $\theta = \frac{1-\gamma}{1-1/\rho}$, ρ is the elasticity of intertemporal substitution, and $R_{c,t+1}$ is the unobservable return on an asset that pays out aggregate consumption as its dividend, where now the state-variable vector is $\mathbf{f_t} = \big(\ln(C_t/C_{t-1}), \ln R_{c,t} \big)'$.

We first consider the scalar state variable case, that is $S = 1$, together with conditionally constant moments, that is $\mathbb{E}_t(m_{t+1}) = \mathbb{E}(m_{t+1}) = \mu_m$, and then discuss the case with time-varying moments, as well as the case of multiple state variables.

Theorem 6.1 (SDF for equilibrium models)**.** *Assume that f_{t+1} has a standard normal distribution. If*

$$\mathbb{E}(m^2(f_{t+1})) < \infty$$

the following orthogonal decomposition holds:

$$m_{t+1}^{eq} = m_{t+1}^{\alpha} + m_{t+1}^{\beta}, \qquad with \qquad \mathbb{E}(m_{t+1}^{\alpha} m_{t+1}^{\beta}) = 0,$$

setting

$$m_{t+1}^{\beta} = \mu_m + b f_{t+1},$$

$$m_{t+1}^{\alpha} = \sum_{h=2}^{\infty} \alpha_h H_h(f_{t+1}),$$

where $H_k(x) = (-1)^h e^{x^2/2} d^k e^{-x^2/2}/dx^k$ *denote the h-th Hermite polynomial* ($h = 0, 1, \ldots$) *and* $\mu_m = a_0, b = a_1$, *for*

$$\alpha_h = \frac{1}{h!} \int_{-\infty}^{\infty} m(f) H_h(f) \frac{e^{-f^2/2}}{\sqrt{2\pi}} df = \mathbb{E}(m(f) H_h(f)), \;\; h = 0, 1, \ldots .$$

Remark 6.1.1. This result demonstrates the roles of the alpha and beta SDFs, respectively: the latter captures the linear part of the model-implied nonlinear SDF whereas the former captures the remaining, nonlinear, component, through the (possibly infinite) set of missing factors, given by $f_{t+1}^2, f_{t+1}^3, \ldots$. In fact, the first few Hermite polynomials are:

$$H_0(x) = 1; \; H_1(x) = x; \; H_2(x) = x^2 - 1; \; H_3(x) = x^3 - 3x; \;\; \cdots .$$

Therefore m_{t+1}^{α} is a (infinite-order) polynomial in the state variable f_{t+1} (skipping the terms of order zero and one). Given that it represents a convergence series, it can be approximated by the finite sum $m_{t+1}^{\alpha J} = \sum_{h=2}^{J} \alpha_h H_h(f_{t+1})$, which is makes it computationally easier to be handle and, at the same time, one can make the approximation error arbitrarily small by taking J large enough.

Remark 6.1.2. Note that no differentiability assumption is required on the (nonlinear) SDF. This is relevant when applying our result to highly non-smooth SDF. However, it turns out that if $m(.)$ is n-th order differentiable, then one obtains a much simplified expression for the m_{t+1}^{α} coefficients α_h (see Hannan (1970)[Chapter II.7, p.86])

$$\alpha_h = \frac{1}{h!} \int_{-\infty}^{\infty} m^{(h)}(f) \frac{e^{-f^2/2}}{\sqrt{2\pi}} df,$$

setting $m^{(h)}(f) = d^h m(f)/df^h$.

Remark 6.1.3. The dependence structure for the SDF m_{t+1} follows. In particular, one obtains (see Hannan (1970, Chapter II.7, p.83)):

$$cov(m_t, m_{t+u}) = \alpha_1^2\gamma(u) + \sum_{h=2}^{\infty} h!\alpha_h^2\gamma^h(u) \text{ for any lag } u = 0, \pm 1, \pm 2, \cdots,$$

setting by $\gamma(u) = cov(f_t, f_{t+u})$ the autocovariance function of the state variable. Moreover, one can also obtain the spectral density of the SDF[20]

$$p_m(\lambda) = \alpha_1^2 p_f(\lambda) + \sum_{h=2}^{\infty} h!\alpha_h^2 p_f^{*h}(\lambda) \text{ for any frequency } -\pi \leq \lambda \leq \pi,$$

where $p_f^{*h}(\lambda)$ is the h-fold convolution of the spectral density of the state variable f_t.[21] The spectral density decomposition allows to quantify the importance of the linear term, as well as of the various nonlinear terms, in terms of the cyclical characteristics of the SDF.

Time-variation of conditional moments is allowed for (see, for instance, Aït-Sahalia (2002) for another example of a conditional Hermite expansion). Assume that $\mathbb{E}_t(f_{t+1}) = \mu_{f,t}$ and $var_t(f_{t+1}) = \sigma_{f,t}^2$. Then $f_{t+1} = \sigma_{f,t}\tilde{f}_{t+1} + \mu_{f,t}$, for a standard normal iid $\tilde{f}_{t+1}$. The previous result then applies with respect to the function $\tilde{m}_t(f) = m(\sigma_{f,t}f + \mu_{f,t})$, yielding:

$$m_{t+1}^{\beta} = \mu_{t,m} + b_t\left(\frac{f_{t+1} - \mu_{f,t}}{\sigma_{f,t}}\right),$$

$$m_{t+1}^{\alpha} = \sum_{h=2}^{\infty} \alpha_{t,h} H_h\left(\frac{f_{t+1} - \mu_{f,t}}{\sigma_{f,t}}\right),$$

$$\text{for} \quad \alpha_{t,h} = \frac{1}{h!}\int_{-\infty}^{\infty} \tilde{m}_t(f)H_h(f)\frac{e^{-f^2/2}}{\sqrt{2\pi}}df$$

$$= \mathbb{E}(\tilde{m}_t(f)H_h(f)) = \mathbb{E}_t\left(m(f_{t+1})H_h\left(\frac{f_{t+1} - \mu_{f,t}}{\sigma_{f,t}}\right)\right), \quad h = 0, 1, \ldots,$$

setting $\mu_{t,m} = \alpha_{t,0}$ and $b_t = \alpha_{t,1}$.

A multivariate extension of our result can be obtained as follows (see Ait-Sahalia (2008) for a multivariate Hermite expansion for the log-likelihood function of multivariate diffusions

[20] The spectral density for a covariance stationary stochastic process is formally the Fourier transform of its autocovariance function and characterizes the dynamic properties of the stochastic process across frequencies.

[21] For example, when $h = 2$ one obtains $p_f^{*h}(\lambda) = \int_{-\infty}^{\infty} p_f(\mu)p_f(\lambda - \mu)d\mu$.

sampled at discrete time intervals). Define the multivariate Hermite polynomials for an S-dimensional vector $\mathbf{x}$:

$$H_{\mathbf{h}}(\mathbf{x}) = (-1)^{tr(\mathbf{h})}\phi^{-1}(\mathbf{x})\frac{\partial^{tr(\mathbf{h})}\phi(\mathbf{x})}{\partial x_1^{h_1}\cdots\partial x_S^{h_S}} \text{ for every vector } \mathbf{h} = (h_1,\ldots,h_S)' \in N^S,$$

where $tr(\mathbf{h}) = h_1 + \cdots h_S$, $\phi(\mathbf{x}) = (-2\pi)^{S/2}e^{-\mathbf{x}'\mathbf{x}/2}$ and $N = \{0,1,2,\cdots\}$. Then, when the state vector $\mathbf{f}_{t+1}$ is multivariate normal with conditional mean $\boldsymbol{\mu}_{f,t}$ and covariance $\boldsymbol{\Sigma}_{f,t}$,

$$m_{t+1}^{\beta} = \mu_{t,m} + \mathbf{b}_t'\left(\boldsymbol{\Sigma}_{f,t}^{-1/2}(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{f,t})\right),$$

$$m_{t+1}^{\alpha} = \sum_{k=2}^{\infty}\left(\sum_{\mathbf{h}=(h_1,\cdots,h_S)' \text{ such that } tr(\mathbf{h})=k} \alpha_{t,\mathbf{h}} H_{\mathbf{h}}\left(\boldsymbol{\Sigma}_{f,t}^{-1/2}(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{f,t})\right)\right),$$

with

$$\alpha_{t,\mathbf{h}} = \frac{1}{h_1!\cdots h_S!} = \mathbb{E}_t\left(m(\mathbf{f}_{t+1})H_{\mathbf{h}}\left(\boldsymbol{\Sigma}_{f,t}^{-1/2}(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{f,t})\right)\right) \quad \text{for every } \mathbf{h} = (h_1,\cdots,h_S)' \in N^S,$$

setting $\mu_{t,m} = \alpha_{t,0} = \mathbb{E}_t(m(\mathbf{f}_{t+1}))$, $\mathbf{b}_t = (\alpha_{t,(1\ldots0)}, \alpha_{t,(0,1,0\ldots0)}, \cdots, \alpha_{t,(0\ldots1)})'$ setting

$$\alpha_{t,\underbrace{(0\ldots1\ldots0)}_{\text{in the s-th position}}} = \mathbb{E}_t\left[m(\mathbf{f}_{t+1})H_{(0\ldots1\ldots0)}\left(\boldsymbol{\Sigma}_{f,t}^{-1/2}(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{f,t})\right)\right] = \mathbb{E}_t\left(m(\mathbf{f}_{t+1})\tilde{f}_{st}\right) \text{ for every } 1 \leq s \leq S,$$

setting $\tilde{\mathbf{f}}_{t+1} = (\tilde{f}_{1t+1}\ldots\tilde{f}_{st+1}\ldots\tilde{f}_{St+1})' = \left(\boldsymbol{\Sigma}_{f,t}^{-1/2}(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{f,t})\right)$, given $H_{(0\ldots1\ldots0)}(\mathbf{x}) = -\phi^{-1}(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial x_s} = x_s$, where $tr(0\ldots1\ldots0) = 1$, for every $1 \leq s \leq S$.

7 Analogies to Related Work on the SDF

We now clarify the analogies of our approach with the literature. We start with the finite-N case, and then discuss the large-N case.

7.1 Analogy to Hansen and Jagannathan (1997)

Hansen and Jagannathan (1997) consider the problem in which y_{t+1} is the possibly misspecified SDF that has been adopted, while m_{t+1} is an admissible but unknown SDF. They wish

to solve the following optimization problem, leading to the so-called Hansen-Jagannathan distance δ_t^{HJ}:

$$\delta_t^{HJ} = \min_{m_{t+1}} \Big(\mathbb{E}_t[y_{t+1} - m_{t+1}]^2 \Big)^{1/2},$$

$$\text{such that } \ \mathbb{E}_t(m_{t+1}\mathbf{R}^e_{N,t+1}) = \mathbf{0}_N \ \text{ and } \ \mathbb{E}_t(m_{t+1}) = \mathbb{E}_t(y_{t+1}).$$

The first constraint says that m_{t+1} is admissible and the second constraint says that we believe that y_{t+1} has the correct mean. This, for instance, follows if a risk-free asset is traded, implying $\mathbb{E}_t(y_{t+1}) = R_{ft}^{-1}$.

Hansen and Jagannathan (1997) show that the solution to the above problem is

$$m_{t+1}^{HJ} = y_{t+1} - \mathbf{c}_t' \Big[\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1}) \Big],$$

where $\mathbf{c}_t = \mathbf{V}_{N,t}^{-1} \mathbb{E}_t(y_{t+1}\mathbf{R}^e_{N,t+1})$. Therefore, like us, Hansen and Jagannathan (1997), provide a *linear* adjustment to the possibly misspecified SDF.[22] Without the (semiparametric) APT assumptions, the expression for the correction required to make y_{t+1} admissible, is extremely difficult to estimate, unless N is small and T is large, because the correction term is *nonparametric*: it requires one to compute $\mathbf{V}_{N,t}$ and the, possibly wrong, prices $\mathbb{E}_t(y_{t+1}\mathbf{R}^e_{N,t+1})$, which in turn will depend also on the parameters determining expected returns, notably another quantity very difficult to estimate accurately regardless of the size of N. This was know to Hansen and Jagannathan (1997), whose main objective was to develop an asymptotic distribution theory for δ_t^{HJ}, for any candidate SDF y_{t+1}.

Importantly, by evaluating the above expression under our APT assumptions, we can show that

$$\mathbf{c}_t' \Big[\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1}) \Big] = -m_{t+1}^{\alpha *},$$

where the Hansen-Jagannathan distance satisfies the APT bound in (1):

$$\delta_t^{HJ} = \Big(\mathbb{E}_t[m_{t+1}^{\alpha *}]^2 \Big)^{1/2} = \frac{\big(\boldsymbol{\alpha}_{N,t}' \mathbf{V}_{N,t}^{-1} \boldsymbol{\alpha}_{N,t} \big)^{1/2}}{R_{ft}} \leq \frac{\big(\boldsymbol{\alpha}_{N,t}' \boldsymbol{\Sigma}_{N,t}^{-1} \boldsymbol{\alpha}_{N,t} \big)^{1/2}}{R_{ft}} \leq \delta_{\text{apt}}.$$

[22] Hansen and Jagannathan (1997) demonstrate how construct the correction term also when ensuring non-negativity of the so-obtained admissible SDF, by means of option pricing mathematics.

7.2 Analogy to Ghosh, Julliard, and Taylor (2017)

Ghosh, Julliard, and Taylor (2017) show how to construct an admissible SDF m_{t+1}^{GJT} starting from a possibly misspecified SDF y_{t+1}:

$$m_{t+1}^{GJT} = y_{t+1}\psi_{t+1},$$

where the *multiplicative* term ψ_{t+1} ensures that m_{t+1}^{GJT} is admissible. In particular, Ghosh, Julliard, and Taylor (2017) show that (up to a positive constant scale factor):[23]

$$\psi_t = \frac{e^{y_t \boldsymbol{\gamma}' \mathbf{R}_{N,t}^e}}{\sum_{s=1}^{T} e^{y_s \boldsymbol{\gamma}' \mathbf{R}_{N,s}^e}},$$

for a set of coefficients $\boldsymbol{\gamma} = \text{argmin}_{\mathbf{c}} \frac{1}{T}\sum_{s=1}^{T} e^{y_s \mathbf{c}' \mathbf{R}_{N,s}^e}$. This correction guarantees non-negativity of the admissible SDF. Moreover, just like Hansen and Jagannathan (1997), the correction is nonparametric because its implementation depends on the sample moments of excess returns $\mathbf{R}_{N,t}^e$, as can be seen by expanding $e^{y_s \mathbf{c}' \mathbf{R}_{N,s}^e}$ around the zero vector $\mathbf{0}_N$.

7.3 Analogy to Kozak, Nagel, and Santosh (2018)

Kozak, Nagel, and Santosh (2018) show that the SDF is spanned by a small number of the dominant PCA estimates of the risk factors, regardless of whether the underlying asset-pricing model is behavioral or rational. Because they estimate the latent risk factors by PCA they automatically solve the problem of missing pervasive factors. A special case of our approach also allows for having latent factors that entirely span the SDF. In particular, for the case where $\mathbf{a}_N = \mathbf{0}_N$, there are no observed factor ($K = 0$), and there might exist some latent factors ($p \geq 0$), we recover from Theorem 4.7 the admissible SDF

$$m_{t+1}^{*} = \mu_{m,t} - \mu_{m,t}\boldsymbol{\alpha}_N' \mathbf{V}_N^{-1}(\mathbf{R}_N^e - \mathbb{E}_t(\mathbf{R}_N^e)) \to_p \mu_{m,t} - \mu_{m,t}\boldsymbol{\lambda}_{miss,t}'(\mathbf{f}_{miss,t+1} - \mathbb{E}_t(\mathbf{f}_{miss,t+1})),$$

where $\mathbf{f}_{miss,t}$ denotes the vector of missing pervasive factors arising when $p \geq 1$. Note that our methodology allows one to identify m_{t+1}^{*} *without* having to first estimate the latent factors, even though the latent factors can be estimated if needed. More importantly, our approach allows one to include in the SDF observed risk factors (such as the market factor or

[23] Ghosh, Julliard, and Taylor (2017) show that other formulations for the correction term ψ_t exist, depending on how the Kullback-Leibler Information Criterion is formulated.

the Fama-French factors) as well as deviations from exact pricing arising from firm-specific characteristics; that is, $\mathbf{a}_N \neq \mathbf{0}_N$

8 Estimation of the Extended APT

In this section, we explain how to estimate the extended APT model of returns, based on the (pseudo) Gaussian maximum likelihood (ML) estimation principle. The Gaussian ML estimator is a natural estimator for our model when the first two moments of asset returns are specified correctly, although distributional assumptions (such as normality) are not required, except for efficiency; hence, the use of pseudo ML. Our ML estimator allows to impose the APT no-arbitrage constraint in a very natural way, leading to identification of model parameters. Moreover, it permits to disentangle the effect of the large pricing errors associated with unobserved (missing) factors, allowing for possible correlation between missing and observed factors, from the effect of the small pricing errors, unrelated to common sources of risk. Moreover, the estimation of the risk premia associated with either non-traded and unobserved factors, take the form of the classical GLS two-pass estimator. Finally, our ML estimator easily permits to handle time-varying parameters by means of state-variables.[24]

For simplicity let us assume that *all* conditional moments are constant, and then we discuss how to handle time-variation. Moreover, assume that the number of missing factors p is known.

Assume the following general form of the extended APT, that includes observed, traded and non-traded, and latent factors, as well as idiosyncratic pricing errors:

$$\mathbf{R}^e_{N,t+1} = \mathbf{a}_N + \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{B}_{1N}(\boldsymbol{\lambda}_1 + \mathbf{f}_{1t+1} - \mathbb{E}(\mathbf{f}_{1t+1})) + \mathbf{B}_{2N}\mathbf{f}^e_{2t+1} + \boldsymbol{\varepsilon}_{N,t+1}, \quad \text{with}$$

$$\boldsymbol{\varepsilon}_{N,t+1} = \mathbf{A}_N(\mathbf{f}_{miss,t+1} - \mathbb{E}(\mathbf{f}_{miss,t+1})) + \boldsymbol{\eta}_{N,t+1},$$

[24] An alternative, popular, approach to estimate the SDF of factor asset pricing models is by means of GMM. In particular, there are $2N$ moment conditions, which enable us to estimate the $N+K$ parameters $\mathbf{b}_t$ and $\mathbf{c}_{N,t}$. However,the *structural* parameters of the extended APT, even for the simplest case when $p = 0$, $K = 1$ and $\mathbf{C}_{N,t} = \sigma^2\mathbf{I}_N$, cannot be estimated, unless suitable instruments are used to formulate *conditional* moment conditions, because $(\sigma^2, \mathbf{a}'_{N,t}, \mathbf{B}'_{N,t}, \boldsymbol{\lambda}'_t, vech'(\boldsymbol{\Omega}_t)'$ totals $2N + K^2 + K/2 + 1$ parameters. Moreover, the GMM estimator does not allow one to disentangle the differential effect of small pricing errors, large pricing errors (latent factors), and observed risk factors, because it will only estimate the (reduced-form) parameters of the SDF ($\mathbf{b}_t$ and $\mathbf{c}_{N,t}$) and not the (structural) parameters governing the extended APT.

where the unobserved innovation $\boldsymbol{\eta}_{N,t+1}$ has mean zero, covariance $\mathbf{C}_N$ and is uncorrelated with all the common factors, $\mathbf{f}_{1t+1}$ is a $K_1 \times 1$ vector of observed non-traded factors, with corresponding risk premia $\boldsymbol{\lambda}_1$, $\mathbf{f}^e_{2t+1}$ is a $K_2 \times 1$ vector of observed traded factors, expressed as an excess portfolio return, with corresponding risk premia $\boldsymbol{\lambda}_2 = \mathbb{E}(\mathbf{f}^e_{2t+1})$ and $\mathbf{f}_{miss,t+1}$ denotes a $p \times 1$ vector of latent factors. Moreover, the first and second moment of the assets' excess returns satisfy:

$$\mathbb{E}(\mathbf{R}^e_{N,t+1}) = \mathbf{a}_N + \mathbf{A}_N\boldsymbol{\lambda}_{miss} + \mathbf{B}_{1N}\boldsymbol{\lambda}_1 + \mathbf{B}_{2N}\boldsymbol{\lambda}_2,$$

$$\text{var}(\mathbf{R}^e_{N,t+1}) = (\mathbf{A}_N, \mathbf{B}_N) \begin{pmatrix} \mathbf{I}_p & \mathbf{Q} \\ \mathbf{Q}' & \boldsymbol{\Omega} \end{pmatrix} (\mathbf{A}_N, \mathbf{B}_N)' + \mathbf{C}_N,$$

where we allow for the observed and latent factors to be correlated, that is $\text{cov}(\mathbf{f}_{m,t+1}, \mathbf{f}'_{t+1}) = \mathbf{Q}$ for a possibly non-zero matrix $\mathbf{Q}$, where $\boldsymbol{\Omega} = \text{var}(\mathbf{f}_{t+1})$ is the covariance matrix corresponding to the $K = K + 1 + K_2$ observed factors $\mathbf{f}_{t+1} = (\mathbf{f}'_{1t+1}, \mathbf{f}^{e\prime}_{2t+1})'$ with mean $(\boldsymbol{\mu}'_1, \boldsymbol{\lambda}'_2)' = \mathbb{E}(\mathbf{f}_{t+1})$ and risk premia $(\boldsymbol{\lambda}'_1, \boldsymbol{\lambda}'_2)'$, where by tradability of the $\mathbf{f}^e_{2t+1}$ one obtains $\mathbb{E}\mathbf{f}^e_{2t+1} = \boldsymbol{\lambda}_2$. Finally, note that we have assumed that the latent factors are rotated such that their covariance matrix is the identity matrix $\mathbf{I}_p$. This rotation is arbitrary but convenient as it ensures the interpretation of the missing factors risk premia as Sharpe ratios. It turns out that imposing an identification assumption, i.e. a rotation, is necessary too, as discussed below.

