Factor Models for Conditional Asset Pricing*

PAOLO ZAFFARONI
Imperial College Business School

This Draft

November 12, 2019

Abstract

This paper develops a methodology for inference on conditional asset pricing models linear in latent risk factors, valid when the number of assets diverges but the time series dimension is fixed, possibly very small. We show that the no-arbitrage condition permits to identify the risk premia as the expectation of the latent risk factors. This result paves the way to an inferential procedure for the factors' risk premia and for the stochastic discount factor, spanned by the latent risk factors. In our set up every feature of the asset pricing model is allowed to be time-varying including loadings, risk premia, idiosyncratic risk and the number of risk factors. Monte Carlo experiments corroborate our theoretical findings. Several empirical applications based on individual asset returns data demonstrate the power of the methodology, allowing to tease out the empirical content of the time-variation stemming from asset pricing theory.

Keywords: Conditional Asset Pricing Models; No-Arbitrage; Latent Factor Model; Risk Premia; Stochastic Discount Factor; Principal Component Analysis.

*I thank Claudia Custodio, Victor DeMiguel, Gabaix Xavier, Marcin Kacperczyk, Frank Kleibergen, Ralph Koijen, Oliver Linton, Stefan Nagel, Olivier Scaillet, Raman Uppal, Dacheng Xiu, Guofu Zhou, audiences at Cambridge, Imperial College London, Luxembourg School of Finance, Yale and Tinbergen Institute (Amsterdam) for comments and suggestions. I am very grateful to Massimo Dello Preite for developing the simulation and empirical exercises of this paper.
1 Introduction

This paper develops a formal methodology to conduct inference on no-arbitrage factor conditional asset pricing models. Our method does not require to specify in any way how the conditional distribution of asset returns changes over time, allowing to leave the form of time-variation of loadings, risk premia and even of the number of risk factors completely unspecified. Moreover, it does not require to pre-select any candidate, observed, risk factor and instead it treats all the risk factors as unobserved. These two features imply that our method mitigates completely the risk of model misspecification arising from either postulating the incorrect dynamics and from relying on the wrong or incomplete set of risk factors. The salient feature of our methodology, that enables to achieve these features, is that it works when the time-series dimension of the data $T$ is fixed, in practice (almost) arbitrarily small, whereas it demands the number of assets $N$ to diverge.

Linear factor models represent the workhorse of asset pricing, whereby the stochastic discount factor (SDF) is assumed to be a linear function of a set of systematic risk factors, common across assets. From a theoretical perspective, this approach has been legitimized by the CAPM of [Sharpe (1964)] and [Lintner (1965)], the Intertemporal CAPM of [Merton (1973)] and the Arbitrage Pricing Theory (APT) of [Ross (1976)]. Empirically, it has proven much more arduous to identify the complete set of risk factors that span the SDF. Although the [Fama & French (1993)] three-factor model and, especially, the [Fama & French (2015)] five-factor model appear successful, with respect to several metrics, there is no consensus about which, among the hundreds of candidate risk factors documented in the empirical asset pricing literature (see [Harvey et al. (2016)]), are really “...important and which are....subsumed by others” (Cochrane (2011)).

At the same time, tens of thousands of financial assets are traded every day in financial markets. However, most of the econometric methods developed to test and estimate asset pricing models are designed for when the time-series dimension $T$ diverges and the number of assets $N$ is fixed. There are many instances when allowing a large $T$ is not beneficial or just not feasible. For instance, economic and financial phenomena are plagued by structural breaks (geopolitical crises and wars, energy crises, financial crises) and, sometimes, a long time series of data simply is not available, such as when considering financial data on emerging markets or data on new financial instruments. An arbitrarily large number of assets $N$ is also dictated by asset pricing theory, in particular when the idiosyncratic component of asset returns is assumed correlated across assets. For example the CAPM hinges on the notion of full diversifiability of idiosyncratic risk, achieved when $N$ diverges.⁠

⁠

More in general, exact pricing (i.e. when expected returns are exactly linear in betas with null pricing errors) requires existence of a well-diversified portfolio on the mean-variance frontier (see [Chamberlain (1983)] [Corollary 1]). On the other hand, when exact
Moreover, dynamic asset pricing models characterize expected returns, prices and risk premia as conditional (hence time-varying) moments and such time-variation has to be explicitly parameterized, in a tractable way, when dealing with any large-$T$ estimation procedure. However, with tractability comes the risk of model misspecification, rendering any large-$T$ inferential procedure potentially invalid. More in general, estimating the first moment of asset returns by means of large-$T$ techniques is notoriously challenging, as elegantly elicited by Merton (1980).

Therefore, empirical and theoretical considerations strongly suggest the need for an inferential methods for conditional asset pricing models that work when $N$ is large and $T$ is small. This is the achievement of the paper. More formally, this paper provides the empirical asset pricing methodology that mirrors the theoretical asset pricing contributions of Chamberlain (1983), Chamberlain & Rothschild (1983a) and Hansen & Richard (1987). In particular, whereas the first two seminal works combine no-arbitrage factor asset pricing models with population Principal Component Analysis (PCA) in a large-$N$ environment, ours develops the inferential theory of the sample PCA for no-arbitrage factor asset pricing models, also in a large-$N$ environment. This implies that our inferential procedure is particularly suited for conditional asset pricing models, formalized upon the conditional no-arbitrage condition of Hansen & Richard (1987), because parameterizing time-variation becomes much less critical, if any at all, when $T$ can be taken sufficiently small, as the assumption of mild or no time-variation represents a plausible approximation over a short time interval.

Our novel PCA asymptotic results are instrumental for developing an inferential procedure for our objects of interest, namely the time-varying risk premia and the time-varying SDF. Our results are first developed for the case when a risk free asset is traded, and then generalized to the case when the zero-beta rate is unknown and needs to be estimated. The usefulness of the PCA estimator of the risk factors, from an asset pricing perspective, naturally stems from its close analogy with the mimicking portfolio estimator: we demonstrate that these two estimators, although not identical, are asymptotically equivalent, implying that they allow to conduct the same inference on the set of true latent risk factors. Our methodology for conditional asset pricing models represents a unique tool to address important empirical questions in empirical asset pricing, and beyond, unthinkable with the usual PCA methodology, valid under double-asymptotic (i.e. when both $N$ and $T$ diverge).

More specifically, we show how to estimate consistently the number of (latent) risk factors and the associated pricing does not hold, such as for the APT, an arbitrarily large $N$ is required for all the theoretical predictions to hold. More specifically, Chamberlain & Rothschild (1983a) state that "... Ross (1976) has argued that the apparent empirical success of the CAPM is due to three assumptions which are more plausible than the assumptions needed to derive the CAPM. These assumptions are first, that there are many assets; second, that the market permits no arbitrage opportunities; and third, that asset returns have a factor structure with a small number of factors..."
factors’ risk premia, and how to conduct inference on them, such as e.g. constructing asymptotically valid confidence intervals. Interestingly, as a by-product of the no-arbitrage condition, our risk premia estimator exhibits two, equivalent, representations: first, as the time-series sample mean of the PCA estimated risk factors, and, second, as the traditional OLS two-pass estimator of Fama & MacBeth (1973), obtained by projecting sample average returns on the PCA estimated loadings. It turns out that our risk premia estimator exhibits the same rate of convergence characterizing risk premia estimators for the case when the risk factors are observed. In other words, somewhat surprising, no loss of information arises from the latent nature of the risk factors within our large-$N$ fixed-$T$ sampling scheme, the only difference being that risk premia are now identified up to an (unknown) rotation. Notice that our focus on the large-$N$ and fixed-$T$ scheme necessarily implies that we can identify the so-called ex-post risk premia, which equals the (ex-ante) risk premia plus the risk factors’ innovation (difference between the risk factors’ sample and population means). For the case of observed risk factors, Gagliardini et al. (2016) and Raponi et al. (2018) demonstrate how valid inference on correct specification of the asset pricing model (i.e. on the asset pricing restriction) can be conducted by relying on the ex-post risk premia, without any loss of power, when $N,T$ are both diverging and when $N$ diverges but with fixed-$T$, respectively.

We then study estimation of the SDF spanned by the latent risk factors. Unlike the risk premia, the SDF is unaffected by the rotation that plagues the PCA estimator of factors and risk premia. We derive the rate of convergence and the asymptotic distribution of our SDF estimator, and show how to construct asymptotically valid confidence intervals. This is important because the limiting statistical properties of the SDF, spanned by latent risk factors, were hitherto unknown, when taking into account the sampling variability of the risk factors’ estimates. In analogy to risk premia, as a consequence of our sampling scheme, note that the object of inference is necessarily the so-called ex-post SDF, which is a (linear) function of the ex-post risk premia. To justify the notion of the ex-post SDF, we formally show that the pricing errors associated with it are almost undistinguishable from the ones associated with the conventional SDF, based on the ex-ante risk premia, even for a moderate $T$.

Finally, we show how to consistently estimate expected returns of large portfolios, based on our risk premia estimator. Just like for the SDF, portfolios expected returns are immune to the rotation affecting the risk factors and their risk premia. Consistent estimation follows because expected portfolio returns depend on the

\[^2\]Recall that equivalence between the sample mean and the two-pass estimator is not warranted for the case of observed tradeable risk factors.

\[^3\]The notion of ex-post risk premia was by coined by Shanken (1992).
(weighted) first moment of the loadings, which can be recovered despite the noisiness of the PCA-estimated loadings themselves.

We develop several empirical applications, based on a data set of individual U.S. monthly stock returns, that demonstrate the strength of our large-\(N\) methodology. First, we show empirically how the estimated number of risk factors varies across time, ranging from one to 10 and increasing during booms and sharply decreasing during financial crises: the estimated correlation between the S&P 500 index and our estimated number of factors is above 60\% in our sample. This is a marked rejection of the conventional approach of constancy of the number of latent factors. Second, we found robust evidence according to which the dominant estimated factors (measured by the size of the corresponding eigenvalues, sorted in descending order) are strongly related to the Fama & French (2015) five factors. In particular, the market factor, the small-minus-big and high-minus-low factors appear paired with the first three PCA-estimated factors, respectively, whereas the profitability and the investment factors appear jointly related to the fourth and fifth PCA-estimated factors. The strength of such relationships varies with time and seems to weaken after the year 2000. Third, we demonstrate how the time-variation of the estimated risk factors’ loadings appear driven by the interest rate spread variables, dividend yield and earnings variables, especially during financial crises. Fourth, we evaluate the economic value of the estimated asset pricing model in terms of pricing performance, i.e. magnitude of the associated pricing errors, and in terms of out-of-sample Sharpe ratios associated with mean-variance portfolios built using the model’s estimated parameters. It emerges that the factor model, based on the time-varying estimated number of risk factors, provides the best performance across both metrics. Noticeably, in terms of out of sample Sharpe ratios, our estimated model, with a time-varying number of risk factors, not only dominates the performance of a model with a constant number of factors but dominates also the celebrated equally-weighted portfolio strategy. Fifth, the time series of the estimated SDF exhibits a prominent, and statistically significant, anti-cyclical behavior, in particular with respect to the NBER recession indicator. Moreover, the estimated SDF appears markedly spanned by the Fama & French (2015) factors until the late 1990s, and rather tenuously after that, except in the aftermath of the subsequent financial crises.

The paper proceeds as follows. Section 2 describes a literature review. Section 3 presents the factor conditional asset pricing model, sets the notation and describes the regularity assumptions. Section 4 presents our methodology: we show how to conduct inference on the latent factors, including consistent estimation of the true number of latent factors; we formalize how to conduct inference on risk premia and the SDF, and how to generalize our results to the case when the zero-beta rate has to be estimated; we establish the asymptotic
equivalence between PCA estimator of the risk factors and the mimicking portfolio estimator. Monte Carlo experiments are presented in Section 5. Section 6 contains the five empirical applications based on the data set of individual US stock returns. Section 7 concludes. Formal proofs are reported in the final appendixes.

2 Literature Review

This paper advances the empirical asset pricing literatures that exploits the available large cross-sections of individual assets returns and the one on estimation of conditional asset pricing models and the econometrics literature on PCA estimation.

Giglio & Xiu (2017) recognize that, by extracting the complete set of systematic risk factors, PCA solves the problems arising from estimating risk premia associated with observed factors, when, in particular, some of these factors are omitted or even mismeasured, leading to the classical omitted variable bias. Gagliardini, Ossola and Scaillet (2018) propose a methodology to detect if the residuals of a potentially misspecified beta-pricing model, with time-varying coefficients, exhibit a factor structure. Although PCA is not directly employed, their method is based on the asymptotic behaviour of the maximum eigenvalue of the residuals’ sample covariance matrix. Lettau & Pelger (2018a) study a version of the PCA estimator designed to also minimize the mispricing in terms of expected return, with the noticeably feature of estimating consistently weak factors (i.e. factors associated with non-diverging eigenvalues), and Lettau & Pelger (2018b) provide several applications of this methodology to characteristics portfolios and individual stock returns. Kozak et al. (2018) demonstrate empirically that a small number of the dominant PCA estimates satisfactorily quantifies the SDF associated with anomalies portfolios returns, regardless of whether the underlying asset pricing model is deemed as behavioural (i.e. risk driven by say stock characteristics) or rational (i.e. risk driven by loadings). Kim & Korajczyk (2019) study estimation of the SDF by minimizing the average mispricing in terms of prices, where the latent factors, when present, have been pre-estimated by PCA, building on the insight of Pukthuanthong & Roll (2017).

Conditional linear factor models have been typically specified by assuming that loadings and risk premia are linear functions of (lagged) observed state variables. Gagliardini et al. (2016) provides a formal statistical analysis for an inferential method of this type of conditional linear factor models, in particular characterized by observed risk factors. In contrast, Connor & Linton (2007), Connor et al. (2012), Fan et al. (2016), Kelly et al. (2017) and Kelly et al. (2018) study, and apply, estimation procedures for conditional models with latent risk

---

factors models whereby the loadings and, when applicable, the mispricing (i.e. the intercept) are instrumented by observed state variables. An alternative approach, closer to the perspective of this paper, is to leave the dynamics of the conditional asset pricing model unspecified, and use nonparametric techniques for estimation. Building on the insight of French et al. (1987), Andersen et al. (2006), Lewellen & Nagel (2006) and Ang & Kristensen (2012) develop several variations of this approach. Whereas the asymptotic analysis of the time-varying loading’s estimator pose no particular problems, reliable nonparametric estimation of the alphas, and thus of the time-varying risk premia, is particularly challenging with nonparametric techniques, as detailed by Ang & Kristensen (2012) who build on the arguments of Merton (1980).

Seminal, econometrics, results for inference on factor models are Bai & Ng (2002), Stock & Watson (2002a), Stock & Watson (2002b) and Bai (2003), whereby the observable variable is a static function of a finite set of latent common factors. However, note that this model is not static, in the sense that both the common factors and the idiosyncratic component are allowed to be time-dependent. In this strand of the literature, time-domain inferential methods and assumptions are have been adopted. In contrast, Forni et al. (2000) and Forni & Lippi (2001) postulate that the observable is a dynamic function of a finite set of latent factors, their approach being denominated as the generalized dynamic factor model. Frequency domain techniques have been adopted here to analyze the statistical properties of the dynamic PCA estimator. Although a more general form of time-dependence is allowed for, this frequency-domain approach is less suitable for forecasting because it necessarily leads to the observable being a function of past and future innovations. Forni et al. (2015) and Forni et al. (2017) show how to reconcile the advantages of both approaches. All the aforementioned work requires double asymptotics, namely that both $N$ and $T$ diverge to infinity. Moreover, with the exception of Bai (2003), none of these works developed inferential methods, namely the asymptotic distribution and the associated standard errors for the PCA estimates of loadings and factors, but focused on estimating the number of factors and deriving consistency (sometimes with rates of convergence) of the PCA estimator of loadings and common factors.

A large literature focused on procedures for estimating the number of latent common factors, since the seminal work of Connor & Korajczyk (1993). Bai & Ng (2002) provide the theory, under double asymptotics, for consistent estimation of the number of factors in a static model. Amengual & Watson (2007) show how their approach can be extended to the case of finite-dimensional vector autoregression latent factors. Onatski (2010).

\footnote{Although we denote these models as conditional, in the sense that the corresponding parameters are function of observables, only Kelly et al. (2017) and Kelly et al. (2018) allow explicitly for dynamics. Brillinger (2001) introduce the dynamic PCA estimator, based on the eigenvalues and eigenvectors of a consistent estimator of the observables’ spectral density matrix.}
allows for an arbitrary speed of divergence of the dominant eigenvalues and \cite{Ahn&Horenstein} relax the dependence from a finite upper bound for the true number of latent factors. \cite{Ait-Sahalia&Xiu} establish consistent estimation of the number of latent factors (and estimation of the factor structure itself) within an in-fill asymptotics framework when high-frequency data are available.\footnote{For the case of dynamic factor models a la \cite{Forni&etal}, \cite{Hallin&Liska} provide a consistent estimator for the number of factors and \cite{Onatski} establishes the asymptotic distribution of a test for the number of factors. As for estimation, these advances have been developed in a double asymptotics setting.}

For the case of dynamic factor models a la \cite{Forni&etal}, \cite{Hallin&Liska} provide a consistent estimator for the number of factors and \cite{Onatski} establishes the asymptotic distribution of a test for the number of factors. As for estimation, these advances have been developed in a double asymptotics setting.

Building on \cite{Dahlhaus} and \cite{Dahlhaus}, factor models with time-varying loadings of unspecified form, are considered by \cite{Motta&etal}, \cite{Brandon&etal}, \cite{Su&Wang} and \cite{Barigozzi&etal}, among others. All these developments are based on double asymptotics.

Very little work has been established within the large-\(N\) fixed-\(T\) sampling scheme, especially when compared with the double asymptotics setting cited above, and moreover the few existing results are scattered across various papers, and depend on different set of regularity conditions. In particular, \cite{Connor&Korajczyk} establish consistency of the PCA estimator for the latent factors when \(T\) is fixed. Subsequently \cite{Bai}[Theorem 4] establish the necessary and sufficient conditions for the \cite{Connor&Korajczyk} large-\(N\) consistency result, namely that the cross-sectional averages of the idiosyncratic components’ variances and autocovariances are, respectively, constant and null. A related, statistical literature, focused on the large-\(N\), small-\(T\) environment, denominated as the high-dimension-low- sample-size (HDLSS) situation, where rates of convergence of the PCA estimator of the latent factors, and variations of, are studied under various regularity conditions; see \cite{Jung&Marron} and \cite{Shen&etal} among others.\footnote{\cite{Ait-Sahalia&Xiu} establish the asymptotic distribution theory of the PC estimator in a continuous time setting but without assuming necessarily a factor structure, hence \(N\) is assumed fixed. \cite{Fan&etal}, cited above, reports consistency results for a variation of the PCA, both in the HDLSS environment and under double asymptotics.}

Our work complements these empirical asset pricing and econometrics literatures in an important way because we specifically focus on the case when \(N\) is large and \(T\) is fixed, developing tools that allow to perform inference, such as confidence intervals and hypothesis testing for the estimated factors, as well establishing consistent estimation of the true number of latent factors. Moreover, we show how our econometrics results are instrumental for developing a formal methodology for inference on risk premia and the SDF associated with the linear asset pricing model, which are our ultimate objects of interest.

Noticeably, individual asset returns, in particular equity returns, exhibit a very limited time-dependence but,
at the same time, are characterized by strong cross-sectional dependence and heterogeneity. As our interest lies predominantly in factor models for asset returns, these empirical features motivate our large-$N$ fixed-$T$ sampling scheme and explain our focusing on static factor models, in particular in light of Bai (2003) findings, as opposed to dynamic factor models a la Forni et al. (2000), although our approach could in principle be applied to the latter too.

3 Factor Conditional Asset Pricing Model: Set-Up and Assumptions

We assume that the $i$-th individual asset return, $x_{it}$, in excess of the risk free rate, satisfies the finite-dimensional conditional factor model, in the sense that each $x_{it}$ is a linear function of a finite number $r_{t-1}$ of latent common factors, and of a (latent) idiosyncratic component:

$$x_{it} = \alpha_{it-1} + \lambda_{it-1}' f_t + e_{it} \text{ for every } i = 1, \cdots, N \text{ and } t \in \mathcal{Z},$$

where $f_t$ defines the $r_{t-1} \times 1$ vector of latent risk factors, $\lambda_{it-1}$ the $r_{t-1} \times 1$ vector of latent time-varying loadings, $e_{it}$ is the idiosyncratic component and $\mathcal{Z} = \{\cdots, -1, 0, 1, \cdots\}$. In representation (1) the parameter $\alpha_{it-1}$ is just a time-varying intercept, not necessarily equal to the conditional expected excess return $E_{t-1}(x_{it})$ (i.e. the risk premium associated with asset $i$). Therefore, (1) does not imply that $E_{t-1}(f_t) = 0$, although we will assume $E_{t-1}(e_{it}) = 0$ in our regularity assumptions. Finally, we assume that a risk-free asset is tradable. The case when any risk-free asset is not tradeable is described subsequently.

It is important to make sure that model (1) does not allow for a ‘free lunch’, formalized as follows (see Chamberlain (1983) [Condition A] and the generalization of Hansen & Richard (1987) [Definition 2.4] to a conditional setting). This requires to define an arbitrary portfolio strategy $p$ with weights $w^p_{Nt-1} = (w^p_{1t-1}, w^p_{2t-1}, \cdots, w^p_{Nt-1})'$ of the $N$ risky assets’ excess returns $x_t \equiv (x_{1t}, \cdots, x_{Nt})'$, leading to the portfolio $p$ excess return $r^p_{N,t} \equiv w^p_{Nt-1}' x_t$. As we explained below, the no-arbitrage assumption turns out to be instrumental for the interpretation of the latent factors, and for derivation of the statistical properties of the PCA estimator.

**Assumption 1 (conditional no-arbitrage)** There is no sequence of portfolios along some subsequence $N'$ for which, for an arbitrary positive scalar $C$:

$$\text{var}_{t-1}(r^a_{N',t}) \to 0 \text{ a.s., when } N' \to \infty, \text{ and } E_{t-1}(r^a_{N',t}) \geq C > 0 \text{ a.s for every } t \in \mathcal{Z}.$$
Ingersoll (1984) clarifies that the above definition of no-arbitrage is equivalent to assume that \( \mathcal{E}_{t-1}(r_{Nt,t}^\prime) \to \infty \) (a.s.) by suitably leveraging, i.e. scaling up, the portfolio weights. Under the above no-arbitrage condition, a limited degree of cross-sectional dependence and a sufficient degree of smoothness of the loadings (see Proposition 4 in Appendix B), there exists a \( r_{t-1} \times 1 \) vector \( \gamma_{t-1} \), namely the time-varying risk premia associated with the risk factors \( f_t \), such that

\[
\mu_{it-1} \equiv \mathcal{E}_{t-1}(x_{it}) = a_{i-1} + \lambda'_{it-1} \gamma_{t-1} \quad \text{a.s. for every } i = 1, \cdots, N \text{ and } t \in \mathbb{Z},
\]

(2)

where the pricing errors \( a_{it-1} \) must satisfy, as a consequence of no-arbitrage (see Chamberlain (1983)[Theorem 1] and Stambaugh (1983)[Theorems 1 and 2] for the static and conditional settings, respectively),

\[
\sum_{i=1}^{\infty} a_{it-1}^2 < \infty \quad \text{a.s. for every } t \in \mathbb{Z}.
\]

(3)

Combining (1) and (2):

\[
x_{it} = a_{it-1} + \lambda'_{it-1} F_t + e_{it},
\]

setting

\[
F_t \equiv f_t + \gamma_{t-1} - \mathcal{E}_{t-1}(f_t).
\]

(5)

A re-writing of (4), such as \( x_{it} = \mu_{it-1} + \lambda'_{it-1} (f_t - \mathcal{E}_{t-1}(f_t)) + e_{it} \), leads to the conditional version of the APT formulation in Ross (1976)[Eq (15)], Chamberlain (1983)[Eq. (4.1)] and Chamberlain & Rothschild (1983a)[Eq. (1.1)], among others. However, (4) provides the ultimate specification that PCA will estimate, in particular obtaining (consistent) estimates of the risk factors \( F_t \) (that is, the \( f_t \) adjusted by \( \gamma_{t-1} - \mathcal{E}_{t-1}(f_t) \)). Note that \( \mathcal{E}_{t-1}(F_t) = \gamma_{t-1} \), that is the no-arbitrage condition leads to a change-of-measure of the latent risk factors \( f_t \), so that the \( F_t \) are centered precisely around the risk premia \( \gamma_{t-1} \). This result does not follow if the \( x_{it} \) are de-meaned beforehand, suggesting the importance of applying PCA directly to the \( x_{it} \) from (4) without any prior de-meaning of the data.
Although the emphasis of this paper is on asset returns, if one wishes to apply model (1) to data other than asset returns, then the restriction (2) does not necessarily hold, and one simply re-arranges (1) as

\[ x_{it} = (\alpha_{it-1}, \lambda_{it-1})' (1, f'_t)' + e_{it} = (\alpha_{it-1}, \lambda_{it-1})' G_t + e_{it}, \]

where \( G_t \) denotes the \((r_t - 1 + 1) \times 1\) vector \((1, f'_t)'\). When \( \alpha_{it-1} = \alpha_i, \lambda_{it-1} = \lambda_i \) and \( r_t-1 = r \), representation (6) is the static latent factor formulation of (1) adopted, among others, by Stock & Watson (2002) and Bai & Ng (2002). It follows that by applying PCA to model (6), with constant parameters, one captures \( r + 1 \) factors in total, one of which being the constant factor (equal to 1 for all periods), as long as such (constant) factor is pervasive, namely \( N^{-1} \sum_{i=1}^{N} \alpha_i^2 \to C > 0 \), for some constant \( C \). Alternatively, focusing again on the constant loadings case of Eq. (6), one can subtract the (time series) sample mean from both sides of (6), and apply the PCA to the \( x_{it} - \bar{x}_i = \lambda'_i (f_t - \bar{f}) + e_{it} - \bar{e}_i \), denoting \( \bar{z}_i \equiv T^{-1} \sum_{t=1}^{T} z_{it} \) for a generic sequence \( z_{it} \). In this case only the \( r \) (de-meaned) factors, \( f_t - \bar{f} \), will be captured by PCA.

Things become more complicated for the case of time-varying parameters, as for instance de-meaning is not even meaningful because it will not eliminate the time-varying intercept term anyway. This is where the no-arbitrage assumption becomes extremely useful, because PCA will detect the \( r_{t-1} \) factors without the need to de-mean the asset returns data, because the constant factor is not pervasive. In other words, for asset returns one can safely estimate the risk factors ignoring completely the intercept \( a_{it-1} \) (i.e. the pricing errors) without inducing any bias in the estimated factors and loadings, an especially advantageous feature when considering conditional factor models.

Summarizing, throughout this paper, we will always implement PCA, without allowing for any intercept term, regardless of whether one envisages to fit the factor model to asset returns or not, as this will always lead to accurate (i.e. consistent) estimates of the common latent factors, under our regularity assumptions. This is methodologically relevant because, as explained in details below, allowing explicitly for an intercept term or, alternatively, de-meaning the data, induces time-dependence in the idiosyncratic component of the fitted model, violating one of the regularity assumptions needed in our large-N fixed-T environment, as clarified by Bai (2003).

\[^{10}\text{Obviously, in practice, such constant factor will be masked by the rotation that affects loadings and factors. It is well-known that this does not lead to any relevant consequence for practical use of the model.}\]
3.1 Econometric Quantities: Factors and Loadings

To define the PCA estimator, it is useful to adopt a matrix formulation, setting the $T \times N$ matrix of observed excess asset returns $X \equiv (x_1 \cdots x_i \cdots x_N)' \equiv (x_{1i} \cdots x_{iT})'$, where $x_i \equiv (x_{1i} \cdots x_{Ti})'$ and $x_t \equiv (x_{1t} \cdots x_{Nt})'$, the $T \times r_{t-1}$ matrix of factors $F \equiv (F_1 \cdots F_T)'$, the $N \times r_{t-1}$ matrix of loadings $\Lambda_{t-1} \equiv (\lambda_{t-1} \cdots \lambda_{Nt-1})'$ and the $T \times N$ matrix of idiosyncratic components $e \equiv (e_1 \cdots e_i \cdots e_N) \equiv (e_{1i} \cdots e_{it} \cdots e_{Nt})'$, setting $e_i \equiv (e_{1i} \cdots e_{iT})'$ and $e_t \equiv (e_{1t} \cdots e_{Nt})'$. Note that factors, loadings and residuals are always latent.

We assume that the $T$ observations used in the PCA are extracted from a longer time-series sample of size $T_0 > T$ and in particular, without loss of generality, that they are made by the observations from $t - T + 1$ to $t$, for any $t$ satisfying $T - 1 < t \leq T_0$. Therefore, whereas $X$ and $F$ contains exclusively $T$ rows (referring to $T$ time series observations), without loss of generality one can always assume that these matrices are extracted from the larger matrixes $X_0$ and $F_0$ with $T_0$ rows. Although the allowed degree of time-variation will be formalized below, for derivation of the PCA it suffices for now to assume that the loadings $\Lambda_{t-1}$ and the number of factors $r_{t-1}$ remain constant within the time interval from $t - T + 1$ to $t$, but can vary freely across the $T_0$ observations.

We present two, equivalent, ways to derive the PCA estimator of factors, loadings and number of factors correspondingly to the time-varying model (1), namely as the nonparametric Nadaraya-Watson estimator, in addition to the classical PCA formulation. For $t \equiv [uT_0]$, corresponding to some $u$ satisfying $T/T_0 \leq u \leq 1$, implying $T \leq t \leq T_0$, set $K_{h}() \equiv h^{-1} K(\cdot)/h$, with $K(\cdot) \equiv 1(\cdot)$, the indicator function, and bandwidth parameter $h \equiv T/T_0$. The $K_{h}(s/T_0 - u)$, for every $1 \leq s \leq T_0$, make the diagonal $T_0 \times T_0$ matrix $K_u \equiv diag(K_{h}(1/T_0 - u), \cdots, K_{h}(1 - u))$.

Without loss of generality for $N$, assume that $N > T \geq r_{t-1}$, for every $T - 1 < t \leq T_0$, as we will only allow $N$ to diverge. As it will be explained, one needs $T \geq r_{t-1}$ to rule out existence, and multiplicity, of the zero eigenvalue of certain relevant matrixes. The PCA estimators for the loadings and the factors are the solution of the least squares problem:

$$\tilde{F}(T_t), \tilde{\Lambda}(T_t) = \text{argmin}_{F_{t}, \Lambda_{t-1}} \frac{1}{NT_T} \sum_{s=t-T+1}^{t} (x_s - \Lambda_{t-1}F_s)'(x_s - \Lambda_{t-1}F_s)$$

$$= \text{argmin}_{F_{t}, \Lambda_{t-1}} \frac{1}{NT_0} \text{trace}\left((X_0 - F_0\Lambda_{t-1}')K_u(X_0 - F_0\Lambda_{t-1}')\right),$$

where the PCA estimators, corresponding to $r_{t-1}$ factors, are denoted by

$$\tilde{F}(T_t) \equiv (\tilde{F}_{T-T+1:t}, \cdots, \tilde{F}_{s:t}, \cdots, \tilde{F}_{t:t})' \text{ and } \tilde{\Lambda}(T_t) \equiv (\tilde{\lambda}_{1t-1}, \cdots, \tilde{\lambda}_{Nt-1})'.$$
to indicate dependence on the time interval $T_t \equiv (t - T + 1, t)$.

Indicating $\tilde{F}(T_t) = \tilde{F}$ and $\tilde{A}(T_t) = \tilde{A}_{t-1}$, that is omitting reference to the interval $T_t$ for the sake of simplicity, $\tilde{F}$ denotes the set of $r_{t-1}$ eigenvectors, multiplied by $\sqrt{T}$, corresponding to the largest $r_{t-1}$ eigenvalues of $XX'/NT$, and correspondingly $\tilde{A}_{t-1} \equiv X'\tilde{F}(\tilde{F}'\tilde{F})^{-1} = X'\tilde{F}/T$. See Magnus & Neudecker (2007), Theorem 7 of Chapter 17, for an elegant proof.

In particular, setting $\tilde{V}_{t-1}$ equal to the diagonal $r_{t-1} \times r_{t-1}$ matrix with the largest $r_{t-1}$ eigenvalues of $XX'/NT$, one obtains:

$$\frac{XX'}{NT}\tilde{F} = \tilde{F}\tilde{V}_{t-1} \text{ where } \frac{\tilde{F}'\tilde{F}}{T} = I_{r_{t-1}}.$$  \hspace{1cm} (7)

Crucially, we do not explicitly allow for an intercept term when implementing PCA, for the reasons discussed above. In the next sections, we will establish the asymptotic properties of $\tilde{F}$ and $\tilde{A}_{t-1}$ as $N \to \infty$ and, more in general, how to conduct inference on model (11), including estimating the true (but unknown) number of factors $r_{t-1}$.

### 3.2 Analogies between PCA and Mimicking Portfolios

The estimated factors $\tilde{F}$ have the interpretation of portfolio (excess) returns, corresponding to the $r_{t-1}$ vectors of portfolio weights $\tilde{w}^{\text{pca}}_{Nt-1} = \tilde{A}_{t-1}(\tilde{A}_{t-1}'\tilde{A}_{t-1})^{-1}$. In fact, by (7) and recalling that $N^{-1}\tilde{A}_{t-1} = \tilde{V}_{t-1}$,

$$\tilde{F} = \frac{XX'}{NT}\tilde{F}\tilde{V}_{t-1}^{-1} = \frac{X'\tilde{F}}{T} (N\tilde{V}_{t-1})^{-1} = X\tilde{A}_{t-1}(\tilde{A}_{t-1}'\tilde{A}_{t-1})^{-1} = X\tilde{w}^{\text{pca}}_{Nt-1}.$$  \hspace{1cm} (*)

The population counterpart of the PCA portfolio weights $\tilde{w}^{\text{pca}}_{Nt-1}$ is clearly $w^{\text{pca}}_{Nt-1} = \Lambda_{t-1}(\Lambda_{t-1}'\Lambda_{t-1})^{-1}$. However, it is more relevant to compare the PCA estimated risk factors $\tilde{F}$ with the notion of mimicking portfolios, coined by Huberman et al. (1987) and Breeden et al. (1989). In particular, setting $F^{mp}_{t} \equiv \mathcal{E}_{t-1}(F_t x_t') (\mathcal{E}_{t-1}(x_t,x_t))^{-1} x_t = w^{mp}_{Nt-1} x_t$ as the mimicking portfolios corresponding the latent factors $F_t$, we will study the limiting behaviour of both $F^{mp}_{t}$ and of $w^{pca}_{Nt-1} x_t$, as $N \to \infty$, where by direct calculations:

$$w^{mp}_{Nt-1} = \left( (a_{t-1} + \Lambda_{t-1} \gamma_{t-1})(a_{t-1} + \Lambda_{t-1} \gamma_{t-1})' + \Lambda_{t-1} \Omega_{t-1} \Lambda_{t-1}' + \Sigma_{et-1} \right)^{-1} \left( a_{t-1} \gamma_{t-1} + \Lambda_{t-1} \gamma_{t-1} \gamma_{t-1}' + \Lambda_{t-1} \Omega_{t-1} \right),$$

11 It is well-known (see the discussion in Bai & Ng (2002)) that an asymptotically equivalent PCA estimator of factors and loadings can be obtained by setting $\tilde{A}_{t-1}$ equal to the eigenvectors of $XX'/NT$, multiplied by $\sqrt{N}$, corresponding to the $r_1$ largest eigenvalues, and $\tilde{F} = \tilde{A}_{t-1}/\sqrt{N}$. As $N$ is much larger than $T$, the latter approach is computationally less convenient than the one described above. Recall that the non-zero eigenvalues of $XX$ and $XX'$ coincide.
setting $\Omega_{t-1} \equiv \text{var}_{t-1}(F_t)$, $\Sigma_{et-1} \equiv \text{var}_{t-1}(e_t)$ and $a_{t-1} \equiv (a_{it-1}, \cdots, a_{Nt-1})'$. More importantly, we establish the limiting behaviour of the portfolio excess returns based on the feasible PCA-based estimator of $w_{Nt-1}^{mp}$, and show it close analogies with the PCA estimator $\bar{F}$.

3.3 Asset Pricing Quantities: Risk Premia and SDF

Conditional expected excess returns $\mu_{it-1}$ and risk premia $\gamma_{t-1}$ in particular, as from equation (2), are of fundamental importance for many asset pricing problems, such as for identification of the SDF implied by (1) and for portfolio construction. In particular, under the conditional no-arbitrage assumption, one can define the (conditional or time-varying) risk premia, when a riskless asset with time-varying (gross) rate $r_{ft-1}$, is traded, as:

$$\gamma_{t-1} \equiv -r_{ft-1} \left( \mathcal{P}_{t-1}(f_t) - \frac{\mathcal{E}_{t-1}(f_t)}{r_{ft-1}} \right) \text{ for every } t \in \mathcal{Z},$$

where $\mathcal{P}_{t}(\cdot)$ denotes the conditional pricing functional, formally defined as $\mathcal{P}_{t}(\cdot) \equiv \mathcal{E}_{t}(m_{t,t+1})$ where $m_{t,t+1}$ is the (conditional) SDF, whose existence is ensured by no-arbitrage (see Chamberlain (1983)[Theorem 1] and the generalization to a conditional setting of Hansen & Richard (1987)[Theorem 2.1]).

In our context, the price $\mathcal{P}_{t-1}(f_t)$ is only notional because the $f_t$ are latent, and hence non-tradeable. However, the formula above applies to any risk factors so, for instance, when $f_t$ represent observed portfolio (gross) returns, then $\mathcal{P}_{t-1}(f_t) = 1_{T_{t-1}}$, and the risk premia coincides with the expected value of the factors in excess of the risk free rate, as (8) simplifies to $\gamma_{t-1} = \mathcal{E}_{t-1}(f_t - r_{ft-1}1_{T_{t-1}})$. More in general, expression (8) defines the risk premia $\gamma_{t-1}$ as the price of the factors $f_t$ less their risk-neutral valuation.

When $T$ is fixed, an important identification issue related to the risk premia $\gamma_{t-1}$ arise. To simplify arguments, consider the static version of the asset pricing model (1), that is assume for now $a_{it-1} = a_i$, $\lambda_{it-1} = \lambda_i$, $r_{t-1} = r$. Then, following Shanken (1992), by taking the average of (1) across the time interval $T_t$, yields:

$$\bar{x}_i = \frac{1}{T} \sum_{s \in T_t} x_{is} = a_i + \lambda_i' \bar{F} + \bar{e}_i = a_i + \lambda_i' \gamma^P + \bar{e}_i,$$

setting $\bar{F} \equiv F'T/T$ and $\bar{e}_i \equiv e_{i}' 1_T/T$. Equation (9) shows how the (sample) average excess returns are linear in the so-called ex-post risk premia, coined by Shanken (1992), namely

$$\gamma^P \equiv \bar{F} = \gamma + \bar{f} - \mathcal{E}(f_t).$$

Expression (8) follows by generalizing Chamberlain (1983)[Theorem 1] to the conditional setting. An alternative, equivalent, expression for the risk premia vector $\gamma_{t-1}$ is based on the limit, as $N$ diverges, of the linear projection with respect to (2), namely $(\Lambda_{t-1}^{'} \Lambda_{t-1})^{-1} \Lambda_{t-1}^{'} \mu_{t-1}$, setting $\mu_{t-1} \equiv (\mu_{it-1}, \cdots, \mu_{Nt-1})'$; see Ingersoll (1984)[Theorem 3].
Interestingly, the ex-post risk premia vector $\gamma^P$ coincides with the sample mean of the (population) $F_t$, although the latter are not observed, and thus not tradeable. The ex-post risk premia exhibit many attractive properties, providing essentially the same informations on the asset pricing model (1) as the ex-ante risk premia $\gamma$. For instance, $\gamma^P$ is centered around (i.e. is unbiased for) $\gamma$ and, as $T$ increases, $\gamma^P \rightarrow_p \gamma$. Moreover, the asset pricing model (1) is linear $\gamma^P$. More importantly, $\gamma^P$ allows to conduct inference on $\gamma - E(f_t)$ which is exactly as informative as $\gamma$ on the asset pricing implications of the model, a feature exploited by Gagliardini et al. (2016).

Note that the discrepancy between the ex-ante and ex-post risk premia is not due to the latency of the risk factors in (1) but is a necessary consequence of keeping $T$ fixed. Raponi et al. (2018) develop a complete methodology for inference on linear asset pricing models, such as (1), but when the $f_t$ are (tradeable or non-tradeable) observed factors. Their object of inference consists necessarily the ex-post risk premia and, as here, they assume that only the number of assets, $N$, diverges.

Therefore, by exploiting the restriction (2) arising from the no-arbitrage assumption, a natural estimator for the ex-post risk premia, $\gamma^P$, emerges, namely:

$$\tilde{\gamma} = T^{-1} \sum_{s \in T} \tilde{F}_{s,t},$$

(10)

as the PCA estimator for the $\tilde{F}_{s,t}$ accurately quantifies the $F_s$. Representation (9) suggests a different, perhaps more traditional, way to estimate the ex-post risk premia $\gamma^P$, namely the two-pass estimator:

$$\tilde{\gamma}_{\text{twopass}} = (\tilde{\Lambda}' \tilde{\Lambda})^{-1} \tilde{\Lambda}' \bar{x}_N,$$

(11)

setting $\bar{x}_N = (\bar{x}_1, \cdots, \bar{x}_N)' = X'1_T/T$. We formally show that (see Proposition 3 in Appendix B)

$$\tilde{\gamma}_{\text{twopass}} = \tilde{\gamma}.$$

(12)

For the case of observed risk factors, this readily follows by simply noticing that as $\bar{f}$ is observed, then an estimator for $\gamma - E(f_t)$ can be immediately obtained by netting out $\bar{f}$ from an estimator for $\gamma^P$. Some care is required when using such estimators for testing the asset pricing restriction as, for instance, when the factors are excess returns from tradeable assets, then the null hypothesis of correct model specification corresponds to a zero value for $\gamma - E(f_t)$ in population.

The two-pass estimator, popularized by Black et al. (1972) and Fama & MacBeth (1973), provides the most popular estimator for the exante risk premia $\gamma$ when the risk factors are observed in (1) and exact pricing holds, i.e. $a_i = 0$ in (2). Shanken (1992) provides the large-$T$ asymptotics for the two-pass estimator, further generalized by Jagannathan & Wang (1998), among others. More recent work on the two-pass methodology, allowing for both $N$ and $T$ to diverge, has been developed by Bai & Zhou (2015) and Gagliardini et al. (2016).

The expression of estimator $\tilde{\gamma}_{\text{twopass}}$ is very similar to the one studied in Giglio & Xiu (2017), the main difference being that Giglio & Xiu (2017) estimate the loadings by PCA after de-meaning model (1), that is by applying PCA to the $x_{it} - \bar{x}_i$. This is not feasible in our fixed-$T$ context as de-meaning induces dynamic serial dependence in the idiosyncratic component, and thus violating our regularity assumptions. However, de-meaning does not cause any bias (asymptotically) in Giglio & Xiu (2017) because they let both $N$ and $T$ to diverge. For the same reason, no (asymptotic) difference arises in their setup between the ex-ante and ex-post risk premia, as $\gamma^P - \gamma \rightarrow_p 0$ as $T \rightarrow \infty$. 

13For the case of observed risk factors, this readily follows by simply noticing that as $\bar{f}$ is observed, then an estimator for $\gamma - E(f_t)$ can be immediately obtained by netting out $\bar{f}$ from an estimator for $\gamma^P$. Some care is required when using such estimators for testing the asset pricing restriction as, for instance, when the factors are excess returns from tradeable assets, then the null hypothesis of correct model specification corresponds to a zero value for $\gamma - E(f_t)$ in population.

14The two-pass estimator, popularized by Black et al. (1972) and Fama & MacBeth (1973), provides the most popular estimator for the exante risk premia $\gamma$ when the risk factors are observed in (1) and exact pricing holds, i.e. $a_i = 0$ in (2). Shanken (1992) provides the large-$T$ asymptotics for the two-pass estimator, further generalized by Jagannathan & Wang (1998), among others. More recent work on the two-pass methodology, allowing for both $N$ and $T$ to diverge, has been developed by Bai & Zhou (2015) and Gagliardini et al. (2016).

15The expression of estimator $\tilde{\gamma}_{\text{twopass}}$ is very similar to the one studied in Giglio & Xiu (2017), the main difference being that Giglio & Xiu (2017) estimate the loadings by PCA after de-meaning model (1), that is by applying PCA to the $x_{it} - \bar{x}_i$. This is not feasible in our fixed-$T$ context as de-meaning induces dynamic serial dependence in the idiosyncratic component, and thus violating our regularity assumptions. However, de-meaning does not cause any bias (asymptotically) in Giglio & Xiu (2017) because they let both $N$ and $T$ to diverge. For the same reason, no (asymptotic) difference arises in their setup between the ex-ante and ex-post risk premia, as $\gamma^P - \gamma \rightarrow_p 0$ as $T \rightarrow \infty$. 

16For the case of observed risk factors, this readily follows by simply noticing that as $\bar{f}$ is observed, then an estimator for $\gamma - E(f_t)$ can be immediately obtained by netting out $\bar{f}$ from an estimator for $\gamma^P$. Some care is required when using such estimators for testing the asset pricing restriction as, for instance, when the factors are excess returns from tradeable assets, then the null hypothesis of correct model specification corresponds to a zero value for $\gamma - E(f_t)$ in population.
This result differs starkly from the observed risk factors case, where the sample mean of the realized return of a traded factor is not necessarily identical to the two-pass CSR OLS estimator, unless exact pricing is imposed (see Shanken (1992, Section 2)). The above equivalence holds under Assumption 1, much milder than exact pricing. Another way to understand our equality result is that, unlike the observed factors case, in our latent factor cases, both estimators use the same information, namely a panel of excess returns. In contrast, for the observed factor case, the sample mean of the factor realized return ignores completely all the information stemming from the panel of returns on the test assets. This feature has been often advocated as a limitation of the sample mean estimator as its use is valid only under the (strong) assumption of exact pricing. This criticism does not apply to our framework, as we allow for pricing errors of unspecified form in (2).

Turning now to the general, time-varying, case, one defines the ex-post risk premia as:

$$\gamma_{t-1}^P \equiv \tilde{F}_t = \gamma_{t-1} + f_t - E_{t-1}(f_t) \text{ for every } t \in \mathbb{Z}. \quad (13)$$

Raponi et al. (2018) introduce the notion of time-varying ex-post risk premia (13) and develop a large-N inferential methodology for the case of observed risk factors. Here we complement these results to the case of latent risk factors. In light of (13), we will provide the limiting distribution theory for the estimator of $$\gamma_{s-1}^P$$:

$$\tilde{\gamma}_{s-1} \equiv \tilde{F}_{s,t} \text{ for every } s \in T_t \text{ and } t \in \mathbb{Z}. \quad (14)$$

As for the static case discussed above, an alternative, numerically equivalent, estimator for the time-varying risk premia is:

$$\tilde{\gamma}_{s-1}^{\text{twopass}} \equiv (\tilde{\Lambda}_{t-1}^t \tilde{\Lambda}_{t-1})^{-1} \tilde{\Lambda}_{t-1}^t x_s \text{ for every } s \in T_t \text{ and } t \in \mathbb{Z}, \quad (15)$$

whose limiting properties immediately follow from the ones of $$\tilde{\gamma}_{s-1}$$ in light of the equivalence $$\tilde{\gamma}_{s-1}^{\text{twopass}} = \tilde{\gamma}_{s-1}$$.

However, when time-variation holds, (10) remains a meaningful quantity, especially when $$T$$ is small, as $$\tilde{\gamma}$$ will estimate the average ex-post risk premia

$$\bar{\gamma}^P \equiv \bar{F} = \frac{1}{T} \sum_{s \in T_t} \gamma_{s-1} + \bar{f} - \frac{1}{T} \sum_{s \in T_t} E_{s-1}(f_s). \quad (16)$$

When $$T$$ is small, $$\bar{\gamma}^P$$ represents a local average of the time-varying $$\gamma_{s-1}^P$$ which, on one hand, can accurately capture the dynamics of the true risk premia (depending on their degree of time-variation) and, on the other,
3.3 Asset Pricing Quantities: Risk Premia and SDF

...can be more accurately estimated than $\gamma_{s-1}^P$, due to the averaging. These considerations prompt us to study also the asymptotic properties of $\hat{\gamma}$.

We now illustrate how to estimate the SDF associated with the asset pricing model (1). It turns out that the PCA estimates provide exactly the necessary ingredients for consistent estimation of the SDF. Uppal et al. (2018) [Theorem 1 and 2] establish that the SDF implied by the linear factor asset pricing models when a risk-free asset is traded and all risk factors $f_t$ are observed (tradeable or non-tradeable), is given by:

$$ m_{t,t+1} = \frac{1}{r_{ft}} - \frac{1}{r_{ft}} \gamma_t^{-1}(f_{t+1} - E_t(f_{t+1})) - \frac{1}{r_{ft}} a_t^T \Sigma^{-1}(x_{t+1} - E_t(x_{t+1})) $$

setting the asset excess returns’s conditional covariance matrix $\Sigma_t \equiv \Lambda_t \Omega_t \Lambda_t' + \Sigma_{et}$. In fact,

$$ P_t(1) = E_t(m_{t,t+1}) = \frac{1}{r_{ft}} $$

and $P_t(x_{i,t+1}) = E_t(m_{t,t+1}x_{it+1}) = 0$ for every $i = 1, \cdots, N$ and $t \in Z$,

as the $x_{it}$ are excess returns.

We depart from Uppal et al. (2018) in several ways. First, the risk factors $f_t$ are never observed in our framework. Second, the true pricing errors $a_{it}$ cannot be identified, and thus cannot be estimated, by the PCA procedure, leading us to ignore the term in the $a_{it}$s in (17). Third, and foremost, as discussed above, in a fixed-$T$ environment one cannot identify the risk premia $\gamma_{t-1}$ but only $\gamma_{s-1}^P$ and its sample average $\bar{\gamma}^P$. Along the same lines, when $T$ remains fixed, we cannot identify the population quantities $\Omega_t$ and $f_{t+1} - E_t(f_{t+1}) = F_{t+1} - E_t(F_{t+1})$ but only their sample counterparts $\hat{\Omega} \equiv T^{-1} F'M_{1,N} F$ and $\hat{f}_{t+1} - \bar{\gamma}^P$.