The joint Gaussian log-likelihood function of the observables $(\mathbf{R}^{e\prime}_{N,t}, \mathbf{f}'_t)'$ equals:[25]

$$\begin{aligned} L(\breve{\boldsymbol{\theta}}) = &-\frac{1}{2}\log(\det(\breve{\mathbf{A}}_N(\mathbf{I}_p - \breve{\mathbf{Q}}'\breve{\boldsymbol{\Omega}}^{-1}\breve{\mathbf{Q}})\breve{\mathbf{A}}'_N + \breve{\mathbf{C}}_N)) \qquad (18) \\ &-\frac{1}{2T}\sum_{t=0}^{T-1}\left(\mathbf{R}^e_{N,t+1} - \breve{\mathbf{a}}_N - \breve{\mathbf{A}}_N\breve{\boldsymbol{\lambda}}_{miss} - (\breve{\mathbf{B}}_{1N}, \breve{\mathbf{B}}_{2N})\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 + \breve{\boldsymbol{\lambda}}_1 \\ \mathbf{f}_{2t+1} \end{pmatrix} - \breve{\mathbf{A}}_N\breve{\mathbf{Q}}'\breve{\boldsymbol{\Omega}}^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_2 \end{pmatrix}\right) \\ &\times (\breve{\mathbf{A}}_N(\mathbf{I}_p - \breve{\mathbf{Q}}'\breve{\boldsymbol{\Omega}}^{-1}\breve{\mathbf{Q}})\breve{\mathbf{A}}'_N + \breve{\mathbf{C}}_N)^{-1} \\ &\times \left(\mathbf{R}^e_{N,t+1} - \breve{\mathbf{a}}_N - \breve{\mathbf{A}}_N\breve{\boldsymbol{\lambda}}_{miss} - (\breve{\mathbf{B}}_{1N}, \breve{\mathbf{B}}_{2N})\begin{pmatrix} \mathbf{f}_{1t+1} - \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_1 \\ \mathbf{f}_{2t+1} \end{pmatrix} - \breve{\mathbf{A}}_N\breve{\mathbf{Q}}'\boldsymbol{\Omega}^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_2 \end{pmatrix}\right) \\ &-\frac{1}{2}\log(\det(\breve{\boldsymbol{\Omega}})) - \frac{1}{2T}\sum_{t=0}^{T-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_2 \end{pmatrix}'\boldsymbol{\Omega}^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_2 \end{pmatrix}, \end{aligned}$$

[25] Note that $\det(\cdot)$ denotes the determinant, $\text{vec}(\cdot)$ denotes the operator that stacks the columns of a matrix into a single column vector, and $\text{vech}(\cdot)$ denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.

for any generic vector $\breve{\boldsymbol{\theta}}$ that collects all parameters values, where we factorized the joint distribution as the product of a conditional distribution and a marginal distribution.[26] Relaxing the i.i.d. assumption requires specification of time-varying conditional means, conditional variances, and conditional covariances: below we extend it by means of introducing dependence of the conditional first and second-moments from observed state-variables.

We now derive the closed-form expression for the constrained maximum likelihood estimator (henceforth MLC), feasible for some of the parameters, formalizing the crucial role of the APT constraint, which is relevant for practical implementation of our estimation procedure. A formal analysis of the statistical properties of the MLC is relegated to a companion technical paper.[27]

Theorem 8.1 (Parameter estimates of extended APT)**.** *Suppose that the vector of asset returns,* $\mathbf{R}_{N,t}$*, satisfies Assumption 3.1, that* p *is known and that* $\boldsymbol{\Sigma}_{f_2^e f_2^e} - \bar{\mathbf{f}}_2^{\,e}\bar{\mathbf{f}}_2^{\,e\prime}$ *is nonsingular, where* $\boldsymbol{\Sigma}_{f_2^e f_2^e} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{2t}^e\mathbf{f}_{2t}^{e\prime}$ *and* $\bar{\mathbf{f}}_2^{\,e} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{2t}^e$*. Then the penalized-MLE is defined as:*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\breve{\boldsymbol{\theta}}}{\text{argmax}} \;\; L(\breve{\boldsymbol{\theta}}) \quad \textit{subject to} \quad \breve{\mathbf{a}}_N'\breve{\mathbf{C}}_N^{-1}\breve{\mathbf{a}}_N \leq \delta,$$

where $L(\breve{\boldsymbol{\theta}})$ *is defined in* (18)*, and* $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\mathbf{a}}_{N,MLC}', \hat{\boldsymbol{\lambda}}_{miss,MLC}', \hat{\boldsymbol{\lambda}}_{1,MLC}', \hat{\boldsymbol{\lambda}}_{2,MLC}', \hat{\boldsymbol{\mu}}_{1,MLC}', \text{vec}(\hat{\mathbf{A}}_{N,MLC})', \text{vec}(\hat{\mathbf{B}}_{N,MLC})', \text{vech}(\hat{\mathbf{C}}_{N,MLC})', \text{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$.

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

$$\mathbf{D}_N = (\mathbf{A}_N, \mathbf{B}_{1N}), \;\; \boldsymbol{\lambda} = (\boldsymbol{\lambda}_{miss}', \boldsymbol{\lambda}_1')',$$

[26] We decompose the excess returns into its linear projection onto the space spanned by $\mathbf{F} = (\mathbf{f}_1, \cdots, \mathbf{f}_T)'$ as:

$$\mathbf{R}_{N,t}^e = proj(\mathbf{R}_{N,t}^e|\mathbf{1}_T, \mathbf{F}) + (\mathbf{R}_{N,t}^e - proj(\mathbf{R}_{N,t}^e|\mathbf{1}_T, \mathbf{F}),$$

where

$$proj(\mathbf{R}_{N,t}^e|\mathbf{1}_T, \mathbf{F}) = \breve{\mathbf{a}}_N + \breve{\mathbf{A}}_N\breve{\boldsymbol{\lambda}}_{miss} + (\breve{\mathbf{B}}_{1N}, \breve{\mathbf{B}}_{2N})\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 + \breve{\boldsymbol{\lambda}}_1 \\ \mathbf{f}_{2t+1} \end{pmatrix} + \breve{\mathbf{A}}_N\breve{\mathbf{Q}}'\breve{\boldsymbol{\Omega}}^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_1 \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_2 \end{pmatrix},$$

and

$$var((\mathbf{R}_{N,t}^e - proj(\mathbf{R}_{N,t}^e|\mathbf{1}_T, \mathbf{F})) = (\breve{\mathbf{A}}_N(\mathbf{I}_p - \breve{\mathbf{Q}}'\breve{\boldsymbol{\Omega}}^{-1}\breve{\mathbf{Q}})\breve{\mathbf{A}}_N' + \breve{\mathbf{C}}_N).$$

When the $(\mathbf{R}_{N,t}^e, \mathbf{f}_t)$ are jointly Gaussian the linear projections always coincide with the conditional first moment so we use them interchangeably, aligned with the PMLE principle.

[27] The of case N fixed and large T is standard and one can rely on existing results whereas the cases when N diverges, with either T fixed or diverging, are not-standard and require a separate methodological analysis.

then

$$\text{vec}(\hat{\mathbf{B}}_{2N,MLC}) = \Big((\boldsymbol{\Sigma}_{f_2^e f_2^e} \otimes \mathbf{I}) - (\bar{\mathbf{f}}_2^e \bar{\mathbf{f}}_2^{e\prime} \otimes (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N))\Big)^{-1} \text{vec}\Big(\boldsymbol{\Sigma}_{hf_2^e} - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)\bar{\mathbf{h}}_N \bar{\mathbf{f}}_2^{e\prime}\Big), \tag{19}$$

$$\hat{\boldsymbol{\lambda}}_{MLC} = (\hat{\mathbf{D}}'_{N,MLC}\hat{\boldsymbol{\Sigma}}^{-1}_{N,MLC}\hat{\mathbf{D}}_{N,MLC})^{-1}\hat{\mathbf{D}}'_{N,MLC}\hat{\boldsymbol{\Sigma}}^{-1}_{N,MLC}\Big(\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,MLC}\,\bar{\mathbf{f}}_2^e\Big),$$

$$\hat{\mathbf{a}}_{N,MLC} = \frac{1}{\hat{\kappa}+1}\Big(\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,MLC}\bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,MLC}\hat{\boldsymbol{\lambda}}_{MLC}\Big), \tag{20}$$

where $\hat{\boldsymbol{\Sigma}}_{N,MLC} = \hat{\mathbf{A}}_{N,MLC}\hat{\mathbf{A}}'_{N,MLC} + \hat{\mathbf{C}}_{N,MLC}$, $\boldsymbol{\Sigma}_{hf_2^e} = \frac{1}{T}\sum_{t=1}^T \mathbf{h}_t\mathbf{f}_{2t}^{e\prime}$, $\bar{\mathbf{h}}_N = \frac{1}{T}\sum_{t=1}^T \mathbf{h}_t$ *with* $\mathbf{h}_t = \mathbf{R}^e_{N,t} - \hat{\mathbf{B}}_{1N,MLC}(\mathbf{f}_{1t} - \bar{\mathbf{f}}_{1t})$, *and*

$$\mathbf{G}_N = \frac{1}{(\hat{\kappa}+1)}\mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)}\hat{\mathbf{D}}_{N,MLC}(\hat{\mathbf{D}}'_{N,MLC}\hat{\boldsymbol{\Sigma}}^{-1}_{N,MLC}\hat{\mathbf{D}}_{N,MLC})^{-1}\hat{\mathbf{D}}'_{N,MLC}\hat{\boldsymbol{\Sigma}}^{-1}_{N,MLC}.$$

Note that $\hat{\mathbf{D}}_{N,MLC} = (\hat{\mathbf{A}}_{N,MLC}, \hat{\mathbf{B}}_{1N,MLC})$ *and* $\hat{\mathbf{C}}_{N,MLC}$ *do not admit a closed-form solution and* $(\hat{\boldsymbol{\mu}}'_{1,MLC}, \hat{\boldsymbol{\lambda}}'_{2,MLC})'$ *and* $\hat{\boldsymbol{\Omega}}_{MLC}$ *coincide with the sample mean and sample covariance of the observed factors* $\mathbf{f}_t = (\mathbf{f}'_{1t}, \mathbf{f}^{e\prime}_{2t})'$.

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

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \bar{\mathbf{R}}^e_N - \hat{\mathbf{B}}_{2N,MLC}\bar{\mathbf{f}}_2^e,$$

and the expression for $\text{vec}(\hat{\mathbf{B}}_{2N,MLC})$ *can be obtained by setting* $\hat{\kappa} = 0$ *in the terms that appear in* (19). *The expressions for* $(\hat{\boldsymbol{\mu}}'_{1,MLC}, \hat{\boldsymbol{\lambda}}'_{2,MLC})'$ *and* $\hat{\boldsymbol{\Omega}}_{MLC}$ *are unchanged, and, as for case (i), the expressions for the estimators of* $\hat{\mathbf{D}}_{N,MLC}$ *and* $\hat{\mathbf{C}}_{N,MLC}$ *do not admit a closed-form solution.*

Although our estimation procedure is essentially standard, being based on the ML principle, it raises many delicate issues, especially with respect to its practical implementation, which we now discuss in detail.

8.1 Estimation Strategy

For practical implementation of the MLE, one needs to implement a procedure that also entails to select the correct number of missing factors p. We propose the following strategy:

(i) Estimate the model *without* the APT constraint, where we indicate the estimates as $\hat{\boldsymbol{\theta}}_{MLE}$. When no missing factors are assumed, that is $p = 0$, this coincides with the OLS estimator unless restrictions are imposed on $\mathbf{C}_N$.

(ii) Estimate p based on an analysis of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{N,MLE} = \hat{\mathbf{A}}_{N,MLE}(\mathbf{I}_p - \hat{\mathbf{Q}}'_{MLE}\hat{\boldsymbol{\Omega}}^{-1}_{MLE}\hat{\mathbf{Q}}_{MLE})\hat{\mathbf{A}}'_{N,MLE} + \hat{\mathbf{C}}_{N,MLE}$. This can be done in various was, using either a statistical or an economic approach. In particular, one can analyze the relative magnitude of the eigenvalues associated with $\hat{\boldsymbol{\Sigma}}_{N,MLE}$ or, more formally, applying the criterion of Gagliardini, Ossola, and Scaillet (2019). Alternatively, given the ultimate scope of constructing an admissible SDF, one can select the p that corresponds to the minimized empirical Hansen-Jagannathan distance associated with the estimated SDF, defined as:[28]

$$HJ = \sqrt{T}\left((\frac{1}{T}\sum_{t=1}^{T}\hat{m}_t\mathbf{X}_t - \mathbf{p})'(\frac{1}{T}\sum^{T}\mathbf{X}_t\mathbf{X}'_t)^{-1}(\frac{1}{T}\sum_{t=1}^{T}\hat{m}_t\mathbf{X}_t - \mathbf{p})\right)^{\frac{1}{2}},$$

where $\hat{m}_t = m_t(\hat{\boldsymbol{\theta}}_{MLC})$ denotes the corrected SDF (either based on m_t^* or m_t^{*+}), estimated with our MLC estimator, and $\mathbf{X}_t = (R_f, \mathbf{R}^{e\prime}_{N,t})'$ is the vector of payoffs with prices $\mathbf{p} = (1, 0, ..., 0)'$.

(iii) Having selected p from the previous step, re-estimate the model using the MLC estimator described in Theorem 8.1. Notice that the APT theory is silent on δ. As various δ lead to different parameter estimates, and thus to different estimated SDF, we propose to identify δ by minimizing the Hansen-Jagannathan distance corresponding to the estimated, corrected, SDF. One can show that the Hansen-Jagannathan distance, based on the corrected estimated SDF, is small when true and parameter estimates are close one another. Therefore, an excessively small δ will be as harmful as a big

[28] We adopt, for simplicity, the first HJ distance, which ignores the non-negativity constraint. The first and second HJ distances developed by Hansen and Jagannathan (1991, 1997) measure specification errors of SDF models by least-squares distances between an SDF model and the set of admissible SDFs that can correctly price a set of test assets. The first HJ distance considers the set of all admissible SDFs, which we denote as M. The second HJ distance considers only the smaller set of strictly positive admissible SDFs . The positivity constraint of the second HJ distance guarantees the admissible SDFs to be arbitrage-free and is important for pricing derivatives associated with the test assets. Hansen and Jagannathan (1997) show that, while the first HJ distance represents the maximum pricing error of a portfolio of the test assets with a unit norm, the second represents the minimax bound of the pricing errors of a portfolio of both the test assets and their related derivatives with a unit norm. The second HJ distance represents a more stringent criterion for evaluating asset pricing models and is generally larger than the first.

one, because it will leads to significant estimation error. In practice, this entails doing a grid search over several values of δ from 0 to $\delta_{max} = \hat{\mathbf{a}}'_{N,MLE}\hat{\mathbf{C}}_{N,MLE}\hat{\mathbf{a}}_{N,MLE}$, which is the value corresponding to the (unconstrained) MLE derived in step (i). In fact, for any $\delta \geq \delta_{max}$, the constrained and unconstrained estimator will coincide, as the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} = 0$.

8.2 Sparsity Parameterizations for $\mathbf{C}_N$

The theoretical predictions of the extended APT require N to be large. For instance, the alpha SDF permits one to disentangle the pricing effects of the idiosyncratic pricing errors from the effect of missing pervasive factors, precisely when N diverges. At the same time, the rich cross-sectional dependence, characterizing asset returns, is captured by the factor structure which depends on a number of parameters of the order $O(N)$. This implies one can tightly parameterize the (purely) idiosyncratic covariance matrix $\mathbf{C}_N$, in particular imposing some sparsity condition, without affecting the flexibility of the APT to explain asset returns. Although the APT (see Theorem 3.1) simply implies that $\mathbf{C}_N$ has bounded eigenvalues, uniformly in N, this does not necessarily rule out that $\mathbf{C}_N$ is governed by $O(N^2)$ parameters. Therefore, without further restriction on the form of $\mathbf{C}_N$, there would not be computational advantages of our approach over the nonparametric approaches, such as Hansen and Jagannathan (1991, 1997) and Ghosh, Julliard, and Taylor (2017), which leave the form of the first- and second-moments of returns completely unspecified. For this reason, for practical estimation, we advocate to parameterize $\mathbf{C}_N = \mathbf{C}(\boldsymbol{c}_N)$, for some vector of parameters $\boldsymbol{c}_N$ of order $O(N)$ and a parametric function $\mathbf{C}(\cdot)$. Important special cases are $\mathbf{C}_N = \sigma^2\mathbf{I}_N$ implying $c_N = \sigma^2$, that is when the idiosyncratic innovation is cross-sectionally uncorrelated and homoskedastic, or $\mathbf{C}_N = diag(\sigma_1^2, \cdots, \sigma_N^2)$ implying $\boldsymbol{c}_N = (\sigma_1^2, \cdots, \sigma_N^2)'$, that is, when the idiosyncratic innovation is cross-sectionally uncorrelated and heteroskedastic; of course, other sparsity assumptions are also possible.

Once suitable parameterizations for $\mathbf{C}_N$ are identified, substantial computational gains can be made for computing the loglikelihood, and its first- and second-derivatives, necessary to compute standard errors analytically, making use of the special structure of the matrix $\breve{\mathbf{\Sigma}}_N = (\breve{\mathbf{A}}_N(\mathbf{I}_p - \breve{\mathbf{Q}}'\breve{\mathbf{\Omega}}^{-1}\breve{\mathbf{Q}})\breve{\mathbf{A}}'_N + \breve{\mathbf{C}}_N)$. Assume for simplicity that $\breve{\mathbf{C}}_N = \breve{\sigma}^2\mathbf{I}_N$. Then, by

the Sherman-Morrison formula

$$\breve{\mathbf{\Sigma}}_N^{-1} = \breve{\sigma}^{-2}\left(\mathbf{I}_N - \breve{\mathbf{A}}_N(\breve{\sigma}^2(\mathbf{I}_p - \breve{\mathbf{Q}}'\breve{\mathbf{\Omega}}^{-1}\breve{\mathbf{Q}})^{-1} + \breve{\mathbf{A}}_N'\breve{\mathbf{A}}_N)^{-1}\breve{\mathbf{A}}_N'\right),$$

implying that one needs to invert a low-dimensional matrix, of size $p \times p$, as opposed to a large-dimensional matrix of size $N \times N$. Moreover, for N large,

$$\breve{\mathbf{G}}_N'\breve{\mathbf{\Sigma}}_N^{-1}\breve{\mathbf{A}}_N \approx \breve{\sigma}^{-2}(\breve{\mathbf{G}}_N'\breve{\mathbf{A}}_N)(\breve{\mathbf{A}}_N'\breve{\mathbf{A}}_N)^{-1}(\mathbf{I}_p - \breve{\mathbf{Q}}'\breve{\mathbf{\Omega}}^{-1}\breve{\mathbf{Q}})^{-1},$$

setting $\breve{\mathbf{G}}_N$ to be equal to either $\breve{\mathbf{a}}_N, \breve{\mathbf{A}}_N$ or $\breve{\mathbf{B}}_N$, leading to substantial computational gains, as there are at least eight terms like $\breve{\mathbf{G}}_N'\breve{\mathbf{\Sigma}}_N^{-1}\breve{\mathbf{A}}_N$ in the log-likelihood expression.

8.3 Identification of the Missing Factors.

Dealing with latent factors necessarily implies that such factors, and their moments, are identified up to an unknown rotation. In turn, this asks for an identification assumption, As explained above, we advocate to consider the identification that leads to a standardization, in terms of unit variances and zero covariances, of the latent (true) factors $\mathbf{f}_{miss,t+1}^{\dagger}$, which moreover still allows for orthogonal rotations as

$$\mathbf{H}var(\mathbf{f}_{miss,t+1}) = \mathbf{H}\mathbf{I}_p\mathbf{H}' = \mathbf{I}_p,$$

for any arbitrary orthogonal matrix $\mathbf{H}$. More generally, this implies that, in practice, every element of $\boldsymbol{\lambda}_{miss}$ is a linear combination of *all* true risk premia corresponding to the true latent factors and, even more importantly, $m_{t+1}^{\alpha*}$ is a linear combination of a constant and of the set of p true missing factors $\mathbf{f}_{miss,t+1}^{\dagger}$, *regardless* of whether the correct number of missing factors p is considered to construct $m_{t+1}^{\alpha*}$.

To further demonstrate the necessity of an identification assumption, consider the simplest version of the extended APT, without any observed factors ($K = 0$) and with $p = 1$ setting $\mathbf{a}_N = \mathbf{0}_N$ (and thus without the APT constraint), $\mathbf{C}_N = \sigma^2\mathbf{I}_N$, without imposing the restrictions $var(\mathbf{f}_{miss,t+1}) = \omega_m = 1$. One can then show that the first-order conditions for $\sigma^2, \lambda_{miss}, \omega_m$ and $\mathbf{A}_N$, setting $\hat{\mathbf{M}} = T^{-1}\sum_{t=1}^{T}(\mathbf{R}_{N,t}^e - \hat{\mathbf{A}}_{N,MLC}\hat{\boldsymbol{\lambda}}_{miss,MLC})(\mathbf{R}_{N,t}^e - \hat{\mathbf{A}}_{N,MLC}\hat{\boldsymbol{\lambda}}_{miss,MLC})'$ are:

$$\text{for } \sigma^2: \quad trace(\hat{\mathbf{\Sigma}}_{N,MLC}^{-1}) = trace(\hat{\mathbf{\Sigma}}_{N,MLC}^{-1}\hat{\mathbf{M}}\hat{\mathbf{\Sigma}}_{N,MLC}^{-1});$$

$$\text{for } \lambda_{miss}: \quad \hat{\lambda}_{miss,\text{MLC}} = \frac{\hat{\mathbf{A}}'_{N,\text{MLC}}\,\hat{\mathbf{\Sigma}}^{-1}_{N,\text{MLC}}\bar{\mathbf{R}}^{e}_{N}}{(\hat{\mathbf{A}}'_{N,\text{MLC}}\,\hat{\mathbf{\Sigma}}^{-1}_{N,\text{MLC}}\,\hat{\mathbf{A}}_{N,\text{MLC}})};$$

$$\text{for } \omega_m: \quad \hat{\omega}_{m,MLC} = \left(\frac{(\hat{\mathbf{A}}'_{N,\text{MLC}}\hat{\mathbf{M}}\hat{\mathbf{A}}_{N,\text{MLC}})}{(\hat{\mathbf{A}}'_{N,\text{MLC}}\hat{\mathbf{A}}_{N,\text{MLC}})^2} - \frac{\hat{\sigma}^2_{MLC}}{(\hat{\mathbf{A}}'_{N,\text{MLC}}\hat{\mathbf{A}}_{N,\text{MLC}})}\right);$$

$$\text{for } \mathbf{A}_N: \quad \hat{\mathbf{A}}_{N,\text{MLC}} = \left(\frac{\hat{\omega}_{m,MLC}}{\hat{\lambda}_{miss,\text{MLC}}}(\hat{\mathbf{M}}^{-1}\hat{\mathbf{\Sigma}}_{N,MLC} - \mathbf{I}_N) - \hat{\lambda}_{miss,\text{MLC}}\mathbf{I}_N\right)^{-1}\bar{\mathbf{R}}^{e}_{N}.$$

Setting $N = 1$, one can immediately see that the first-order conditions for σ^2 and ω_m are identical, but one can derive the ratio λ_m/ω_m, hence our identification strategy to set $\omega_m = 1$.

Notice that for constructing any of the SDF correction terms, such as $m^{\alpha}_{t+1}, m^{\alpha+}_{t+1}$ and their projection-versions, we do not need to estimate the (normalized) latent factors themselves but we only need their loadings and risk premia. In other words, one does need to estimate the latent factors in order to estimate an admissible SDF even if the latter, in population, depends on such latent factors. However, if one is interested in backing out the estimates of the latent factors, having estimated the APT parameters, one can derive such estimates by means of OLS cross-sectional regressions such as:

$$\hat{\mathbf{f}}^{*}_{miss,t+1} = (\hat{\mathbf{A}}'_{N,MLC}\hat{\mathbf{A}}_{N,MLC})^{-1}\hat{\mathbf{A}}'_{N,MLC}\hat{\mathbf{p}}_{t+1,MLC}$$

setting $\hat{\mathbf{p}}_{t+1,MLC} = (\mathbf{R}^{e}_{N,t+1} - \hat{\mathbf{a}}_{N,MLC} - \hat{\mathbf{A}}_{N,MLC}\hat{\boldsymbol{\lambda}}_{miss,MLC} - \hat{\mathbf{B}}_{1N,MLC}(\hat{\boldsymbol{\lambda}}_{1,MLC} + \mathbf{f}_{1t+1} - \hat{\boldsymbol{\mu}}_{1,MLC}) - \hat{\mathbf{B}}_{2N}\mathbf{f}^{e}_{2t+1})$ and where for simplicity we denote the demeaned latent factors as $\mathbf{f}^{*}_{miss,t+1} = \mathbf{f}_{miss,t+1} - \mathbb{E}(\mathbf{f}_{miss,t+1})$. A GLS estimator of the latent factors is also available, such as

$$(\hat{\mathbf{A}}'_{N,MLC}\hat{\mathbf{C}}^{-1}_{N,MLC}\hat{\mathbf{A}}_{N,MLC})^{-1}\hat{\mathbf{A}}'_{N,MLC}\hat{\mathbf{C}}^{-1}_{N,MLC}\hat{\mathbf{p}}_{t+1,MLC},$$

which of course is identical to $\hat{\mathbf{f}}^{*}_{miss,t+1}$ when $\mathbf{C}_N$ is assumed spherical.

8.4 Time-Variation of Conditional Moments

Although various approaches are possible, we advocate to use observe state variables to capture time-variation of risk premia, loadings and factors' conditional expectations. In particular, following the formulation of Gagliardini, Ossola, and Scaillet (2016) and Giglio

and Xiu (2017), assume that there are $K_c \times 1$ common observed state variables, $\mathbf{z}_t$, which includes the unit constant and the observed factors $\mathbf{f}_t = (\mathbf{f}_{1t}', \mathbf{f}_{2t}^{e\prime})'$, and $N \times K_s$ asset-specific state variables, $\mathbf{Z}_{N,t} = (\mathbf{z}_{1,t}, \cdots, \mathbf{z}_{N,t})'$ observed at every period t. Although our formulation is completely general, typical examples of state-variables commonly used are the dividend yield, bond spreads, CAPE with respect to common variates, and firms' characteristics with respect to asset-specific variates.

Then assume the following specifications for the parameters of the (time-varying) extended APT:

$$\boldsymbol{\lambda}_{miss,t} = \mathbf{G}_m \mathbf{z}_t \text{ for a } p \times K_c \text{ matrix } \mathbf{G}_m,$$

$$\boldsymbol{\lambda}_{1,t} = \mathbf{G}_1 \mathbf{z}_t \text{ for a } K_1 \times K_c \text{ matrix } \mathbf{G}_1,$$

$$\mathbb{E}_t(\mathbf{f}_{t+1}) = \begin{pmatrix} \boldsymbol{\mu}_{1,t} \\ \boldsymbol{\lambda}_{2,t} \end{pmatrix} = \mathbf{G}_f \mathbf{z}_t \text{ for a } K \times K_c \text{ matrix } \mathbf{G}_f,$$

$$\mathbf{a}_{N,t} = \begin{pmatrix} \mathbf{z}_{1t}' & 0 & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & 0 & \mathbf{z}_{Nt}' \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_N \end{pmatrix} \text{ for } K_s \times 1 \text{ vectors } \mathbf{a}_i,$$

$$\mathbf{A}_{N,t} = (\mathbf{I}_N \otimes \mathbf{z}_t') \begin{pmatrix} \mathbf{A}_{c1}' \\ \vdots \\ \mathbf{A}_{cN}' \end{pmatrix} + \begin{pmatrix} \mathbf{z}_{1t}' & 0 & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & 0 & \mathbf{z}_{Nt}' \end{pmatrix} \begin{pmatrix} \mathbf{A}_{s1}' \\ \vdots \\ \mathbf{A}_{sN}' \end{pmatrix}$$

for $p \times K_c$ and $p \times K_s$ matricies $\mathbf{A}_{ci}$ and $\mathbf{A}_{si}$,

$$\mathbf{B}_{N,t} = (\mathbf{B}_{1N,t}, \mathbf{B}_{2N,t}) = (\mathbf{I}_N \otimes \mathbf{z}_t') \begin{pmatrix} \mathbf{B}_{c1}' \\ \vdots \\ \mathbf{B}_{cN}' \end{pmatrix} + \begin{pmatrix} \mathbf{z}_{1t}' & 0 & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & 0 & \mathbf{z}_{Nt}' \end{pmatrix} \begin{pmatrix} \mathbf{B}_{s1}' \\ \vdots \\ \mathbf{B}_{sN}' \end{pmatrix}$$

for $K \times K_c$ and $K \times K_s$ matricies $\mathbf{B}_{ci}$ and $\mathbf{B}_{si}$,

and where time-variation of the covariance matrices $\mathbf{Q}_t = \mathbf{Q}(\mathbf{z}_t), \boldsymbol{\Omega}_t = \boldsymbol{\Omega}(\mathbf{z}_t)$ and $\mathbf{C}_{N,t} = \mathbf{C}(\mathbf{c}_N(\mathbf{Z}_t))$ is also permitted, for some given parametric functions $\mathbf{Q}(\cdot), \boldsymbol{\Omega}(\cdot)$ and $\mathbf{c}_N(\cdot)$, such as for example ARCH and GARCH specifications.[29] Finally, inserting the above expressions into the log-likelihood yields a suitable generalization of (18) that permits time-variation.[30]

[29] The formulae for the MLC parameter estimates established in Theorem 8.1 are not valid but can be generalized.

[30] Using the prediction decomposition of the joint density function, $pdf(\mathbf{X}_1, \cdots, \mathbf{X}_T) = pdf(\mathbf{X}_1) pdf(\mathbf{X}_2|\mathbf{X}_1) \cdots pd(\mathbf{X}_T|\mathbf{X}_{T-1}, ..., \mathbf{X}_1)$ for a generic stochastic process $\{\mathbf{X}_t\}$, the joint Gaussian log-

8.5 Roles of the APT Constraint

Imposing the APT constraint on the idiosyncratic pricing errors $\mathbf{a}_N$ serves several important purposes in our estimation strategy. First, it is theoretically justified because, when not imposed, its violation could be lead to arbitrage opportunities. Second, it possibly leads to a more precise estimator of $\mathbf{a}_N$ compared to the *unconstrained* estimator; in particular (20) has the expression of a ridge estimator. Third, it provides exactly the condition required to econometrically identify the extended APT, in the sense that $\boldsymbol{\lambda} = (\boldsymbol{\lambda}'_{miss}, \boldsymbol{\lambda}'_1)'$ and $\boldsymbol{\alpha}_{N,t}$ *cannot* be identified separately unless the APT restriction is imposed.[31] In contrast, when $\hat{\kappa} = 0$, and identification fails, obviously formula (20) continue to hold providing the (classical) interpretation of the estimated $\mathbf{a}_N$ as the empirical residual of a cross-sectional regression, namely the difference between the sample excess returns and the estimated risk premia multiplied by their corresponding loadings. Fourth, when the APT constraint binds, $\hat{\mathbf{a}}_{N,MLC}$ obviously represents the solution to the first-order conditions of a ML problem, and therefore one can derive its standard errors using the ML mathematics; that is, combining the estimated Hessian and covariance matrix of the score. In turn, as an important by-product, this would lead to a test for correct model specification, with respect to the null hypothesis $\mathbf{a}_N = \mathbf{0}_N$.

likelihood function of the observables $(\mathbf{R}^{e\prime}_{N,t}, \mathbf{f}'_t)'$, when time-variation is allowed for, equals

$$
\begin{aligned}
L(\breve{\boldsymbol{\theta}}) = & -\frac{1}{2T}\sum_{t=1}^{T-1}\log(\det(\breve{\mathbf{A}}_{N,t}(\mathbf{I}_p - \breve{\mathbf{Q}}_t{}'\breve{\boldsymbol{\Omega}}_t^{-1}\breve{\mathbf{Q}}_t)\breve{\mathbf{A}}'_{N,t} + \breve{\mathbf{C}}_{N,t})) \\
& - \frac{1}{2T}\sum_{t=1}^{T-1}\left(\mathbf{R}^e_{N,t+1} - \breve{\mathbf{a}}_{N,t} - \breve{\mathbf{A}}_{N,t}\breve{\boldsymbol{\lambda}}_{miss,t} - (\breve{\mathbf{B}}_{1N,t}, \breve{\mathbf{B}}_{2N,t})\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_{1,t} + \breve{\boldsymbol{\lambda}}_{1,t} \\ \mathbf{f}_{2t+1}\end{pmatrix} - \breve{\mathbf{A}}_{N,t}\breve{\mathbf{Q}}'_t\breve{\boldsymbol{\Omega}}_t^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_{1,t} \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_{2t}\end{pmatrix}\right)' \\
& \times (\breve{\mathbf{A}}_{N,t}(\mathbf{I}_p - \breve{\mathbf{Q}}'_t\breve{\boldsymbol{\Omega}}_t^{-1}\breve{\mathbf{Q}}_t)\breve{\mathbf{A}}'_{N,t} + \breve{\mathbf{C}}_{N,t})^{-1} \\
& \times \left(\mathbf{R}^e_{N,t+1} - \breve{\mathbf{a}}_{N,t} - \breve{\mathbf{A}}_{N,t}\breve{\boldsymbol{\lambda}}_{miss,t} - (\breve{\mathbf{B}}_{1N,t}, \breve{\mathbf{B}}_{2N,t})\begin{pmatrix} \mathbf{f}_{1t+1} - \boldsymbol{\mu}_{1,t} + \boldsymbol{\lambda}_{1,t} \\ \mathbf{f}_{2t+1}\end{pmatrix} - \breve{\mathbf{A}}_{N,t}\breve{\mathbf{Q}}'_t\boldsymbol{\Omega}_t^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_{1,t} \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_{2,t}\end{pmatrix}\right) \\
& - \frac{1}{2T}\sum_{t=1}^{T-1}\log(\det(\breve{\boldsymbol{\Omega}}_t)) - \frac{1}{2T}\sum_{t=1}^{T-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_{1,t} \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_{2,t}\end{pmatrix}'\boldsymbol{\Omega}_t^{-1}\begin{pmatrix} \mathbf{f}_{1t+1} - \breve{\boldsymbol{\mu}}_{1,t} \\ \mathbf{f}_{2t+1} - \breve{\boldsymbol{\lambda}}_{2,t}\end{pmatrix},
\end{aligned}
$$

where, unlike the iid case (18), to compute $L(\breve{\boldsymbol{\theta}})$ we skipped the term corresponding to the marginal log-density of $(\mathbf{R}^{e\prime}_{N,1}, \mathbf{f}'_1)'$, that is for time $t = 1$, as it is asymptotically irrelevant.

[31] Incidentally, regarding the MLC estimator for $\boldsymbol{\lambda} = (\boldsymbol{\lambda}'_{miss}, \boldsymbol{\lambda}'_1)'$, that is the risk premia associated with both missing factors and non-traded factors, it is interesting to note that coincide with the GLS CSR estimator. Although this is well-known with respect to non-traded factors, we are the first to establish this result with respect to the risk premia of the missing pervasive factors. In other words, from the point of view of estimation of risk premia, missing factors and observed non-traded factors receive the same treatment from the MLC estimator.

We have expressed the APT constraint as $\mathbf{a}_N'\mathbf{C}_N^{-1}\mathbf{a}_N \leq \delta$ but it is equivalent to express it as $\mathbf{a}_N'\mathbf{a}_N \leq \delta$ (the constant δ can change), as $\mathbf{C}_N$ has bounded eigenvalues. One might wonder whether, in the absence of the $\mathbf{a}_N$ and when missing factors are allowed for, one should still impose the APT constraint in terms of the *large* pricing errors $\boldsymbol{\alpha}_N = \mathbf{A}_N\boldsymbol{\lambda}_{miss}$, such as $\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N \leq \delta$. However, it turns out that this constraint is always satisfied, for some finite δ, and therefore it is not necessary to impose it in the estimation. In fact,

$$\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N \to \boldsymbol{\lambda}_{miss}'\boldsymbol{\lambda}_{miss} \text{ as } N \to \infty.$$

This means that the APT restriction is always satisfied for the case of only missing pervasive factors, for any $\delta \geq \boldsymbol{\lambda}_{miss}'\boldsymbol{\lambda}_{miss}$ as N diverges. In turn, this result stems from recognizing that both $\boldsymbol{\Sigma}_N$ and $\boldsymbol{\alpha}_N$ contain the loadings of the missing factors, $\mathbf{A}_N$, inducing a compensation in $\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N$. It is also interesting to note that this feature, namely that $\mathbf{A}_N$ appears in both the (conditional) mean and covariance matrix of returns, bears an important implication in terms of precision of the MLC estimator, as both the first and second moments of returns contribute to the estimation of the $\mathbf{A}_N$.[32]

9 Simulation Experiment

In this section, we evaluate our theoretical results in a controlled environment using simulated data. In particular, we check whether we can take an SDF based on a misspecified return-generating model and correct the SDF so that it is admissible.

We simulate data according to the Fama and French (2015) model with five factors: Market, Size, Value, Profitability, and Investment. To set the parameters for our model, we estimate the five-factor model on monthly returns from 1963:07 to 2019:02 for $N = 96$ asset portfolios that are formed by sorting stocks based on: (1) Size-BM-Operating Profitability; (2) Size-BM-Investment; and (3) Size-Operating Profitability-Investment. Consequently, the simulated data matches the properties of the empirical returns data. The estimated Sharpe ratio for $\mathbf{a}_N$ is 2.29 and $\sigma^2 = 0.0037$ per year.

[32] MacKinlay and Pástor (2000) firstly pointed out this insight for improving the precision of the estimated $\mathbf{A}_N$ parameters. However, MacKinlay and Pástor (2000) consider a different identification assumption. For $p = 1$, they estimate $\boldsymbol{\alpha}_N$ without distinguishing between $\mathbf{A}_N$ and $\boldsymbol{\lambda}_{miss}$, implying that the contribution of the single missing factor to the return variance equals $\boldsymbol{\alpha}_N\boldsymbol{\alpha}_N'/(\text{SR}^{\text{miss}})^2$, where SR^{miss} is the Sharpe ratio of the missing factor.

Table 1: Pricing errors using true parameter values

This table reports three measures of the pricing error: the GMM J test where the weighting matrix is the matrix of second moments of the pricing errors, the GMM J test where the weighting matrix is the covariance matrix of the pricing errors, and the Hansen-Jagannathan distance.

	GMM (second mom.)	GMM (cov.)	HJ dist.
Panel A: Using both the alpha and beta SDFs			
Statistic	106.551	126.808	170.146
p-value	0.572	0.572	0.406
quantile 95%	140.178	177.476	215.085
Panel B: Using only the beta SDF when $KK = 5, P = 0$			
Statistic	265.838	442.000	437.664
p-value	0.000	0.000	0.000
quantile 95%	122.302	149.763	149.854
Panel C: Using only the beta SDF when $KK = 4, P = 1$			
Statistic	271.596	458.151	468.333
p-value	0.000	0.000	0.000
quantile 95%	121.953	149.240	147.469
Panel D: Using only the beta SDF when $KK = 3, P = 2$			
Statistic	273.930	464.830	481.342
p-value	0.000	0.000	0.000
quantile 95%	120.786	147.496	146.427
Panel E: Using only the beta SDF when $KK = 2, P = 3$			
Statistic	276.649	472.716	482.542
p-value	0.000	0.000	0.000
quantile 95%	119.800	146.028	143.845
Panel F: Using only the beta SDF when $KK = 1, P = 4$			
Statistic	280.646	484.508	486.006
p-value	0.000	0.000	0.000
quantile 95%	119.387	145.416	143.758
Panel G: Using only the beta SDF when $KK = 0, P = 5$			
Statistic	284.645	496.549	492.783
p-value	0.000	0.000	0.000
quantile 95%	117.229	142.226	144.861

We undertake our empirical analysis in two steps. In the first step, we consider the case where the parameter values are known. In the second step, we consider the case where the parameter values are not known and need to be estimated. In each of these two steps, we consider a variety of models. (1) The full model with no misspecification. (2) The model where misspecification arises because $\mathbf{a}_N$ is omitted. (3) The model with omitted $\mathbf{a}_N$ and also missing factors, where the number of missing factors ranges from one to five. In both steps, to evaluate the pricing errors we report three metrics: the Hansen-Jagannathan

distance (HJ), the GMM J statistic with the weighting matrix being the matrix of second moments of the pricing errors, and the J statistic with the weighting matrix being the covariance matrix of the pricing errors.

In Table 1, Panel A gives the pricing errors for the full model with no misspecification. Panels B to G report the pricing errors for misspecified models: in Panel B, we consider the model where only $\mathbf{a}_N$ is omitted, while in Panels C to G, we consider models where in addition to $\mathbf{a}_N$ being omitted also $P = \{1, 2, \ldots, 5\}$ risk factors are missing. The pricing errors for the models considered in Panels B to G, *after* correction for misspecification, is given again by Panel A. That is, when misspecification is corrected in the models considered in Panels B to G, one recovers exactly the true model reported in Panel A. Therefore, for the case where one knows the true parameters of the model, no matter what is the source of misspecification—omitted $\mathbf{a}_N$ and P missing factors—the alpha-SDF is able to fully correct for the misspecification.

Next, in Table 2, we consider the setting where one does not know the true parameter values, and therefore, these need to be estimated. We estimate the model over $T = 1,000$ monthly return observations using maximum-likelihood estimation.[33] When estimating the model, we impose the APT restriction by setting $\delta_{\text{apt}} = 0.44$, which corresponds to the Sharpe ratio of $\mathbf{a}_N$ being 2.29. The six panels in Table 2 are grouped in pairs, with the first panel of each pair reporting the pricing-error statistics for the *corrected* model and the second panel of the pair reporting the statistics for the model based only on the beta SDF. In Panel A, we consider the model where $\mathbf{a}_N$ is omitted. In Panels B to F, we consider models where $\mathbf{a}_N$ is omitted and there are $P = \{1, 2, \ldots, 5\}$ missing factors. Looking at the pricing errors reported in this table, we draw the same conclusion as for Table 1: even for the case where the parameters of the return-generating model need to be estimated, the alpha SDF is remarkably effective in reducing the pricing error and correcting the model for the various sources of misspecification considered in these panels.

Table 3 illustrates how the alpha SDF spans the space of missing pervasive factors (after adjusting for the effect of $\mathbf{a}_N$). In this table, we report the R^2 from regressing the alpha SDF on a number of missing factors. For instance, in the first row of the table the number

[33] The large value for T ensures convergence of the empirical Hansen-Jagannathan distance to its population counterpart but it is not required for our theory.

Table 2: Pricing errors using estimated parameter values

This table reports three measures of the pricing error for the setting where the parameters of the model have to be estimated: the GMM J test where the weighting matrix is the matrix of second moments of the pricing errors, the GMM J test where the weighting matrix is the covariance matrix of the pricing errors, and the Hansen-Jagannathan distance.

	GMM (second mom.)	GMM (cov.)	HJ dist.
Panel A1: Using both the alpha and beta SDFs when $KK = 5, P = 0$			
Statistic	61.441	67.674	79.670
p-value	0.998	0.998	1.000
quantile 95%	141.439	179.503	214.686
Panel A2: Using only the beta SDF when $KK = 5, P = 0$			
Statistic	265.189	440.210	430.984
p-value	0.000	0.000	0.000
quantile 95%	124.488	153.054	152.204
Panel B1: Using both the alpha and beta SDFs when $KK = 4, P = 1$			
Statistic	61.269	67.467	78.898
p-value	1.000	1.000	1.000
quantile 95%	142.533	181.268	218.867
Panel B2: Using only the beta SDF when $KK = 4, P = 1$			
Statistic	274.469	466.386	460.891
p-value	0.000	0.000	0.000
quantile 95%	120.869	147.619	145.566
Panel C1: Using both the alpha and beta SDFs when $KK = 3, P = 2$			
Statistic	116.307	140.871	159.376
p-value	0.326	0.326	0.495
quantile 95%	138.438	174.696	206.284
Panel C2: Using only the beta SDF when $KK = 3, P = 2$			
Statistic	277.420	474.970	478.330
p-value	0.000	0.000	0.000
quantile 95%	120.609	147.232	144.569
Panel D1: Using both the alpha and beta SDFs when $KK = 2, P = 3$			
Statistic	58.697	64.361	76.256
p-value	1.000	1.000	1.000
quantile 95%	140.283	177.645	214.088
Panel D2: Using only the beta SDF when $KK = 2, P = 3$			
Statistic	278.971	479.535	480.059
p-value	0.000	0.000	0.000
quantile 95%	119.370	145.389	142.988
Panel E1: Using both the alpha and beta SDFs when $KK = 1, P = 4$			
Statistic	58.270	63.848	75.804
p-value	1.000	1.000	1.000
quantile 95%	142.861	181.800	219.342
Panel E2: Using only the beta SDF when $KK = 1, P = 4$			
Statistic	283.819	494.042	481.023
p-value	0.000	0.000	0.000
quantile 95%	119.459	145.522	143.488
Panel F1: Using both the alpha and beta SDFs when $KK = 0, P = 5$			
Statistic	111.062	133.249	147.388
p-value	0.452	0.452	0.733
quantile 95%	141.255	179.207	211.817
Panel F2: Using only the beta SDF when $KK = 0, P = 5$			
Statistic	284.644	496.548	492.782
p-value	0.000	0.000	0.000
quantile 95%	116.935	141.793	140.788

Table 3: Spanning the effect of missing pervasive factors

This table measures how the alpha-SDF spans the space of missing pervasive factors (after adjusting for the effect of $\mathbf{a}_N$). The table reports the R^2 from regressing the alpha-SDF on a number of missing factors. In the first row of the table, the number of observed factors is $K = 4$ and the number of missing factors is 1. In the second row, the number of observed factors is $K = 3$ and the number of missing factors is 2.

Experiment	Accounting for missing factors				
	1	**2**	**3**	**4**	**5**
$K = 4$	0.94				
$K = 3$	0.42	0.92			
$K = 2$	0.03	0.31	0.95		
$K = 1$	0.06	0.10	0.34	0.95	
$K = 0$	0.01	0.03	0.18	0.21	0.92

of observed factors is $K = 4$ and the number of missing factors is 1, while in the second row the number of observed factors is $K = 3$ and the number of missing factors is 2. The table shows that as we complete the space of missing factors, the R^2 goes to 1, confirming the result in Theorem 4.8.