This implies that, within our fixed-$T$ environment, the object of inference consists necessarily of the ex-post SDF:

$$ m^P_{t,t+1} = \frac{1}{r_{ft}} - \frac{1}{r_{ft}} \gamma^P \Omega^{-1}(F_{t+1} - \bar{\gamma}^P) $$

for every $t \in Z$.

Note that $m^P_{t,t+1}$ is not feasible and needs to be estimated as the $F_t$ are not observed or, in other words, $m^P_{t,t+1}$ represents the population ex-post SDF, from a fixed-$T$ perspective, even though $m^P_{t,t+1}$ is a function of the time-series sample moments of the $F_t$.

Clearly, $m_{t,t+1}$ and $m^P_{t,t+1}$ differ but, under mild assumptions on the latent risk factors, the ex-post SDF $m^P_{t,t+1}$ satisfies the pricing conditions (18) with a (pricing) error of order $O(T^{-1})$, namely (see Proposition 2 for a formal proof):

$$ E_t(m^P_{t,t+1}x_{it+1}) = O(T^{-1}) $$

for every $i = 1, \cdots, N$ and $t \in Z$.

\footnote{Although this is not considered in our large-$N$ and fixed-$T$ setting, if one considers the nonparametric sampling scheme by which $T_0^{-1} + T/T_0 \to 0$, that is when the length of the sample becomes arbitrarily small as $T_0$ increases, then $\bar{\gamma}^P$ would identify the time-varying risk premia $\gamma_{t-1}$. See Raponi et al. (2018) Internet Appendix, for details on nonparametric estimation of risk premia. This provides a further justification for studying $\bar{\gamma}$ as an estimator of $\bar{\gamma}^P$.}
This fast rate is achieved because, although \( m_{t,t+1} \) and \( m^P_{t,t+1} \) differ by an order of magnitude of \( O_p(T^{-\frac{1}{2}}) \), due to the difference between \( \Omega_t, F_{t+1} - \gamma_t F_{t+1} \) and \( \hat{\Omega}, F_{t+1} - \hat{\gamma} F_{t+1} \), the averaging embedded in expectation operator involved when using the SDFs for pricing (e.g. Eq. (18), leads to the faster \( O(T^{-1}) \) rate.\(^{18}\)

As the PCA procedure provides the estimator \( \tilde{F}_{s,t} \) of the \( F_s \), the natural estimator for \( m^P_{s,s+1} \) is:

\[
\tilde{m}_{s,s+1} \equiv \frac{1}{r_f} - \frac{1}{r_f} \hat{\gamma}' \hat{\Omega}^{-1} (\tilde{F}_{s+1,t} - \hat{\gamma}) \text{ for every } s \in \mathcal{T}_t \text{ and } t \in \mathcal{Z},
\]

(20)

setting \( \hat{\Omega} \equiv T^{-1} \tilde{F}' \tilde{M} \tilde{F} = I_{r_t} - \hat{\gamma} \hat{\gamma}' \). We will illustrate below the limiting properties, as \( N \to \infty \), for our SDF estimator, which will take into account the sampling variability stemming from the PCA estimates of the risk factors.

### 3.4 Assumptions

We now present our regularity assumptions needed to deriving the limiting properties of the PCA estimator for the number of factors and for the factors themselves. Note that all convergences hold for \( N \to \infty \) and conditionally on \( F = (F_1, \cdots, F_T)' \) unless stated otherwise. Without loss of generality, we can follow Bai & Ng (2002) and assume that the loadings \( \lambda_i \) are non random but, given our large-\( N \) approach, it seems justified to allow them to be random. We report our notation in Appendix A.

**Assumption 2 (factors)** For any \( T \) such that \( r_t \leq T < \infty \),

\[
\hat{\Sigma}_F = \frac{F'F}{T} > 0 \text{ a.s.}
\]

If \( r_{t-1} > T \) then \( \hat{\Sigma}_F \) will be singular, with at least \( r_{t-1} - T \) zero eigenvalues, a case that we rule out as it induces multiple (and, moreover, zero) eigenvalues.

**Assumption 3 (loadings time-varying)** For every \( t \in \mathcal{Z} \):

\[
\frac{\Lambda_{t-1} A_{t-1}}{N} \to_p \Sigma_{A_{t-1}} > 0.
\]

**Assumption 4 (eigenvalues)** For every \( t \in \mathcal{Z} \), the eigenvalues of the \( r_{t-1} \times r_{t-1} \) matrix \( (\Sigma_{A_{t-1}} \hat{\Sigma}_F) \) are distinct a.s.

---

\(^{18}\)Proposition 2 represents the only result of this paper where we allow \( T \) to diverge, throughout this paper, and its only purpose is to establish that using the ex-post SDF \( m^P_{t,t+1} \) leads to pricing errors of almost identical magnitude to \( m_{t,t+1} \).
Assumption 5 (idiostatic component - temporal and cross-sectional dependence) Set \( \sigma_{ij,uv} = \mathcal{E}(e_{iu}e_{ju}) \) and \( \kappa_{h,i_1...i_h,t_1...t_h} = \text{cumulant}(e_{i_1t_1}, e_{i_2t_2}, \ldots, e_{i_ht_h}) \), with \( \mathcal{E}_s(e_{iu}e_{ju}) = \mathcal{E}(e_{iu}e_{ju}) \) for every \( s < u,v \).

For every \( 1 \leq i,j \leq N \), every \( s,v \in T_i \) and \( t \in \mathbb{Z} \):

\[
\mathcal{E}(e_{is}) = 0, \mathcal{E}(e_{it}e_{is}e_{ju}) = 0, |\sigma_{ij,sv}| \leq C, |\kappa_{iijj,ttss}| \leq C
\]

\[
\frac{1}{N} \sum_{i=1}^{N} (e_{is}e_{iv} - \sigma_{si,sv}) \rightarrow_p 0,
\]

\[
\frac{1}{N} \sum_{i=1}^{N} (e_{is}^4 - \kappa_{4,iii,ssss} - 3\sigma_{si,sv}^2) \rightarrow_p 0,
\]

with

\[
\frac{1}{N} \sum_{i=1}^{N} (\sigma_{si,sv}^2 - \sigma_{4t}) = o\left(\frac{1}{\sqrt{N}}\right), \frac{1}{N} \sum_{i=1}^{N} |\sigma_{ij,sv}| \rightarrow 0 \quad \text{and} \quad \sup_{t \in \mathbb{Z}} g_t(\Sigma_{et-1}) = o(N^{\frac{1}{4}}) \quad a.s.,
\]

\[
\frac{1}{N} \sum_{i=1}^{N} (\sigma_{si,sv}^2 - \sigma_{4t}) = o\left(\frac{1}{\sqrt{N}}\right),
\]

\[
\frac{1}{N} \sum_{i=1}^{N} (\kappa_{4,iii,ttss} - 1(s=v)\sigma_{4t-1}) = o\left(\frac{1}{\sqrt{N}}\right) \quad \text{and} \quad \frac{1}{N} \sup_{t \in \mathbb{Z}} \sum_{i=1}^{N} |\kappa_{h,i_1...i_h,t_1...t_h}| \rightarrow 0
\]

for every \( 3 \leq h \leq 8 \), every \( t_j \in T_i \ (1 \leq j \leq h) \) and for at least one \( i_j \ (2 \leq j \leq h) \) different from \( i_1 \), for some \( \sigma_{4t}, \sigma_{4t}, \kappa_{4t} \) satisfying

\[
0 < \inf_{t \in \mathbb{Z}} \sigma_{4t}^2 \leq \sup_{t \in \mathbb{Z}} \sigma_{4t}^2 < \infty, 0 < \inf_{t \in \mathbb{Z}} \sigma_{4t} \leq \sup_{t \in \mathbb{Z}} \sigma_{4t} < \infty, 0 \leq \inf_{t \in \mathbb{Z}} |\kappa_{4t}| \leq \sup_{t \in \mathbb{Z}} |\kappa_{4t}| < \infty.
\]

Our assumptions of (local in time but asymptotic in \( N \)) dynamic homoskedasticity and uncorrelatedness of the \( e_{it}s \) are suggested by [Bai (2003)], who establishes that these conditions are necessary and sufficient for consistency of the PCA factors when \( T \) is fixed and the factor model is static, strengthening [Connor & Korajczyk (1980)] sufficiency result. We are able to somewhat relax these in the sense to impose them locally within any given time-interval of length \( T \). At any rate, asset returns, especially at the level of individual assets, appear to be empirically only mildly autocorrelated across time. Our assumption imposes equality between conditional and unconditional moments of the \( e_{it}s \). However, this is not as restrictive as it seems given that we are not imposing constancy but are instead allowing for (local) time-variation of the unconditional moments, a form of local stationarity along the lines of [Dahlhaus (1997)]. As \( N \) diverges, the number of parameters say in \( \Sigma_{et-1} \) increases, a manifestation of the curse of dimensionality and it becomes necessary to limit the cross-sectional heterogeneity of the moments of the idiosyncratic component \( e_{it} \). The rate on the maximum eigenvalue of \( \Sigma_{et-1} \) does not follow from the previous condition on the behaviour of the sum of the cross-covariances, and is
necessary in order to deal with the effect of the pricing errors $a_{it}$ in the asymptotics. This eigenvalue condition is not needed when $a_{it-1} = 0$ for every $i$ and $t$.

Our assumptions are extremely mild in terms of the permitted degree of cross-sectional dependence. For instance assume that the $e_{it}$ satisfy a factor structure, imposing constant moments for simplicity, such as:

$$e_{i,t} = \delta_i u_t + \eta_{i,t},$$

where, for some positive finite constants $C_1 \leq C_2$ and $0 < \epsilon_1 \leq 2\epsilon_2 < 1/2$, for any $N$,

$$C_1 N^{\epsilon_1} \leq \sum_{i=1}^{N} \delta_i^2 \text{ and } \sum_{i=1}^{N} |\delta_i| \leq C_2 N^{\epsilon_2},$$

with $u_t$ i.i.d. $(0, 1)$ and $\eta_{i,t}$ i.i.d. $(0, \sigma^2_\eta)$ over time and across units, where the $u_t$ and the $\eta_{i,s}$ are mutually independent for every $i, s, t$. It follows that the second and fourth condition (involving the second moments) of Assumption 5 hold and, more in general all the other conditions will follow by making suitable higher-moments assumptions on the $\eta_{it}$. At the same time, the maximum eigenvalue of the covariance matrix $\Sigma_e = \text{var}(e_{it})$,

$$\sum_{i=1}^{N} \delta_i^2 \text{ and } \sum_{i=1}^{N} |\delta_i| \leq C_2 N^{\epsilon_2},$$

with $u_t$ i.i.d. $(0, 1)$ and $\eta_{i,t}$ i.i.d. $(0, \sigma^2_\eta)$ over time and across units, where the $u_t$ and the $\eta_{i,s}$ are mutually independent for every $i, s, t$. It follows that the second and fourth condition (involving the second moments) of Assumption 5 hold and, more in general all the other conditions will follow by making suitable higher-moments assumptions on the $\eta_{it}$. At the same time, the maximum eigenvalue of the covariance matrix $\Sigma_e = \text{var}(e_{it})$, given by $\sum_{i=1}^{N} \delta_i^2 + \sigma^2_\eta$, diverges at a rate not slower than $O(N^{\epsilon_1})$, for some $0 < \epsilon_1 < 1/2$. This implies, for example, that the the row- and column-norm of $\Sigma_e$, namely $\sup_j \sum_{i=1}^{N} |\sigma_{ij,tt}|$ also necessarily diverges with $N$, manifesting a substantial degree of cross-sectional dependence.

The zero third-moment assumption is only made to simplify exposition and can be relaxed. The cumulants’ assumption implies that the fourth-order cumulant $\kappa_{4,iijj,tstv} = \text{cumulant}(e_{it}, e_{is}, e_{jt}, e_{jv})$ satisfies $N^{-1} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \kappa_{4,iijj,tstv} \to 0$.

**Assumption 6 (mixed moments)** For every $t \in \mathbb{Z}$:

$$\frac{1}{N} \sum_{i=1}^{N} \lambda_{it-1} e'_{i} \to_p 0_{r_{t-1} \times T},$$

$$\frac{1}{N} \sum_{i=1}^{N} (\lambda_{it-1} \lambda'_{it-1} \otimes \mathcal{E}(e_{i} e'_{j})) \to_p (\Sigma_{\Lambda_{t-1}} \otimes \sigma^2_{t-1} I_T),$$

Condition [24] can be generalized to

$$\frac{1}{N} \sum_{i=1}^{N} (\lambda_{it-1} \lambda'_{it-1} \otimes \mathcal{E}(e_{i} e'_{j})) \to_p \Gamma_{t-1},$$

as in [Bai 2003], without imposing $\Gamma_{t-1} = (\Sigma_{\Lambda_{t-1}} \otimes \sigma^2_{t-1} I_T)$\footnote{Consistent estimation of $\Gamma_{t-1}$ is possible although with a more elaborate proof.}. In contrast, under our assumptions, it is possible to estimate the asymptotic covariance matrix of $1/\sqrt{N} \sum_{i=1}^{N} (e_{i} e_{is} - \mathcal{E}(e_{i} e_{is}))$, as explained below.
Assumption 7 (joint convergence in distribution) For every \( s \in T_t \) and \( t \in \mathbb{Z} \):

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ (\mathbf{e}_i e_{is} - \mathcal{E}(\mathbf{e}_i e_{is})) \right] \rightarrow_d \mathcal{N} \left( 0_{T(r_t-1+1)}, \begin{bmatrix} (\kappa_{t-1} + \sigma_{t-1}) I_t & 0_{T \times r_{t-1}} \\ 0_{T \times r_{t-1}} & 0_{T \times T} \end{bmatrix} \right). \tag{25}
\]

Our theorems require the joint distribution of

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ (\mathbf{e}_i e_{is} - \mathcal{E}(\mathbf{e}_i e_{is})) \right],
\]

which can be derived by premultiplying the right hand side of (25) by the \((T + r_{t-1} + r_{t-1}) \times (T + r_{t-1} T)\) matrix:

\[
A_{s,t} \equiv \begin{bmatrix} I_T & 0_T \times r_{t-1} \times (T-s) \\ 0_{r_{t-1} T \times T} & I_{r_{t-1} T \times r_{t-1} T} \end{bmatrix},
\]

one for every \( s \in T_t \) and \( t \in \mathbb{Z} \).

Primitive conditions for Assumption 7 can be derived but at the cost of raising the level of complexity of our proofs. For instance, when (21) with (22) hold,

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{e}_i \rightarrow_d \mathcal{N} \left( 0_T, \sigma^2 \eta I_T \right) \text{ as } N \rightarrow \infty,
\]

by Theorem 2 of Kuersteiner & Prucha (2013) when the \( \eta_{it} \) satisfy their martingale difference assumption (see their Assumptions 1 and 2.) This result extends easily to (25) under suitable additional assumptions. (Details are available upon request.)

The implicit assumption that justifies this new angle is that the time-variation affecting model (1), in particular regarding the loadings \( \Lambda_{t-1} \) and the idiosyncratic variances \( \sigma_{ii,tt} \), is sufficiently smooth within the time window of size \( T \). The strength and generality afforded by such argument relies on the possibility of taking \( T \) sufficiently small, which the asymptotic theory of the preceding sections allows, as long as \( r_{t-1} < T \) for every \( T \leq t \leq T_0 \). In particular, the following three smoothness assumptions suffice.

Assumption 8 (number of factors - smoothness) For every \( s \in T_t \) and \( t \in \mathbb{Z} \),

\[
r_s = r_{t-1}.
\]

Our fixed-\( T \) large-\( N \) approach permits time-variation of virtually any parameter of interest. Even if \( r_{t-1} \) is allowed to change over time, we are ruling out that such change occurs within the given interval \( T_t \).
Assumption 9 (loadings - smoothness) For every $s \in T_t$ and $t \in \mathbb{Z}$, the $\Lambda_{t-1}$ are differentiable functions of $t$ such that their first-derivative, $\Lambda_{t-1}^{(1)}$, satisfy

$$\Lambda_{t-1}^{(1)} = o_p(N), \ e_i^{(1)} \Lambda_{t-1} = o_p(N).$$

As an example of Assumption 9, one can have that only a fraction $N_0/N$ of the units has a time-varying loadings, with $N_0/N \to 0$. An important, special, case is when the loading matrix $\Lambda_{t-1}$ is a differentiable function of a set of observed state variables, say represented by the finite-dimensional vector $z_{t-1}$, that is $\Lambda_{t-1} = \Lambda(z_{t-1})$.

Assumption 10 (idiosyncratic component - smoothness) The $e_i$ satisfy

$$e_i = \sum_{i} \eta_{i1} \sigma_{i1}^{(1)} \text{ with } \Sigma = \mathbb{E}(e_i e_j^\prime) = \sigma_{i1,uv} = 0 \text{ for } i = 1, \ldots, N \text{ and } u, v \in T_t, t \in \mathbb{Z},$$

and $\eta_{i1} = (\eta_{i1}, \ldots, \eta_{iT})^\prime$ satisfying $\mathbb{E}(\eta_{i1}) = 0, \mathbb{E}(\eta_{i1} \eta_{i1}^\prime) = \mathbf{I}_T$ with $\eta_{i1}$ and $\Sigma_{i1}$ mutually independent for any $i, j$. Then, for every $s \in T_t$ and $t \in \mathbb{Z}$, assume that the $\sigma_{i1,s}$ and $\kappa_{i1,i1,s}$ are differentiable functions of $s$ such that they first-derivative, $\sigma_{i1,s}^{(1)}$ and $\kappa_{i1,i1,s}^{(1)}$, satisfy, for every $a, b, c, d \in T_t$ and $\sigma_{i1,a} \leq b \leq \sigma_{i1,b} \leq t, c \leq \sigma_{i1,c} \leq t, d \leq \sigma_{i1,d} \leq t$:

$$\mathop{\sum}\limits_{i=1}^{N} |\eta_{i1} \sigma_{i1,a}^{(1)} \sigma_{i1,b}^{(1)}| = o_p(N), \mathop{\sum}\limits_{i=1}^{N} |\eta_{i1} \sigma_{i1,a}^{(1)} \sigma_{i1,tt}| = o_p(N),$$

$$\mathop{\sum}\limits_{i=1}^{N} |\sigma_{i1,s}^{(1)}| = o(N), \mathop{\sum}\limits_{i=1}^{N} |\sigma_{i1,tt}^{(1)}| = o(N), \mathop{\sum}\limits_{i=1}^{N} |\sigma_{i1,a}^{(1)} \sigma_{i1,b}^{(1)}| = o(N),$$

$$\mathop{\sum}\limits_{i=1}^{N} (\Lambda_{it-1} - \Lambda_{it}) |\sigma_{i1,a}^{(1)}| = o_p(N), \mathop{\sum}\limits_{i=1}^{N} |\kappa_{i1,i1,s}^{(1)}| = o(N).$$

4 Inference

4.1 Inference on Risk Factors

In this section we present the limiting distribution theory for the PCA-estimated risk factors.

Theorem 1 (rate of convergence and limiting distribution of the risk factors estimator) Under Assumptions 1-10, as $N \to \infty$, $\hat{F}(T_t) = (\hat{F}_{t-T_t+1}, \ldots, \hat{F}_{s,t}, \ldots, \hat{F}_{t,t})'$ and $\hat{\Lambda}(T_{t-1}) = (\hat{\Lambda}_{1t-1}, \ldots, \hat{\Lambda}_{Nt-1})'$ satisfy:

(i) $\|\hat{F}_{s,t} - \hat{F}_{s,t}^*\| = O_p(N^{-1/2})$ for every $s \in T_t$ and $t \in \mathbb{Z}$. 
(ii) \[
\sqrt{N}(\tilde{F}_{s,t} - \tilde{H}_{t-1}F_s) \xrightarrow{d} N(0_{r_t-1}, A_{r_t-1}B_{s,t-1}A_{t-1}) \text{ for every } s \in T_t \text{ and } t \in \mathcal{Z},
\]
setting
\[
A_{s,t} = \begin{pmatrix}
\sigma_t^2 U_t^{-2} + (\kappa_t + \sigma_t^4) & 0_{r_t \times r_t} & 0_{r_t \times r_t} \\
0_{r_t \times r_t} & \sigma_t^2 U_t^{-1} & 0_{r_t \times r_t} \\
0_{r_t \times r_t} & \sigma_t^2 U_t^{-1} & \sigma_t^2 U_t^{-1} Q_t \Sigma A_t F_s F'_s H_t U_t^{-1} \\
\end{pmatrix},
\]
where \(Q_t\) and \(U_t\) are defined in Lemma 3.

**Proof.** See Appendix D.

When \(T\) is also allowed to diverge, the asymptotic distribution of the estimated factors simplifies significantly. In particular, the result of Bai (2003) Theorem 5, is re-obtained for our static case, except that we obtain a much simplified expression for the asymptotic covariance matrix: thank to our Assumption 6, adopting Bai (2003) notation, \(V_{t-1}^{-1}Q_{t-1} \Gamma_{t-1} Q_{t-1} V_{t-1}^{-1} = \sigma_t^2 U_t^{-1}\). The latter formula carries a very neat interpretation: it states that the (asymptotic) precision of the estimated factors increases proportionally with the magnitude of the associated (normalized) eigenvalue. At the same time, the precision diminishes, equally across the \(r_{t-1}\) factors, when the idiosyncratic average variance \(\sigma_t^2 U_t^{-2}\) increases.

In order to construct confidence intervals and to perform hypothesis testing, for the estimated factors, one needs to consistently estimate the asymptotic covariance matrix reported in Theorem 1(ii). This entails estimating \(\sigma_t^2 U_t^{-1}\) and the matrices \(\Sigma A_{t-1}, Q_{t-1}, U_{t-1}\), among others. For the static case Bai (2003) shows that this is feasible when both \(N\) and \(T\) diverge, despite the rotation indeterminacy of the estimated factors and loadings associated with the rotation matrix \(\tilde{H}_{t-1}\). However, in our sampling framework, estimation of \(\Sigma A_{t-1}\) appears particularly problematic because the \(\lambda_{it-1}\) are not consistently estimated when \(T\) is fixed, regardless of the rotation. In particular, following Bai (2003) [proof of Theorem 2] and using by Theorem 1(i), the following decomposition and limit hold:
\[
\lambda_{it-1} = \tilde{H}_{t-1}^{-1} \lambda_{it-1} + \frac{\tilde{F}' e_i}{T} + \frac{1}{T} \tilde{F}' (F - \tilde{F} \tilde{H}_{t-1}^{-1}) \lambda_{it-1} \rightarrow_p \tilde{H}_{t-1}^{-1} \lambda_{it-1} + \frac{\tilde{H}_{t-1}^{-1} \tilde{F}' e_i}{T}.
\]
Note that the lack of consistency of the loadings is unrelated to latency of the factors: it will still hold when the \(F_t\) are observed. However, by a careful analysis, we are still able to estimate consistently the second moment matrix of the \(\lambda_{it-1}\) for \(T\) fixed. This is accomplished in the next theorem.
Theorem 2 (consistent estimation of asymptotic covariance matrix) Under Assumptions 1-10 and the identification condition $\kappa_{4t-1} = 0$, as $N \to \infty$,

$$\| \hat{B}_{s,t} - B_{s,t} \|_p \to 0 \text{ for every } s \in \mathcal{T}_t \text{ and } t \in \mathcal{Z},$$

where

$$
\begin{bmatrix}
\hat{\sigma}_t^2 T^{-1} \hat{U}_t^{*\dagger} (I_{r_t} + T^{-1} \tilde{F}_{s,t} \tilde{F}_{s,t}') \hat{U}_t^{*\dagger} & 0_{r_t \times r_t} & 0_{r_t \times r_t} \\
0_{r_t \times r_t} & \hat{\sigma}_t^2 T^{-1} \hat{U}_t^{*\dagger} \tilde{F}_{s,t} \tilde{F}_{s,t}' \Sigma_{\lambda_t} \hat{U}_t^{*\dagger} & 0_{r_t \times r_t} \\
0_{r_t \times r_t} & 0_{r_t \times r_t} & \hat{\sigma}_t^2 T^{-1} \hat{U}_t^{*\dagger} \Sigma_{\lambda_t} \tilde{F}_{s,t} \tilde{F}_{s,t}' \hat{U}_t^{*\dagger}
\end{bmatrix},
$$

setting

$$
\hat{\sigma}_t^2 \equiv \frac{T}{T - r_t} V_{\mathcal{T}_t}(r_t) \text{ with } V_{\mathcal{T}_t}(r_t) \equiv \frac{1}{N_T} \text{trace} \left( (X - \tilde{F}\tilde{A})'(X - \tilde{F}\tilde{A})' \right),
$$

$$
\hat{\sigma}_{4t} \equiv \left( 3 + \frac{27}{T} \left( \sum_{s \in \mathcal{T}_t} \left( \frac{\tilde{F}_{s,t}' \tilde{F}_{s,t}}{T} \right)^2 \right) + \frac{18r_t}{T} \right)^{-1} \left( \frac{1}{N_T} \sum_{i=1}^{N} \sum_{s \in \mathcal{T}_t} \bar{e}_{is}^4 \right),
$$

$$
\hat{U}_t^* = \tilde{V}_t - \frac{\hat{\sigma}_t^2}{T} I_{r_t} = \tilde{\Sigma}_{\lambda_t}.
$$

Proof. See Appendix D.