10 Empirical Analysis

In this section, we illustrate how our methodology can be used to study stock-return data. The design of our empirical analysis is motivated by the work in Ghosh, Julliard, and Taylor (2017). Just like them, we use quarterly returns on 26 Fama-French portfolios over the postwar period: 1947:Q2 to 2015:Q4. The 26 portfolios we study consist of 6 size and book-to-market portfolios, 10 industry portfolios, and 10 momentum portfolios. We also use quarterly data on per capita real personal consumption expenditures on nondurable goods for a total of 275 quarterly observations.

We consider a setting where the candidate beta SDF is the consumption CAPM; that is, the beta SDF is given by the single-factor ($K = 1$) consumption growth. Then, we estimate the extended-APT factor model and find that the best admissible SDF (the one with the smallest Hansen-Jagannathan distance) is one with 5 missing factors; that is, $P = 5$. For this case, the Hansen-Jagannathan distance is 80 and the optimal $\delta = 0.054$. The mean and standard deviation of the beta SDF, the alpha SDF, and the admissible SDF are given in Table 4.

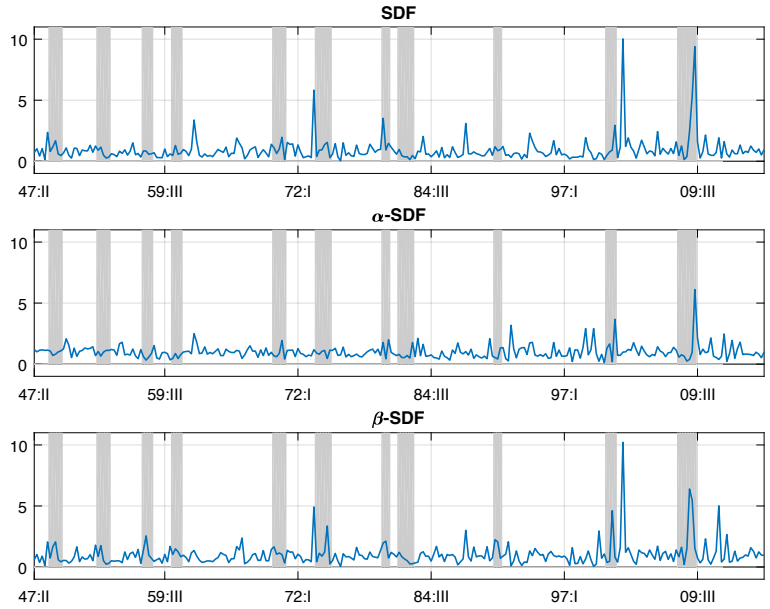
Table 4: Moments of the admissible SDF and its components

This table reports the mean and standard deviation of the α-SDF, the β-SDF, and the admissible SDF.

Quantity	Mean	Std. Dev.
$\log(m^\beta)$	-0.2311	0.7093
$\log(m^\alpha)$	-0.1277	0.5022
$\log(m)$	-0.3588	0.7023
m^β	1.0232	0.9696
m^α	0.9983	0.5827
m	0.9189	1.0116

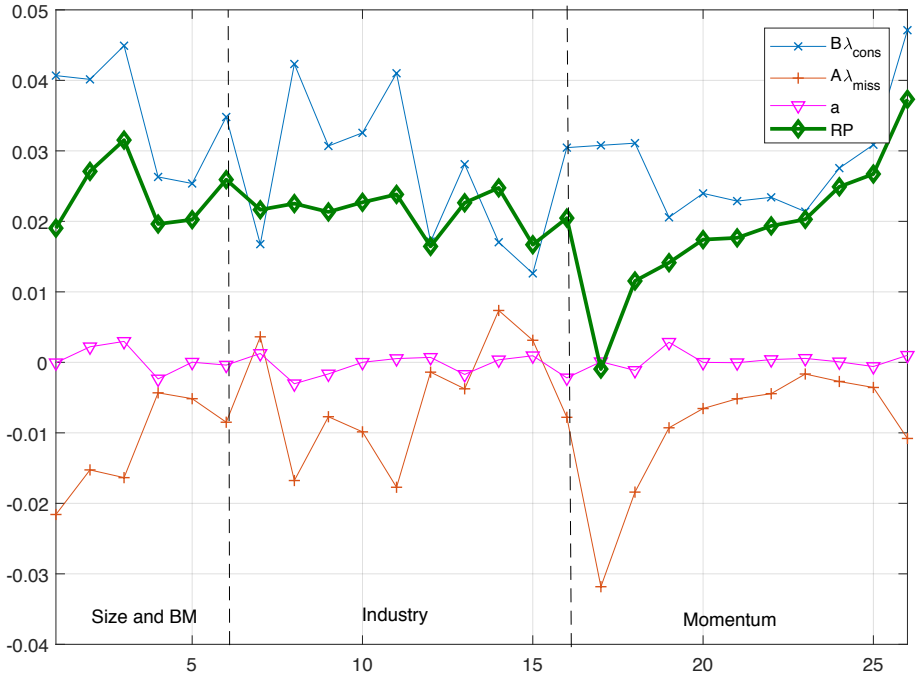
Figure 1: Time series of the admissible SDF and its components

This figure plots the admissible SDF along with the α-SDF and the β-SDFs. The grey bars on the plots indicate NBER recessions.



To get a sense of how the the admissible SDF and its alpha and beta components vary with the business cycle, we estimate these stochastic discount factors and they are plotted in Figure 1. We see that there is considerable variation in the admissible SDF and its components and that this variation has increased over time. We also notice that the level and volatility of the SDF and its components is higher during recessionary periods of the business cycle.

Figure 2: Components of the risk-premia for the 26 portfolios
This figure plots the components of the risk premia for the 26 Fama-French portfolios. The figure includes plots for the (i) the risk premium for consumption growth; (ii) the risk premia for the 5 missing factors; (iii) the risk premium arising from firm-specific (idiosyncratic) risk; and, (iv) the total risk premium.



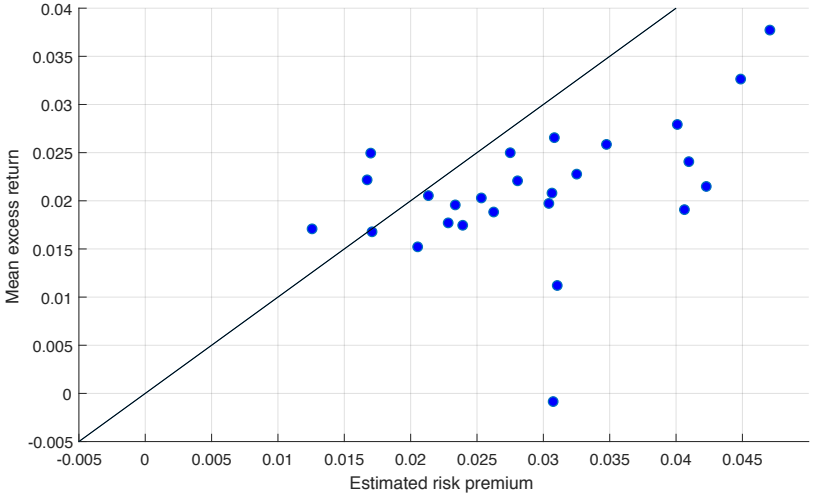
To interpret the risk-premia, we decompose the total risk-premia that we have estimated for the 26 portfolios into the following components: (i) the risk premium for consumption growth (blue line with crosses); (ii) the risk premia for the 5 missing factors (orange line with plus signs); (iii) the risk premium arising from firm-specific (idiosyncratic) risk (pink line with triangles); and, (iv) the total risk premium (bold green line with diamonds). These risk-premia are plotted in Figure 2. We see from the figure that if one were to consider only the risk premium for consumption growth (the blue line), one would over-estimate the true risk premium. The reason for this is that the risk premia for the missing factors are negative for most of the portfolios. The risk premium for idiosyncratic risk are close to zero, but positive for some portfolios and negative for others.

To understand the importance of the risk premia for the 5 missing factors and for idiosyncratic risk, we first plot mean excess returns against just the risk premium for con-

Figure 3: Plot of mean excess returns on risk premia

Panel A in this figure contains a plot of mean excess returns against the risk premium for consumption growth, while Panel B contains a plot of mean excess returns against the total risk premium (that is, the risk premium for consumption growth, the missing factors, and idiosyncratic risk).

Panel A: Plot of mean excess returns on consumption-growth risk premium



Panel B: Plot of mean excess returns on total risk premia

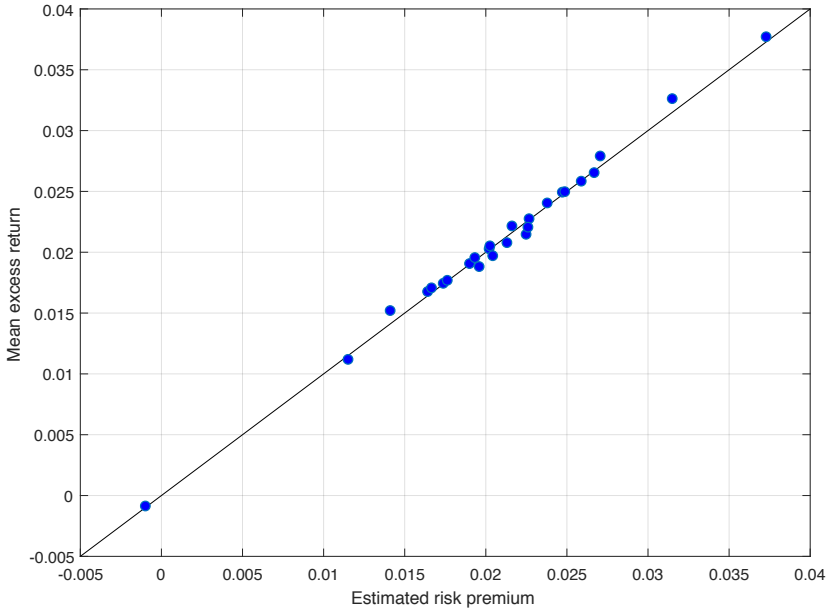


Table 5: Regression of mean excess returns on risk premia

This table reports the results of regressing $\log(m^\alpha)$ on various common factors considered in the empirical asset-pricing literature: market (Mkt.), size (SMB), value (HML), momentum (Mom.), profitability, measured as Robust minus Weak (RMW), investment, measured Conservative minus Aggressive (CMA), and intermediary capital (Interm.). The table reports the coefficient estimate, its t-value, and the *cumulative* R^2 (adjusted) as each new factor is added.

	1947:Q2–2015:Q4 275 qrtly. obs.				1963:Q2 210 qrtly. obs.		1970:Q1 184 qrtly. obs.	
Factor	Mkt.	SMB	HML	Mom.	RMW	CMA	Interm.	Const.
estimate	−1.33	4.88	1.02	−0.74	−5.13	−3.12	0.02	−0.01
t-value	−2.63	10.37	1.66	−2.51	−9.33	−3.66	0.06	−0.41
cumulative R^2 (adj.)	0.04	0.32	0.32	0.32	0.55	0.57	0.57	

sumption growth. We see from Panel A of Figure 3 that the risk-premium for consumption growth does not line up very well with mean excess returns. In contrast, once we consider the total risk premia that corrects the risk premium for consumption growth with the risk premia for the missing factors and idiosyncratic risk, mean excess returns line up almost perfectly with the total risk premium, as can be seen in Panel B of Figure 3.

Finally, we show how our methodology can be used to understand the factors that drive the α-SDF. To do this, we regress $\log(m^\alpha)$ on common factors that have been considered in the empirical asset-pricing literature. These results are presented in Table 5. From the row reporting the t-value for each factor, we see that size (SMB) and profitability (RMW) are the most prominent of the missing factors with t-values of 10.37 and −9.33, respectively. This can also be observed from the last row that reports the *cumulative* R^2 (adjusted) for each of the factors. On the other hand, value (HML) and intermediary capital (Interm.) have t-values that are not significant.

11 Conclusion

In this paper, we have shown how, given a misspecified stochastic discount factor (SDF), one can construct an *admissible* SDF, namely an SDF that prices assets correctly. There is a large literature that builds on the work of Hansen and Jagannathan (1991, 1997) to

provide bounds that an admissible SDF must satisfy and to characterizes the distance between a given, potentially misspecified, SDF and an admissible SDF. We first extend the classical Arbitrage Pricing Theory (APT) so that it allows not only for idiosyncratic pricing errors but also for pervasive pricing errors that are related to factors. Then, using the extended APT, we show how to construct an *admissible* SDF, which prices assets correctly, given a misspecified SDF. Our approach can handle misspecification arising from a number of sources, such as, missing factors, mismeasured factors, incorrect specification of the distribution of the factors, and idiosyncratic pricing errors unrelated to factors.

We show that the admissible SDF is on the mean-variance efficient frontier, and thus, satisfies the Hansen and Jagannathan (1991) bound *exactly*. We also show how this admissible SDF can be decomposed into two orthogonal components: one that corresponds to the misspecified SDF based on the chosen set of factors, which we label the *beta* SDF, and the other that corresponds to the required correction, which we call the *alpha* SDF. The alpha SDF corrects various prominent sources of misspecification considered in the literature: for example, if one were working with a linear factor model, then the misspecification could arise from missing or mismeasured factors; alternatively, if one were working with a representative-agent model, then the misspecification could arise from an erroneous specification of the utility function or the state variables.

For the case where the number of assets, N, is asymptotically large, we obtain results that are even stronger, in contrast to the existing literature that requires N to be small. For the case of large N, we show that our admissible SDF recovers exactly the contribution of any missing pervasive factors *without requiring one to identify which factors are missing*. Moreover, because of the structure imposed by the extended APT, estimation of our admissible SDF does not suffer from the curse of dimensionality that typically arises when the number of assets is large.

We illustrate the implications of our theoretical results using both simulated data and U.S. stock-return data. These illustrations demonstrate that the theory we develop is remarkably effective in correcting various sources of model misspecification.

A Proofs for Theorems

Proof of Theorem 3.1

By Chamberlain and Rothschild (1983, Theorem 4) the residual covariance matrix satisfies

$$\mathbf{\Sigma}_{N,t} = \mathbf{A}_{N,t}\mathbf{A}'_{N,t} + \mathbf{C}_{N,t},$$

where $\mathbf{C}_N$ is a positive definite matrix with eigenvalues uniformly bounded by $g_{p+1\,N}(\mathbf{\Sigma}_{N,t})$.[34] By the Sherman-Morrison-Woodbury decomposition,

$$\mathbf{\Sigma}^{-1}_{N,t} = \mathbf{C}^{-1}_{N,t} - \mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}.$$

Therefore, by substitution,

$$\begin{aligned}
\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t} &= \boldsymbol{\alpha}'_{N,t}\mathbf{C}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t} - \boldsymbol{\alpha}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t} \\
&= (\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} + \mathbf{a}_{N,t})'\mathbf{C}^{-1}_{N,t}(\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} + \mathbf{a}_{N,t}) \\
&\quad - (\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} + \mathbf{a}_{N,t})'\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}(\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} + \mathbf{a}_{N,t}) \\
&= \boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} \\
&\quad - \boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} \\
&\quad + \mathbf{a}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{a}_{N,t} - \mathbf{a}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{a}_{N,t} \\
&\quad + 2\mathbf{a}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} - 2\mathbf{a}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}.
\end{aligned}$$

We now show that $\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}$ is bounded even as N diverges. We look each of the term on the right-hand side of the last equality sign, one by one. Thus,

$$\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} - \boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}$$

$$\begin{aligned}
&= \boldsymbol{\lambda}'_{miss,t}(\mathbf{I}_N - \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1})\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} \\
&= \boldsymbol{\lambda}'_{miss,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} \leq \boldsymbol{\lambda}'_{miss,t}\boldsymbol{\lambda}_{miss,t},
\end{aligned}$$

because $\mathbf{I}_p - (\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t}$ is positive semidefinite. Next, for the third term,

$$\mathbf{a}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{a}_{N,t} \leq \mathbf{a}'_{N,t}\mathbf{a}_{N,t}g^{-1}_{NN}(\mathbf{C}_{N,t}).$$

[34] We differ slightly from Chamberlain and Rothschild (1983, Theorem 4) as our model is conditional, where all quantities are time-varying, but such time-variation plays no role in our proof.

Now, the jth element of $\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}$, obtained by considering the jth column of $\mathbf{A}_{N,t}$, for every $1 \leq j \leq p$, satisfies

$$|\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}g_{jN}^{\frac{1}{2}}\mathbf{v}_{jN}| \leq g_{jN}^{\frac{1}{2}}(\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{a}_{N,t})^{\frac{1}{2}}(\mathbf{v}'_{jN}\mathbf{C}_{N,t}^{-1}\mathbf{v}_{jN})^{\frac{1}{2}} \leq g_{jN}^{\frac{1}{2}}g_{NN}^{-1}(\mathbf{C}_{N,t})(\mathbf{a}'_{N,t}\mathbf{a}_{N,t})^{\frac{1}{2}},$$

recalling that $\mathbf{v}'_{jN}\mathbf{v}_{jN} = 1$, where for simplicity we set $\mathbf{v}_{jN} = \mathbf{v}_{jN}(\mathbf{\Sigma}_{N,t})$, and $g_{jN} = g_{jN}(\mathbf{\Sigma}_{N,t})$. Moreover, the (i,j)th element, for every $1 \leq i,j \leq p$, of $(\mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})$ is equal to $g_{iN}^{\frac{1}{2}}g_{jN}^{\frac{1}{2}}\mathbf{v}'_{iN}\mathbf{C}_{N,t}^{-1}\mathbf{v}_{jN}$. Therefore, assuming without loss of generality that $g_{1N} = \max[g_{1N}, \cdots, g_{pN}]$ for N large enough, then $(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})^{-1}$ decreases at rate g_{1N}^{-1}. On the other hand, for the same reason, the elements of the vector $\mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{a}_{N,t}$ diverge at most at rate $g_{1N}^{\frac{1}{2}}$. Thus, the fourth term satisfies:

$$|\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{a}_{N,t}| \leq \delta g_{1N}^{\frac{1}{2}}g_{1N}^{\frac{1}{2}}g_{1N}^{-1} = \delta.$$

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

$$\begin{aligned}
&\left|2\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t} - 2\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}\boldsymbol{\lambda}_{miss,t}\right| \\
&= 2\left|\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})^{-1}\boldsymbol{\lambda}_{miss,t}\right| \\
&\leq (\mathbf{a}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})^{-1}\mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{a}_{N,t})^{\frac{1}{2}}(\boldsymbol{\lambda}'_{miss,t}(\mathbf{I}_p + \mathbf{A}'_{N,t}\mathbf{C}_{N,t}^{-1}\mathbf{A}_{N,t})^{-1}\boldsymbol{\lambda}_{miss,t})^{\frac{1}{2}} \\
&\leq \delta g_{pN}^{-\frac{1}{2}} \to 0.
\end{aligned}$$

Proof of Theorem 4.1

In general, the SDF can always be re-written as *linear* in payoffs or excess returns. Without loss of generality we can assume that the factors are traded, implying $\mathbb{E}_t\mathbf{f}_{t+1} = R_{ft}\mathbf{1}_K + \boldsymbol{\lambda}_t$. In fact, if the factors are non traded, by standard arguments one replaces them with the corresponding (traded) mimicking portfolios.

$$\begin{aligned}
\mathbf{0}_N &= \mathbb{E}_t(m_{t+1}(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N)) \\
&= \mathbb{E}_t(m_{t+1}(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N)) \\
&= \mathbb{E}_t\big(\operatorname{proj}(m_{t+1}|\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N)(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N)\big).
\end{aligned}$$

Therefore, given that in the APT payoffs (excess returns) are linear in $\mathbf{f}_t$, $\boldsymbol{\varepsilon}_{N,t}$, and $\mathbf{1}$, then the SDF under the APT must satisfy:

$$m_{t+1} = \mathbb{E}_t(m_{t+1}) + \mathbf{b}'_t(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}'_{N,t}\boldsymbol{\varepsilon}_{N,t+1}, \tag{A1}$$

for some given coefficient vector $\mathbf{b}_t$, which is $K \times 1$, and coefficient vector $\mathbf{c}_{N,t}$, which is $N \times 1$. We determine $\mathbf{b}_t$ and $\mathbf{c}_{N,t}$ below whereas $\mathbb{E}_t(m_{t+1}) = \mu_{m,t}$.