When estimating the asymptotic covariance matrix, one needs necessarily to set $\kappa_{4t-1} = 0$ for identification purposes, as evinced from Lemma 3 in Appendix C. However, higher-order cumulants are not constrained to be zero, implying that $\kappa_{4t-1} = 0$ is not equivalent to Gaussianity. Moreover, heterogeneity of the $\sigma_{it,tt}$ across assets induces (potentially) a very large (average) volatility of the idiosyncratic component $e_{it}$, quantified by the difference between $\sigma_{4t-1}$ and $\sigma_{t-1}^2$. Note that $\Sigma_{\lambda_{t-1}}$ is identical to $\hat{U}_{t-1}^*$ as $\tilde{V}_{t-1} = \tilde{\Lambda}_{t-1}^{-1} \tilde{\Lambda}_{t-1} \hat{U}_{t-1}^* N_T$, but, for expositional clarity, we prefer to refer to them with a different notation, in particular to facilitate the construction of the standard errors for the PCA-estimated risk factors.

We now illustrate estimation of the number of latent factors. Hence, assume that a $k$ factor model is estimated, where $k$ can be smaller, equal or larger than the true number of factors $r_{t-1}$. In particular, $\tilde{F}^k$ denotes the $T \times k$ matrix of the $k$ eigenvectors associated with $\tilde{V}_{t-1}^k$, the diagonal matrix containing the $k$ largest eigenvalues of $XX'/NT$ and, similarly, $\tilde{\Lambda}_{t-1}^k = T^{-1} X' \tilde{F}^k$ denotes the $N \times k$ matrix of loadings Set, for some function $g(N)$,

$$
PC_{\mathcal{T}_t}(k) \equiv V_{\mathcal{T}_t}^2(k) + kg(N)
$$

with $V_{\mathcal{T}_t}^2(k) \equiv \left( \frac{T}{T-K} \right) V_{\mathcal{T}_t}(k)$ and

$$
V_{\mathcal{T}_t}(k) \equiv \frac{1}{N_T} \text{trace} \left( (X - \tilde{F}^k \tilde{\Lambda}_{t-1}') (X - \tilde{F}^k \tilde{\Lambda}_{t-1}')' \right).
$$
Theorem 3 (consistent estimation of the number of factors) Suppose that Assumptions 1-10 hold and that \( g(N) \to 0 \) together with \( \sqrt{N} g(N) \to \infty \). For every \( t \in \mathbb{Z} \), set

\[
\tilde{k}_{t-1} \equiv \arg\min_{0 \leq k \leq T} PC_T(k).
\]

Then

\[
\text{Prob}(\tilde{k}_{t-1} = r_{t-1}) \to 1 \quad \text{as} \quad N \to \infty.
\]

**Proof.** See Appendix D.

As for the double asymptotic result of Bai & Ng (2002), the result follows by showing that \( V^*_{T_t}(k) \) is (a.s.) up-ward biased when \( k < r_{t-1} \) although it is not (asymptotically) biased whenever \( k \geq r_{t-1} \). The penalization through the \( g(N) \) function, on the other hand, is asymptotically negligible precisely when is not needed, that is when \( k < r_{t-1} \), and instead it induces a bias, as desired, whenever \( k > r_{t-1} \), the bigger the slower is \( g(N) \) converging towards zero, leading to consistency of \( \tilde{k}_{t-1} \) for \( r_{t-1} \).

The strength of Theorem 3 is that it represents the only existing method to consistently estimate the number of latent factors for model (1) when \( T \) is fixed and \( N \) diverges. All other existing methods surveyed in Section 2 require double asymptotics.

Theorem 3 relies on studying the behaviour of \( \tilde{F}^k \) including the case \( k > r_{t-1} \) (see Lemma 5). In particular, we show that when an erroneously larger number of risk factors are considered, that is \( k > r_{t-1} \), the corresponding risk premia consist of a linear combination of the risk premia associated with the correct number (viz. \( r_{t-1} \)) of risk factors. This suggests that the use of PCA for risk premia estimation could potentially avoid the problems arising from including too many factors, such as when useless, or spurious, factors (i.e. factors that are uncorrelated with the test assets) are included.

4.2 Inference on Risk Premia and the SDF

The previous results permit to derive the limiting properties of the risk premia and SDF estimators. In particular, in view of (14), Theorem 1 provides directly the limiting distribution theory for the estimator of

\cite[Jagannathan & Wang (1998)], \cite[Kan & Zhang (1999a, b)], \cite[Gospodinov et al. (2017)] and \cite[Anatolyev & Mikusheva (2018) and Raponi & Zaffaroni (2018)] pointed out how inclusion of useless factors leads to serious problems when conducting inference on beta-pricing models, and various methods have been proposed to tackle them; see \cite[Kleibergen (2009), Gospodinov et al. (2014), Bryzgalova (2014), Burnside (2016), Anatolyev & Mikusheva (2018) and Raponi & Zaffaroni (2018)] among others.
\( \gamma_{s-1}^P \), namely:
\[
\tilde{\gamma}_{s-1} = \tilde{F}_{s,t} \quad \text{for every } s \in \mathcal{T}_t \text{ and } t \in \mathcal{Z},
\]
and for the numerically equivalent estimator \( \tilde{\gamma}_{s-1}^{\text{pass}} \).

It is important to recognize that no gain appears with respect to estimation of the time-varying risk premia \( \gamma_{s-1} \) from taking \( T \) large. This represents a key motivation for using our large-\( N \) results. More specifically,
\[
\tilde{\gamma}_{s-1} - \tilde{H}_{t-1}^T \gamma_{s-1} = \tilde{H}_{t-1}^T (f_s - \mathcal{E}_{s-1}(f_s)) + \tilde{\Gamma}_{s,N_T} = O_p(1) + O_p(N^{-\frac{1}{2}}),
\]
for a random quantity satisfying \( \tilde{\Gamma}_{s,N_T} = O_p(N^{-\frac{1}{2}}) \), by Theorem 1. The only effect of allowing also a large \( T \) is that the asymptotic covariance matrix of \( \tilde{\gamma}_{s-1} \) will be smaller (in the matrix sense) without affecting the rate of convergence, which is still \( O_p(N^{-\frac{1}{2}}) \).

Summarizing, when time-variation is allowed for, the ex-ante risk premia \( \tilde{H}_{t-1}^T \gamma_{s-1} \) cannot be consistently estimated even when \( N \) and \( T \) both diverge. Instead, the ex-post risk premia \( \tilde{H}_{t-1}^T \gamma_{s-1}^P \) can be consistently estimated only when \( N \) diverges, with \( T \) affecting the asymptotic precision. The latter effect could be relevant in finite, small, samples.

Considering now the average risk premia estimator \( \tilde{\gamma} \) in (10), it follows that
\[
\tilde{\gamma} = T^{-1} \sum_{s \in \mathcal{T}_t} \tilde{F}_{s,t} \to_p T^{-1} \sum_{s \in \mathcal{T}_t} \tilde{H}_{t-1}^T \gamma_{s-1}^P = \tilde{H}_{t-1} \left( \tilde{\gamma} + \tilde{f} - T^{-1} \sum_{s \in \mathcal{T}_t} \mathcal{E}_{s-1}(f_s) \right) \equiv \tilde{H}_{t-1} \tilde{\gamma}^P,
\]
setting \( \tilde{\gamma} = T^{-1} \sum_{s \in \mathcal{T}_t} \gamma_{s-1} \) and \( \tilde{f} = T^{-1} \sum_{s \in \mathcal{T}_t} f_s \), that is \( \tilde{\gamma} \) now captures accurately the local average of the time-varying risk premia \( \tilde{\gamma}^P \), for which inference can be made by generalizing our PCA results as follows.

**Theorem 4 (rate of convergence and limiting distribution of the risk premia estimator)** Under Assumptions 1-10, as \( N \to \infty \),

(i) The risk premia estimator \( \tilde{\gamma} \) identifies the population risk premia \( \gamma^P \) multiplied by the rotation matrix \( \tilde{H}_{t-1} \).

(ii)
\[
\| \tilde{\gamma} - \tilde{H}_{t-1} \tilde{\gamma}^P \| = O_p(N^{-\frac{1}{2}}) \text{ for every } t \in \mathcal{Z}.
\]

21In particular:
\[
\gamma_{s-1} = \tilde{H}_{t-1} \gamma_{s-1}^P + \tilde{U}_{t-1} \tilde{F} \left( \mathcal{E}_{s-1} \frac{\mathcal{F}}{N} - \mathcal{E}(\frac{\mathcal{F}}{N}) \right) + \tilde{U}_{t-1} \tilde{F} \mathcal{E}_{s-1} \frac{\mathcal{F}}{N} + \tilde{U}_{t-1} \tilde{F} \mathcal{E}_{s-1} \frac{\mathcal{F}}{N} \mathcal{E}_{s-1} \tilde{F}_s
\]
\[
= \tilde{H}_{t-1} \gamma_{s-1} + \tilde{H}_{t-1} \mathcal{E}_{s-1} \tilde{F}_s + \tilde{\Gamma}_{s,N_T},
\]
where \( \Gamma_{s,N_T} \equiv T^{-1} \tilde{U}_{t-1} \tilde{F} \left( \mathcal{E}_{s-1} \frac{\mathcal{F}}{N} - \mathcal{E}(\frac{\mathcal{F}}{N}) \right) + T^{-1} \tilde{U}_{t-1} \tilde{F} \mathcal{E}_{s-1} \frac{\mathcal{F}}{N} + T^{-1} \tilde{U}_{t-1} \tilde{F} \mathcal{E}_{s-1} \tilde{F}_s. \)
(iii) $$\sqrt{N}(\hat{\gamma} - \hat{H}_{t-1}'\tilde{\gamma}^F) \to_d N(0_{r_1 \times 1}, A_{t-1}'C_{t-1}A_{t-1})$$ for every $$t \in \mathbb{Z}$$, setting

$$C_t \equiv \begin{bmatrix} \left(\frac{\kappa_{4t} + \sigma_{4t}}{T}\right)U_t^{-2} + \frac{\sigma_{4t}}{T}U_t^{-1}H_t'\bar{F}\bar{F}'H_tU_t^{-1} & 0_{r_1 \times r_t} & 0_{r_1 \times r_t} \\ 0_{r_t \times r_1} & \frac{\sigma_t^2}{T}U_t^{-1} & \frac{\sigma_t^2}{T}U_t^{-1}Q_t\Sigma_t\bar{F}\bar{F}'H_tU_t^{-1} \\ 0_{r_t \times r_1} & 0_{r_t \times r_1} & \frac{\sigma_t^2}{T}U_t^{-1}Q_t\Sigma_t\bar{F}\bar{F}'H_tU_t^{-1} \end{bmatrix},$$

where $$Q_t$$ and $$U_t$$ are defined in Lemma 3.

(iv) When, in addition, $$\kappa_{4t} = 0$$,

$$\|\hat{C}_t - C_t\| \to_p 0$$ for every $$t \in \mathbb{Z}$$, setting

$$\hat{C}_t \equiv \begin{bmatrix} \frac{\sigma_t^2}{T}\bar{U}_t^{-2} + \frac{\sigma_t^2}{T}U_t^{-1}\bar{F}\bar{F}'U_t^{-1} & 0_{r_1 \times r_t} & 0_{r_1 \times r_t} \\ 0_{r_t \times r_1} & \frac{\sigma_t^2}{T}U_t^{-1} & \frac{\sigma_t^2}{T}U_t^{-1}\Sigma_t\bar{F}\bar{F}'U_t^{-1} \\ 0_{r_t \times r_1} & 0_{r_t \times r_1} & \frac{\sigma_t^2}{T}U_t^{-1}\Sigma_t\bar{F}\bar{F}'U_t^{-1} \end{bmatrix},$$

where $$\tilde{\sigma}_t^2, \tilde{\sigma}_{4t}, \tilde{U}_t$$ and $$\tilde{\Sigma}_t$$ are defined in Theorem 2.

**Proof.** See Appendix D.

Theorem 4 is in agreement with Raponi et al. (2018), who noticed that their large-$$N$$ fixed-$$T$$ estimation procedure, designed for constant risk premia, would still lead to a meaningful estimator for local averages of the time-varying risk premia. Recall that $$T$$ is fixed and possibly very small so such local averages might still display a pronounced time-variation when evaluated over a sequence of rolling samples. In contrast, if one lets $$T$$ to diverge, then $$\tilde{\gamma}$$ will capture the integrated risk premia over the entire time line, that is it will converge to (net of a rotation) $$\int_{-\infty}^{\infty} \tilde{\gamma}_s ds$$, given that $$\tilde{f} - T^{-1}\sum_{s=t-T+1}^{t} \tilde{\varepsilon}_{s-1}(f_s) = o_p(1)$$ under mild conditions. Therefore, no information on time-variation can be recovered from $$\tilde{\gamma}$$ if one lets $$T$$ to diverge.

As the vector of estimated risk premia is a sample average of the latent factors, its precision increases with $$T$$, in particular at rate $$O(T^{-\frac{1}{2}})$$. Indeed, although this is deliberately not explored in this paper, our result shows that $$\tilde{\gamma}$$ is consistent at rate $$O((NT)^{\frac{1}{2}})$$ when both $$N$$ and $$T$$ diverge. A closer analysis explains the reasons behind such fast rate of convergence. In fact, averaging (28) across $$T_t$$ gives

$$\tilde{\gamma} = \tilde{H}'_{t-1}(\frac{1}{T} \sum_{s \in T_t} \gamma_{s-1}) + \tilde{H}'_{t-1}(\tilde{f} - \frac{1}{T} \sum_{s \in T_t} \tilde{\varepsilon}_{s-1}(f_s)) + \tilde{\Gamma}_{NT},$$

where we set $$\tilde{\Gamma}_{NT} \equiv T^{-1}\sum_{s \in T_t} \tilde{\Gamma}_{s,NT}$$. By Theorem 4, it follows that $$\tilde{\Gamma}_{NT} = o_p((NT)^{-\frac{1}{2}})$$ and, in fact, Dello Preite & Zaffaroni (2019) establish asymptotic normality of $$\sqrt{NT}\tilde{\Gamma}_{NT}$$, and thus of $$\sqrt{NT}(\gamma - \tilde{H}'_{t-1}\tilde{\gamma}^F),$$
as both \( N, T \) diverge. This fast rate seems at odds with the usual rate of convergence for estimated risk premia, in the presence of observed risk factors. For instance, the sample mean of a traded factor (i.e., a portfolio excess return) converges at rate \( O(\sqrt{T}) \), and the same rate applies to the traditional two-pass OLS estimator. The reason for this apparent contradiction is that \( \tilde{\gamma} - \tilde{H}'(\frac{1}{T}\sum_{s \in T_i} \nu_{s-1}) = \tilde{\gamma}^{P} \) is, implicitly, capturing the randomness (i.e., the asymptotic distribution) around \( \frac{1}{T}\sum_{s \in T_i} \nu_{s-1} \), where \( \nu_{s-1} \equiv \gamma_{s-1} - \mathcal{E}_{s-1}(f_s) \), namely the average risk premia netted out by the factor average conditional expectation or, in other words, the portion of the risk premia that is not linearly dependent on the factors’ expectation. Besides the rotation by the matrix \( \tilde{H}'(\frac{1}{T}\sum_{s \in T_i} \nu_{s-1}) \), which is just a by-product of having unobserved risk factors, such fast rate is identical to the one found by Gagliardini et al. (2016), in a double asymptotic setting, and when the risk factors are assumed to be observed.

In contrast, if inference on the average (ex-ante) average risk premia \( \frac{1}{T}\sum_{s \in T_i} \gamma_{s-1} \) is aimed for, then

\[
\tilde{\gamma} - \tilde{H}'(\frac{1}{T}\sum_{s \in T_i} \nu_{s-1}) = \tilde{\gamma}^{P} + \tilde{\Gamma}_{NT} = O_p(T^{-\frac{1}{2}}) + O_p((NT)^{-\frac{1}{2}}),
\]

and the slower, traditional, \( \sqrt{T} \)-rate of convergence emerges.

Regarding the two-pass estimator \( \tilde{\gamma}_{t-1}^{\text{twopass}} \), Theorem 4 applies, in view of (12), but one can also use direct arguments to show \( \sqrt{N} \)-rate of convergence:

\[
\tilde{\gamma}_{t-1}^{\text{twopass}} = \left( \tilde{A}'_{t-1}\tilde{A}_{t-1} \right)^{-1}\tilde{A}'_{t-1}\tilde{X}_{N} \to p \left( H^{-1}_{t-1} \Sigma A_{t-1} H^{-1}_{t-1} + \frac{\sigma^2_{t-1}}{T} I_{t-1} \right)^{-1} \left( H^{-1}_{t-1} \Sigma A_{t-1} H^{-1}_{t-1} \gamma^{P} + \frac{\sigma^2_{t-1}}{T} H'_{t-1} \tilde{F} \right) = H_{t-1} \gamma^{P}.
\]

Interestingly, no-bias adjustment is required for \( \tilde{\gamma}_{t-1}^{\text{twopass}} \), unlike the case when the risk factors are observed. Given (12), inference on \( \tilde{\gamma}_{t-1}^{\text{twopass}} \) can be readily conducted by using Theorem 4.

Giglio & Xiu (2017) illustrate a generalization of their method to time-varying loadings whereby the loadings are function of observed state variables. Similarly, but in the context of observed risk factors, Gagliardini et al. (2016) allow for observed state variables in loadings and risk premia. Our approach differs from both contributions because we do not need to make any assumption on the form of time-variation of risk premia, such as dependence from some state-variables.

Our asymptotic distribution theory delivers the limiting properties of the SDF estimator \( \tilde{m}_{t,t+1} \) of Eq. (20).

---

22 Obviously, with tradeable, and thus observed, factors \( \nu_{s-1} = 0_{s-1} \) under correct model specification.

23 See Raponi et al. (2018) for the analysis of the two-pass risk premia estimation when only \( N \) diverges.
Theorem 5 (rate of convergence and limiting distribution of the SDF estimator) Under Assumptions 1-10, as \( N \to \infty \),

(i) \[ \| \hat{m}_{s,s+1} - m^{P}_{s,s+1} \| = O_p(N^{-\frac{1}{2}}) \text{ for every } s \in T_t \text{ and } t \in \mathbb{Z}. \]

(ii) \[ \sqrt{N} \left( \hat{m}_{s,s+1} - m^{P}_{s,s+1} \right) \to_d N(0, D'_{s,t-1} E_{t-1} D_{s,t-1}) \text{ for every } s \in T_t \text{ and } t \in \mathbb{Z}, \]

setting

\[
D_{s,t} \equiv (I_T \otimes H_t^{-1}) \frac{1}{r_{fs}} \left( -((\frac{1}{T}) \otimes I_{r}) \Omega^{-1}(F'(\tau_{s+1} - \frac{1}{T})) - ((\tau_{s+1} - \frac{1}{T}) \otimes I_{r}) \Omega^{-1}(F'\frac{1}{T}) \right) \\
+ (M_L \tilde{F} \Omega^{-1} F' \otimes \Omega^{-1} F') \left( \frac{1}{T} \otimes (\tau_{s+1} - \frac{1}{T}) \right) + ((\tau_{s+1} - \frac{1}{T}) \otimes \frac{1}{T}) \right),
\]

\[
E_t \equiv (I_T \otimes U_t^{-1} H_t' F' \otimes \Omega) \Upsilon_t (I_T \otimes \frac{F}{T} H_t' H_t') + (\sigma_t^2 I_T \otimes U_t^{-1}) + (F \Sigma_\lambda F' \otimes \sigma_t^2 U_t^{-2})
\]

where \( \Upsilon_t \) is defined in Appendix 8.3 and \( K_{ab} \) defines the commutation matrix of order \((a,b)\) (see [Magnus & Neudecker (2007)](Chapter 3, Section 7))

(iii) \[ \| \hat{D}_{s,t} - D_{s,t} \| \to_p 0, \| \hat{E}_t - E_t \| \to_p 0 \text{ for every } s \in T_t \text{ and } t \in \mathbb{Z}, \]

setting

\[
\hat{D}_{s,t} = \frac{1}{r_{fs}} \left( -((\frac{1}{T}) \otimes I_{r}) \Omega^{-1}(F'(\tau_{s+1} - \frac{1}{T})) - ((\tau_{s+1} - \frac{1}{T}) \otimes I_{r}) \Omega^{-1}(F'\frac{1}{T}) \right) \\
+ (M_L \tilde{F} \Omega^{-1} F' \otimes \Omega^{-1} F') \left( \frac{1}{T} \otimes (\tau_{s+1} - \frac{1}{T}) \right) + ((\tau_{s+1} - \frac{1}{T}) \otimes \frac{1}{T}) \right),
\]

\[
\hat{E}_t \equiv (I_T \otimes \hat{U}_t^{-1} \hat{F}' \otimes \Omega) \Upsilon_t (I_T \otimes \frac{\hat{F}}{T} \hat{U}_t^{-1}) + (\hat{\sigma}_t^2 I_T \otimes \hat{U}_t^{-1}) + (\hat{F} \Sigma_\lambda \hat{F}' \otimes \hat{\sigma}_t^2 \hat{U}_t^{-2})
\]

where \( \hat{\sigma}_t^2, \hat{\sigma}_t, \hat{U}_t^* \) are defined in Theorem 4 and \( \Upsilon_t \) is defined in Appendix 8.

It is interesting to compare the limiting behaviour of the SDF estimator \( \hat{m}_{s,s+1} \) as \( N \to \infty \), established above, with its behaviour under double asymptotics, that is when both \( N,T \to \infty \). It turns out that in the double asymptotics setting, \( \hat{m}_{s,s+1} \) converges to the (ex-ante) SDF \( m_{s,s+1} \) in Eq. (17) although at the slower
rate \( \delta_{NT} \equiv \min[\sqrt{T}, \sqrt{N}] \). Although different techniques, from the one developed in this paper, have to be used, as the latent factors \( F \), and their PCA estimates, become high-dimensional as \( T \) diverges, it can be shown that \( \delta_{NT}(\tilde{m}_{s,s+1} - m_{s,s+1}) \) is asymptotically normal and that asymptotically valid standard errors can be derived (see Dello Preite & Zaffaroni (2019) for details).

The time-varying risk premia result leads immediately to the limiting behaviour for (conditional) portfolio expected returns, yielding

\[
\tilde{\mu}_{t-1}^a \equiv w_{Nt-1}^\prime \tilde{A}_{t-1} \tilde{\gamma} \to_p \mu_{t-1}^a \equiv \mu_{Nt-1}^\prime H_{t-1}^{-1} \tilde{\gamma}^P = \mu_{Nt-1}^\prime \tilde{\gamma}^P,
\]

whenever \( \tilde{A}_{t-1} w_{Nt-1}^a \to_p H_{t-1}^{-1} \mu_{Nt-1}^a \). This condition easily follows under mild conditions on the weights \( w_{Nt-1}^a \).

4.3 Inference with Mimicking Portfolios

Here we examine the limiting behaviour of the mimicking portfolios excess returns, both in population and using its PCA-based estimator. It turns out that the population mimicking portfolio excess returns \( F_{mp}^t = w_{Nt-1}^\prime x_t \), reproduce accurately the true (latent) risk factors \( F_t \) as \( N \) diverges. This result does not formally depend on whether \( F_t \) is observed or not, demonstrating that the notion of mimicking portfolios is useful also in an APT framework, namely in the presence of pricing errors. However, precisely due to the presence of pricing errors, the mimicking portfolios require \( N \) large in order to be effective. It follows that, in an APT framework, one cannot estimate the mimicking portfolios by means of the sample moments, as \( T^{-1} \sum_{s=1}^T x_s x_s' \) is singular for \( N < T \) (precisely at most of rank \( N - T \)). However, a consistent estimator for \( F_{mp}^t \), valid when \( N \to \infty \) for every \( t \in \mathbb{Z} \), can be derived, using our PCA procedure, yielding the mimicking portfolio estimator:

\[
\tilde{F}_{mp}^t = \left( \tilde{\gamma}^P + \tilde{\Omega} \right) \tilde{A}_{t-1} \left( (\tilde{A}_{t-1} \tilde{\gamma})(\tilde{A}_{t-1} \tilde{\gamma})' + \tilde{A}_{t-1} \tilde{\Omega} \tilde{A}_{t-1} + \tilde{\sigma}_{t-1}^2 I_N \right)^{-1} x_t = \tilde{w}_{Nt-1}^\prime x_t \text{ for every } t \in \mathbb{Z}. \quad (30)
\]

Moreover, by means of some algebraic manipulations, one can show that the mimicking portfolio estimator \( \tilde{F}_{mp}^t \) satisfies:

\[
\tilde{F}_{mp}^t = \left( \tilde{\sigma}_{t-1}^2 I_{r_{t-1}} + \tilde{A}_{t-1}^\prime \tilde{A}_{t-1} \right)^{-1} \tilde{A}_{t-1}^\prime x_t \text{ for every } t \in \mathbb{Z}.
\]

It follows that, although not identical to the PCA-estimator, the mimicking portfolio estimator has the same limiting properties of \( \tilde{F}_{t,t} \) as \( N \) diverges, and thus it will converge to a suitable rotation of the true (latent) risk factors, as enunciated below.
Theorem 6 (mimicking portfolios) As \( N \to \infty \),

(i) Under Assumptions 1, 3 and \( g_N(\Sigma_{et-1}) \geq C > 0 \) a.s.,

\[
\| \hat{F}_t^{mp} - F_t \| = o_p(1) \quad \text{for every } t \in \mathbb{Z},
\]

(ii) Under Assumptions 1-10:

\[
\| \hat{F}_t^{mp} - \hat{F}_{t,t} \| = O_p(N^{-1}) \quad \text{for every } t \in \mathbb{Z}. \tag{31}
\]

By part (i), the "unfeasible" (population) mimicking portfolio returns converge to the true (latent) factors. A rate of convergence can be derived by strengthening Assumption 2, in particular is by defining the rate of convergence of

\[
N^{-1/2} \Lambda_{t-1}^{-1} \Sigma_{et-1}^{-1} (\hat{F}_t - F_t).\]

The closeness between the mimicking portfolio estimator \( \hat{F}_t^{mp} \) and the PCA estimator \( \hat{F}_{t,t} \), stated in part (ii), implies that they have the same rate of convergence and asymptotic distribution, centered around \( \hat{H}_{t-1} F_t \), as spelled out in Theorem 1. In fact

\[
\sqrt{N} (\hat{F}_t^{mp} - \hat{H}_{t-1} F_t) = O_p(N^{-1}) + \sqrt{N} (\hat{F}_{t,t} - \hat{H}_{t-1} F_t),
\]

4.4 Inference in the Absence of the Risk Free Asset

We now describe how our results need to be modified when a risk free asset is not tradable. All our results still applies, except that one now needs to estimate an additional quantity, namely the zero-beta rate, which in our large-\( N \) environment requires some further details. Throughout this section \( x_{it} \) denotes the gross return for asset \( i \), still assumed to satisfy (1). However, the conditional asset pricing restriction (2) is now replaced by

\[
\mu_{it-1} \equiv E_{t-1}(\hat{F}_t) = \alpha_{it-1} + \gamma_{0t-1} + \lambda_{it-1}' \gamma_{t-1} a.s. \quad \text{for every } i = 1, \ldots, N \text{ and } t \in \mathbb{Z},
\]

where \( \gamma_{0t-1} \) indicates the (unknown) zero-beta rate, implying

\[
x_{it} = \alpha_{it-1} + \gamma_{0t-1} + \lambda_{it-1}' \gamma_{t-1} + e_{it} \quad \text{for every } i = 1, \ldots, N \text{ and } t \in \mathbb{Z},
\]

where as before \( F_t = f_t + \gamma_{t-1} - E_t(\hat{f}_t) \). In vector form one obtains:

\[
x_t = a_{t-1} + \gamma_{0t-1} + \lambda_{t-1}' \gamma_{t-1} + F_t + e_t \quad \text{for every } t \in \mathbb{Z}.
\]
However, unlike the previous case when \( \gamma_{0t-1} = r_{it-1} + 1 \), one cannot apply PCA to the row data \( \mathbf{X} \), even in the static case \( \gamma_{0t-1} = \gamma_0 \), because \( \gamma_0 \) is the risk premium associated with the constant (unit) factor and its presence is not asymptotically negligible.

Giglio & Xiu (2017), who focus on the static case, circumvent this difficulty by applying PCA to the time-series de-meaned data \( x_{it} - \bar{x}_i \), and then estimate \( \gamma_0 \) in a second stage. This avenue is ruled out in our framework, because it induces serial correlation in the \( e_{it} \), but one can still eliminate the influence of \( \gamma_0 \) by de-meaning with respect to the cross-sectional sample averages:

\[
x_{it}^* \equiv \mathbf{M}_{1N} \mathbf{x}_i = \mathbf{M}_{1N} \mathbf{a}_{t-1} + \mathbf{M}_{1N} \Lambda_{t-1} \mathbf{F}_t + \mathbf{M}_{1N} \mathbf{e}_t.
\]

as \( \gamma_{0t-1} \mathbf{M}_{1N} \mathbf{1}_N = \mathbf{0} \). In other words, one applies the PCA to the \( x_{it}^* = x_{it} - \bar{x}_i = x_{it} - (N^{-1} \sum_{j=1}^{N} x_{jt}) \), where the same definition applies for the other starred quantities defined above. A crucial feature of such cross-sectional de-meaning is that it does not alter the degree of serial (temporal) dependence.

Thus, application of PCA to the \( \mathbf{X}^* \) yields

\[
\hat{\mathbf{F}}^*(T_t), \hat{\Lambda}^*(T_t) = \text{argmin}_{\mathbf{F}, \Lambda_{t-1}} \frac{1}{NT} \text{trace} \left( (\mathbf{X}^* - \mathbf{F} \Lambda_{t-1}^*)^T (\mathbf{X}^* - \mathbf{F} \Lambda_{t-1}^*) \right)
\]

where, as before, \( \hat{\mathbf{F}}^* = \hat{\mathbf{F}}^*(T_t) \) is given by the set of \( r_{t-1} \) eigenvectors, multiplied by \( \sqrt{T} \), corresponding to the largest \( r_{t-1} \) eigenvalues of \( \mathbf{X}^* \mathbf{X}^{**}/NT \), and correspondingly \( \hat{\Lambda}^*(T_t) = \hat{\Lambda}_{t-1}^* = \mathbf{X}^{*'} \hat{\mathbf{F}}^* (\hat{\mathbf{F}}^* \hat{\mathbf{F}}^*)^{-1} = \mathbf{X}^{*'} \hat{\mathbf{F}}^*/T \).

Notice that the true risk factors \( \mathbf{F} \) are not affected by pre-multiplication with the matrix \( \mathbf{M}_{1N} \) but we denote the PCA estimator as \( \hat{\mathbf{F}}^* \) to emphasize that these are obtained from the cross-sectionally de-meaned data.

It turns out that all our results apply once \( \lambda_{it-1} \) and \( \Sigma_{\Lambda_{t-1}} \) are replaced by \( \lambda_{it-1}^* \) and \( \Sigma_{\Lambda_{t-1}}^* \) throughout our assumptions. Notice that, by construction, \( \sum_{i=1}^{N} \lambda_{it-1}^* = \mathbf{0}_{r_{t-1}} \) and that \( \Sigma_{\Lambda_{t-1}}^* \) is the covariance matrix of the \( \lambda_{it} \) (and not the second moment matrix as before), given that:

\[
\frac{\Lambda_{t-1}^* \Lambda_{t-1}^{'}}{N} = \frac{\Lambda_{t-1} \mathbf{M}_{1N} \Lambda_t}{N} = \frac{\Lambda_{t-1}^* \Lambda_{t-1}^{'}}{N} - \mu_{\Lambda_{t-1}} \hat{\mu}_{\Lambda_{t-1}}^* \rightarrow_p \Sigma_{\Lambda_{t-1}}^* > 0,
\]

setting \( \hat{\mu}_{\Lambda_{t-1}}^* \equiv N^{-1} \Lambda_{t-1}^* \mathbf{1}_N \).

As the true risk factors are not affected by the (cross-sectional) de-meaning, their estimator and thus the estimators for the risk premia and the SDF, will not be affected either, except of course for the rotation that will necessarily differ from the case when the risk free asset is tradable. However, estimation of portfolios risk premia will be affected by the de-meaning as one cannot recover the (weighted) first-moment of the loadings.
from the $\tilde{\Lambda}_{t-1}$, as these are now centered around zero. Moreover, one still needs to estimate the zero-beta rates $\gamma_{0t-1}$. For the static case, Giglio & Xiu (2017) derive the asymptotics for the estimator

$$\hat{\gamma}_{0}^{GX} \equiv (1_N^\prime M_A^\prime 1_N)^{-1} 1_N^\prime M_A^\prime \bar{x}_N.$$  

However, in our setting, $\hat{\gamma}_{0}^{GX} \to p \gamma_0 + \mu_0^\prime \gamma^P$ as $N \to \infty$, because $1_N^\prime \tilde{A}^* = 0_t^\prime$, implying that $\hat{\gamma}_{0}^{GX}$ is not consistent for $\gamma_0^{24}$. Further complications arise when the zero-beta rate is time varying, as in our framework.

Instead, one can construct a consistent estimator for both $\gamma_{0t-1}$ and (a rotation of) $\mu_{A_{t-1}}$ jointly, as follows. By taking the cross-sectional average of both sides in (34), and stacking across $t$, one obtains

$$\bar{x}_T = \bar{a}_T + \gamma_0 + F \mu_{A_{t-1}} + \bar{e}_T = \bar{a}_T + M_{1_T} \gamma_0 + D \left( \frac{\tilde{\gamma}_0}{\mu_{A_{t-1}}} \right) + \bar{e}_T,$$

setting $\bar{x}_T \equiv X_{1N}/N$, $\bar{e}_T \equiv e_{1N}/N$, $\bar{a}_T \equiv (a_0^\prime 1_N/N, \cdots, a_{T-1}^\prime 1_N/N)^\prime$, $\gamma_0 \equiv (\gamma_{00}, \cdots, \gamma_{0T-1})^\prime$, $\tilde{\gamma}_0 \equiv \gamma_0^T/T$ and $D \equiv (1_T, F)$. Representation (35) naturally suggests the two-pass OLS estimator for $(\tilde{\gamma}_0, \tilde{\mu}_{A_{t-1}})^\prime$:

$$\left( \begin{array}{c} \tilde{\gamma}_0 \\ \tilde{\mu}_A \end{array} \right) \equiv (D^\prime D)^{-1} D^\prime \bar{x}_T,$$

based on the (first-pass) estimator of the risk factors $\tilde{F}^\ast$, where we set $D \equiv (1_T, \tilde{F}^\ast)$. Noticeably, the two-pass estimator $(\tilde{\gamma}_0, \tilde{\mu}_A)^\prime$ is exactly the mirror image of the traditional two-pass estimator. Whereas the latter is based on a single cross-sectional regression of $N$ time-series averaged returns on the estimated loadings, evaluated as the $T$ diverges (see Black et al. (1972), Fama & MacBeth (1973) and Shanken (1992), among others), our two-pass estimator is obtained from a single time-series regression of $T$ cross-sectional averaged returns on the estimated factors, evaluated as $N$ diverges.

The estimators $(\tilde{\gamma}_0, \tilde{\mu}_A)^\prime$ are clearly meaningful if both $\bar{a}_T$ and $M_{1_T} \gamma_0$ are (asymptotically) negligible as $N$ diverges. Indeed, these conditions hold as a consequence of no-arbitrage (for the vector of pricing errors $\bar{a}_T$) and by suitable smoothing assumptions (for the $\gamma_0$), as specified below. The next theorem provides an inferential procedure for $(\tilde{\gamma}_0, \tilde{\mu}_A)^\prime$.

**Theorem 7 (rate of convergence, limiting distribution, consistent estimation of the asymptotic covariance matrix of $\tilde{\gamma}_0$ and $\tilde{\mu}_A$)**

*Under Assumptions 1-10, as $N \to \infty$,*

\footnote{Giglio & Xiu (2017) do not encounter this problem because, as they consider a double asymptotic setting, they apply PCA to the time-series de-meaned data as already discussed above, for which their estimator $\hat{\gamma}_{0}^{GX}$ is valid.}
(i) 
\[
\left( \begin{array}{c}
\hat{\gamma}_0 \\
\hat{\mu}_\Lambda 
\end{array} \right) - \left( \begin{array}{c}
\tilde{\gamma}_0 \\
\tilde{\mu}_{\Lambda_{t-1}} 
\end{array} \right) = \left( \begin{array}{c}
\tilde{\gamma}' \\
\tilde{\mu}' 
\end{array} \right) \Omega^{-1} \overline{\text{Cov}}(\tilde{\mu}', \gamma_0) + O_p(N^{-\frac{1}{2}}),
\]

setting 
\[
\overline{\text{Cov}}(\tilde{\mu}', \gamma_0) = \frac{1}{T} \tilde{\mu}'M_1 \gamma_0.
\]

and \( \tilde{F} \equiv T^{-1}\tilde{F}'T, \tilde{\Omega} \equiv T^{-1}\tilde{\mu}'M_1 \tilde{F} \).

When, in addition, \( \mu_{\Lambda_{t-1}} \rightarrow p \mu_{\Lambda_{t-1}}, \tilde{a}_T = o(N^{-\frac{1}{2}}) \) and \( \gamma_0n = \gamma_0t-1 \) for every \( s \in T_t \) and \( t \in \mathbb{Z} \):

(ii) 
\[
\left( \begin{array}{c}
\hat{\gamma}_0 \\
\hat{\mu}_\Lambda 
\end{array} \right) - \left( \begin{array}{c}
\gamma_{0t-1} \\
\tilde{\mu}_{\Lambda_{t-1}} 
\end{array} \right) = O_p(N^{-\frac{1}{2}}).
\]

(iii) 
\[
\sqrt{N} \left( \begin{array}{c}
\hat{\gamma}_0 \\
\hat{\mu}_\Lambda 
\end{array} \right) - \left( \begin{array}{c}
\gamma_{0t-1} \\
\tilde{\mu}_{\Lambda_{t-1}} 
\end{array} \right) \rightarrow_d N(0_{r_{t-1}+1}, G_{t-1}'L_{t-1}G_{t-1}),
\]

setting 
\[
G_t = \frac{1}{T} \left( \begin{array}{c}
(1_T, FH_t)(1_{\gamma P})^{-1} \\
-(1_T, FH_t)(1_{\gamma P})^{-1} \otimes \gamma \tilde{\mu}_\Lambda 
\end{array} \right), L_t = \left( \begin{array}{c}
L_{11t} \\
L_{21t} \\
L_{22t} 
\end{array} \right),
\]

with 
\[
L_{11t} = \sigma_1^2 1_T, \quad L_{21t} = \sigma_2^2 (I_T \otimes U_t^{-1} H_t^{-1} \mu_{\Lambda_t}) + \sigma_2^2 (F \mu_{\Lambda_t} \otimes U_t^{-1} H_t F' T),
\]
\[
L_{22t} = (I_T \otimes U_t^{-1} H_t F' T) \mu_{\Lambda_t} (I_T \otimes FH_t T U_t^{-1}) + \sigma_2^2 (I_T \otimes U_t^{-1}) + (F \Sigma_{\Lambda_t} F' \otimes \sigma_2^2 U_t^{-2})
\]
\[
+ \sigma_2^2 (F H_t T U_t^{-1} \otimes H_t F') K_{Tt} + K_{Tt} (\sigma_2^2 U_t^{-1} H_t F' T \otimes FH_t).
\]

(iii) When, in addition, \( \kappa_{4t} = 0 \),

\[ || \tilde{G}_t - G_t || \rightarrow_p 0, \quad || \tilde{L}_t - L_t || \rightarrow_p 0, \]

setting 
\[
\tilde{G}_t = \frac{1}{T} \left( \begin{array}{c}
(1_T, \tilde{\mu}) (1_{\tilde{\gamma}'})^{-1} \\
-(1_T, \tilde{\mu}) (1_{\tilde{\gamma}'})^{-1} \otimes \tilde{\mu}_\Lambda 
\end{array} \right), \tilde{L}_t = \left( \begin{array}{c}
\tilde{L}_{11t} \\
\tilde{L}_{21t} \\
\tilde{L}_{22t} 
\end{array} \right),
\]
4.4 Inference in the Absence of the Risk Free Asset

with

\[
\begin{align*}
\tilde{L}_{11t} & \equiv \tilde{\sigma}_t^2 I_T, \quad \tilde{L}_{21t} \equiv \tilde{\sigma}_t^2 (I_T \otimes \tilde{U}_t^{*-1} \tilde{\mu}_\Lambda) + \tilde{\sigma}_t^2 (\tilde{F}^s \tilde{\mu}_\Lambda \otimes \tilde{U}_t^{*-1} \tilde{F}^s/\hat{T}), \\
\tilde{L}_{22t} & \equiv (I_T \otimes \tilde{U}_t^{*-1} \tilde{F}^s/\hat{T}) \tilde{\mu}_\Lambda (I_T \otimes \tilde{F}^s/\hat{T}) \tilde{U}_t^{*-1} + (\tilde{F}^s \tilde{\Sigma}_\Lambda \tilde{F}^s/\hat{T}) \tilde{U}_t^{*-2} + (\tilde{\sigma}_t^2 \tilde{U}_t^{*-1} \tilde{F}^s/\hat{T}) K_{rt} + K_{Tr}(\tilde{\sigma}_t^2 \tilde{U}_t^{*-1} \tilde{F}^s/\hat{T} \otimes \tilde{F}^s),
\end{align*}
\]

where \(\tilde{\sigma}_t^2, \tilde{\sigma}_t^2, \tilde{U}_t^*, \tilde{\Sigma}_\Lambda^*\) are defined in Theorem 2 replacing \(\tilde{F}, \tilde{\Lambda}_t\) with \(\tilde{F}^*, \tilde{\Lambda}_t^*\) and \(K_{rt}\) is defined in Theorem 2 and \(\tilde{U}_t\) is defined in Appendix E.

(iv) When, in addition, \(\tilde{\mu}_{\Lambda_{t-1}} - \mu_{\Lambda_{t-1}} = o_p(N^{-\frac{1}{2}})\), then (i)-(ii)-(iii) also apply replacing \(\tilde{\mu}_{\Lambda_{t-1}}\) with \(\mu_{\Lambda_{t-1}}\).

Although our focus is on the large-\(N\) finite-\(T\) case, by analyzing the asymptotic covariance matrix it follows that our estimators for \(\gamma_{0t-1}\) and \(H_{t-1}^{-1} \mu_{\Lambda_{t-1}}\) are, in fact, converging at the fast rate \(O_p((NT)^{-\frac{1}{2}})\) if one allows both \(N\) and \(T\) to diverge. This rate resembles the rate that Gagliardini et al. (2016) obtain for their estimator of the (nonlinear component of the) risk premia associated with the risk factors. The analysis of the zero-beta rate is excluded from their analysis but we conjecture that the same rate as ours, when the factors are observed, applies.

Our result shows that, when the risk-free rate is not sufficiently smooth, our risk premia estimator would quantify the (local) average of the time-varying zero beta rates, namely \(\bar{\gamma}_0\), although a bias emerges, that depends on the sample covariance between the risk factors and the time-varying zero-beta rates. A similar bias affects our estimator of the loadings’ first moment \(\mu_{\Lambda_{t-1}}\). However, as \(T\) can be arbitrarily small, the smoothness assumption on the \(\gamma_{0t-1}\) is extremely mild in practice.

In the absence of the risk free asset, the population ex-post SDF becomes:

\[
m_{t,t+1}^P = \frac{1}{\gamma_0} - \frac{1}{\gamma_0} \gamma^P \Omega^{-1}(F_{t+1} - \gamma^P) \text{ for every } t \in \mathcal{Z}, \tag{41}
\]

which satisfies the pricing conditions \(\mathbb{E}_t(m_{t,t+1}^P x_{it+1}) = 1 + O(T^{-1})\), again with a (pricing) error of order \(O(T^{-1})\) by an easy extension of Proposition 2. Our PCA-based estimator for \(\tilde{m}_{t,t+1}^P\) is then:

\[
\tilde{m}_{s,s+1} = \frac{1}{\gamma_0} - \frac{1}{\gamma_0} \tilde{\gamma}^* \tilde{\Omega}^{-1}(\tilde{F}^*_{s+1,t} - \tilde{\gamma}^*) \text{ for every } s \in \mathcal{T}_t \text{ and } t \in \mathcal{Z},
\]

setting \(\tilde{\gamma}^* = \tilde{F}^* 1_T/T\), and the analogue of Theorem 3 applies.\(^{25}\)

\(^{25}\)Details are available upon request.
Having obtained a consistent estimator for $H_{t-1}^{-1}\mu_{\Lambda_{t-1}}$, we can re-construct the estimator for the loadings as $\hat{\Lambda}_{t-1}^\dagger \equiv \hat{\Lambda}_{t-1}^\ast + 1_N\hat{\mu}_{\Lambda_{t-1}}$. In turn, this allows to estimate consistently the expected portfolio returns corresponding to any generic portfolio weights $w^a_N$, namely

$$\hat{\mu}_{a}^t \equiv \underbrace{w^a_N}_{Nt-1} \hat{\mu}^t_{\Lambda_{t-1}} = \underbrace{w^a_N}_{Nt-1} \hat{\gamma}^P = \underbrace{w^a_N}_{Nt-1} \hat{\gamma}^P,$$

under mild assumptions on the portfolio weights $w^a_N$ such as $\hat{\Lambda}_{t-1}^\dagger w^a_N \rightarrow_p H_{t-1}^{-1} \mu_{\Lambda_{t-1}}^a$.

5 Monte Carlo Experiments

We illustrate the finite-sample performance of the PCA estimators for the number of factors and for the factors themselves when $N$ is allowed to diverge but $T$ is fixed and, moreover, assumed to be very small. We compare the results from using our methodology with the ones that are valid when $N$ and $T$ both diverge. Given the scope of the Monte Carlo exercise, we focus on the static model. We consider cases when $T$ is very small, such as $T = 2, 5$ and also the cases when $T = 50, 100$. We combine these values of $T$ with $N$ ranging from 10 up to 3,000. Data are generating according to model (1), imposing zero pricing errors,

$$x_{it} = \lambda_i^t F_t + e_{it}, i = 1, \ldots, N, \ t = 1, \ldots, T,$$

where the common factors $F_t$ are assumed iid (independent and identically distributed) across $t$ and $N(0, I_r)$, corresponding to a given number of factors $r$, the loadings $\lambda_i$ are assumed iid across $i$ and $N(0, I_r)$ and the the idiosyncratic components $e_{it}$ are assumed iid across both $i$ and $t$ standard normal, unless we say otherwise. Moreover $F_t, \lambda_i$ and $e_{js}$ are assumed mutually independent for every $i, j, t, s$.

All the results reported below are based on 1,000 Monte Carlo iterations where the common factors $F_t$ are not re-sampled at every iteration but only sampled once at the outset of every Monte Carlo exercise.

5.1 Estimation of the Number of Factors

Our estimator is based on the criterion $PC(k) = V^*(k) + kg(N)$ where $V^*(k) = \frac{T-k}{T} V(k)$ and the penalization function is constructed:

$$g(N) = \frac{(\log(N))^{\epsilon_1} N^{\epsilon_2}}{\sqrt{N}} \text{ for } \epsilon_1 \geq 0 \text{ and } 0 < \epsilon_2 < 1/2,$$

which only depends on $N$. We compare our results with some of the Bai & Ng (2002) [Section 5] criteria, asymptotically valid when $N$ and $T$ both diverge, namely the $PC_p$ and $IC_p$ criteria. We also consider, as a

\[\text{Results for the case of cross-sectional dependent } e_{js} \text{ are available upon request.}\]

\[\text{Variations of our criterion were tried. A similar, good, performance to ours has been obtained also for } V(k) + kg(N).\]
5.2 Estimation of the Asymptotic Distribution of Factors

In this section we always assume \( r = 1 \) for simplicity. We follow Bai (2003) [Section 6] in terms of the various analysis developed to demonstrate the accuracy of the asymptotic distribution of the PCA estimated factors, illustrated in our Theorems 1 and 2 respectively. Table IV reports the correlation, averaged across Monte Carlo iterations, between true and estimated factors for \( T = 2, 5, 50 \) and \( N = 10, 100, 1000, 3000 \). We also report the correlation corresponding to the loadings, as a comparison, in Table V. The results show that the factors are accurately estimated, even when \( T \) is very small, once \( N \) is sufficiently large. In contrast, the loadings require a sufficiently large \( T \) in order to obtain a similar accuracy, appearing poorly estimated when \( T = 2, 5 \) and, instead, well estimated when \( T = 50 \). Moreover, the quality of the estimates for the loadings does not change very much as one varies \( N \) from 10 to 3000. We also report in Figure I and II the histogram of the estimated correlations for the factors and the loadings, respectively, when \( T = 10, N = 100 \) and when \( T = 100, N = 100 \). Whereas the correlations are all concentrated around unity, for factors, they are instead concentrated around

---

Footnote: Further results are available upon request.
0.95 for the loadings, markedly away from unity, when $T$ is much smaller than $N$. Instead, the histograms have the same shape when $T$ and $N$ are both equal to 100, as indicated in Figure II.

**TABLE IV and V HERE**

**FIGURE I HERE**

**FIGURE II HERE**

We then consider the the histograms of factors’ estimates, across the Monte Carlo iterations. In particular we report the histograms associated with the studentized PCA estimates, satisfying by Theorem 1, for every $1 \leq s \leq T$,

$$f_s^{\text{small}-T} \equiv (A'\tilde{B}_sA)^{-\frac{1}{2}}N^\frac{1}{2}(\tilde{F}_s - \tilde{H}'F_s) \rightarrow_d N(0_r, I_r) \text{ as } N \rightarrow \infty,$$

where $A$ and $\tilde{B}_s$ are defined in Theorem 2, setting $B_s = B_{s,t}$ in view of the static formulation adopted here.