Given that we assumed the existence of a risk-free asset, $R_{ft} = 1 + r_{ft}$, it must be that:

$$\begin{aligned}\mathbf{0}_K &= \mathbb{E}_t(m_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)\\ \mathbf{0}_N &= \mathbb{E}_t(m_{t+1}(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N)),\end{aligned}$$

leading to a total of $K + N$ constraints. Substituting m_{t+1} from (A1) one gets:

$$\begin{aligned}\mathbf{0}_K &= \mathbb{E}_t\left[\left(\mathbb{E}_t(m_{t+1}) + \mathbf{b}_t'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\right)(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)\right]\\ &= \mathbb{E}\left[\left(\mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\right)(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)\right]\\ &= \mu_{m,t}\mathbb{E}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K) + \mathbb{E}\left((\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda})'\mathbf{b}_t\right)\\ &\quad + \mathbb{E}\left((\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)\boldsymbol{\varepsilon}_{N,t+1}'\right)\mathbf{c}_{N,t}\\ &= \mu_{m,t}\boldsymbol{\lambda}_t + \boldsymbol{\Omega}_t\mathbf{b}_t,\end{aligned}$$

implying that

$$\mathbf{b}_t = -\mu_{m,t}\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t.$$

Next,

$$\begin{aligned}\mathbf{0}_N &= \mathbb{E}_t\left[\left(\mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\right)\left(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N\right)\right]\\ &= \mathbb{E}_t\Big[\left(\mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\right) \times\\ &\qquad \left(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \mathbf{B}_{N,t}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \boldsymbol{\varepsilon}_{N,t+1}\right)\Big]\\ &= \mu_{m,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) + \mathbf{B}_{N,t}\boldsymbol{\Omega}_t\mathbf{b}_t + \boldsymbol{\Sigma}_{N,t}\mathbf{c}_{N,t}\\ &= \mu_{m,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) - \mu_{m,t}\mathbf{B}_{N,t}\boldsymbol{\Omega}_t\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t + \boldsymbol{\Sigma}_{N,t}\mathbf{c}_{N,t}\\ &= \mu_{m,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) - \mu_{m,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_{N,t} + \boldsymbol{\Sigma}_{N,t}\mathbf{c}_{N,t}\\ &= \mu_{m,t}\boldsymbol{\alpha}_{N,t} + \boldsymbol{\Sigma}_{N,t}\mathbf{c}_{N,t},\end{aligned}$$

implying that

$$\mathbf{c}_{N,t} = -\mu_{m,t}\boldsymbol{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t}.$$

We now establish the result for the projection SDF, m_{t+1}^*. By construction, setting $\mathbf{X}_{N,t+1} = (1, \mathbf{R}_{N,t+1}')'$ and $\boldsymbol{\mu}_{N,t} = \mathbb{E}_t(\mathbf{R}_{N,t+1})$,

$$m_{t+1}^* = \mathbb{E}_t(m_{t+1}\mathbf{X}_{N,t+1}')(\mathbb{E}_t(\mathbf{X}_{N,t+1}\mathbf{X}_{N,t+1}')^{-1}\mathbf{X}_{N,t+1}$$

$$= (\mu_{m,t}, \mu_{m,t}\boldsymbol{\mu}'_{N,t} + \mathbf{b}'_t\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{c}'_{N,t}\boldsymbol{\Sigma}_{N,t}) \begin{pmatrix} 1 + \boldsymbol{\mu}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\mu}_{N,t} & -\boldsymbol{\mu}'_{N,t}\mathbf{V}^{-1}_{N,t} \\ -\mathbf{V}^{-1}_{N,t}\boldsymbol{\mu}_{N,t} & \mathbf{V}^{-1}_{N,t} \end{pmatrix} \mathbf{X}_{t+1}$$

$$= (\mu_{m,t} - (\mathbf{b}'_t\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{c}'_{N,t}\boldsymbol{\Sigma}_{N,t})\mathbf{V}^{-1}_{N,t}\boldsymbol{\mu}_{N,t}, (\mathbf{b}'_t\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{c}'_{N,t}\boldsymbol{\Sigma}_{N,t})\mathbf{V}^{-1}_{N,t})\mathbf{X}_{t+1}$$
$$= \mu_{m,t} + (\mathbf{b}'_t\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{c}'_{N,t}\boldsymbol{\Sigma}_{N,t})\mathbf{V}^{-1}_{N,t}(\mathbf{R}_{N,t+1} - \boldsymbol{\mu}_{N,t}),$$

where we use the block formula for the inverse of a square matrix to simplify $\mathbb{E}_t(\mathbf{X}_{t+1}\mathbf{X}'_{t+1})$.

Proof of Theorem 4.2

The decomposition of m_{t+1} into m^{α}_{t+1} and m^{β}_{t+1} follows from Theorem 4.1. Moreover,

$$\mathbb{E}_t(m^{\alpha}_{t+1}m^{\beta}_{t+1}) = \mathbb{E}_t\left((-\mathbf{c}'_{N,t}\boldsymbol{\varepsilon}_{N,t+1})(\mu_{m,t} - \mathbf{b}'_t(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})))\right)$$
$$= -\mu_{m,t}\mathbf{c}'_{N,t}\mathbb{E}_t(\boldsymbol{\varepsilon}_{N,t+1}) + \mathbf{b}'_t\mathbb{E}_t((\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1}))\boldsymbol{\varepsilon}'_{N,t+1})\mathbf{c}_{N,t} = 0.$$

Proof of Theorem 4.3

Without loss of generality, assume that the factors are traded, implying that $\mathbb{E}_t\mathbf{f}_{t+1} = r_f\mathbf{1}_K + \boldsymbol{\lambda}_t$ and

$$\mathbf{R}^e_{t+1} = \boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \mathbf{B}_{N,t}(\mathbf{f}_{t+1} - r_f\mathbf{1}_K - \boldsymbol{\lambda}_t) + \boldsymbol{\varepsilon}_{N,t+1}.$$

Given this, let the candidate non-negative SDF be:

$$m^+_{t+1} = exp(\mu^+_{m,t} + (\mathbf{b}^+_t)'(\mathbf{f}_{t+1} - r_f\mathbf{1}_K - \boldsymbol{\lambda}_t) + (\mathbf{c}^+_{N,t})'\boldsymbol{\varepsilon}_{N,t+1}),$$

which implies that to identify m^+_{t+1}, we need to find: $\mu^+_{m,t}$, $\mathbf{b}^+_t$, and $\mathbf{c}^+_{N,t}$.

Imposing the following $1 + K + N$ constraints,

$$\mu_{m,t} = \mathbb{E}_t(m^+_{t+1}),$$
$$\mathbf{0}_K = \mathbb{E}_t(m^+_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)),$$
$$\mathbf{0}_N = \mathbb{E}_t(m^+_{t+1}\mathbf{R}^e_{N,t+1}),$$

allows one to identify m^+_{t+1}, as we show below. Starting with the first restriction, using Lemma A.1 below, we get:

$$\mu_{m,t} = \mathbb{E}_t(m^+_{t+1}) = \mathbb{E}_t(\exp[\mu^+_{m,t} + (\mathbf{b}^+_t)'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}^{+\prime}_{N,t}\boldsymbol{\varepsilon}_{N,t+1}])$$

$$= \exp[\mu^+_{m,t}] \exp[(\frac{1}{2}\mathbf{b}_t^{+\prime}\boldsymbol{\Omega}_t\mathbf{b}_t^+ + \frac{1}{2}\boldsymbol{c}^{+\prime}_{N,t}\boldsymbol{\Sigma}_{N,t}\mathbf{c}^+_{N,t}]$$

implying

$$\exp[\mu^+_{m,t}] = \mu_{mt} \exp[-(\frac{1}{2}\mathbf{b}_t^{+\prime}\boldsymbol{\Omega}_t\mathbf{b}_t^+ + \frac{1}{2}\boldsymbol{c}^{+\prime}_{N,t}\boldsymbol{\Sigma}_{N,t}\mathbf{c}^+_{N,t})].$$

Next, considering the K restrictions and using Lemma A.1 again, we obtain:

$$\begin{aligned}
\mathbf{0}_K &= \mathbb{E}_t(m^+_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)) \\
&= \mathbb{E}(m^+_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t)) + \boldsymbol{\lambda}_t\mathbb{E}_t(m^+_{t+1}) \\
&= \boldsymbol{\lambda}_t\mathbb{E}_t(m^+_{t+1}) + e^{\mu^+_{m,t}}\mathbb{E}_t(e^{\mathbf{c}^{+\prime}_{N,t}\boldsymbol{\varepsilon}_{N,t+1}})\mathbb{E}_t(e^{\mathbf{b}_t^{+\prime}(\mathbf{f}_{t+1}-R_{ft}\mathbf{1}_K-\boldsymbol{\lambda}_t)})(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t)) \\
&= \boldsymbol{\lambda}_t\mu_{m,t} + e^{(\mu^+_{m,t}+\frac{1}{2}\mathbf{c}^{+\prime}_{N,t}\boldsymbol{\Sigma}_{N,t}\mathbf{c}^+_{N,t}+\frac{1}{2}\mathbf{b}_t^{+\prime}\boldsymbol{\Omega}_t\mathbf{b}_t^+)}\boldsymbol{\Omega}_t\mathbf{b}_t^+ \\
&= \boldsymbol{\lambda}_t\mu_{m,t} + \mu_{m,t}\boldsymbol{\Omega}_t\mathbf{b}_t^+
\end{aligned}$$

yielding

$$\mathbf{b}_t^+ = -\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t.$$

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

$$\begin{aligned}
\mathbf{0}_N &= \mathbb{E}_t(m^+_{t+1}\mathbf{R}^e_{t+1}) \\
&= \mathbb{E}_t\big(m^+_{t+1}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \mathbf{B}_{N,t}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \boldsymbol{\varepsilon}_{N,t+1})\big) \\
&= (\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\mathbb{E}_t(m^+_{t+1}) + \mathbb{E}_t(m^+_{t+1}\mathbf{B}_{N,t}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t)) + \mathbb{E}_t(m^+_{t+1}\boldsymbol{\varepsilon}_{N,t+1}) \\
&= (\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\mathbb{E}(m^+_{t+1}) - \mathbf{B}_{N,t}\boldsymbol{\lambda}_t\mathbb{E}_t(m^+_{t+1}) + \mu_{m,t}\boldsymbol{\Sigma}_{N,t}\mathbf{c}_{N,t},
\end{aligned}$$

implying that

$$\mathbf{c}_{N,t} = -\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t},$$

where we used

$$\mathbb{E}_t(m^+_{t+1}\boldsymbol{\varepsilon}_{N,t+1}) = e^{(\mu^+_{m,t}+\frac{1}{2}\mathbf{b}_t^{+\prime}\boldsymbol{\Omega}_t\mathbf{b}_t^+)}\mathbb{E}_t(e^{\mathbf{c}^{+\prime}_{N,t}\boldsymbol{\varepsilon}_{N,t+1}}\boldsymbol{\varepsilon}_{N,t+1}) = \mu_{m,t}\boldsymbol{\Sigma}_{N,t}\mathbf{c}_{N,t}.$$

Putting the terms together

$$m^+_{t+1} = \mu_{m,t}e^{-(\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}(\mathbf{f}^e_{t+1}-\boldsymbol{\lambda}_t)+\frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t)}e^{-(\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\varepsilon}_{N,t+1}+\frac{1}{2}\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t})}.$$

Lemma A.1. *For the vector of random variables* $\mathbf{z} \sim N(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$*, and any constant vector* $\mathbf{d}$*, one gets:*

(i)

$$E(e^{\mathbf{d}'\mathbf{z}}) = e^{\mathbf{d}'\boldsymbol{\mu}_z+\frac{1}{2}\mathbf{d}'\boldsymbol{\Sigma}_z\mathbf{d}}.$$

(ii)

$$E(\mathbf{z}e^{\mathbf{d}'\mathbf{z}}) = \boldsymbol{\mu}^* e^{\frac{1}{2}(\boldsymbol{\mu}^{*\prime}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\mu}^* - \boldsymbol{\mu}_z'\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\mu}_z)},$$

setting

$$\boldsymbol{\mu}^* = (\boldsymbol{\mu}_z + \boldsymbol{\Sigma}_z\mathbf{d}).$$

An alternative expression is

$$E(\mathbf{z}e^{\mathbf{d}'\mathbf{z}}) = (\boldsymbol{\mu}_z + \boldsymbol{\Sigma}_z\mathbf{d})e^{(\frac{1}{2}\mathbf{d}'\boldsymbol{\Sigma}_z\mathbf{d} + \boldsymbol{\mu}_z'\mathbf{d})}.$$

Proof. (i) is well-known. For (ii), denoting by n_z the dimensionality of the vector $\mathbf{z}$,

$$E(\mathbf{z}e^{\mathbf{d}'\mathbf{z}}) = \frac{1}{(\sqrt{2\pi})^{n_z}|\boldsymbol{\Sigma}_z|^{\frac{1}{2}}} \int \mathbf{z}e^{\mathbf{d}'\mathbf{z}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_z)'\boldsymbol{\Sigma}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)} d\mathbf{z}.$$

Then

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

implying

$$E(\mathbf{z}e^{\mathbf{d}'\mathbf{z}}) = e^{-\frac{1}{2}\boldsymbol{\mu}_z'\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\mu}_z + \frac{1}{2}\boldsymbol{\mu}^{*\prime}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\mu}^*} \left(\frac{1}{(\sqrt{2\pi})^{n_z}|\boldsymbol{\Sigma}_z|^{\frac{1}{2}}} \int \mathbf{z}e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_z^*)'\boldsymbol{\Sigma}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z^*)} d\mathbf{z} \right).$$

Lemmata

Below, we list a series of lemmas that will be useful for proving the next set of results, valid when $N \to \infty$. These lemmas are proved under assumptions that are to be added.

Lemma A.2. *Let* $\mathbf{V} = \mathbf{B}\boldsymbol{\Omega}\mathbf{B}' + \boldsymbol{\Sigma}$ *with* $N \times N$ *and* $K \times K$ *matrices* $\boldsymbol{\Sigma} > 0$ *and* $\boldsymbol{\Omega} > 0$ *for any* N*, and a full-column rank* $N \times K$ *matrix* $\mathbf{B}$ *satisfying* $\mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B}/N \to \mathbf{D} > 0$ *for some non-singular* $\mathbf{D}$*. Then:*

$$\mathbf{B}'\mathbf{V}^{-1}\mathbf{B} \to \boldsymbol{\Omega}^{-1}.$$

Proof: The result follows from

$$\mathbf{V}^{-1} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1},$$

and pre-multiplying by $\mathbf{B}'$, and re-arranging terms, yields

$$\begin{aligned}
\mathbf{B}'\mathbf{V}^{-1} &= \mathbf{B}'\mathbf{\Sigma}^{-1} - \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1} \\
&= (I_K - \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1})\mathbf{B}'\mathbf{\Sigma}^{-1} \\
&= ((\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}) - \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1} \\
&= \mathbf{\Omega}^{-1}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1}.
\end{aligned}$$

Post-multiplying by $\mathbf{B}$ and taking the limit as $N \to \infty$ gives

$$\mathbf{B}'\mathbf{V}^{-1}\mathbf{B} \to \mathbf{\Omega}^{-1},$$

because $(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B} \to I_K$.

Lemma A.3. *Under the assumptions of Lemma A.2 and for a random vector $\boldsymbol{\varepsilon}$ with mean zero and covariance $\mathbf{\Sigma}$:*

$$\mathbf{B}'\mathbf{V}^{-1}\boldsymbol{\varepsilon} = O_p(N^{-\frac{1}{2}}).$$

Proof: Pre-multiplying by $\mathbf{B}'$ and post-multiplying by $\boldsymbol{\varepsilon}$ one obtains:

$$\mathbf{B}'\mathbf{V}^{-1}\boldsymbol{\varepsilon}_t = \mathbf{\Omega}^{-1}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}.$$

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

Lemma A.4. *Under the assumptions of Lemma A.2 and letting $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}' + \mathbf{C}$ for a column full-rank $N \times p$ matrix $\mathbf{A}$ and a $N \times N$ matrix $\mathbf{C}$ non-singular for any N such that $\mathbf{A}'\mathbf{C}^{-1}\mathbf{A}/N \to \mathbf{E} > 0$ and $\mathbf{B}'\mathbf{C}^{-1}\mathbf{B}/N \to \mathbf{F} > 0$. Then:*

$$\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{B} = O(1).$$

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

$$\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{B} = (\mathbf{I}_p + \mathbf{A}'\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{C}^{-1}\mathbf{B} = (\frac{\mathbf{I}_p}{N} + \frac{\mathbf{A}'\mathbf{C}^{-1}\mathbf{A}}{N})^{-1}\frac{\mathbf{A}'\mathbf{C}^{-1}\mathbf{B}}{N} = O(1),$$

where, by Schwartz inequality, $\| \mathbf{A}'\mathbf{C}^{-1}\mathbf{B} \| \leq \| \mathbf{A}'\mathbf{C}^{-1}\mathbf{A} \|^{\frac{1}{2}} \| \mathbf{B}'\mathbf{C}^{-1}\mathbf{B} \|^{\frac{1}{2}}$ with $\| \cdot \|$ denoting the Euclidean norm.

Lemma A.5. *Under the assumptions of Lemma A.4:*

$$\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A} \to \mathbf{I}_p.$$

Proof: This is a special case of Lemma A.2.

Lemma A.6. *Under the assumptions of Lemma A.4 and for a random vector* $\boldsymbol{\eta}$ *with mean zero and covariance* $\mathbf{C}$*: Then:*

$$\mathbf{A}'\mathbf{\Sigma}^{-1}\boldsymbol{\eta} = O_p(N^{-\frac{1}{2}}).$$

Proof: This is a special case of Lemma A.3.

Lemma A.7. *Under the assumptions of Lemma A.4, setting* $\boldsymbol{\alpha} = \mathbf{a} + \mathbf{A}\boldsymbol{\lambda}_m$ *with* $\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} = O(1)$ *and a* $p \times 1$ *vector of constants* $\boldsymbol{\lambda}_m$*, then:*

$$\boldsymbol{\alpha}'\mathbf{V}^{-1}\mathbf{B} = O(N^{-\frac{1}{2}}).$$

Proof: Given

$$\begin{aligned}
\boldsymbol{\alpha}'\mathbf{V}^{-1}\mathbf{B} &= \mathbf{a}'\mathbf{V}^{-1}\mathbf{B} + \boldsymbol{\lambda}_m'\mathbf{A}'\mathbf{V}^{-1}\mathbf{B} \\
&= \mathbf{a}'(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1})\mathbf{B} \\
&\quad + \boldsymbol{\lambda}_m'\mathbf{A}'(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1})\mathbf{B} \\
&= \mathbf{a}'\mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{\Omega}^{-1} + \boldsymbol{\lambda}_m'\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{\Omega}^{-1} + \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{\Omega}^{-1} \\
&= O(N^{-\frac{1}{2}}) + O(N^{-1}),
\end{aligned}$$

by Lemma A.4, the bound $\| \mathbf{a}'\mathbf{\Sigma}^{-1}\mathbf{B} \| \leq \| \mathbf{a}'\mathbf{\Sigma}^{-1}\mathbf{a} \|^{\frac{1}{2}} \| \mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B} \|^{\frac{1}{2}}$ and

$$\begin{aligned}
&|\mathbf{a}'\mathbf{\Sigma}^{-1}\mathbf{a}| = |\mathbf{a}'(\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{A}(\mathbf{I}_p + \mathbf{A}'\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{C}^{-1})\mathbf{a}| \\
&\leq |\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}| + |\mathbf{a}\mathbf{C}^{-1}\mathbf{A}(\mathbf{I}_p + \mathbf{A}'\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{C}^{-1}\mathbf{a}| \\
&\leq |\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}| + |\mathbf{a}\mathbf{C}^{-1}\mathbf{a}|^{\frac{1}{2}} \parallel \mathbf{A}'\mathbf{C}^{-1}\mathbf{A} \parallel^{\frac{1}{2}} \parallel (\mathbf{I}_p + \mathbf{A}'\mathbf{C}^{-1}\mathbf{A})^{-1} \parallel |\mathbf{a}\mathbf{C}^{-1}\mathbf{a}|^{\frac{1}{2}} \parallel \mathbf{A}'\mathbf{C}^{-1}\mathbf{A} \parallel^{\frac{1}{2}} \\
&= |\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}| + |\mathbf{a}\mathbf{C}^{-1}\mathbf{a}| \parallel \mathbf{A}'\mathbf{C}^{-1}\mathbf{A} \parallel \parallel (\mathbf{I}_p + \mathbf{A}'\mathbf{C}^{-1}\mathbf{A})^{-1} \parallel = O(1).
\end{aligned}$$

Proof of Theorem 4.5

The result follow by Lemmas A.2 and A.3.

Proof of Theorem 4.6

From $\mathbf{V}_{N,t}^{-1} = \mathbf{\Sigma}_{N,t}^{-1} - \mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t}(\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1}\mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}$ one obtains

$$m_{t+1}^{\alpha*} = -\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{t+1} - \mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t}(\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1}\mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{t+1} \quad \text{(A2)}$$

$$+ \mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t}(\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1}\mathbf{\Omega}_t^{-1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_N - \boldsymbol{\lambda}_t). \quad \text{(A3)}$$

We now show that the third term on the right hand side is $O_p(N^{-\frac{1}{2}})$. In fact,

$$\| \boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t} \| \leq \| \boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\alpha}_{N,t} \|^{\frac{1}{2}} \| \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t} \|^{\frac{1}{2}} = O_p(N^{\frac{1}{2}}),$$

whereas

$$\| (\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1} \| = O_p(N^{-1}).$$

Remark A.0.1. Given that m_{t+1}^{α} and $m_{t+1}^{\alpha*}$ have the same pricing implications, that is, $E(m_{t+1}^{\alpha}(R_{it} - R_f)) = E(m_{t+1}^{\alpha*}(R_{it} - R_f))$, it follows that the last two terms on the right hand side of (A2) and (A3) induce, in terms of pricing, the same quantity in absolute value but opposite sign.

Proof of Theorem 4.7

Consider

$$m_{t+1}^{\alpha*} = -\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{N,t+1} - \mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t}(\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1}\mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{N,t+1}$$

$$+ \mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t}(\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1}\mathbf{\Omega}_t^{-1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_N - \boldsymbol{\lambda}_t).$$

Setting $\boldsymbol{\varepsilon}_{N,t+1} = \mathbf{A}_{N,t}\mathbf{z}_{miss,t+1} + \boldsymbol{\eta}_{N,t+1}$ and $\mathbf{f}_{m,t+1} = \mathbf{z}_{miss,t+1} + \mathbb{E}_t(\mathbf{f}_{m,t+1})$, the first term on the right-hand side satisfies

$$\begin{aligned}
-\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{N,t+1} &= -\mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}(\mathbf{A}_{N,t}\mathbf{z}_{miss,t+1} + \boldsymbol{\eta}_{N,t+1}) \\
&= -\mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{A}_{N,t}\mathbf{z}_{miss,t+1} - \mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\eta}_{N,t+1} \\
&= -\mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{z}_{miss,t+1} + O_p(N^{-\frac{1}{2}}),
\end{aligned}$$

by Lemmas A.5 and A.6. The second term satisfies

$$\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t}(\mathbf{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\mathbf{B}_{N,t})^{-1}\mathbf{B}'_{N,t}\mathbf{\Sigma}_{N,t}^{-1}\boldsymbol{\varepsilon}_{N,t+1}$$

$$= \mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t}(\boldsymbol{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t})^{-1}\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{A}_{N,t}\mathbf{z}_{miss,t+1}$$
$$+ \mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t}(\boldsymbol{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t})^{-1}\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\eta}_{N,t+1}$$
$$= O_p(N^{-1}) + O_p(N^{-\frac{1}{2}}),$$

by Lemmas A.4, recalling that $\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t} = O_p(N)$ and making use of $\mathbf{B}'_{N,t}\mathbf{C}^{-1}_{N,t}\boldsymbol{\eta}_{N,t+1} = O_p((\mathbf{B}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{B}_{N,t})^{\frac{1}{2}})$ and $\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\boldsymbol{\eta}_{N,t+1} = O_p((\mathbf{A}'_{N,t}\mathbf{C}^{-1}_{N,t}\mathbf{A}_{N,t})^{\frac{1}{2}})$.

Finally, the third term satisfies

$$\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t}(\boldsymbol{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t})^{-1}\boldsymbol{\Omega}_t^{-1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_N - \boldsymbol{\lambda}_t)$$
$$=\mu_{m,t}\boldsymbol{\lambda}'_{miss,t}\mathbf{A}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t}(\boldsymbol{\Omega}_t^{-1} + \mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t})^{-1}\boldsymbol{\Omega}_t^{-1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_N - \boldsymbol{\lambda}_t) = O_p(N^{-1}),$$

by Lemma A.4 and $\mathbf{B}'_{N,t}\boldsymbol{\Sigma}^{-1}_{N,t}\mathbf{B}_{N,t} = O_p(N)$.