We only display case $t = 1$ for simplicity in all the figures. For comparison, we also display the histograms corresponding to the studentized PCA estimates but using the (estimated) asymptotic covariance matrix that is valid when both $N$ and $T$ diverge, namely $\tilde{\sigma}^2\tilde{U}^{s-1}$. The latter formula can be easily obtained by evaluating the limit of $\tilde{B}_s$ as $T \rightarrow \infty$, yielding

$$f_s^{\text{large}-T} \equiv (\tilde{\sigma}^2\tilde{U}^{s-1})^{-\frac{1}{2}}N^\frac{1}{2}(\tilde{F}_s - \tilde{H}'F_s),$$

where we expect tails larger than the standard normal as the incorrect standard errors, used for $f_s^{\text{large}-T}$, are always smaller than the asymptotically correct ones, used for $f_s^{\text{small}-T}$. In the last two graphs we also report the histogram corresponding to the studentized PCA estimates based on the (estimated) asymptotic covariance matrix that is valid when both $N$ and $T$ diverge and robust to heteroskedasticity (see Bai (2003) [Eq.(7), Section 5]).

The results are illustrated in Figures III to X. The finite-sample distribution of the $f_s^{\text{small}-T}$ is remarkably close to the standard normal density when $N$ is large, regardless of $T$, for instance even when $T = 2$. More in details, we always report the histogram for the $f_s^{\text{small}-T}$ in green and for the $f_s^{\text{large}-T}$ in blue. Figure III considers the case $T = 2, N = 10$: although $N$ is very moderate, the $f_s^{\text{small}-T}$’s empirical distribution does not appear too far from the shape of the standard normal although it is wrongly centered around a negative value. Instead, using the distribution of the $f_s^{\text{large}-T}$ appears much more spread out than the standard normal, the result of adopting excessively small standard errors. Figure IV still considers the very small $T = 2$ case but now with $N = 1,000$: now the factors appear extremely well estimated, the empirical distribution of the $f_s^{\text{small}-T}$ is very close to the standard normal whereas, again, the $f_s^{\text{large}-T}$ exhibit much fatter tails.
As expected, as one increases \( T \), the discrepancy between the distribution of the \( f_{s}^{\text{small} - T} \) and \( f_{s}^{\text{large} - T} \) diminishes, and it is evident from Figure V and VI (\( T = 5 \)) and Figures VII and VIII (\( T = 50 \)). In particular, when \( T = 50 \) there is almost no difference between the two histograms, in turn very close to the standard normal density.

Finally, in Figures IX and X we report, together with the histograms of \( f_{s}^{\text{small} - T} \) and \( f_{s}^{\text{large} - T} \), also the histograms of the (studentized) factors’ estimated using the robust standard errors (in red) for the cases \( T = 50, N = 1,000 \) and \( T = 100, N = 1,000 \), respectively. In the latter case, the three histograms are essentially identical, and close to the standard normal density, whereas even for \( T = 50 \) the robust standard errors are still underestimating the true asymptotic covariance matrix.

Next, we evaluate the 95% (asymptotic) confidence intervals for the true factors \( F_{t} \). Again we consider the average intervals across the 1000 Monte Carlo iterations. We follow Bai (2003) and construct the confidence intervals as:

\[
\left( \hat{\beta} F_{s} - 1.96 \frac{\hat{\beta}}{N^{\frac{1}{2}} (A^{\prime} B_{s} A)^{\frac{1}{2}}}, \hat{\beta} F_{s} + 1.96 \frac{\hat{\beta}}{N^{\frac{1}{2}} (A^{\prime} B_{s} A)^{\frac{1}{2}}} \right), \ s = 1, \cdots, T,
\]

where \( \hat{\beta} \equiv (F^{\prime} F)^{-1} F^{\prime} F \), that is the OLS estimator from projecting the estimated factors \( \hat{F} \) on the true factors \( F \) without intercept. This provides an accurate estimation of the (inverse of the) rotation matrix \( \hat{H} \) of Theorem 1. Figures XI and XII report the times series of these confidence intervals for cases \( T = 5, N = 10 \) and \( T = 5, N = 1,000 \) respectively.

\[\text{FIGURE XI HERE}\]

\[\text{FIGURE X HERE}\]

\[\text{We do not report the values corresponding to the robust standard errors for the other combinations of } N \text{ and } T \text{ because the corresponding histograms is too far off from the ones of } f_{s}^{\text{small} - T} \text{ and } f_{s}^{\text{large} - T}.\]
FIGURE XII HERE

The confidence intervals, corresponding to \( T = 5, N = 10 \) are sufficiently accurate to include the true factor’s realization, where the interval becomes much narrower as \( N \) increases, essentially nailing down the time series of the true factor. Note that we set \( T = 5 \) in both figures.

6 Empirical Application

In this section we assess the performance of our method using an (unbalanced) panel of monthly individual assets traded in the NYSE from January 1960 to December 2013 (source CRSP). Our method appears ideal to address important empirical questions related to the possible time-variation characterizing the factor asset pricing model. Our empirical results are all based on adopting a short rolling window of \( T = 12 \) time observations. We have repeated all the exercises for \( T = 24 \) and \( T = 36 \) without noticeably differences of the empirical results.

6.1 Time-Variation of Number of Factors

We investigate the extent to which the number of risk factors is varying across time over long horizons, for example whether they have been increasing along the increased sophistication of financial markets. Moreover, we investigate whether the number of factors increases rapidly over economic expansions and financial market booms and decrease otherwise, thus charactering the time-variation over short horizons.

Figure XIII reports the estimated number of factors from January 1960 to January 2013 using rolling windows of \( T = 12 \) months, leading to a number of individual assets varying from 800 to 5,000 across the rolling samples.

FIGURE XIII HERE

The results clearly shows that the number of risk factors was small, between one and two, until the beginning of the 1980s, possibly a symptom of under-developed financial markets and of financial segmentation. From then on, the number has been steadily increasing until the dot-com bubble of the 2000, reaching 10 risk factors, then decreasing sharply to two risk factors as the financial crises unfolds. More in general, from January 2000 onward, the estimated number of factors tracks the dynamics of the financial market remarkably well, increasing during periods of booms and decreasing during crashes: the correlation between the two series is above 80% between March 2000 and June 2009, with the number of factors ranging from two to 10. Our method
6.2 Identifying Risk Factors

corroborates the observation according to which the correlation across financial markets increases dramatically during crashes, as one obtains when the number of risk factors reduces to one or two. The correlation across the whole period (January 1996 until December 2013) is 58%.

Figure XIV shows the same times series of the estimated number of factors, with the NBER business cycle indicator and various other macroeconomic and financial crises indicators: it is, again, remarkable how the estimated number of factors co-vary negatively with almost every negative macroeconomic event.

FIGURE XIV HERE

Our finding might explain the disagreement emerging in the empirical asset pricing literature on the correct number of risk factors for the universe of stock US equity returns, ranging from the small number of Fama and French factors (three or 5), to the recent advances that suggest up to 10 risk factors. Once time-variation is allowed for, it is clear that the notion of assuming a fixed number of risk factors appears uncorrect. Moreover, our maximum estimated number of factors appears aligned with previous findings.

6.2 Identifying Risk Factors

We describe the dynamic behaviour of the time series of the five dominant estimated factors and, exploiting the time variation of our approach, show how one can identify the extent to which the estimated latent risk factors are related to some of the observed factors proposed in the empirical asset pricing literature. We focus here on the five observed factors associated with the Fama & French (2015) asset pricing model.

Figure XV reports the first five estimated factors using rolling windows of size $T = 12$, where the number of stocks varies from 800 to 5,000 across the rolling samples.

Table VI reports the correlation, across the full sample, between the five estimated factors and the observed factors of the Fama & French (2015) asset pricing model. It emerges that the first factor has the highest correlation with the market excess return (mkt), the second factor with the profitability return (rmw) and with the high-minus-low return (hml), and the third factor with the small-minus-large return (smb) and, in part, with the investment return (cma). The fourth and fifth factors appear mildly correlated with all the five observed returns. It is interesting to compare these correlations with the ones obtained during the sub-prime financial crises crash (identified as Aug 2007 to Feb 2009), reported in Table VII: the first factors is even more prominently related to mkt, the second factor to hml, the third factor to both smb and hml, the fourth factor especially to cma and, to a lesser extent, to hml and rmw, and, finally, the fifth factor to cma.
Therefore, the links of the first three estimated risk factors with the Fama & French (2015) factors appear robust, although varying with time. This can be better appreciated by looking Figures XVI, XVII and XVIII, where we report the time series of the estimated correlations between the PCA estimates and the Fama & French (2015) five factors, over rolling windows of 60 observations. In particular, Figure XVI shows how the market excess return is clearly related to the first PCA, the smb return to the second PCA and the hml return to the third PCA. Although the magnitude of these estimated correlations vary through time, it is evident that in general the link becomes more tenuous from the late 1990s. Figures XVII and XVIII show the link of the rmw and cma returns with the fourth and fifth PCA: it appears that rmw and cma are related to a linear combination of those PCAs. Again the strength of the relationship appears to fade away over the last part of the sample. Our evidence on the larger number of estimated PCA factors explaining the cross-section of returns from the 1990s onwards appears aligned with this attenuation of the estimated correlations. A more economically meaningful measure of the (dynamic) importance of the Fama & French (2015) five factors is obtained by relating them to the estimated SDF, as illustrated below. There, we confirm that, indeed, the relevance of the Fama and French factors appears strong in the first half of the our sample period, until the early 1990s, and much weaker thereafter, except during the most volatile periods of the financial crises. This agrees with our empirical finding, above, suggesting that only a small number of risk factors is relevant during such periods.

FIGURE XV HERE
TABLES VI and VII HERE
FIGURES XVI, XVII, XVIII HERE

6.3 Detecting the State Variables for Loadings

We examine the time variation of the estimated equally-weighted portfolio return:

\[ \tilde{\mu}_{t-1}^{\text{ew}} \equiv w_{\text{ew}}^{\text{est}} \tilde{A}_{t-1} \tilde{\gamma}, \]

setting \( w_{\text{ew}}^{\text{est}} = N^{-1}1_N \). Under our assumptions, \( \tilde{\mu}_{t-1}^{\text{ew}} \xrightarrow{p} \mu_{t-1}^{\text{ew}} \equiv \mu_{\text{ew}}^{\text{est}} \tilde{\gamma}^P \). As already indicated, \( \mu_{t-1}^{\text{ew}} \) is rotation-free, and thus its time-variation must be attributed to the dynamics of the loadings, and not induced by the rotation matrix. In fact, our identification strategy is based on making the expected portfolio return above function of the average risk premia over the \( T \) data points, namely \( \tilde{\gamma}^P \), as opposed to considering the time-varying risk premia \( \gamma_{t-1}^P \).
Figure XIX reports the time series of $\tilde{\mu}_{t-1}$ together with four state-variables often advocated to drive the dynamics of the loadings to the risk factors, namely the market dividend yield (ratio of aggregate dividends to the S&P 500 index; source Robert Shiller’s website), the default spread (Moody’s Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity, Not Seasonally Adjusted; source FRED), the term spread (10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity, Not Seasonally Adjusted; source FRED) and the CAPE ratio (Cyclically Adjusted Price Earnings Ratio P/E10; source Robert Shiller website). Several papers suggest the explanatory power of these state-variables (see Gagliardini et al. (2016) and Kelly et al. (2018) among others). Data are monthly from Dec 1960 until Dec 2013.

FIGURE XIX HERE

The prominent time-variation of the portfolio estimated expected return is evident. We now assess more accurately the dynamic relationships between these variables, in particular trying to quantify the extent to which the portfolio estimated expected return moved together with the four state variables. Figure XX reports the time series of the adjusted-$R^2$ of the regression, over rolling windows of 60 observations, of $\tilde{\mu}_{t-1}$ over the four above described state-variables, both jointly (black line) and also taken one by one (red line for CAPE, green line for term spread, blue line for default spread, light blue for dividend yield).

The results from the regressions reported in figure XX show that the time-variation of the risk factors' loadings appears strongly driven by the four state-variables when taken jointly, except between the late 1990s and the mid 2000s. The CAPE appears to be more relevant in the first part of the sample. Interestingly, the dividend yield seems to explain time-variation of loadings in the aftermath of financial and economic crisis whereas the spread variables appear more relevant during booms and expansions.

FIGURE XX HERE

6.4 Pricing Performance of Risk Premia

We report in Figures XXI, XXII and XXIII the time series of the estimated (ex-post) risk premia for risk factor one, two and three, respectively (black line). When these are statistically significant, positive or negative, at 5% significance value, we report them in green and red, respectively. Notice that, for risk premia two and three, sometimes the lines are missing, corresponding to the rolling windows for which the estimated number of factor is smaller, respectively, than two and three. The grey bars indicate financial crises and recessions. Notice that the confidence bands have a time-varying width because they are based on a time-varying number of stocks (from
It emerges that the estimated risk premia have some tendency for being anti-cyclical and, more importantly, their variability increases substantially in the aftermath of crises. Moreover, as illustrated by our theory, the risk premium associated with the highest-ranked factor (in terms of the corresponding eigenvalues) is more precisely estimated, and the precision progressively diminishes as one considers the second, the third, and so forth, risk premium.

FIGURES XXI,XXII,XXIII HERE.

To quantify the pricing performance of the estimated risk premia, we consider the empirical pricing errors stemming from our PCA procedure:

\[ \hat{\delta}_{t-1} \equiv \bar{x}_i - \hat{\lambda}_{t-1}^\prime \gamma. \]  

Note that these are almost identical to the estimated intercept from the time series regression of the \( i \)-th excess returns \( x_{is} \) on the PCA estimates \( \hat{F}_{s,t} \), namely \( \tilde{\alpha}_{it-1} \equiv \bar{x}_i - \hat{\beta}_{it-1}^\prime \gamma \), with the estimated regression coefficient given by \( \hat{\beta}_{it-1}^\prime \equiv (\hat{F}_{s,t}^\prime \hat{F}_{s,t})^{-1} \hat{F}_{s,t}^\prime x_i \). The difference between the \( \hat{\delta}_{t-1} \) and the \( \tilde{\alpha}_{it-1} \) is that the former rely on the estimated loadings \( \hat{\lambda}_{t-1}^\prime = (\hat{F}_{s,t}^\prime \hat{F}_{s,t})^{-1} \hat{F}_{s,t}^\prime x_i \) whereas the \( \tilde{\alpha}_{it-1} \) rely on the \( \hat{\beta}_{it-1} \). The popular GRS test of correct specification of Gibbons et al. (1989) is based on the \( \tilde{\alpha}_{it-1} \).

We evaluate the following statistic \( \sum_{i=1}^{N} \frac{\hat{\delta}_{it-1}^2}{T} \), corresponding to any given rolling window \( T_t \) of size \( T \). However, under the null hypothesis of correct model specification, namely Eq. (2), it follows that \( \hat{\delta}_{it-1} \rightarrow_p \tilde{\alpha}_i + \tilde{e}_i - (e_i^\prime F/T)H_{t-1}^\prime \hat{F} \), where \( \tilde{\alpha}_i = (a_{i0}, \cdots, a_{i(T-1)})1_T/T \) and \( \tilde{e}_i = e_i^\prime 1_T/T \), implying

\[ \frac{1}{N} \sum_{i=1}^{N} \hat{\delta}_{it-1}^2 \rightarrow_p \sigma_i^2 \frac{1}{T} (1 - \tilde{\gamma}^P H_{t-1} H_{t-1}^\prime \tilde{\gamma}^P) \]

as \( N \rightarrow \infty \).

The above finding, namely that the population pricing errors are not negligible even under the hypothesis of correct pricing when \( T \) is fixed, has been discovered by Raponi et al. (2018) in the context of testing beta-pricing models with observed risk factors. Here we derive the analogue result the latent risk factors case. In view of this, we plot the time series of the centered average squared (sample) pricing errors:

\[ \hat{\Delta}_{t-1} = \frac{1}{N} \sum_{i=1}^{N} \hat{\delta}_{it-1}^2 - \frac{1}{T} \hat{\delta}_{t-1}^2 (1 - \hat{\gamma}^2) \]

in order to quantify the pricing performance of the PCA-estimated asset pricing model. Figure XXIV plots the time series of the \( \hat{\Delta}_{t-1} \) over rolling windows of size \( T = 12 \) corresponding to various cases: when the number of factors equal \( \hat{k}_{t-1} \) (as reported in Figure XIII) which is presumably the correct model (black line), and also

\[ \text{Their result, namely that a suitably standardized quadratic form of the } \hat{\alpha}_{it-1} \text{ is exactly distributed like a chi-square, holds for a fixed } N \text{ and } T. \]
when the number of factors is arbitrarily kept constant, such as \( k = 1 \) (red line), \( k = 2 \) (green line), \( k = 3 \) (blue line) and \( k = 5 \) (light blue). It is evident how the one-factor model leads to the poorest performance whereas the model associated with \( \tilde{k}_{t-1} \) gives the best performance, yet close to the three- and five-factor model.

**FIGURE XXIV HERE.**

Another, alternative, way to assess whether the conditional factor asset pricing model adds economic value is to construct mean-variance efficient portfolios and to assess its *out-of-sample* performance, for instance in terms of Sharpe ratios. Specifically, assuming that the universe of assets is driven by the factor time-varying asset pricing model (1) with \( r_{t-1} \) risk factors, we construct time series of (tangency) portfolio excess returns as:

\[
r_{t}^{\text{mv},k_{t-1}} \equiv \tilde{w}_{t-1}^{\text{mv}}x_{t},
\]

setting

\[
\tilde{w}_{t-1}^{\text{mv}} \equiv \frac{\Sigma_{t-1}^{-1} \Lambda_{t-1} \tilde{\gamma}}{1_{N} \Sigma_{t-1}^{-1} \Lambda_{t-1} \tilde{\gamma}} \text{ with } \Sigma_{t-1} \equiv \left( \tilde{\Lambda}_{t-1} \tilde{\Omega} \tilde{\Lambda}_{t-1} + \tilde{\sigma}_{t-1}^{2} I_{N} \right).
\]

The results are reported in Table VIII. We have evaluated the portfolio weights over rolling time-windows of size \( T = 36 \) where the (time-varying) number of assumed risk factors is either taken from Section 6.1 above (based on our selection criterion), namely \( \tilde{k}_{t-1} \), or assumed fixed throughout the exercise (\( k = 1 \), \( k = 5 \) and \( k = 10 \)). The portfolio returns \( r_{t}^{\text{mv},k_{t}} \) have been carefully constructed in *real time*, meaning that the corresponding portfolio weights only use past information, in particular data from period \( t - 35 \) until \( t - 1 \). We have also evaluated the \( t \)-ratios to compare each of the Sharpe ratios with the one associated with the equally weighted portfolios. Such \( t \)-ratios are asymptotically standard normal under the null hypothesis of equality (see Lo (2003)).

**TABLE VIII**

In summary, the conditional factor asset pricing model (1), corresponding to the estimated number of risk factors obtained using our criteria (namely \( \tilde{k}_{t-1} \)), appears to perform well in terms of out-of-sample Sharpe ratios. In particular, the corresponding optimal portfolio statistically beats the equally weighted portfolio, which is usually a hard benchmark to achieve for any model-based portfolio strategy (see DeMiguel et al. (2009)). Our approach, based on the estimated time-varying number of risk factors \( \tilde{k}_{t-1} \), also beats the cases based on assuming a constant number of factors: for instance, the results are particularly weak when assuming \( k = 10 \) risk factors, a symptom of the over-fitting arising from selecting too many risk factors.
6.5 Dynamic Spanning of the SDF

We examine the dynamic properties of the estimated SDF $\tilde{m}_{t,t+1}$ (black line), whose times series is reported in Figure XXV, together with its 95% confidence band (red lines) where, as before, the grey bands correspond to recessions and financial crises. The results are markedly indicating the anti-cyclicality of the SDF, confirming its interpretation as an *index of bad times*. We consider here rolling windows of size $T = 24$ (two years).

In fact, when $T = 12$, one obtains both large swings in the estimated SDF and, correspondingly, confidence intervals with a large width, especially towards the second half of the time period, due to the proximity of the chosen short time window ($T = 12$) and the relatively large number of estimated factors, with $\hat{k}_{t-1}$ often taking values up to 10. This causes quasi-singularity of $\tilde{\Omega}$ and thus excessively large estimates of the SDF and their standard errors.\[31\]

**FIGURE XXV HERE.**

More importantly, we examine the extent to which the Fama & French (2015) five factors span, in part the SDF. The results are reported in Figure XXVI, where we report the adjusted-$R^2$ from regressing the SDF, respectively, on the market return, the three Fama & French (1993) factors and the five Fama & French (2015) factors, using rolling windows of 60 months. It emerges that all three models explained a sizeable amount of time-variation of the SDF until the 1990s. Thereafter, the CAPM poorly spans the SDF whereas the three- and five-factor models seem more relevant, although the magnitude of explained $R^2$ diminishes, with positive spikes observed in the aftermath of financial and economics crises, for instance especially after the 2007 financial crises. This fact squares with our finding regarding the small number of risk factors determined during market crashes, as opposed to the large number of risk factors that appear relevant during market booms.

**FIGURE XXVI HERE.**

7 Conclusion

This paper develops a methodology for inference on conditional asset pricing models linear in latent risk factors, valid when the number of assets diverges but the time series dimension is fixed, possibly very small. We show that the no-arbitrage condition implies that the PCA estimator of the risk factors is asymptotically equivalent to the mimicking portfolios estimator. Moreover, no-arbitrage permits to identify the risk premia as

\[31\text{Note that when } \hat{k}_{t-1} = T, \text{ which is its largest possible value, then the } T \times T \text{ matrix } \tilde{\Omega} \text{ has rank } T - 1, \text{ and hence becomes singular. Thus, for accurate estimation of the SDF, it is advisable to allow for a } T \text{ much larger than } \hat{k}_{t-1}.\]
the expectation of the latent risk factors. This result paves the way to an inferential procedure, based on the PCA approach, for the factors’ risk premia and for the stochastic discount factor, spanned by the latent risk factors. The strength of our set up is that it naturally handles time-varying factor models, where every feature is allowed to be time-varying including loadings, idiosyncratic risk and the number of risk factors. Several Monte Carlo experiments corroborate our theoretical findings. Our results represent a unique tool to address several empirical questions in empirical asset pricing, and beyond, unthinkable with the usual PCA methodologies valid when under double-asymptotic. For instance, our empirical analysis demonstrates how the estimated number of risk factors varies across time, increasing during booms and sharply decreasing during financial crises. We also show how the time-variation of the risk factors’ loadings appear driven by the interest rate spread variables and dividend and earnings variables, especially during financial crises. The estimates of the SDF implied by our cross-section of individual returns exhibit a substantial anti-cyclicality and appears strongly spanned by the Fama & French (2015) factors during market crashes, than otherwise.

References


REFERENCES


8 Appendixes

This section contains five appendixes: Appendix A sets the mathematical notation adopted in the paper; Appendix B presents four ancillary propositions; Appendix C presents the technical lemmas necessary for the proofs of the main theorems, which are then reported in Appendix D; Appendix E reports the form of the covariance matrix of the \((\textbf{e}_i \otimes \textbf{e}_i)\).

To simplify arguments, we report the proofs for the static case, that is when \(a_{it-1} = a_{i}, \lambda_{it-1} = \lambda_{i}, \sigma_{t-1}^2 = \sigma^2, r_{t-1} = r\). Generalization to the time-varying case is obtained without further difficulties in most cases.\(^{32}\) Likewise, \(\sum_{s \in \mathcal{T}_t} \) will be replaced by \(\sum_{s=1}^{T} \) throughout all the proofs.

8.1 Appendix A: Notation

The following notation is adopted throughout the paper: a.s. means almost surely; \(\mathbb{1}_{\mathcal{A}}\) denotes the indicator function equal to one when event \(A\) holds; \(t_i\) denotes the \(t-th\) column of the \(T \times T\) identity matrix \(\mathbf{I}_T\); \(C\) denotes a finite positive constant, not always the same; \(\mathbf{a}_{b \times c}\) and \(\mathbf{A}_{b \times c}\) denote a generic vector and matrix, respectively, of size \(b \times 1\) and \(b \times c\) implying that \(\mathbf{0}_{a}\) and \(\mathbf{0}_{a \times b}\) are the \(a \times 1\) vector and the \(a \times b\) matrix of zeros matrix, respectively; \(\text{diag}(\mathbf{A})\) is the diagonal matrix with the diagonal elements of the matrix \(\mathbf{A}\); for any full column rank matrix \(\mathbf{A}\) of size \(T \times a\), set the projection matrixes \(\mathbf{P}_{\mathbf{A}} \equiv \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\) and \(\mathbf{M}_{\mathbf{A}} \equiv \mathbf{I}_T - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\); \(g_j(\mathbf{A})\) defines the \(j-th\) eigenvalue in decreasing order of any symmetric \(a \times a\) matrix \(\mathbf{A}\) implying \(g_1(\mathbf{A}) \geq g_a(\mathbf{A})\); \(\otimes\) denotes the Kronecker product; \(\text{vec}(\cdot)\) and \(\text{trace}(\cdot)\) denote the vec and trace operators; \(\mathcal{Z} = \{\cdots, -1, 0, 1, \cdots\}\) denotes the set of relative numbers; the \(o(\cdot), O(\cdot)\) and \(o_p(\cdot), O_p(\cdot)\) notation is adopted for scalars and finite-dimensional vectors and matrixes (whose number of rows and columns are not a function of \(N\)); \(\rightarrow_{p}, \rightarrow_{d}\) denote convergence in probability and distribution, respectively; \(\mathcal{E}(\cdot), \text{Var}(\cdot), \text{Cov}(\cdot)\) and \(\mathcal{E}_t(\cdot), \text{Var}_t(\cdot), \text{Cov}_t(\cdot)\) indicate, respectively, the unconditional mean, variance, covariance and the conditional mean, variance, covariance with respect to all the available information up to time \(t\).