Proof of Theorem 4.8

Proof. Case (i). By Theorem 4.7, $m_t^{\alpha*} \to_p -\mu_{m,t-1}\boldsymbol{\lambda}'_{miss,t-1}(\mathbf{f}_{miss,t} - \mathbb{E}_{t-1}(\mathbf{f}_{miss,t})) = -\xi_{At}$, yielding γ_A and the limit of R^2_{miss}. However, when $\mu_{m,t} = \mu_m, \boldsymbol{\lambda}_{miss,t} = \boldsymbol{\lambda}_{miss}, \mathbb{E}_{t-1}(\mathbf{f}_{miss,t}) = \mathbb{E}(\mathbf{f}_{miss,t})$, then

$$\gamma_A = -(\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss})^{-1}(\sum_{t=1}^{T}\tilde{\mathbf{f}}_{miss,t}(\mathbf{f}_{miss,t} - \mathbb{E}_{t-1}(\mathbf{f}_{miss,t}))')\mu_m\boldsymbol{\lambda}_{miss}$$
$$= -(\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss})^{-1}(\sum_{t=1}^{T}\tilde{\mathbf{f}}_{miss,t}\mathbf{f}'_{miss,t})\mu_m\boldsymbol{\lambda}_{miss}$$
$$= -(\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss})^{-1}\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\mu_m\boldsymbol{\lambda}_{miss}$$
$$= -\mu_m\boldsymbol{\lambda}_{miss}.$$

Regarding the limit of the R^2, its numerator simplifies to

$$\gamma'_A\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\gamma_A = \mu_m^2\boldsymbol{\lambda}'_{miss}\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\boldsymbol{\lambda}_{miss},$$

and, given that $\boldsymbol{\xi}_A = -(\mathbf{F}_{miss} - \mathbf{1}_T\mathbb{E}(\mathbf{f}'_{miss}))\boldsymbol{\lambda}_{miss}\mu_m$, its denominator becomes

$$\boldsymbol{\xi}'_A\mathbf{M}_{1_T}\boldsymbol{\xi}_A = \mu_m^2\boldsymbol{\lambda}'_{miss}(\mathbf{F}_{miss} - \mathbf{1}_T\mathbb{E}(\mathbf{f}'_{miss}))'\mathbf{M}_{1_T}(\mathbf{F}_{miss} - \mathbf{1}_T\mathbb{E}(\mathbf{f}'_{miss}))\boldsymbol{\lambda}_{miss}$$
$$= \mu_m^2\boldsymbol{\lambda}'_{miss}(\mathbf{F}'_{miss}\mathbf{M}_{1_T}\mathbf{F}_{miss})\boldsymbol{\lambda}_{miss} = \mu_m^2\boldsymbol{\lambda}'_{miss}\tilde{\mathbf{F}}'_{miss}\tilde{\mathbf{F}}_{miss}\boldsymbol{\lambda}_{miss},$$

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

Part (ii). By Remark 4.6.2

$$m_t^{\alpha *} \rightarrow_d \xi_{at},$$

yielding convergence in distribution to γ_a. The limiting distribution of the R^2 easily follows both under the case of time-varying and constant $\mu_{m,t}$.

Proof of Theorem 4.9

We start with the conjecture that the SDF is still linear in the observed factors $\mathbf{f}_{t+1}$ and idiosyncratic risk $\boldsymbol{\varepsilon}_{N,t+1}$, although now these can be cross-correlated. Staking the $K+N$ pricing equations:

$$\begin{aligned}\mathbf{0}_K &= \mathbb{E}_t(m_{t+1}(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)\\ \mathbf{0}_N &= \mathbb{E}_t(m_{t+1}(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N)),\end{aligned}$$

yields,

$$\begin{aligned}\mathbf{0}_{N+K} &= \begin{pmatrix} \mathbb{E}_t\left[\left(\mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\right)(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K)\right] \\ \mathbb{E}_t\left[\left(\mu_{m,t} + \mathbf{b}_t'(\mathbf{f}_{t+1} - R_{ft}\mathbf{1}_K - \boldsymbol{\lambda}_t) + \mathbf{c}_{N,t}'\boldsymbol{\varepsilon}_{N,t+1}\right)\left(\mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N\right)\right] \end{pmatrix}\\ &= \mu_{m,t}\begin{pmatrix} \boldsymbol{\lambda}_t \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \boldsymbol{\alpha}_{N,t}\end{pmatrix} + \begin{pmatrix} \boldsymbol{\Omega}_t\mathbf{b}_t + \mathbf{P}_{N,t}\mathbf{c}_{N,t} \\ (\mathbf{B}_{N,t}\boldsymbol{\Omega}_t + \mathbf{P}_{N,t}')\mathbf{b}_t + (\boldsymbol{\Sigma}_{N,t} + \mathbf{B}_{N,t}\mathbf{P}_{N,t})\mathbf{c}_{N,t}\end{pmatrix}\\ &= \mu_{m,t}\begin{pmatrix} \boldsymbol{\lambda}_t \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \boldsymbol{\alpha}_{N,t}\end{pmatrix} + \begin{pmatrix} \boldsymbol{\Omega}_t & \mathbf{P}_{N,t} \\ (\mathbf{B}_{N,t}\boldsymbol{\Omega}_t + \mathbf{P}_{N,t}') & (\boldsymbol{\Sigma}_{N,t} + \mathbf{B}_{N,t}\mathbf{P}_{N,t})\end{pmatrix}\begin{pmatrix} \mathbf{b}_t \\ \mathbf{c}_{N,t}\end{pmatrix}.\end{aligned}$$

Using the block-wise formula for the inverse of a matrix, in view of the lack of perfect correlation between the $\mathbf{f}_{t+1}$ and the $\boldsymbol{\varepsilon}_{N,t+1}$, one derives the solution:

$$\begin{aligned}\begin{pmatrix} \mathbf{b}_t \\ \mathbf{c}_{N,t}\end{pmatrix} &= -\mu_{m,t}\begin{pmatrix} \boldsymbol{\Omega}_t & \mathbf{P}_{N,t} \\ (\mathbf{B}_{N,t}\boldsymbol{\Omega}_t + \mathbf{P}_{N,t}') & (\boldsymbol{\Sigma}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\Omega}_t)\end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\lambda}_t \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \boldsymbol{\alpha}_{N,t}\end{pmatrix}\\ &= -\mu_{m,t}\begin{pmatrix} \boldsymbol{\Omega}_t^{-1} + \boldsymbol{\Omega}_t^{-1}\mathbf{P}_{N,t}\mathbf{H}_{N,t}^{-1}(\mathbf{B}_{N,t} + \mathbf{P}_{N,t}'\boldsymbol{\Omega}_t^{-1}) & -\boldsymbol{\Omega}_t^{-1}\mathbf{P}_{N,t}\mathbf{H}_{N,t}^{-1} \\ -\mathbf{H}_{N,t}^{-1}(\mathbf{B}_{N,t} + \mathbf{P}_{N,t}'\boldsymbol{\Omega}_t^{-1}) & \mathbf{H}_{N,t}^{-1}\end{pmatrix}\begin{pmatrix} \boldsymbol{\lambda}_t \\ \mathbf{B}_{N,t}\boldsymbol{\lambda}_t + \boldsymbol{\alpha}_{N,t}\end{pmatrix}\\ &= -\mu_{m,t}\begin{pmatrix} \boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t - \boldsymbol{\Omega}_t^{-1}\mathbf{P}_{N,t}\mathbf{H}_{N,t}^{-1}(\boldsymbol{\alpha}_{N,t} - \mathbf{P}_{N,t}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t) \\ \mathbf{H}_{N,t}^{-1}(\boldsymbol{\alpha}_{N,t} - \mathbf{P}_{N,t}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\lambda}_t)\end{pmatrix}.\end{aligned}$$

We now establish the result for the projection SDF m^*_{t+1}. By construction, setting $\mathbf{X}_{N,t+1} = (1, \mathbf{R}'_{N,t+1})'$ and $\boldsymbol{\mu}_{N,t} = \mathbb{E}_t(\mathbf{R}_{N,t+1})$,

$$
\begin{aligned}
m^*_{t+1} &= \mathbb{E}_t(m_{t+1}\mathbf{X}'_{N,t+1})(\mathbb{E}_t(\mathbf{X}_{N,t+1}\mathbf{X}'_{N,t+1})^{-1}\mathbf{X}_{N,t+1} \\
&= (\mu_{m,t}, \mu_{m,t}\boldsymbol{\mu}'_{N,t} + \mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}] + \mathbf{c}'_{N,t}[\boldsymbol{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t}]) \begin{pmatrix} 1 + \boldsymbol{\mu}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\mu}_{N,t} & -\boldsymbol{\mu}'_{N,t}\mathbf{V}^{-1}_{N,t} \\ -\mathbf{V}^{-1}_{N,t}\boldsymbol{\mu}_{N,t} & \mathbf{V}^{-1}_{N,t} \end{pmatrix} \mathbf{X}_{t+1} \\
&= (\mu_{m,t} - (\mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}] + \mathbf{c}'_{N,t}[\boldsymbol{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t}])\mathbf{V}^{-1}_{N,t}\boldsymbol{\mu}_{N,t}, (\mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}] + \mathbf{c}'_{N,t}\boldsymbol{\Sigma}_{N,t})\mathbf{V}^{-1}_{N,t})\mathbf{X}_{t+1} \\
&= \mu_{m,t} + (\mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}] + \mathbf{c}'_{N,t}[\boldsymbol{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t}])\mathbf{V}^{-1}_{N,t}(\mathbf{R}_{N,t+1} - \boldsymbol{\mu}_{N,t}),
\end{aligned}
$$

where we use the block formula for the inverse of a square matrix to $E_t(\mathbf{X}_{t+1}\mathbf{X}'_{t+1})$, which exists in view of our assumption of not-perfect correlation between observed factors and idiosyncratic shocks. Finally, by means of algebraic manipulations,

$$
\mathbf{b}'_t[\boldsymbol{\Omega}_t\mathbf{B}'_{N,t} + \mathbf{P}_{N,t}] + \mathbf{c}'_{N,t}[\boldsymbol{\Sigma}_{N,t} + \mathbf{P}'_{N,t}\mathbf{B}'_{N,t}] = \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t} + \boldsymbol{\alpha}'_{N,t}.
$$

Proof of Theorem 5.2

We have that:

$$
\begin{aligned}
0 = E_t(R^e_{n,t+1}m_{t+1}) &= E_t(R^e_{n,t+1})E_t(m_{t+1}) + \mathrm{cov}_t(R^e_{n,t+1}, m_{t+1}) \\
&= E_t(R^e_{n,t+1})\mu_{m,t} + \mathrm{cov}_t(R^e_{n,t+1}, m^\alpha_{t+1} + m^\beta_{t+1}) \\
&= E_t(R^e_{n,t+1})\mu_{m,t} + \mathrm{cov}_t(R^e_{n,t+1}, m^\alpha_{t+1}) + \mathrm{cov}_t(R^e_{n,t+1}, m^\beta_{t+1}),
\end{aligned}
$$

and re-arranging

$$
\begin{aligned}
E_t(R^e_{n,t+1}) &= -\frac{\mathrm{cov}_t(R^e_{n,t+1}, m^\alpha_{t+1})}{\mathrm{var}_t(m^\alpha_{t+1})}\frac{\mathrm{var}_t(m^\alpha_{t+1})}{\mu_{m,t}} - \frac{\mathrm{cov}_t(R^e_{n,t+1}, m^\beta_{t+1})}{\mu_{m,t}} \\
&= -\frac{\mathrm{cov}_t(R^e_{n,t+1}, m^\alpha_{t+1})}{\mathrm{var}_t(m^\alpha_{t+1})}\frac{\mathrm{var}_t(m^\alpha_{t+1})}{\mu_{m,t}} - \mathrm{cov}_t(R^e_{n,t+1}, \mathbf{f}'_{t+1})\mathrm{cov}_t^{-1}(\mathbf{f}_{t+1})\frac{\mathrm{cov}_t(\mathbf{f}_{t+1})\mathbf{b}_t}{\mu_{m,t}}.
\end{aligned}
$$

The result then follows.

Proof of Theorem 6.1

The result follows by applying Hannan (1970, Ch. II.7) to the SDF $m(f_{t+1})$ relying on the square-integrability assumption.

B Decomposition of SDF Return in Terms of α- and β-Portfolio Returns

In the theorem below, we decompose the SDF return in excess of the risk-free rate, R^*_{t+1}, in terms of the return on the $\mathbf{w}^\alpha_N$ and $\mathbf{w}^\beta_N$ portfolios, both of which are inefficient. In particular, we show that we can span the entire SDF frontier based on the return of these two portfolios. The theorem below provides the correction, $m^{\alpha*}_{t+1}$, required to make the possibly wrong $m^{\beta*}_{t+1}$ admissible, as a function only of the return on the $\mathbf{w}^\alpha_N$ portfolio.

Theorem B.1 (Decomposition of the excess return in terms of return on $\mathbf{w}^\alpha_N$ and $\mathbf{w}^\beta_N$ portfolios)**.** *Under Assumptions 3.1 and 3.2, for any $\mu_{m,t}$, R^*_{t+1} satisfies:*

$$R^*_{t+1} - R_{ft} = \phi^\alpha_t \mathbf{w}^{\alpha\prime}_{0,t}\mathbf{R}^e_{N,t+1} + \phi^\beta_t \mathbf{w}^{\beta\prime}_{0,t}\mathbf{R}^e_{N,t+1} + (\frac{1}{\mu_{m,t}} - R_{ft}),$$

where we set $\phi^\alpha_t = \frac{\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}}{R_{ft}\mathbb{E}((m^*_{t+1})^2)}$, $\phi^\beta_t = \frac{\mu_{m,t}\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t}{R_{ft}\mathbb{E}((m^*_{t+1})^2)}$ and $\mathbf{w}^\alpha_{\mu^*,t}$ and $\mathbf{w}^\beta_{\mu^*,t}$ are the α-portfolio and the β-portfolio, for given target mean μ^*, respectively:

$$\mathbf{w}^\alpha_{\mu^*,t} = \frac{(\mu^* - R_{ft})}{\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}, \qquad \mathbf{w}^\beta_{\mu^*,t} = \frac{(\mu^* - R_{ft})}{\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t,$$

with excess return $R^\alpha_{t+1} - R_{ft} = \mathbf{w}^{\alpha\prime}_{0,t}\mathbf{R}^e_{N,t+1}$ and $R^\beta_{t+1} - R_{ft} = \mathbf{w}^{\beta\prime}_{0,t}\mathbf{R}^e_{N,t+1}$.

Proof: Using (15), and subtracting R_{ft} from both sides,

$$\begin{aligned}
R^*_{t+1} - R_{ft} &= \kappa^\alpha\Big(R^{\alpha*}_{t+1} - R_{ft}\Big) + (1-\kappa^\alpha)\Big(R^{\beta*}_{t+1} - R_{ft}\Big) \\
&= \frac{\kappa^\alpha}{\mathbb{E}(m^*_{t+1}m^{\alpha*}_{t+1})}\Big(- \mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}(\mathbf{R}^e_{N,t+1} - \boldsymbol{\alpha}_{N,t} - \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) - R_{ft}\mu^2_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big) \\
&+ \frac{(1-\kappa^\alpha)}{\mathbb{E}(m^*_{t+1}m^{\beta*}_{t+1})}\Big(\mu_{m,t} - \mu_{m,t}\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}(\mathbf{R}^e_{N,t+1} - \boldsymbol{\alpha}_{N,t} - \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) \\
&\quad - R_{ft}\mu^2_{m,t}(1 + \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big) \\
&= \frac{\kappa^\alpha\mu_{m,t}}{\mathbb{E}(m^*_{t+1}m^{\alpha*}_{t+1})}\Big(- \boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}(\mathbf{R}^e_{N,t+1} - \boldsymbol{\alpha}_{N,t} - \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) - R_{ft}\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big) \\
&+ \frac{(1-\kappa^\alpha)\mu_{m,t}}{\mathbb{E}(m^*_{t+1}m^{\beta*}_{t+1})}\Big(1 - \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}(\mathbf{R}^e_{N,t+1} - \boldsymbol{\alpha}_{N,t} - \mathbf{B}_{N,t}\boldsymbol{\lambda}_t) - R_{ft}\mu_{m,t}(1 + \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big)
\end{aligned}$$

$$= \frac{\kappa^\alpha \mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\alpha*}_{t+1})}\Big(- \boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1} - (R_{ft}\mu_{m,t} - 1)\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big)$$

$$+ \frac{(1-\kappa^\alpha)\mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\beta*}_{t+1})}\Big(- \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1} - (R_{ft}\mu_{m,t} - 1)(1 + \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big)$$

$$= \frac{\kappa^\alpha \mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\alpha*}_{t+1})}\Big(- \boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1}\Big) + \frac{(1-\kappa^\alpha)\mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\beta*}_{t+1})}\Big(- \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1}\Big)$$

$$- \frac{\mu_{m,t}(R_{ft}\mu_{m,t} - 1)}{\mathbb{E}((m^*_{t+1})^2)}\Big(1 + (\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)'\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)\Big)$$

$$= \frac{\kappa^\alpha \mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\alpha*}_{t+1})}\Big(- \boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1}\Big) + \frac{(1-\kappa^\alpha)\mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\beta*}_{t+1})}\Big(- \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1}\Big)$$

$$= \frac{\kappa^\alpha \mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\alpha*}_{t+1})}\Big(- \boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1}\Big) + \frac{(1-\kappa^\alpha)\mu_{m,t}}{\mathbb{E}(m^*_{t+1} m^{\beta*}_{t+1})}\Big(- \boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{R}^e_{N,t+1}\Big) + (\frac{1}{\mu_{m,t}} - R_{ft})$$

$$= \frac{\mu_{m,t}\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}}{R_{ft}\mathbb{E}((m^*_{t+1})^2)}\mathbf{w}^{\alpha'}_{0,t}\mathbf{R}^e_{N,t+1} + \frac{\mu_{m,t}\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t}{R_{ft}\mathbb{E}((m^*_{t+1})^2)}\mathbf{w}^{\beta'}_{0,t}\mathbf{R}^e_{N,t+1} + (\frac{1}{\mu_{m,t}} - R_{ft})$$

$$= \phi^\alpha_t \mathbf{w}^{\alpha'}_{0,t}\mathbf{R}^e_{N,t+1} + \phi^\beta_t \mathbf{w}^{\beta'}_{0,t}\mathbf{R}^e_{N,t+1} + (\frac{1}{\mu_{m,t}} - R_{ft})$$

$$= \phi^\alpha_t (R^\alpha_{t+1} - R_{ft}) + \phi^\beta_t (R^\beta_{t+1} - R_{ft}) + (\frac{1}{\mu_{m,t}} - R_{ft}).$$

Remark B.1.1. When a risk-free asset is traded, $\mu_{m,t} = R^{-1}_{ft}$ and one obtains:

$$R^*_{t+1} - R_{ft} = \phi^\alpha_t \mathbf{w}^{\alpha\prime}_{0,t}\mathbf{R}^e_{N,t+1} + \phi^\beta_t \mathbf{w}^{\beta\prime}_{0,t}\mathbf{R}^e_{N,t+1}, \quad \text{with}$$

$$\phi^\alpha_t = \frac{\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\boldsymbol{\alpha}_{N,t}}{1 + (\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)'\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)}, \phi^\beta_t = \frac{\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t}{1 + (\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)'\mathbf{V}^{-1}_{N,t}(\boldsymbol{\alpha}_{N,t} + \mathbf{B}_{N,t}\boldsymbol{\lambda}_t)}.$$

Moreover, as $N \to \infty$, one obtains $\phi^\alpha_t + \phi^\beta_t \to_p 1$ because $\boldsymbol{\alpha}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t} \to_p \mathbf{0}$ by Lemma A.7.

Remark B.1.2. Note that the formulae used for the α- and β-portfolios are slightly different from the ones adopted by Raponi, Uppal, and Zaffaroni (2019). In fact, here we use the *population* values for $\boldsymbol{\alpha}_{N,t}$, as opposed to the finite-N projection. However, as $N \to \infty$, the formulae for the β-portfolio coincide, because, element by element,

$$\mathbf{w}^\beta_{\mu^*,t} = \frac{(\mu^* - R_{ft})}{\boldsymbol{\lambda}'_t\mathbf{B}'_{N,t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t \to_p \frac{(\mu^* - R_{ft})}{\boldsymbol{\lambda}'_t\boldsymbol{\Omega}^{-1}_t\boldsymbol{\lambda}_t}\mathbf{V}^{-1}_{N,t}\mathbf{B}_{N,t}\boldsymbol{\lambda}_t.$$

C Different Forms of SDF Misspecification

In this section, we describe the different forms of misspecification affecting the SDF. Recall from our results above that any admissible SDF can be expressed as the sum of two components:

$$m_{t+1} = m^{\beta}_{t+1} + m^{\alpha}_{t+1},$$

where for simplicity, we have not included an orthogonal component (with zero price), which would arise if markets were not complete. In the equation above, m^{β}_{t+1} represents the conventional SDF that has the factor representation in (16); that is:

$$\begin{aligned} m^{\beta}_{t+1} &= \mathbb{E}_t(m^{\beta}_{t+1}) + \mathbf{b}_t'\big(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})\big) \\ &= \frac{1}{\lambda_{0,t}} - \frac{1}{\lambda_{0,t}}\boldsymbol{\lambda}_{1,t}'\,\boldsymbol{\Omega}_t^{-1}\big(\mathbf{f}_{t+1} - \mathbb{E}_t(\mathbf{f}_{t+1})\big). \end{aligned}$$

The m^{α}_{t+1} component represents the correction required in order to obtain the admissibility of m_{t+1}; only in the case of zero pricing error in the return-generating process would this term equal to zero. Alternatively, when the pricing errors are non-zero , then m^{α}_{t+1} takes the following form:

$$m^{\alpha}_{t+1} = -\frac{1}{\lambda_{0,t}}\boldsymbol{\alpha}_{N,t}'\boldsymbol{\Sigma}^{-1}_{N,t}\boldsymbol{\varepsilon}_{N,t+1}.$$

Below, we describe four forms of model misspecification that are captured by the framework we have described in the paper. The first is related to the "beta" component of the SDF. The second, third, and fourth are related to the "alpha" component of SDF, arising from the presence of: a pricing error that is unrelated to factors, missing factors, and mismeasured factors.

Case 1: Pure pricing errors (unrelated to factors)

Next, consider the case where all factors are observed without error, but expected returns have pricing errors (that is, returns depend also on non-factor-related characteristics), denoted by $\mathbf{a}_{N,t} \neq \mathbf{0}$, implying that

$$m^{\alpha}_{t+1} = -\frac{1}{\lambda_{0,t}}(\mathbf{a}_{N,t})'\boldsymbol{\Sigma}^{-1}_{N,t}\,\boldsymbol{\varepsilon}_{N,t+1}.$$

In contrast to the next case, in this setting $\mathbf{a}_{N,t}$ does not influence the variance-covariance matrix of returns, $\mathbf{V}_{N,t}$.