\(^{32}\)Further details for the time-varying case are provided for the proofs of the main theorem, when not straightforward.
8.2 Appendix B: Propositions

Proposition 1 (alternative representations of the objective function)

(i) For any arbitrary $T \times k$ matrix $A = (a_1, \cdots, a_T)'$ of rank $k$, and arbitrary $N \times k$ matrix $B = (b_1, \cdots, b_N)'$,

\[ V(k, A) \equiv \min_B \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - B'tA_t)^2 \]

\[ = \frac{1}{NT} \min_B \text{trace} \left( (X' - BA')' (X' - BA') \right) \]

\[ = \frac{1}{NT} \min_B \text{trace} \left( (X - AB')(X' - BA') \right) \]

\[ = \frac{1}{NT} \text{trace} \left( MXX' \right). \]

(ii) Setting $\tilde{A}$ equal to the $T \times k$ matrix of eigenvectors of $XX'/NT$, times $\sqrt{T}$, associated with its $k$ largest eigenvalues,

\[ V(k, \tilde{A}) \equiv \min_A \frac{1}{NT} (\text{trace} (XX') - \text{trace}(P_{\tilde{A}}XX')) \]

\[ = \frac{1}{NT} \left( \text{trace} (XX') - \text{trace} \left( \frac{\tilde{A}'XX'}{T} \right) \right) \]

\[ = \sum_{t=1}^T g(t) \frac{XX'}{NT} - \sum_{k=1}^k g(t) \frac{XX'}{NT}) = (\text{trace} (\tilde{V}_T) - \text{trace}(\tilde{V}_k)), \]

setting $\tilde{V}_a = \text{diag}(g_1(XX'/NT) \cdots g_a(XX'/NT))$, for every $1 \leq a \leq T$.

(iii) $V(k)$ is unchanged when replacing $\tilde{A}$ by $\tilde{A}C$ for any non-singular $k \times k$ matrix $C$.

Proposition 2 (pricing errors associated with the ex-post SDF $m_{t-1,t}'$) Let

\[ \Omega \equiv \text{cov}(f_t) \text{ be a positive definite matrix,} \]

\[ g_t \equiv \Omega^{-\frac{1}{2}} (f_t - \mathcal{E}(f_t)), \]

such that:

\[ \sup_{1 \leq t \leq T} \sum_{s=1}^T \| \mathcal{E}(g_t'g_s) \| = O(1), \]

\[ \sup_{1 \leq t, v \leq T} \sum_{s=1}^T \| \mathcal{E}(g_t'g_s g_v) \| = O(1), \]

\[ \sup_{1 \leq t \leq T} \sum_{s=1}^T \| \mathcal{E}(g_t'g_s g_t') - I_r \| = O(1), \]

\[ g_1(T^{-1/2} \sum_{t=1}^T g_t g_t') \geq C > 0 \text{ a.s. when } T \text{ large enough.} \]
Then, when (1) holds with \( a_i = 0 \) for every \( i \), the ex-post SDF \( m^P_{t-1,t} \), defined in (19), satisfies:

\[
E(m^P_{t-1,t} x_{it}) = O(T^{-1}) \text{ for every } i. 
\]

(43)

Proof.

\[
E(m^P_{t-1,t} x_{it}) = E \left[ \left( \frac{1}{r_f} - \frac{1}{T} \gamma' \Omega^{-1} (f_t - \bar{f}) \right) (X'_t \gamma + X'_t (E(f_t)) + e_{it}) \right]
\]

\[
= \frac{X'_t \gamma}{r_f} - \frac{X'_t \gamma}{r_f} E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} \gamma' \right) - \frac{X'_t \gamma}{r_f} E \left( (f_t - \bar{f})' \Omega^{-1} \gamma' \right)
\]

\[
= \frac{X'_t \gamma}{r_f} E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} (\gamma + \bar{f} - E(f_t)) \right) - \frac{X'_t \gamma}{r_f} E \left( (f_t - \bar{f})' \Omega^{-1} (\gamma + \bar{f} - E(f_t)) \right)
\]

For \( I \)

\[
I = -\frac{X'_t \gamma}{r_f} \left(1 - \frac{1}{T}\right) E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} \gamma' \right) - \frac{X'_t \gamma}{r_f} E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} (\bar{f} - E(f_t)) \right)
\]

\[
= \frac{X'_t \gamma}{r_f} \left(1 - \frac{1}{T}\right) E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} (\bar{f} - E(f_t)) \right) - \frac{X'_t \gamma}{r_f} E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} (\bar{f} - E(f_t)) \right)
\]

Term \( I_2 \) satisfies:

\[
E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} (\bar{f} - E(f_t)) \right) = E \left( (f_t - E(f_t)) (f_t - \bar{f})' \Omega^{-1} (\bar{f} - E(f_t)) \right) - E \left( (f_t - E(f_t)) (\bar{f} - E(f_t))' \Omega^{-1} (\bar{f} - E(f_t)) \right)
\]

\[
= O(T^{-1}).
\]

The two term on the right hand side of the last expression are made by a finite number of terms like

\[
T^{-1} \sum_{t=1}^{T} E((f_{at} - E(f_{at}))(f_{at} - E(f_{at}))(f_{cs} - E(f_{cs}))) \text{ and } T^{-2} \sum_{s,t=1}^{T} E((f_{at} - E(f_{at}))(f_{at} - E(f_{at}))(f_{cs} - E(f_{cs})))
\]

respectively, all of which are \( O(T^{-1}) \) in absolute value, whereas for the second term

Term \( I_3 \) is clearly \( o(1) \) as, when \( T \) is large enough,

\[
\| E \left( (f_t - E(f_t)) (f_t - E(f_t))' \Omega^{-1} - \Omega^{-1} \right) \| = \| E \left( (f_t - E(f_t)) (f_t - E(f_t))' \Omega^{-1} \right) - E \left( (f_t - E(f_t)) (\bar{f} - E(f_t))' \Omega^{-1} \right) \|
\]

\[
= O \left( C^{-1} \| E \left( (f_t - E(f_t)) (f_t - E(f_t))' \Omega^{-1} \right) - E \left( (f_t - E(f_t)) (f_t - E(f_t))' \Omega^{-1} \right) - E \left( (f_t - E(f_t)) (\bar{f} - E(f_t))' \Omega^{-1} \right) \| \right)
\]

\[
= O \left( \| \Omega \frac{1}{T} E \left( g_t g_t' \Omega^\frac{1}{2} \Omega^\frac{1}{2} - I_s \right) \Omega^\frac{1}{2} \| \right) = O(T^{-1}).
\]
Terms $I_4$ and $I_5$ are also $O(T^{-1})$ by Holder’s inequality, as they each contain two terms that are of order $O(T^{-\frac{1}{2}})$, namely $\mathcal{E} \| \Omega^{-1} - \Omega^{-1} \| = O(T^{-\frac{1}{2}})$, $\mathcal{E} \| \bar{f} - \mathcal{E}(\bar{f}) \| = O(T^{-\frac{1}{2}})$.

Consider now part II.

$$\mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) = \mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) + \mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\bar{f} - \mathcal{E}(f)) \right)$$

Then

$$\mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) = \mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\bar{f} - \mathcal{E}(f)) \right) + \mathcal{E} \left( (\mathcal{E}(f_t) - \bar{f})'\Omega^{-1}(\bar{f} - \mathcal{E}(f)) \right)$$

and

$$\mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) = \mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) + (\mathcal{E}(f_t) - \bar{f})'\Omega^{-1}(\bar{f} - \mathcal{E}(f))$$

as

$$\mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) = \mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) + O(T^{-1})$$

$$\mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) = \mathcal{E} \left( (f_t - \bar{f})'\Omega^{-1}(\gamma + \bar{f} - \mathcal{E}(f)) \right) + O(T^{-1})$$

where take into account that the terms involving $\bar{g}$ are of smaller order. Collecting terms

$$\mathcal{E}(m_{i-1,i}^T) = \frac{\lambda_{X}'}{r_f} - \frac{\lambda_{X}'}{r_f} (1 - \frac{1}{T}) + O(T^{-1}) + O(T^{-1}) = O(T^{-1}).$$

QED

Remark. Our assumptions on the latent factors $f_t$ are weaker than (temporal) iid and zero mixed-third moments.

Remark. Proposition 2 can be extended to the case $a_i \neq 0$. However, we only present case $a_i = 0$ as our PCA procedure, valid when $N$ becomes large, does not allow to identify the pricing errors $a_i$.

**Proposition 3 (equivalence between risk premia estimators)** Under Assumption [7]:

$$\hat{\gamma}^{\text{twopass}} = \hat{\gamma}.$$

**Proof.** The result follows from the identifies $N^{-1}A'\hat{A} = \hat{V}$ and $(NT)^{-1}F'XX' = \hat{V}F'$, as:

$$\hat{\gamma}^{\text{twopass}} = \left( \frac{\hat{A}'\hat{A}}{N} \right)^{-1} \frac{\hat{A}'\hat{X}_N}{N} = \hat{V}^{-1} \hat{A}'X' 1_T \hat{N} = \hat{V}^{-1} \hat{F}'X' 1_T = \hat{V}^{-1} \hat{F}' 1_T = \hat{\gamma}.$$
Proposition 4 (conditional APT) Under Assumptions [1][2] and [3]:

(i) there exists a $K_{1t-1} < \infty$ a.s such that

$$\sup_N a'_t \Sigma_{et-1} a_{t-1} \leq K_{1t-1} \text{ a.s. for every } t \in \mathbb{Z}. \tag{44}$$

(ii) When, in addition, with $\sup_N g_t(\Sigma_{et-1}) \leq C < \infty$, there exists a $K_{2t-1} < \infty$ a.s such that

$$\sup_N a'_t a_{t-1} \leq K_{2t-1} \text{ a.s. for every } t \in \mathbb{Z}. \tag{45}$$

(iii) When, in addition, $\mathbb{E}(\gamma'_t \gamma_t-1) \leq C < \infty$, $\mathbb{E}(K_{2t-1}) \leq C < \infty$ and the smoothness condition $\sum_{i=1}^{\infty} (\mathbb{E}(\lambda_{it-1} - \mathbb{E}(\lambda_{it-1}) | I_{t-1})^2) < C < \infty$ holds, then for any information set $I_{t-1}$, there exists $K_{3t-1} < \infty$ a.s. such that:

$$\sup_N \sum_{i=1}^{N} \left( \mathbb{E}(x_{it} | I_{t-1}) - \mathbb{E}(\lambda_{it-1} | I_{t-1}) \right)^2 \leq K_{3t-1} \text{ a.s. for every } t \in \mathbb{Z}. \tag{46}$$

**Proof.** Part (i) follows along the lines of [Ingersoll (1984)](Theorem 1) and part (ii) follows along the lines of [Ingersoll (1984)](Theorem 3), applied to the conditional factor model (1).

For part (iii) we generalize [Stambaugh (1983)](Theorem 2) to the case of time-varying loadings. From part (ii)

$$\mathbb{E}(K_{2t-1} | I_{t-1}) \geq \sum_{i=1}^{N} (\mathbb{E}(\mu_{it-1} | I_{t-1}) - \mathbb{E}(\mu_{it-1} | I_{t-1}) | I_{t-1})^2 \geq \sum_{i=1}^{N} \left( \mathbb{E}(\lambda_{it-1} | I_{t-1}) - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) \right)^2$$

$$= \sum_{i=1}^{N} \left( \mathbb{E}(\lambda_{it-1} | I_{t-1}) - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) \right)^2.$$

By assumption it follows that $\mathbb{E}(K_{2t-1} | I_{t-1}) < \infty$ a.s. Therefore, the factor asset pricing model assumed by the investor with the coarser information set $I_{t-1}$ satisfies the pricing errors bound

$$\sum_{i=1}^{N} \left( \mathbb{E}(\mu_{it-1} | I_{t-1}) - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) \right)^2 \leq K_{3t-1} < \infty \text{ a.s.}$$

for some $K_{3t-1}$ if $\sum_{i=1}^{\infty} \mathbb{E}(\lambda_{it-1} - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) < \infty$ a.s.. In turn, this follows by

$$\mathbb{E}\left[ \sum_{i=1}^{\infty} \left( \mathbb{E}(\lambda_{it-1} - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) \right)^2 \right]$$

$$\leq \sum_{i=1}^{\infty} \left( \mathbb{E}(\lambda_{it-1} - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) \right)^2 \leq \mathbb{E}(\gamma'_t \gamma_t-1) \frac{C < \infty}{\frac{1}{2}} \sum_{i=1}^{\infty} \left( \mathbb{E}(\lambda_{it-1} - \mathbb{E}(\lambda_{it-1} | I_{t-1}) | I_{t-1}) \right)^2 \leq C < \infty.$$

QED

Remark. The smoothness assumption on the loadings is trivially satisfied for the static case, namely $\lambda_{it-1} = \lambda_t$, which is the case examined by [Stambaugh (1983)]. However, it will be more generally satisfied in a variety of
dynamic set-ups. For instance, when $\lambda_{is} = \lambda_i(s)$ for some differentiable function $\lambda_i(\cdot)$, as in Assumption [3] and assuming for simplicity that the coarser information set simply means that information is acquired with some delay, so that $\mathcal{E}(\lambda_{it-1} | I_{t-1}) = \lambda_{is-1}$ for some $s < t$, then $\lambda_{it-1} - \mathcal{E}(\lambda_{it-1} | I_{t-1}) = (t-s)\lambda_{is}^{(1)}$ and the smoothness assumption can be expressed as

$$\mathcal{E}\left(\text{tr}(\Lambda_{i-1}^{(1)}\Lambda_{i-1}^{(1)})\right) \leq C < \infty \text{ for every } t \in \mathcal{Z}.$$ 

Remark. As indicated by Stambaugh (1983)[Theorem 3], a special case of a coarser information set consists of no information, leading to a restriction on the unconditional pricing errors $\mathcal{E}(a_{it})$.

Remark. Stambaugh (1983)[Lemma 1] derives condition $\mathcal{E}(K_{2r-1} | I_{t-1}) < \infty$ under distributional assumptions and static loadings.

8.3 Appendix C: Lemmas

Lemma 1 Under Assumption 1 and 3, as $N \to \infty$,

$$\| \Lambda' a \| = O_p(N^{\frac{1}{2}}).$$

Proof.

$$\| \Lambda' a \| \leq \sum_{i=1}^{N} |a_i| \| \lambda_i \| \leq (\sum_{i=1}^{N} a_i^2)^{\frac{1}{2}} (\sum_{i=1}^{N} \| \lambda_i \|) = O_p(N^{\frac{1}{2}}).$$

QED

Lemma 2 Under Assumption 1 and 5, as $N \to \infty$,

$$\| e'a \| = O_p(g_1^2(\Sigma_e)).$$

Proof.

$$e'a = O_p((\sum_{i,j=1}^{N} a_i a_j \mathcal{E}(e_i e_j'))^{\frac{1}{2}}) = O_p((\sum_{i,j=1}^{N} a_i a_j \sigma_{ij})^{\frac{1}{2}}) = O_p((a'a)^{\frac{1}{2}}g_1^2(\Sigma_e)) = O_p(g_1^2(\Sigma_e)).$$

QED

Lemma 3 (central lemma) Under Assumptions 1-6, as $N \to \infty$,

(i)

$$\frac{1}{T} \tilde{F}' \left( \frac{1}{NT} XX' - \frac{1}{T} \sigma^2 I_T \right) \tilde{F} = \tilde{U} \equiv (\tilde{V} - \frac{1}{T} \sigma^2 I_r) \to_p \tilde{U} > 0.$$

where recall that $\tilde{U}$ denotes the diagonal $r \times r$ matrix of eigenvalues of $\left( \frac{1}{NT} XX' - \frac{1}{T} \sigma^2 I_T \right)$ corresponding to $\tilde{F}$, and $U$ denotes the diagonal $r \times r$ matrix of distinct eigenvalues of $\Sigma^2_{\Lambda} \Sigma_F \Sigma^2_{\Lambda}$.
(ii) \[ \frac{\hat{F}F}{T} \to_p Q = U\frac{1}{2}Y\Sigma_\Lambda^{-\frac{1}{2}}, \]

for a non-singular \( Q \), where \( Y \) is the \( r \times r \) eigenvectors matrix of \( \Sigma_\Lambda \Sigma_F \Sigma_\Lambda^\frac{1}{2} \).

**Proof.** Consider case \( a_i = 0 \) first. We will then show that the no-arbitrage condition implies that the \( a_i \)s will be washed out as \( N \) diverges.

(i) The result follows given that \[ \frac{1}{NT}XX' = \frac{1}{T}F(\frac{\Lambda'\Lambda}{N})F' + \frac{ee'}{NT} + o_p(1) \to_p \frac{1}{T}F\Sigma_\Lambda F' + \frac{\sigma^2}{T}I_T. \]

Note that, as highlighted in Bai (2003), Lemma D.1, \( \frac{1}{NT}XX' \) and \( \left( \frac{1}{NT}XX' - \frac{1}{T}\sigma^2I_T \right) \) have the same set of eigenvectors, but different eigenvalues. In particular, the eigenvalues corresponding to \( \hat{F} \) equal \( \hat{V} \) and \( \hat{U} \), respectively, for the two matrices. Continuity of the eigenvalue function together with Slutzky theorem concludes, recalling that the the non-zero eigenvalues of \( T^{-1}F\Sigma_\Lambda F' \) coincide with the entire set of \( r \) eigenvalues of \( \Sigma_\Lambda^\frac{1}{2}\hat{\Sigma}_F \Sigma_\Lambda^\frac{1}{2} \). The second part of (i) follows directly.

Now consider the general case when the \( a_i \neq 0 \). Then, from \( X = 1_Ta' + F\Lambda' + e \),
\[ \frac{1}{NT}XX' = \frac{1}{T}F(\frac{\Lambda'\Lambda}{N})F' + \frac{ee'}{NT} + \frac{1}{NT}1'a1'T + \frac{1}{NT}a'a'e' + \frac{1}{NT}1'\Lambda F' + \frac{1}{NT}F\Lambda'1T + o_p(1) \to_p \frac{1}{T}F\Sigma_\Lambda F' + \frac{\sigma^2}{T}I_T, \]

by Lemmas 1 and 2. The rest of the proof follows along the same lines.

(ii) Consider case \( a_i = 0 \) first. We adapt the proof of Bai (2003), Proposition 1. In fact, one needs to replace \( XX'/NT \) with \( (XX'/NT - \sigma^2I_T/T) \) to take into account the fact that the term \( ee'/NT \) is non-negligible when \( T \) is fixed, and it would contribute to the non-zero eigenvalues (as \( N \to \infty \)) even though it is not related to the common component.

In particular, given
\[ \frac{(XX'/NT - \sigma^2I_T)}{T} \to_F \hat{F}\left(\hat{V}_r - \frac{\sigma^2}{T}I_r\right) = \hat{F}\hat{U}, \]
pre-multiplying both sides by \( T^{-1}(\Lambda'\Lambda/N)^\frac{1}{2}F' \), and re-arranging terms, yields
\[ \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'}{T}\left(\frac{XX'}{NT} - \sigma^2I_T\right) \to_F \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'F}{T} - \frac{\sigma^2}{N}F' + C_N = \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'\hat{F}}{T}\hat{U}, \]
re-written as
\[ (B_N + C_ND_N^{-1})E_N = E_N\hat{U}, \]
setting \( E_N = D_N(diag(D_N^2))^{-\frac{1}{2}} \) and
\[ B_N = \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'F}{T} - \frac{\sigma^2}{N}F' + C_N^2 \]
\[ = \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'F}{T} - \frac{\sigma^2}{N}F' + C_N^2 \]
\[ = \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'F}{T}, \]
and
\[ D_N = \left(\frac{\Lambda'\Lambda}{N}\right)^\frac{1}{2}\frac{F'F}{T}, \]
as $D_N$ is $a.s.$ invertible for $N$ large enough by part (i). One then follows the steps in Bai (2003), but replacing $\Sigma_F$ with $\hat{\Sigma}_F$, obtaining $B_N \rightarrow_p \Sigma^{1/2}_\Lambda \hat{\Sigma}_F \Sigma^{1/2}_\Lambda$ and $E_N \rightarrow_p \Upsilon'$, and thus

$$\frac{F'\tilde{F}}{T} = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1/2} E_N(diag(D'N D_N))^{1/2} \rightarrow_p \Sigma^{1/2}_\Lambda \Upsilon' U^1.$$

The same result is obtained when $a_i \neq 0$, the only difference being that $C_N$ is now given by

$$C_N = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} \left(\frac{F'F \Lambda'F'}{T^2} + F'e \Lambda'F' \frac{e}{N} + \frac{1}{T^2} F' \left(\frac{e}{N} - \sigma^2 I_T\right) F + \frac{F' \Lambda'F'}{T^2} \frac{1}{NT} + \frac{1}{T^2} \frac{a'}{NT} F + \frac{1}{T^2} \frac{a'}{NT} F + \frac{F' \Lambda'F'}{T^2} \frac{1}{NT} F\right) = o_p(1),$$

which is still $o_p(1)$ by Lemmas 1 and 2. QED

Lemma 4 (quantities for asymptotic covariance matrix) Under Assumptions 1-6, as $N \rightarrow \infty$,

(i) $$\tilde{\sigma}^2 = \frac{T}{T - Tr} V(r) \rightarrow_p \sigma^2.$$

(ii) $$\tilde{\sigma}_4 \equiv \left(3 + \frac{2T}{T} \left(\sum_{t=1}^T \left(\frac{F'F}{T}\right)^2\right) + \frac{18r}{T}\right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{e_i^4}{T}\right) \rightarrow_p \sigma_4 + C_\kappa \kappa_4,$$

for a constant $C_\kappa$ (function of $r$, $T$, $F$ and $H$) defined in (50) below.

(iii) $$\tilde{\Sigma}_\Lambda \equiv \frac{\Lambda'\Lambda}{N} - \frac{\sigma^2}{T} I_r \rightarrow_p H^{-1} \Sigma_\Lambda H^{-1}.$$

Proof. Consider case $a_i = 0$ first. We will then show that the no-arbitrage condition implies that the $a_i$s will
be washed out as \( N \) diverges. (i) From

\[
\hat{e}_{it} = x_{it} - \lambda_i \tilde{F}_t = x_{it} - \left( \lambda_i \tilde{H}^{-1} + \frac{e_i^T \tilde{F}}{T} \right) \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right)
\]

yielding

\[
V(r) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 + \frac{1}{T^3} \sum_{t=1}^{T} \tilde{F}_t^T \left( \frac{1}{N} \sum_{i=1}^{N} e_i e_i^T \right) \tilde{F}_t
\]

\[
+ \frac{1}{T} \sum_{t=1}^{T} \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right) \tilde{H}^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \lambda_i \lambda_i^T \right) \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right)
\]

\[
- 2 \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it} \left( \lambda_i \tilde{H}^{-1} \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right) + \left( \frac{e_i^T \tilde{F}}{T} \right) \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right) \right)
\]

\[
+ 2 \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left( \lambda_i \tilde{H}^{-1} \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right) \right) \left( \frac{e_i^T \tilde{F}}{T} \right) \left( \tilde{F}_t - \tilde{H} \tilde{F}_t \right)
\]

Case \( a_i \neq 0 \) now easily follows by Lemmas \( \text{[1]} \) and \( \text{[2]} \) To prove this, simply replace \( e_{it} \) in the above expressions with \( e_{it}^* = e_{it} + a_i \), and notice that all the parts involving the \( a_i \) vanish asymptotically. For instance, the first of these terms satisfies

\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} (e_{it}^*)^2 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} (e_{it} + a_i)^2
\]

\[
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 + \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} a_i^2 + \frac{2}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it} a_i
\]

\[
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 + O(N^{-1}) + O_p(N^{-\frac{1}{2}}).
\]

(ii) Set \( a_i = 0 \). By developing \( N^{-1} \sum_{t=1}^{N} e_{it}^4 \), along the same lines as part (i), one obtains

\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^4 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^4 + \frac{6}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 ((e_i^T \tilde{F}) \tilde{F}_t)^2
\]

\[
+ \frac{4}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 ((e_i^T \tilde{F}) \tilde{F}_t)^3 + O_p(N^{-\frac{1}{2}}).
\]
Set 
\[ c_t = (c_{1t} \cdots c_{Pt})' = \frac{1}{T} \mathbf{F} \mathbf{H} \mathbf{H}' \mathbf{F}_t = \frac{1}{T} \left( (\mathbf{F}'_1 \mathbf{H} \mathbf{H}'_1) \cdots (\mathbf{F}'_T \mathbf{H} \mathbf{H}'_T) \cdots (\mathbf{F}'_T \mathbf{H} \mathbf{H}'_T) \right)' \] .