Case 2: Missing factors

Suppose now that of the true K^0 factors, only K are observed and $p = K^0 - K > 0$ are missing and suppose for simplicity that the observed and missing factors are uncorrelated. For simplicity, we assume $\mathbf{a}_{N,t} = \mathbf{0}$. Then,

$$\begin{aligned} m_{t+1}^{\alpha} &= -\frac{1}{\lambda_{0,t}} \boldsymbol{\alpha}_{N,t}' \boldsymbol{\Sigma}_{N,t}^{-1} \boldsymbol{\varepsilon}_{N,t+1} \\ &= -\frac{1}{\lambda_{0,t}} (\mathbf{A}_{N,t}\, \boldsymbol{\lambda}_{miss,t})' \, (\mathbf{A}_{N,t} \mathbf{A}_{N,t}' + \mathbf{C}_{N,t})^{-1} (\mathbf{a}_{N,t+1} + \mathbf{A}_{N,t}\, \mathbf{z}_{miss,t}). \end{aligned}$$

where $\mathbf{C}_{N,t}$ is the conditional covariance of $\mathbf{u}_{N,t+1}$, $\mathbb{E}_t[\mathbf{z}_{miss,t+1}] = \mathbf{0}$, and $\mathbb{E}_t[\mathbf{z}_{miss,t+1} \mathbf{z}_{miss,t+1}']$ $= \mathbf{I}_p$ to achieve identification.

Case 3: Mismeasured factors

Finally, consider the case where all K^0 factors are measured with error. In particular, the observed factors satisfy $\mathbf{f}_{t+1} = \mathbf{f}_{t+1}^0 + \boldsymbol{\eta}_{t+1}$, where the measurement error $\boldsymbol{\eta}_{t+1}$ has mean $\mathbb{E}_t[\boldsymbol{\eta}_{t+1}] = \boldsymbol{\mu}_{\eta,t}$ and covariance matrix $\mathbb{E}_t[(\boldsymbol{\eta}_{t+1} - \boldsymbol{\mu}_{\eta,t})(\boldsymbol{\eta}_{t+1} - \boldsymbol{\mu}_{\eta,t})'] = \boldsymbol{\Sigma}_{\eta,t}$. Recall that for simplicity we assume $\mathbf{a}_{N,t}^0 = \mathbf{0}$. Then,

$$\begin{aligned} m_{t+1}^{\alpha} &= -\frac{1}{\lambda_{0,t}} \boldsymbol{\alpha}_{N,t}' \boldsymbol{\Sigma}_{N,t}^{-1} \boldsymbol{\varepsilon}_{N,t+1} \\ &= -\frac{1}{\lambda_{0,t}} (-\mathbf{B}_{N,t} \boldsymbol{\mu}_{\eta,t})' (\mathbf{B}_{N,t} \boldsymbol{\Sigma}_{\eta,t} \mathbf{B}_{N,t}' + \mathbf{C}_{N,t})^{-1} \big(\boldsymbol{\varepsilon}_{N,t+1} - \mathbf{B}_{N,t} (\boldsymbol{\eta}_{t+1} - \boldsymbol{\mu}_{\eta,t})\big). \end{aligned}$$

Case 4: Incorrect mean of the SDF

Above, we have looked at the cases where model misspecification is present in the return-generating process for the risky assets, expressed in terms of the excess return of the risky assets net of the zero-beta rate, $\gamma_{0,t}^0$. In practice, unless a risk-free rate is traded, we need to use one of the three alternatives discussed above, each of which could be misspecified; that is, we choose $\gamma_{0,t} \neq \gamma_{0,t}^0$. This would affect the specification of both m_{t+1}^{α} and m_{t+1}^{β}.

Case 5: Incorrect functional form of the SDF

For example, if one were working with a representative-agent model, then the misspecification could arise from an erroneous specification of the utility function or the state variables,

or it could arise from using a Taylor-series expansion of a nonlinear SDF instead of the exact model-implied nonlinear SDF.

D The SDF: Basic Notions

Just as in Hansen and Jagannathan (1997), we consider asset-market transactions that take place at two dates, t and $t+1$. There are N financial assets that are traded at date t. Each asset delivers a payoff at date $t+1$. We let $\mathbf{p}_{N,t} = (p_{1t}, \ldots, p_{Nt})'$ denote the vector of prices and $\mathbf{x}_{N,t+1}$ the corresponding vector of random payoffs on these N assets.[35] This single period between t and $t+1$ is replicated over time in a stationary manner.[36]

In addition to the primitive assets described above, the payoff space also includes new payoffs that can be formed from portfolios of the primitive assets. We assume that investors can form any portfolio of traded assets.

Assumption D.1 (Portfolio formation)**.** *Let $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_{t+1}$ denote the space of portfolio payoffs, which includes the primitive payoffs $\mathbf{x}_{N,t+1}$ along with arbitrary portfolios constructed with these primitive payoffs; that is, $\boldsymbol{\mathcal{X}}$ is a linear space. Moreover, we assume that it is closed.*

Assumption D.2 (Law of one price)**.** *The price $p(\cdot)$ of a portfolio payoff is a linear functional on $\boldsymbol{\mathcal{X}}$, $p(a\, x_{n,t+1} + b\, x_{n',t+1}) = a\, p(x_{n,t+1}) + b\, p(x_{n',t+1})$, which is continuous at every point (including zero).*

In order to rule out nontrivial pricing functions, we impose the following assumption, which implies that there is at least one payoff whose price is non-zero.

Assumption D.3 (Nontrivial price)**.** *There exists a payoff $x_{n,t+1} \in \boldsymbol{\mathcal{X}}$ for which we have $prob_t(p(x_{n,t+1}) = 0) = 0$.*

[35] To make clear the dependence on the number of assets, we index quantities that are N-dimensional by the subscript N.

[36] The *payoff space* $\boldsymbol{\mathcal{X}}$ is the set of all the payoffs that investors can receive at the end of each period. If, for example, there are S discrete states at date $t+1$, then the payoff of asset n in state s is denoted by x_{ns}, in which case

$$\boldsymbol{\mathcal{X}} = \begin{pmatrix} x_{11} & \ldots & x_{1S} \\ \vdots & \vdots & \vdots \\ x_{N1} & \ldots & x_{NS} \end{pmatrix}.$$

Although not made explicit in our main analysis, in order to ensure that prices are well defined, we are implicitly modeling portfolio payoffs that are elements of a Hilbert space, which is included in the space L^2_{t+1} of all random variables with finite second moments conditional on information at date t. Moreover, L^2_{t+1} is endowed with the usual inner product and norm: for any $h_1, h_2 \in L^2_{t+1}$

$$\langle h_1 | h_2 \rangle = \mathbb{E}_t(h_1 h_2) \quad \text{and} \quad \| h_1 \| = \langle h_1 | h_1 \rangle^{1/2}.$$

Definition D.1 (Admissible SDF)**.** *An* admissible *SDF is a random variable m_{t+1} with finite second moment such that the expected price of a payoff $x_{n,t+1}$ can be represented as the inner product of the payoff and m_{t+1}:*

$$p(x_{n,t+1}) = \mathbb{E}_t(x_{n,t+1}\, m_{t+1}) \quad \forall x_{n,t+1} \in \boldsymbol{\mathcal{X}},$$

where $\boldsymbol{\mathcal{M}}$ is the set of all admissible SDFs.

For certain results, it will be important to rule out arbitrage opportunities in an economy with a finite number of risky assets, N. Below, we define a notion of no arbitrage for the price functional, $p(\cdot)$ on $\boldsymbol{\mathcal{X}}$.

Definition D.2 (No arbitrage opportunities; Hansen and Richard (1987, Definition 2.4))**.** *A price functional $p(\cdot)$ has no arbitrage opportunities on $\boldsymbol{\mathcal{X}}$ if for any payoff $x_{n,t+1} \in \boldsymbol{\mathcal{X}}$ for which $prob_t(x_{n,t+1} > 0) = 1$, then $prob_t(\{p(x_{n,t+1}) \leq 0\} \cap \{x_{n,t+1} > 0\}) = 0$, where $prob_t(\cdot)$ denotes the probability conditional on information at date t.*

Given Assumptions D.1 and D.2, Hansen and Richard (1987, Theorem 2.1) and Hansen and Jagannathan (1991) show that there exists an admissible *stochastic discount factor* (SDF), that is, the *unique* payoff $m_t^* \in \boldsymbol{\mathcal{X}}$ such that $\mathbf{p}_{N,t} = p(\mathbf{x}_{N,t+1}) = \mathbb{E}_t(m_{t+1}^* \mathbf{x}_{N,t+1})$ for all $\mathbf{x}_{N,t+1} \in \boldsymbol{\mathcal{X}}$. This m_{t+1}^* is:

$$m_{t+1}^* = \mathbf{p}_{N,t}' \left[\mathbb{E}_t(\mathbf{x}_{N,t+1}\mathbf{x}_{N,t+1}')\right]^{-1} \mathbf{x}_{N,t+1}. \tag{D4}$$

If financial markets are complete, there is no other admissible SDF. On the other hand, if markets are incomplete, then there are an infinite number of SDFs such that $m_{t+1} = m_{t+1}^* + \epsilon_{t+1}$ where $\mathbb{E}_t(\epsilon_{t+1}\mathbf{x}_{N,t+1}) = \mathbf{0}_N$ for all $\mathbf{x}_{N,t+1} \in \boldsymbol{\mathcal{X}}$. Observe that m_{t+1}^* is the projection of any admissible SDF on the space of payoffs $\boldsymbol{\mathcal{X}}$: the pricing implication of any SDF is the same as those of its projection on $\boldsymbol{\mathcal{X}}$:[37]

$$\mathbf{p}_{N,t} = \mathbb{E}_t(m_{t+1}\mathbf{x}_{N,t+1}) = \mathbb{E}_t\Big(\big[\,\mathrm{proj}(m_{t+1}|\boldsymbol{\mathcal{X}}) + \epsilon_{N,t+1}\big]\,\mathbf{x}_{N,t+1}\Big) = \mathbb{E}_t\big(\,\mathrm{proj}(m_{t+1}|\boldsymbol{\mathcal{X}})\,\mathbf{x}_{N,t+1}\big),$$

where $\mathrm{proj}(Y|\boldsymbol{\mathcal{X}}) = [\mathbb{E}_t(\mathbf{x}_{N,t+1}Y)]' \left[\mathbb{E}_t(\mathbf{x}_{N,t+1}\mathbf{x}_{N,t+1}')\right]^{-1} \mathbf{x}_{N,t+1}$, for any $Y \in L_{t+1}^2$. Observe that m_{t+1}^* is the minimum-variance SDF that lies on the boundary of the SDF frontier identified in Hansen and Jagannathan (1991).

[37] Note that when Assumption D.2 holds, then $p(\cdot)$ has no arbitrage opportunities in $\boldsymbol{\mathcal{X}}$ if and only if $\mathrm{prob}_t(m_{t+1}^* > 0) = 1$; see Hansen and Richard (1987, Lemma 2.3).

There are several representations of the SDF m^*_{t+1}, depending on the nature of the payoffs. One is in terms of generic payoffs, as in Hansen and Jagannathan (1991, eq. (11)):

$$m^*_{t+1} = \mathbb{E}_t(m^*_{t+1}) + \big[\mathbf{p}_{N,t} - \mathbb{E}_t(m^*_{t+1})\mathbb{E}_t(\mathbf{x}_{N,t+1})\big]' \operatorname{cov}_t^{-1}(\mathbf{x}_{N,t+1}) \big(\mathbf{x}_{t+1} - \mathbb{E}_t(\mathbf{x}_{N,t+1})\big),$$

where $\operatorname{cov}_t(\mathbf{x}_{N,t+1}) = \mathbb{E}_t\big[\big(\mathbf{x}_{N,t+1} - \mathbb{E}_t(\mathbf{x}_{N,t+1})\big)\big(\mathbf{x}_{N,t+1} - \mathbb{E}_t(\mathbf{x}_{N,t+1})\big)'\big]$ is the covariance matrix of payoffs.

Given our assumption that a risk-free asset is available, then

$$\mathbb{E}_t(m^*_{t+1}) = \frac{1}{R_{ft}},$$

which allows us to obtain the following representation:

$$m^*_{t+1} = \frac{1}{R_{ft}} - \frac{1}{R_{ft}}\, \mathbb{E}_t(\mathbf{R}^e_{N,t+1})' \operatorname{cov}_t^{-1}(\mathbf{R}^e_{N,t+1}) \big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big), \tag{D5}$$

where $\operatorname{cov}_t(\mathbf{R}^e_{N,t+1}) = \mathbb{E}_t\big[\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big)\big(\mathbf{R}^e_{N,t+1} - \mathbb{E}_t(\mathbf{R}^e_{N,t+1})\big)'\big]$ is the covariance matrix of excess returns, $\mathbf{R}^e_{N,t+1} = \mathbf{R}_{N,t+1} - R_{ft}\mathbf{1}_N$.[38] Given that gross returns are given by $\mathbf{R}_{N,t+1} = \mathbf{R}^e_{N,t+1} + R_{ft}\mathbf{1}_N$, then by the law of one price, the SDF in (D5) can also price gross returns. That is,

$$p(\mathbf{R}_{N,t+1}) = p(\mathbf{R}^e_{N,t+1}) + R_{ft}\, p(\mathbf{1}_N) = R_{ft}\, p(\mathbf{1}_N) = R_{ft}\, \frac{1}{R_{ft}} \mathbf{1}_N = \mathbf{1}_N.$$

[38] Note that when studying excess returns, one needs to consider the projection on the payoff space of excess returns as well as 1; otherwise, equation (D4) would imply that $m^*_{t+1} = 0$.

E Estimation of the Extended APT

In this section, we explain how to estimate the extended APT model of returns. Our arguments apply to virtually any (parametric) estimation procedure, but we will illustrate it with respect to the (pseudo) Gaussian ML estimator; the estimation could also be done using a Bayesian approach. The Gaussian ML estimator is a natural estimator for our model when the first two moments of asset returns are specified correctly, although distributional assumptions (such as normality) are not required; hence, the use of pseudo ML.

However, because the APT restriction could be violated leading to arbitrage opportunities, one needs to consider the maximum-likelihood estimator subject to the APT restriction. Moreover, not only does the APT restriction lead to a more precise estimator of $\boldsymbol{\alpha}_N$ compared to the *unconstrained* estimator, but it provides exactly the condition required to econometrically identify the extended APT, as demonstrated in the theorem below; that is, $\boldsymbol{\lambda}_{miss,t}$ and $\boldsymbol{\alpha}_{N,t}$ *cannot* be identified separately unless the APT restriction is imposed.

Finally, for simplicity let us assume that *all* conditional moments are constant.

Assume that

$$\mathbf{R}^e_{N,t+1} = \boldsymbol{\alpha}_N + \mathbf{B}_{1N}(\boldsymbol{\lambda}_1 + \mathbf{f}_{1t+1} - \mathbb{E}(\mathbf{f}_{1t+1})) + \mathbf{B}_{2N}\mathbf{f}^e_{2t+1} + \boldsymbol{\varepsilon}_{t+1},$$

with

$$\boldsymbol{\alpha}_N = \mathbf{a}_N + \mathbf{A}_N\boldsymbol{\lambda}_{miss}\,\mathrm{var}(\mathbf{R}^e_{N,t+1}) = \mathbf{V}_N = \mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}'_N + \mathbf{A}_N\mathbf{A}'_N + \mathbf{C}_N,$$

where we set $\mathbf{B}_N = (\mathbf{B}_{1N}, \mathbf{B}_{2N})$, $\boldsymbol{\Omega} = \mathrm{var}(\mathbf{f}_{t+1})$, $\mathbf{f}_{t+1} = (\mathbf{f}'_{1t+1}, \mathbf{f}^{e\prime}_{2t+1})'$, with $\mathbf{f}_{1t+1}$ denoting the set of K_1 non-traded observed factors and $\mathbf{f}^e_{2t+1}$ the set of K_2 traded observed factors, expressed as excess returns, where $K = K_1 + K_2$. Note that, for simplicity, we initially assume that the missing factors are uncorrelated with the observed factors, and later discuss the case where they are correlated. Given that $\mathbf{f}^e_{2t}$ are excess returns on traded assets, their risk premia satisfy $\boldsymbol{\lambda}_2 = \mathbb{E}(\mathbf{f}^e_{2t+1})$ and, to avoid confusion with the risk premia of the non-traded assets, we will use the expectation formulation for $\boldsymbol{\lambda}_{2t}$.

The joint log-likelihood function takes the following form:[39]

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = &-\frac{1}{2}\log(\det(\tilde{\mathbf{A}}_N\tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)) \\ &-\frac{1}{2T}\sum_{t=1}^{T}\Big(\mathbf{R}^e_{N,t} - \tilde{\mathbf{A}}_N\tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_{1N}(\tilde{\boldsymbol{\lambda}}_1 + \mathbf{f}_{1t} - \widetilde{\mathbb{E}(\mathbf{f}_{1t})}) - \tilde{\mathbf{B}}_{2N}\mathbf{f}^e_{2t}\Big)' \\ &\times(\tilde{\mathbf{A}}_N\tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)^{-1}\Big(\mathbf{R}^e_{N,t} - \tilde{\mathbf{A}}_N\tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_{1N}(\tilde{\boldsymbol{\lambda}}_1 + \mathbf{f}_{1t} - \widetilde{\mathbb{E}(\mathbf{f}_{1t})}) - \tilde{\mathbf{B}}_{2N}\mathbf{f}^e_{2t}\Big) \end{aligned} \tag{E6}$$

[39] Note that $\det(\cdot)$ denotes the determinant, $\mathrm{vec}(\cdot)$ denotes the operator that stacks the columns of a matrix into a single column vector, and $\mathrm{vech}(\cdot)$ denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.

$$-\frac{1}{2}\log(\det(\tilde{\boldsymbol{\Omega}}))-\frac{1}{2T}\sum_{t=1}^{T}\Big((\mathbf{f}_{1t}',\mathbf{f}_{2t}^{e\prime})-(\widetilde{\mathbb{E}(\mathbf{f}_{1t})}',\widetilde{\mathbb{E}(\mathbf{f}_{2t}^{e})}')\Big)\tilde{\boldsymbol{\Omega}}^{-1}\Big((\mathbf{f}_{1t}',\mathbf{f}_{2t}^{e\prime})-(\widetilde{\mathbb{E}(\mathbf{f}_{1t})}',\widetilde{\mathbb{E}(\mathbf{f}_{2t}^{e})}')\Big)'.$$

Notice that we have expressed the joint distribution as the product of a conditional distribution and a marginal distribution. Relaxing the i.i.d. assumption requires specification of time-varying conditional means, conditional variances, and conditional covariances.

Theorem E.1 (Parameter estimation of extended APT)**.** *Suppose that the vector of asset returns, $\mathbf{R}_{N,t}$, satisfies Assumption 3.1 and that $\boldsymbol{\Sigma}_{f_2^e f_2^e}-\bar{\mathbf{f}}_2^{\,e}\bar{\mathbf{f}}_2^{\,e\prime}$ is nonsingular, where $\boldsymbol{\Sigma}_{f_2^e f_2^e}=T^{-1}\sum_{t=1}^{T}\mathbf{f}_{2t}^{e}\mathbf{f}_{2t}^{e\prime}$ and $\bar{\mathbf{f}}_2^{\,e}=T^{-1}\sum_{t=1}^{T}\mathbf{f}_{2t}^{e}$. Then*

$$\hat{\boldsymbol{\theta}}_{MLC}=\underset{\tilde{\boldsymbol{\theta}}}{\operatorname{argmax}}\;\; L(\tilde{\boldsymbol{\theta}})\quad \textit{subject to}\quad \tilde{\mathbf{a}}_N'\tilde{\boldsymbol{\Sigma}}_N^{-1}\tilde{\mathbf{a}}_N\le\delta,$$

where $L(\tilde{\boldsymbol{\theta}})$ is defined in (E6)*, and $\hat{\boldsymbol{\theta}}_{MLC}=(\hat{\mathbf{a}}_{N,MLC}',\ \hat{\boldsymbol{\lambda}}_{miss,MLC}',\ \hat{\boldsymbol{\lambda}}_{1,MLC}',\ \widehat{E(\mathbf{f}_{1t})}_{MLC}'\ ,\ \widehat{E(\mathbf{f}_{2t}^{e})}_{MLC}',\ \mathrm{vec}(\hat{\mathbf{A}}_{N,MLC})',\ \mathrm{vec}(\hat{\mathbf{B}}_{N,MLC})',\ \mathrm{vech}(\hat{\mathbf{C}}_{N,MLC})',\ \mathrm{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$.*

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

$$\mathbf{D}_N=(\mathbf{A}_N,\mathbf{B}_{1N}),\ \ \boldsymbol{\lambda}=(\boldsymbol{\lambda}_{miss}',\boldsymbol{\lambda}_1')',$$

then

$$\begin{aligned}
\mathrm{vec}(\hat{\mathbf{B}}_{2N,MLC})&=\Big((\boldsymbol{\Sigma}_{f_2^e f_2^e}\otimes\mathbf{I})-(\bar{\mathbf{f}}_2^{\,e}\bar{\mathbf{f}}_2^{\,e\prime}\otimes(2\mathbf{G}_N-\mathbf{G}_N\mathbf{G}_N))\Big)^{-1}\mathrm{vec}\Big(\boldsymbol{\Sigma}_{hf_2^e}-(2\mathbf{G}_N-\mathbf{G}_N\mathbf{G}_N)\bar{\mathbf{h}}_N\bar{\mathbf{f}}_2^{\,e\prime}\Big),\\
\hat{\boldsymbol{\lambda}}_{MLC}&=(\hat{\mathbf{D}}_{N,MLC}'\hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}\hat{\mathbf{D}}_{N,MLC})^{-1}\hat{\mathbf{D}}_{N,MLC}'\hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}\Big(\bar{\mathbf{h}}_N-\hat{\mathbf{B}}_{2N,MLC}\,\bar{\mathbf{f}}_2^{\,e}\Big),\\
\hat{\mathbf{a}}_{N,MLC}&=\frac{1}{\hat{\kappa}+1}\Big(\bar{\mathbf{h}}_N-\hat{\mathbf{B}}_{2N,MLC}\bar{\mathbf{f}}_2^{\,e}-\hat{\mathbf{D}}_{N,MLC}\hat{\boldsymbol{\lambda}}_{MLC}\Big),
\end{aligned}\tag{E7}$$

where $\hat{\boldsymbol{\Sigma}}_{N,MLC}=\hat{\mathbf{A}}_{N,MLC}\hat{\mathbf{A}}_{N,MLC}'+\hat{\mathbf{C}}_{N,MLC}$, $\boldsymbol{\Sigma}_{hf_2^e}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t\mathbf{f}_{2t}^{e\prime}$, $\bar{\mathbf{h}}_N=\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t$ with $\mathbf{h}_t=\mathbf{R}_{N,t}^{e}-\hat{\mathbf{B}}_{1N,MLC}(\mathbf{f}_{1t}-\bar{\mathbf{f}}_{1t})$, and

$$\mathbf{G}_N=\frac{1}{(\hat{\kappa}+1)}\mathbf{I}_N+\frac{\hat{\kappa}}{(\hat{\kappa}+1)}\hat{\mathbf{D}}_{N,MLC}(\hat{\mathbf{D}}_{N,MLC}'\hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}\hat{\mathbf{D}}_{N,MLC})^{-1}\hat{\mathbf{D}}_{N,MLC}'\hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}.$$

Note that $\hat{\mathbf{D}}_{N,MLC}=(\hat{\mathbf{A}}_{N,MLC},\hat{\mathbf{B}}_{1N,MLC})$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution and, as before, $\widehat{E(\mathbf{f}_t)}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ coincide with the sample mean and sample covariance of the observed factors $\mathbf{f}_t=(\mathbf{f}_{1t}',\mathbf{f}_{2t}')'$.

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

$$\hat{\boldsymbol{\alpha}}_{N,MLC}=\bar{\mathbf{R}}_N^{e}-\hat{\mathbf{B}}_{2N,MLC}\bar{\mathbf{f}}_2^{\,e},$$

and the expression for $\text{vec}(\hat{\mathbf{B}}_{2N,MLC})$ *can be obtained by setting* $\hat{\kappa} = 0$ *in the terms that appear in* (E7)*. The expressions for* $\widehat{E(\mathbf{f}_t)}_{MLC}$ *and* $\hat{\mathbf{\Omega}}_{MLC}$ *are unchanged, and, as before, the expressions for the estimators of* $\hat{\mathbf{D}}_{N,MLC}$ *and* $\hat{\mathbf{C}}_{N,MLC}$ *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 $\hat{\boldsymbol{\theta}}$ the MLC corresponding to $\tilde{\kappa} = 0$, this is unfeasible whenever we have that $\hat{\mathbf{a}}'_{N,MLC}\hat{\mathbf{\Sigma}}^{-1}_{N,MLC}\hat{\mathbf{a}}_{N,MLC} > \delta$. Similarly, case $\tilde{\kappa} > 0$ is unfeasible whenever, for every $\tilde{\kappa} > 0$,

$$\Big(\bar{\mathbf{R}}^e_N - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}^e_2 - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\Big)'\hat{\mathbf{\Sigma}}^{-1}_{N,\text{MLC}}\Big(\bar{\mathbf{R}}^e_N - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}^e_2 - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\Big) < \delta,$$

because $(1+\hat{\kappa})^2 = \frac{\left[\bar{\mathbf{R}}^e_N - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}^e_2 - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\right]'\hat{\mathbf{\Sigma}}^{-1}_{N,\text{MLC}}\left[\bar{\mathbf{R}}^e_N - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}^e_2 - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\right]}{\delta}$. When both cases are feasible, the optimal value for the Karush-Kuhn-Tucker multiplier $\tilde{\kappa}$ will be greater, or equal to zero, depending on which case maximizes the log-likelihood, namely depending on whether $L(\hat{\boldsymbol{\theta}}_{\text{MLC}})$ or $L(\hat{\boldsymbol{\theta}})$ is largest, respectively. Note that when $\tilde{\kappa} > 0$ then $\hat{\mathbf{a}}'_{N,\text{MLC}}\hat{\mathbf{\Sigma}}^{-1}_{N,\text{MLC}}\hat{\mathbf{a}}_{N,\text{MLC}} = \delta$ 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 $\boldsymbol{\lambda}$, $\mathbf{a}_N$, and the Lagrange multiplier κ, the first $K + N$ equations (after some algebra) are:

$$\begin{pmatrix} \mathbf{D}_N{}'\mathbf{\Sigma}^{-1}_N \\ \mathbf{I}_N \end{pmatrix}\Big(\bar{\mathbf{R}}^e_N - \mathbf{B}_{2N}(\bar{\mathbf{f}}^e_2)\Big) = \begin{pmatrix} \mathbf{D}'_N\mathbf{\Sigma}^{-1}_N\mathbf{D}_N & \mathbf{D}'_{N,\text{MLC}}\mathbf{\Sigma}^{-1}_{N,\text{MLC}} \\ \mathbf{D}_N & (1+\hat{\kappa})\mathbf{I}_N \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\lambda}}_{\text{MLC}} \\ \hat{\mathbf{a}}_{N,\text{MLC}} \end{pmatrix},$$

where recall that $\mathbf{\Sigma}_N = \mathbf{A}_N\mathbf{A}'_N + \mathbf{C}_N$. It is straightforward to see that, because of the APT restriction, $\boldsymbol{\lambda}$ and $\mathbf{a}_N$ 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 $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\mathbf{a}}_{N,\text{MLC}}$ is non-singular for every $\hat{\kappa} > 0$, leading to the closed-form solution:

$$\hat{\boldsymbol{\lambda}}_{\text{MLC}} = (\mathbf{D}'_N\mathbf{\Sigma}^{-1}_N\mathbf{D}_N)^{-1}\mathbf{D}'_N\mathbf{\Sigma}^{-1}_N\Big(\bar{\mathbf{R}}^e_N - \mathbf{B}_{2N}\bar{\mathbf{f}}^e_2\Big),$$

$$\hat{\mathbf{a}}_{N,\text{MLC}} = \frac{1}{\hat{\kappa}+1}\Big(\bar{\mathbf{R}}^e_N - \mathbf{B}_{2N}\bar{\mathbf{f}}^e_2 - \mathbf{D}_N\hat{\boldsymbol{\lambda}}_{\text{MLC}}\Big).$$

Turning now to the first-order condition with respect to the generic (a, b)th element of $\mathbf{B}_{2N}$, denoted by B_{2ab}, one obtains,

$$-\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}'_t\mathbf{\Sigma}^{-1}_N(-\frac{\partial\mathbf{B}_{2N}}{\partial B_{2ab}}\,\mathbf{f}^e_{2t} + \mathbf{G}_N\frac{\partial\mathbf{B}_{2N}}{\partial B_{2ab}}\,\bar{\mathbf{f}}^e_2) = 0, \text{ with } 1 \le a \le N,\ 1 \le b \le K,$$

setting, for simplicity,

$$\mathbf{g}_t = \Big(\mathbf{h}_{N,t} - \mathbf{G}_N\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,\text{MLC}}\mathbf{f}^e_{2t} + \mathbf{G}_N\hat{\mathbf{B}}_{2N,\text{MLC}}\bar{\mathbf{f}}^e_2\Big) \quad \text{and} \quad \mathbf{g} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t,$$

with

$$\mathbf{G}_N = \frac{1}{(\hat{\kappa}+1)}\mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)}\mathbf{D}_N(\mathbf{D}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}'_N\boldsymbol{\Sigma}_N^{-1}.$$

Taking the vec operator for both sides of the first-order condition above gives

$$\frac{1}{T}\sum_{t=1}^{T}(\mathbf{f}^{e\prime}_{2t}\otimes\mathbf{g}'_t\boldsymbol{\Sigma}_N^{-1})\text{vec}\Big(\frac{\partial\mathbf{B}_{2N}}{\partial B_{2ab}}\Big) = (\bar{\mathbf{f}}^{e\prime}_2\otimes\mathbf{g}'\boldsymbol{\Sigma}_N^{-1}\mathbf{G}_N)\text{vec}\Big(\frac{\partial\mathbf{B}_{2N}}{\partial B_{2ab}}\Big), \text{ with } 1\le a\le N,\ 1\le b\le K,$$

which can be rewritten more succinctly as

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}^e_{2t}\mathbf{g}'_t = \bar{\mathbf{f}}^e_2\mathbf{g}'\boldsymbol{\Sigma}_N^{-1}\mathbf{G}_N\boldsymbol{\Sigma}_N.$$

Next, recalling that $\boldsymbol{\Sigma}_{hf^e} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_{N,t}\mathbf{f}^{e\prime}_{2t}$, and $\boldsymbol{\Sigma}_{f^e_2f^e_2} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}^e_{2t}\mathbf{f}^{e\prime}_{2t}$, with $\boldsymbol{\Sigma}_{f^e_2h} = \boldsymbol{\Sigma}'_{hf^e_2}$, one obtains $\boldsymbol{\Sigma}_N^{-1}\mathbf{G}_N\boldsymbol{\Sigma}_N = \frac{1}{(\hat{\kappa}+1)}\mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa}+1)}\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N(\mathbf{D}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}'_N = \mathbf{G}'_N$ and rearranging the above first order-condition gives

$$\boldsymbol{\Sigma}_{f^e_2h} - \bar{\mathbf{f}}^e_2\,\bar{\mathbf{h}}'_N\mathbf{G}'_N - \boldsymbol{\Sigma}_{f^e_2f^e_2}\hat{\mathbf{B}}'_{N,\text{MLC}} + \bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\hat{\mathbf{B}}'_{N,\text{MLC}}\mathbf{G}'_N - (\bar{\mathbf{f}}^e_2\,\bar{\mathbf{h}}'_N - \bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\mathbf{B}'_{2N,\text{MLC}})(\mathbf{I}_N - \mathbf{G}'_N)\mathbf{G}'_N$$

$$= \boldsymbol{\Sigma}_{f^e_2h} - \bar{\mathbf{f}}^e_2\,\bar{\mathbf{h}}'_N(2\mathbf{G}'_N - \mathbf{G}'_N\mathbf{G}'_N) - \boldsymbol{\Sigma}_{f^e_2f^e_2}\hat{\mathbf{B}}'_{N,\text{MLC}} + \bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\hat{\mathbf{B}}'_{2N,\text{MLC}}(2\mathbf{G}'_N - \mathbf{G}'_N\mathbf{G}'_N) = \mathbf{0}.$$

Transposing both sides, taking the vec, and solving for $\text{vec}(\hat{\mathbf{B}}_{2N,\text{MLC}})$ gives

$$\text{vec}(\hat{\mathbf{B}}_{2N,\text{MLC}}) = \Big((\boldsymbol{\Sigma}_{f^e_2f^e_2}\otimes\mathbf{I}_N) - (\bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\otimes(2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N))\Big)^{-1}\text{vec}\Big(\boldsymbol{\Sigma}_{hf^e_2} - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)\bar{\mathbf{h}}^e_N\,\bar{\mathbf{f}}^{e\prime}_2\Big).$$

We need to show that a solution for $\hat{\mathbf{B}}_{2N,\text{MLC}}$ exists. This requires one to establish that the matrix $\Big((\boldsymbol{\Sigma}_{f^e_2f^e_2}\otimes\mathbf{I}_N) - (\bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\otimes(2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N))\Big)$ is invertible. This matrix can be written as

$$\Big((\boldsymbol{\Sigma}_{f^e_2f^e_2}\otimes\mathbf{I}_N) - (\bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\otimes(2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N))\Big) = \Big(((\boldsymbol{\Sigma}_{f^e_2f^e_2} - \bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2)\otimes\mathbf{I}_N) + (\bar{\mathbf{f}}^e_2\,\bar{\mathbf{f}}^{e\prime}_2\otimes(\mathbf{I}_N - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)))\Big).$$

The first matrix on the right hand side is non-singular. One then just needs to show that the second matrix is positive semi-definitive. This follows because, $\mathbf{I}_N - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N) = (\mathbf{I}_N - \mathbf{G}_N)(\mathbf{I}_N - \mathbf{G}_N)$, and we show below that $(\mathbf{I}_N - \mathbf{G}_N)$ is positive semi-definite.

$$\mathbf{I}_N - \mathbf{G}_N = \mathbf{I}_N - \frac{1}{(\hat{\kappa}+1)}\mathbf{I}_N - (\frac{\hat{\kappa}}{1+\hat{\kappa}})\mathbf{D}_N(\mathbf{D}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}'_N\boldsymbol{\Sigma}_N^{-1}$$

$$= (\frac{\hat{\kappa}}{1+\hat{\kappa}})(\mathbf{I}_N - \mathbf{D}_N(\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1})$$
$$= (\frac{\hat{\kappa}}{1+\hat{\kappa}})\boldsymbol{\Sigma}_N(\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N(\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1})$$
$$= (\frac{\hat{\kappa}}{1+\hat{\kappa}})\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^{-1/2}(\mathbf{I}_N - \boldsymbol{\Sigma}_N^{-1/2}\mathbf{D}_N(\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1/2})\boldsymbol{\Sigma}_N^{-1/2}.$$

The right-hand side is the product of positive-definite matrices, namely $\boldsymbol{\Sigma}_N$ and $\boldsymbol{\Sigma}_N^{-1/2}$, and of the matrix $\mathbf{I}_N - \boldsymbol{\Sigma}_N^{-1/2}\mathbf{D}_N(\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1/2}$, which is a projection matrix orthogonal to $\boldsymbol{\Sigma}_N^{-1/2}\mathbf{D}_N$, and therefore, positive semidefinite.

Therefore, plugging $\hat{\mathbf{B}}_{2N,\text{MLC}}$ into $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\mathbf{a}}_{N,\text{MLC}}$, one obtains that

$$\hat{\boldsymbol{\lambda}}_{\text{MLC}} = \hat{\boldsymbol{\lambda}}(\mathbf{D}_N, \mathbf{C}_N), \quad \hat{\mathbf{a}}_{N,\text{MLC}} = \hat{\mathbf{a}}_N(\mathbf{D}_N, \mathbf{C}_N) \quad \text{and} \quad \hat{\kappa} = \hat{\kappa}(\mathbf{D}_N, \mathbf{C}_N).$$

Substituting them, together with $\hat{\mathbf{B}}_{2N,\text{MLC}}$, into $L(\boldsymbol{\theta}) - \kappa(\mathbf{a}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{a}_N - \delta)$, gives the concentrated likelihood function, which is a function of only $\mathbf{D}_N$ and $\mathbf{C}_N$ which will be maximized numerically, providing $\mathbf{D}_{N,\text{MLC}}$ and $\mathbf{C}_{N,\text{MLC}}$.

Suppose now that $\hat{\kappa} = 0$ holds, and recall that in this case the MLC is indicated by $\hat{\boldsymbol{\theta}}$. One can clearly obtain a unique solution for $(\mathbf{D}_N, \mathbf{I}_N)\begin{pmatrix} \boldsymbol{\lambda} \\ \hat{\mathbf{a}}_N \end{pmatrix} = \mathbf{D}_N\boldsymbol{\lambda} + \hat{\mathbf{a}}_N$. However, to solve for $\boldsymbol{\lambda}$ and $\hat{\mathbf{a}}_N$ separately, one needs to invert the matrix

$$\begin{pmatrix} \mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix}(\mathbf{D}_N, \mathbf{I}_N) = \begin{pmatrix} \mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N & \mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1} \\ \mathbf{D}_N & \mathbf{I}_N \end{pmatrix},$$

which is not possible because it is of dimension $(N+K)\times(N+K)$ but of rank N, as the left-hand side shows that it is obtained from the product of two matrices of dimension $(N+K)\times N$. All the other parameters are identified separately, and their expressions follow from differentiating $L(\boldsymbol{\theta})$ and solving the resulting first-order conditions. For instance, the formula for $\hat{\mathbf{B}}_{2N}$ follow by setting $\mathbf{G}_N = \mathbf{I}_N$ into (E7).

The constrained estimator $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$ turns out to be precisely the ridge estimator for $\boldsymbol{\alpha}_N$, unless $\hat{\kappa} = 0$, in which case the OLS estimator is re-obtained. Besides $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$, the estimators of $\hat{\mathbf{B}}_{2N,\text{MLC}}$ and $\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}$ are also functions of $\hat{\kappa}$ because of the APT constraint, in contrast to $\widehat{E(\mathbf{f}_t)}_{\text{MLC}}$ and $\hat{\boldsymbol{\Omega}}_{\text{MLC}}$, which are simply the sample mean and sample covariance of $\mathbf{f}_t$ because the APT constraint does not affect the distribution of the observed factors $\mathbf{f}_t$.

F P-values for Weighted Sum of Squared Pricing Errors

Define the pricing errors:

$$e(m) = \mathbb{E}(m_t X_t) - p.$$

When the SDF depends on a vector of parameters θ, $m_t(\hat{\theta})$ the estimated SDF with estimated pricing errors $\hat{e} = T^{-1}\sum_{t=1}^{T} m_t(\hat{\theta})X_t - p$. By the Delta method:

$$\sqrt{T}\hat{e} \to_d N(0, V_e) \text{ where } V_e = \frac{\partial e}{\partial \theta'} V_\theta \frac{\partial e'}{\partial \theta} \text{ and } \frac{\partial e}{\partial \theta'} = \mathbb{E}(X_{t+1}\frac{\partial m_{t+1}(\theta)}{\partial \theta'}).$$

Consider three different metrics (with some abuse of notation since HJ is the squared Hansen-Jagannathan first-distance multiplied by T):

$$\begin{aligned} HJ &= T\hat{e}'(X'X/T)^{-1}\hat{e}, \\ SS &= T\hat{e}'\hat{e}, \\ J &= T\hat{e}'\hat{V}_e^{-1}\hat{e}. \end{aligned}$$

Theorem F.1 (Pricing errors)**.** *For a $N \times 1$ vector of standard normal random variable z, and a sequence of iid χ^2_{1i} chi-squared random variable with 1 degree of freedom, as $T \to \infty$,*

$$\begin{aligned} HJ &\to_d z' A_{hj} z = \sum_{i=1}^{N} \lambda_{hj,i}\chi^2_{1i}, \\ SS &\to_d z' A_{SS} z = \sum_{i=1}^{N} \lambda_{ss,i}\chi^2_{1i}, \\ J &\to_d z'z = \chi^2_N \end{aligned}$$

where

$$\begin{aligned} &\lambda_{hj,1}, \cdots, \lambda_{hj,N} \text{ are the eigenvalues of } A_{hj} = V_e^{\frac{1}{2}}(EX_tX_t')^{-1}V_e^{\frac{1}{2}}, \\ &\lambda_{ss,1}, \cdots, \lambda_{ss,N} \text{ are the eigenvalues of } A_{ss} = V_e, \end{aligned}$$

Proof. Follows from Jagannathan and Wang (1996, Thm. 3).

References

Aït-Sahalia, Y. (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70(1), 223–262.

Ait-Sahalia, Y. (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *The Annals of Statistics*, 36(2), 906–937.

Al-Najjar, N. I. (1998): "Factor Analysis and Arbitrage Pricing in Large Asset Economies," *Journal of Economic Theory*, 78(2), 231–262.

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

——— (2017): "Economic Implications of Nonlinear Pricing Kernels," *Management Science*, 63(10), 3361–3380.

Alvarez, F., and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73(6), 1977–2016.

Arrow, K. J. (1964): "The Role of Securities in the Optimal Allocation of Risk-Bearing," *Review of Economic Studies*, 31, 91–96.

Back, K. (2017): *Asset Pricing and Portfolio Choice Theory*. Oxford University Press, 2nd edn.

Backus, D., M. Chernov, and S. Zin (2014): "Sources of Entropy in Representative Agent Models," *Journal of Finance*, 69(1), 51–99.

Bansal, R., and B. N. Lehman (1997): "Growth-Optimal Portfolio Restrictions on Asset Pricing Models," *Macroeconomic Dynamics*, 1, 333–354.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Box, G. E., and G. C. Tiao (1973): *Bayesian Inference in Statistical Analysis*. Addision-Wesley, Reading, MA.

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

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

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

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

Cochrane, J. H. (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104, 572–621.

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

Connor, G., L. Goldberg, and R. A. Korajczyk (2010): *Portfolio Risk Analysis*. Princeton University Press, Princeton, New Jersey.

Delbaen, F., and W. Schachermayer (2006): *The Mathematics of Arbitrage*. Springer, Berlin.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal Versus Naïve Diversification: How Inefficient is the 1/N Portfolio Strategy?," *Review of Financial Studies*, 22, 1915–1953.

Dybvig, P. H., and J. E. Ingersoll, Jr. (1982): "Mean-Variance Theory in Complete Markets," *The Journal of Business*, 55(2), 233.

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

Feng, G., S. Giglio, and D. Xiu (2019): "Taming the Factor Zoo," WOrking Paper, Univeristy of Chicago.

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

——— (2017): "A Diagnostic Criterion for Approximate Factor Structure," Working Paper, University of Lugano.

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

Ghosh, A., C. Julliard, and A. P. Taylor (2017): "What Is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models," *The Review of Financial Studies*, 30(2), 442–504.

Giglio, S., and D. Xiu (2017): "Asset Pricing with Omitted Factors," Discussion paper, Working paper.

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

Hannan, E. (1970): *Multiple Time Series*. New York: Wiley.

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.

Hansen, L. P., and S. F. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

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.

Jagannathan, R., and Z. Wang (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *The Journal of finance*, 51(1), 3–53.

——— (1998): "An asymptotic theory for estimating beta-pricing models using cross-sectional regression," *Journal of Finance*, 53(4), 1285–1309.

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.

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

Lettau, M., and S. Ludvigson (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time Varying," *Journal of Political Economy*, 109(1238-1287).

Liu, Y. (2015): "Index Option Returns and Generalized Entropy Bounds," Working Paper, Texas A&M University.

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

MacKinlay, A. C., and Ľ. Pástor (2000): "Asset Pricing Models: Implications for Expected Returns and Portfolio Selection," *Review of Financial Studies*, 13(4), 883–916.

Madan, D. B., and E. Seneta (1990): "The variance gamma (VG) model for share market returns," *Journal of Business*, 63(4), 511–524.

Orlowski, P., A. Sali, and F. Trojani (2016): "Arbitrage Free Dispersion," Working Paper, University of Lugano.

Raponi, V., R. Uppal, and P. Zaffaroni (2019): "Portfolio Choice with Model Misspecification: A Foundation for Alpha and Beta Portfolios," Working paper, Imperial College.

Renault, E., T. van der Heijden, and B. Werker (2017): "Arbitrage Pricing Theory for Idiosyncratic Variance Factors," Available at SSRN: `https://ssrn.com/abstract=3065854`.

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

Ross, S. A. (1978): "Mutual Fund Separation in Financial Theory – The Separating Distributions," *Journal of Economic Theory*, 17(2), 254–286.

Sandulescu, M., F. Trojani, and A. Vedolin (2017): "Model-Free International SDFs in Segmented Markets," Working Paper, University of Lugano.

Snow, K. N. (1991): "Diagonosing Asset Pricing Models Using the Distribution of Asset Returns," *Journal of Finance*, 46(3), 955–983.

Stutzer, M. (1995): "A Bayesian Approach to Diagonosis of Asset Pricing Models," *Journal of Econometrics*, 68, 367–397.