Then, the second term on the right hand side of (47) can be re-written as \((NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} (e_{is} c_{st})^4 + o_p(1)\), with mean satisfying:
\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left( \sum_{s=1}^{T} e_{is} c_{st} \right)^4 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i_1=1}^{T} \sum_{s_1, s_2, s_3, s_4 = 1}^{T} \kappa_4, i_{i_1} s_1, i_{i_2} s_2, i_{i_3} s_3, s_4 = 1 \sigma_{i_{i_1} s_1 s_2} \sigma_{i_{i_2} s_3 s_4} + \sigma_{i_{i_1} s_1 s_3} \sigma_{i_{i_2} s_2 s_4} + \sigma_{i_{i_1} s_1 s_4} \sigma_{i_{i_2} s_2 s_3} c_{s_1 t} c_{s_2 t} c_{s_3 t} c_{s_4 t}
\]
\[
+ \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{s=1}^{T} \sum_{i_1=1}^{T} \sum_{s_1, s_2, s_3, s_4 = 1}^{T} \kappa_4, i_{i_1} s_1, i_{i_2} s_2, i_{i_3} s_3, s_4 = 1 \sigma_{i_{i_1} s_1 s_2} \sigma_{i_{i_2} s_3 s_4} + \sigma_{i_{i_1} s_1 s_3} \sigma_{i_{i_2} s_2 s_4} + \sigma_{i_{i_1} s_1 s_4} \sigma_{i_{i_2} s_2 s_3} c_{s_1 t} c_{s_2 t} c_{s_3 t} c_{s_4 t}
\]
\[
\rightarrow \kappa_4^4 \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{s=1}^{T} c_{st}^4 + \frac{3\sigma_4^4}{T} \sum_{t=1}^{T} \left( \sum_{s=1}^{T} c_{st}^2 \right)^2,
\]
and variance
\[
\text{Var} \left( \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{u=1}^{T} e_{iu} c_{ut} \right)^4 = \frac{1}{N^2 T^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \text{Cov} \left( \sum_{u=1}^{T} e_{iu} c_{ut} \right)^4, \left( \sum_{v=1}^{T} e_{ju} c_{vs} \right)^4
\]
\[
= \frac{1}{N^2 T^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1, u_2, v_1, v_2, v_3, v_4 = 1}^{T} \sum_{u_3, u_4 = 1}^{T} c_{u_1 t} c_{u_2 t} c_{u_3 t} c_{u_4 t} c_{v_1 s} c_{v_2 s} c_{v_3 s} c_{v_4 s}
\]
\[
\times \text{Cov} \left( e_{iu_1}, e_{iu_2}, e_{iu_3}, e_{iu_4}, e_{jv_1}, e_{jv_2}, e_{jv_3}, e_{jv_4} \right)
\]
\[
= \frac{1}{N^2 T^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1, u_2, v_1, v_2, v_3, v_4 = 1}^{T} \sum_{u_3, u_4 = 1}^{T} c_{u_1 t} c_{u_2 t} c_{u_3 t} c_{u_4 t} c_{v_1 s} c_{v_2 s} c_{v_3 s} c_{v_4 s}
\]
\[
\times \left( \kappa_8 \left( e_{iu_1}, e_{iu_2}, e_{iu_3}, e_{iu_4}, e_{jv_1}, e_{jv_2}, e_{jv_3}, e_{jv_4} \right) \right)
\]
\[
+ \sum_{(6,2)}^{(6,2)} \kappa_6 \left( e_{iu_1}, e_{iu_2}, e_{iu_3}, e_{iu_4}, e_{jv_1}, e_{jv_2}, e_{jv_3}, e_{jv_4} \right) \text{Cov} \left( e_{jv_3}, e_{jv_4} \right)
\]
\[
+ \sum_{(4,4)}^{(4,4)} \kappa_4 \left( e_{iu_1}, e_{iu_2}, e_{jv_1}, e_{jv_2} \right) \kappa_4 \left( e_{iu_3}, e_{iu_4}, e_{jv_3}, e_{jv_4} \right)
\]
\[
+ \sum_{(4,2,2)}^{(4,2,2)} \kappa_4 \left( e_{iu_1}, e_{iu_2}, e_{jv_1}, e_{jv_2} \right) \text{Cov} \left( e_{iu_3}, e_{iu_4} \right) \text{Cov} \left( e_{jv_3}, e_{jv_4} \right)
\]
\[
+ \sum_{(2,2,2)}^{(2,2,2)} \text{Cov} \left( e_{iu_1}, e_{iu_2} \right) \text{Cov} \left( e_{iu_3}, e_{iu_4} \right) \text{Cov} \left( e_{jv_3}, e_{jv_4} \right),
\]
where \(\kappa_4(\cdot), \kappa_6(\cdot),\) and \(\kappa_8(\cdot)\) denote the fourth-, sixth-, and eighth-order mixed cumulants, respectively. Expression (48) is a consequence of the cumulants’ theorem (see Brillinger (2001)), whereby \(\sum_{(\nu_1, \nu_2, \ldots, \nu_k)}\) denotes the sum over all possible partitions of a group of \(K\) random variables into \(k\) subgroups of size \(\nu_1, \nu_2, \ldots, \nu_k\), respectively. As an example, \(\sum_{(6,2)}\) defines the sum over all possible partitions of the group of eight random variables \(\{e_{iu_1}, e_{iu_2}, e_{iu_3}, e_{iu_4}, e_{jv_1}, e_{jv_2}, e_{jv_3}, e_{jv_4}\}\) into two subgroups of size six and two, respectively. Moreover,
since $\mathcal{E}(e_{it}) = \mathcal{E}(e_{it}^3) = 0$, we do not need to consider further partitions in the above relation. Under our assumptions it follows that the number of indecomposable partitions, in (48), is of order $O(N)$, implying

$$\text{Var}\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\sum_{s=1}^{T} e_{is}c_{st})^4\right) = O\left(\frac{1}{N}\right).$$

(49)

Therefore, it follows that

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\sum_{s=1}^{T} e_{it}^4 \tilde{F}_t)^4 \to_p \kappa_4 \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} c_{it}^4 + \frac{3\sigma_4}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} c_{st}^2)^2.$$

For the third term of (47) one obtains

$$6(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 ((e_{it}/T) \tilde{F}_t)^2 = 6(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 (\sum_{s=1}^{T} e_{is}c_{st})^2 + o_p(1),$$

and, along the same lines,

$$\frac{6}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} E(\sum_{s=1}^{T} e_{is}c_{st})^2) \to \frac{6}{T}(\kappa_4 + 2\sigma_4)(\sum_{t=1}^{T} c_{tt}^2) + \frac{6r}{T}\sigma_4,$$

where by easy calculations

$$\sum_{t=1}^{T} \sum_{s=1}^{T} c_{st}^2 = \sum_{t=1}^{T} c_{tt} = r, \quad \sum_{t=1}^{T} \sum_{s=1}^{T} c_{tt}^2 = \sum_{t=1}^{T} c_{tt}^2.$$

Using again the cumulants’ theorem, it follows that

$$\text{Var}\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2 (\sum_{s=1}^{T} e_{is}c_{st})^2\right) = \frac{1}{N^2T^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \text{Cov}\left(\sum_{u=1}^{T} e_{iu}^2 e_{ut}^2, \sum_{v=1}^{T} e_{jv}^2 e_{uv}^2\right),$$

implying

$$\frac{6}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 ((e_{it}/T) \tilde{F}_t)^2 \to \frac{6}{T}(\kappa_4 + 2\sigma_4)(\sum_{t=1}^{T} c_{tt}^2) + \frac{6r}{T}\sigma_4.$$

Along the same lines, for the fourth and fifth terms on the right hand side of (47) one obtains

$$\frac{4}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^3 ((e_{it}/T) \tilde{F}_t)^3 \to \frac{4}{T}(\kappa_4 + 3\sigma_4)(\sum_{t=1}^{T} c_{tt}^3) + \frac{4r}{T}\sigma_4.$$
Remark. The following identity holds:
\[ \sum_{i=1}^{N} \hat{\lambda}_i \hat{\lambda}_i' = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{\mathbf{H}}^{-1} \lambda_i + \frac{\hat{\mathbf{F}}' e_i}{T} + \frac{1}{T} \hat{\mathbf{F}}' (\hat{\mathbf{F}} - \hat{\mathbf{F}} \hat{\mathbf{H}}^{-1}) \lambda_i \right) ' \]

implying
\[ \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_i' \rightarrow_p \mathbf{H}^{-1} \mathbf{\Sigma}_A \mathbf{H}^{-1} + \frac{\sigma^2}{T} \mathbf{I}. \]

Case \( a_t \neq 0 \) follows by replacing the \( e_{it} \) with the \( e_{it} = e_{it} + a_t \), implying \( e_{it}' = e_{it} + a_t \mathbf{1}_T \), and using Lemmas [1] and [2]. For instance,
\[ \frac{1}{N} \sum_{i=1}^{N} \mathbf{e}_i^* \mathbf{e}_i'' = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{e}_i + a_t \mathbf{1}_T) (\mathbf{e}_i + a_t \mathbf{1}_T)' = \frac{1}{N} \sum_{i=1}^{N} \mathbf{e}_i \mathbf{e}_i' + \frac{1}{T} \sum_{i=1}^{N} a_t^2 + \frac{1}{N} \sum_{i=1}^{N} a_t \mathbf{1}_T \mathbf{e}_i' + \frac{1}{N} \sum_{i=1}^{N} a_t \mathbf{1}_T \mathbf{e}_i \]
\[ = \frac{1}{N} \sum_{i=1}^{N} \mathbf{e}_i \mathbf{e}_i' + O(N^{-1}) + O_p(N^{-\frac{1}{2}}). \]

QED

Remark. The following identity holds:
\[ \mathbf{\Sigma}_A = \hat{\mathbf{U}}^*. \]

Remark. Based on Lemma [4] one can construct a plug-in consistent estimator of the asymptotic covariance matrix.

**Lemma 5 (generalization of consistency with rates for every \( k \))** Under Assumptions 1-6, as \( N \to \infty \),
\[ \| \hat{\mathbf{F}}^k_t - \mathbf{F}_t \| = O_p(N^{-\frac{1}{2}}), \text{ for every } 1 \leq k, t \leq T, \]
setting \( \hat{\mathbf{F}}^k_t \equiv \hat{\mathbf{U}}_k \hat{\mathbf{F}}^k_t \), with
\[ \hat{\mathbf{U}}_k^t \equiv \hat{\mathbf{V}}_k - \frac{\sigma^2}{T} \mathbf{I}_k, \text{ and } \mathbf{H}_k \equiv \mathbf{H}_k \hat{\mathbf{U}}_k = (\frac{\mathbf{A}' \mathbf{A}}{N}) (\frac{\mathbf{F}' \hat{\mathbf{F}}_t}{T}) \text{ being a full-rank } r \times k \text{ matrix.} \]

**Proof.** This follows from the proof to Theorem [1](i). Pre-multiplying (52) by \( \hat{\mathbf{U}}_k \), where we express all in terms of a generic \( 1 \leq k \leq T \),
\[ \hat{\mathbf{F}}^k_t - \mathbf{H}_k \mathbf{F}_t = T^{-1} \sum_{s=1}^{T} \hat{\mathbf{F}}^k_{s, \eta_t} + T^{-1} \sum_{s=1}^{T} \hat{\mathbf{F}}^k_{s, \xi_t} + T^{-1} \sum_{s=1}^{T} \hat{\mathbf{F}}^k_{s, \eta_s}, \]
and then apply Lemma [3] given that \( \| \mathbf{F}' \mathbf{F}/T \| = O_p(1) \) and \( \| \hat{\mathbf{F}}^k_t \hat{\mathbf{F}}_t^k/T \| = O(1) \). QED
Lemma 6 (behaviour of \( V(k) \)) Under Assumptions 1-6,

(i) When \( k \leq r \):

\[
V(k) = \sigma^2 - \sigma^2 \frac{K}{T} + \sum_{s=k+1}^{r} u_s + O_p(N^{-\frac{1}{2}}).
\]

where the \( u_s \) are the random (diagonal) elements of \( U \), satisfying \( u_1 \geq u_2 \geq \cdots \geq u_r > 0 \) a.s..

(ii) When \( k > r \):

\[
V(k) = \sigma^2 - \sigma^2 \frac{K}{T} + O_p(N^{-\frac{1}{2}}).
\]

**Proof.** Consider case \( a_i = 0 \), as the proof follows along the same lines when \( a_i \neq 0 \) by Lemmas 1 and 2

(i) Case \( k \leq r \):

Given \( \lambda^k = X'\tilde{F}^k(\tilde{F}^k\tilde{F}^k)^{-1} = X'\tilde{F}^k/T \), one gets:

\[
\tilde{\lambda}_t^k = \frac{e_i'\tilde{F}^k}{T} + \lambda_t^i(\frac{F^t\tilde{F}^k}{T}),
\]

setting the \( r \times k \) matrix \( \tilde{H}_k \equiv (\Lambda'\Lambda/N)(F'\tilde{F}^k/T)\tilde{U}_k^{-1} \) of rank \( k \),

yielding

\[
\hat{c}_{it}^k \equiv x_{it} - \tilde{\lambda}_t^k\tilde{F}_t^k = e_{it} + \lambda_t^iF_t - (e_i'\tilde{F}^k + \lambda_t^i(\frac{F^t\tilde{F}^k}{T})(\tilde{H}_kF_t + (\tilde{F}_t^k - \lambda_t^iF_t)))
\]

\[
= e_{it} + \lambda_t^iF_t - e_i'\tilde{F}^kF_t^k - \lambda_t^i(\frac{F^t\tilde{F}^k}{T})(\tilde{H}_kF_t + O_p(N^{-1/2})).
\]

By Lemma 3 above, recalling that \( \tilde{F}_r = \tilde{F} \) with \( r \geq k \),

\[
\frac{F^t\tilde{F}^k}{T} = \frac{F^t\tilde{F}}{T}[I_k,0_{k\times r-k}]' \rightarrow_p Q'[I_k,0_{k\times r-k}]',
\]

recalling that \( Q = U\frac{1}{2}\Upsilon\Sigma^{\frac{1}{2}}_\Lambda \). Then

\[
\left(\frac{F^t\tilde{F}^k}{T}\right)H_k \rightarrow_p Q'[I_k,0_{k\times r-k}]'U_k^{-1}[I_k,0_{k\times r-k}]Q\Sigma_\Lambda = \Sigma^{\frac{1}{2}}_\Lambda \Upsilon'(\begin{bmatrix} I_k & 0_{k\times r-k} \\ 0_{r-k\times k} & 0_{r-k\times r-k} \end{bmatrix})\Sigma^{\frac{1}{2}}_\Lambda,
\]

obtaining

\[
\hat{c}_{it}^k = e_{it} + \lambda_t^i(I_r - \Sigma^{\frac{1}{2}}_\Lambda \Upsilon'(\begin{bmatrix} I_k & 0_{k\times r-k} \\ 0_{r-k\times k} & 0_{r-k\times r-k} \end{bmatrix})\Sigma^{\frac{1}{2}}_\Lambda)F_t - e_i'\tilde{F}^kF_t^k + O_p(N^{-1/2}),
\]

and implying

\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} (\hat{c}_{it}^k)^2 = V(k) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it}^2 + \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{F}_t^k\tilde{F}_t^k e_{it} + \frac{e_i'\tilde{F}^k}{T} - 2 \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} e_{it} e_i'\tilde{F}^k
\]

\[
+ \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{F}_t'(I_r - \Sigma^{\frac{1}{2}}_\Lambda \Upsilon'(\begin{bmatrix} I_k & 0_{k\times r-k} \\ 0_{r-k\times k} & 0_{r-k\times r-k} \end{bmatrix})\Sigma^{\frac{1}{2}}_\Lambda)\lambda_t^i(I_r - \Sigma^{\frac{1}{2}}_\Lambda \Upsilon'(\begin{bmatrix} I_k & 0_{k\times r-k} \\ 0_{r-k\times k} & 0_{r-k\times r-k} \end{bmatrix})\Sigma^{\frac{1}{2}}_\Lambda)F_t + O_p(N^{-1/2})
\]

\[
\rightarrow_p \sigma^2 + \sigma^2 \frac{K}{T} - 2\sigma^2 \frac{K}{T} + \text{trace} \left( (I_r - \Upsilon'(\begin{bmatrix} I_k & 0_{k\times r-k} \\ 0_{r-k\times k} & 0_{r-k\times r-k} \end{bmatrix})\Sigma^{\frac{1}{2}}_\Lambda \Sigma^{\frac{1}{2}}_\Lambda \Sigma^{\frac{1}{2}}_\Lambda) (I_r - \Upsilon'(\begin{bmatrix} I_k & 0_{k\times r-k} \\ 0_{r-k\times k} & 0_{r-k\times r-k} \end{bmatrix})\Sigma^{\frac{1}{2}}_\Lambda \Sigma^{\frac{1}{2}}_\Lambda) \right)
\]

\[
= \sigma^2 - \sigma^2 \frac{K}{T} + \left( \text{trace}[ ((0_{k\times r-k} & 0_{k\times r-k} \\ 0_{r-k\times k} & 1_{r-k}) U)] \right)
\]

\[
= \sigma^2 - \sigma^2 \frac{K}{T} + \sum_{s=k+1}^{r} u_s,
\]
(ii) Case \( k > r \):

In this case \( \hat{H}_k = (A'A)^{-1} (F'H_k) \) has full-row rank matrix, implying that its Moore-Penrose satisfies \( \hat{H}_k \hat{H}_k^+ = I_r \) and \( (\hat{H}_k^+) \hat{H}_k' = I_r \). By Proposition 1 and Lemma 3, \( V(k) = (NT)^{-1} \sum_{t=1}^T \sum_{l=1}^N (\hat{c}_{il})^2 = (NT)^{-1} \sum_{t=1}^T \sum_{l=1}^N (\hat{e}_{it})^2 \) for

\[
\hat{c}_{il} = \hat{x}_{it} - \lambda_i \hat{F}_t 
\]

Then

\[
e_{it} = e_{it} - \frac{e_i' F_k}{T} \hat{F}_l + \lambda_i (\hat{H}_k') \hat{H}_k F_l - \lambda_i \frac{F_i' F_k}{T} \hat{F}_l 
\]

\[
e_{it} = e_{it} - \frac{e_i' F_k}{T} \hat{F}_l + \lambda_i (\hat{H}_k') \hat{H}_k F_l - \lambda_i \frac{F_i' F_k}{T} \hat{F}_l 
\]

\[
e_{it} = e_{it} - \frac{e_i' F_k}{T} \hat{F}_l + \lambda_i (\hat{H}_k') \hat{H}_k F_l - \lambda_i \frac{F_i' F_k}{T} \hat{F}_l 
\]

\[
e_{it} = e_{it} - \frac{e_i' F_k}{T} \hat{F}_l + \lambda_i \times O_p(N^{-\frac{1}{2}})
\]

where the \( O_p(N^{-\frac{1}{2}}) \) term does not depend of \( i \). Thus

\[
V(k) = \frac{1}{NT} \sum_{t=1}^T \sum_{l=1}^N \hat{c}_{il}^2 + \frac{1}{T^3} \sum_{l=1}^T \hat{F}_l F_k (\frac{1}{N} \sum_{i=1}^N e_i e_i') \hat{F}_l F_k - \frac{2}{T^2} \sum_{l=1}^T \hat{F}_l F_k (\frac{1}{N} \sum_{i=1}^N e_i e_{it}) + O_p(N^{-\frac{1}{2}})
\]

\[
= \frac{\sigma^2}{T} + \frac{\sigma^2}{T} (\frac{\hat{F}_l F_k}{T})^2 - \frac{2\sigma^2}{T^2} \sum_{l=1}^T \hat{F}_l F_k \hat{F}_l + O_p(N^{-\frac{1}{2}})
\]

\[-\rho \sigma^2 + \frac{\sigma^2}{T} - \frac{2\sigma^2}{T} = \sigma^2 (1 - \frac{k}{T})
\]

QED

8.4 Appendix D: Proofs of Theorems

To simplify arguments, we first report the proofs for the static case, that is when \( a_{is} = a_i, \lambda_{is} = \lambda_i, \sigma_i^2 = \sigma^2, r_l = r \), and the illustrate their generalization to the time-varying case only when this adds further complexity to the arguments.

**Proof of Theorem 1** Consider case \( a_i = 0 \), as the proof follows along the same lines when \( a_i \neq 0 \) by Lemmas 1 and 2.

The following identity holds (see Bai (2003), equation (D.1)):

\[
\hat{F}_l - \hat{H} F_l = \hat{U}^{-1} T^{-1} \sum_{s=1}^T \hat{F}_s \zeta_{st} + \hat{U}^{-1} T^{-1} \sum_{s=1}^T \hat{F}_s \eta_{st} + \hat{U}^{-1} T^{-1} \sum_{s=1}^T \hat{F}_s \eta_{st}.
\]

(52)

where, by Assumptions 6 and 7, for every given \( s, t = 1, \cdots, T \),

\[
\zeta_{st} = \frac{e_i' e_t}{N} - \frac{e_i' e_t}{N} = O_p(N^{-\frac{1}{2}}), \quad \eta_{st} = \frac{F_s A_i e_t}{N} = O_p(N^{-\frac{1}{2}}).
\]

(53)
The result follows noting that, by Lemma 8.3, \( \| \tilde{U}^{-1} \| = O_p(1) \), \( \| \tilde{F}F/T \| = O_p(1) \) for \( N \) large enough, and \( \| \tilde{F}F/T \| = O(1) \) by construction.

(ii) By part (i), the first term on the right hand side of (52) satisfies

\[
\tilde{U}^{-1} \frac{\sum_{s=1}^{T} \tilde{F}_s \zeta_{st}}{T} = \tilde{U}^{-1} \frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s \left( \frac{e_{s,t} e_{t}}{N} - \mathcal{E} \left( \frac{e_{s,t}}{N} \right) \right) = \tilde{U}^{-1} \frac{1}{T} \tilde{F}' \left( \frac{e_{s,t}}{N} - \mathcal{E} \left( \frac{e_{s,t}}{N} \right) \right) = \tilde{U}^{-1} \frac{1}{T} \tilde{H}' \tilde{F}' \left( \frac{e_{s,t}}{N} - \mathcal{E} \left( \frac{e_{s,t}}{N} \right) \right) + o_p(1),
\]

where, by Lemma 8.3,

\[
\tilde{H} \rightarrow_p \Sigma \Lambda Q^T U^{-\frac{1}{2}} = \Sigma^\frac{1}{2} \Lambda^\frac{1}{2} Q^T U^{-\frac{1}{2}} = Q^{-1}.
\]

Noticing that

\[
\tilde{H}' \tilde{\Sigma} \tilde{F} \tilde{H} = I_r,
\]

one obtains

\[
\sqrt{N} \tilde{U}^{-1} \frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} \rightarrow_d N(0_r, \sigma_d^2 \frac{U^{-2}}{T} + \frac{(\mu_4 - 2\sigma_4)}{T^2} U^{-1} \tilde{H}' \tilde{F}' \tilde{F} \tilde{H} U^{-1}),
\]

as, by having limited the degree of cross-sectional heterogeneity of the moments of the \( e_{it} \) as formalized in Assumption 5, an identical limiting distribution is obtained when centering with respect to \( E(\frac{e_{i,t}}{N}) \) or to its large-\( N \) limit. The same reasoning applies to all other limiting distributions that follow.

In particular, to derive the asymptotic covariance matrix of \( \sqrt{N} \tilde{U}^{-1} \frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} \), consider that the \((s,v)\)th element of \( \mathcal{E}(e_{i,t} e_{s,t} e'_{j,t} e_{j,t}) = \mathcal{E}(e_{i,t} e_{j,t}) \mathcal{E}(e'_{j,t} e_{j,t}) \), for every \( 1 \leq s, v \leq T \), satisfies

\[
\text{Cov}(e_{i,s}, e_{j,v} e_{j,t}) = \kappa_{4,i,j,j,s,v} e_{i,s} e_{j,v} e_{j,t} + \mathcal{E}(e_{i,s} e_{j,v} e_{j,t}) = \kappa_{4,i,j,j,s,v} \sigma_{i,s} \sigma_{j,v} + \sigma_{i,s} \sigma_{j,v} \sigma_{i,t} \sigma_{j,t},
\]

yielding, in matrix form,

\[
\text{Cov}(e, e, e', e) = \begin{pmatrix} \kappa_{4,i,i,j,j,s,v} & \cdots & \kappa_{4,i,i,j,j,s,v} \\ \kappa_{4,i,i,j,j,s,v} & \cdots & \kappa_{4,i,i,j,j,s,v} \\ \vdots & \cdots & \vdots \\ \kappa_{4,i,i,j,j,s,v} & \cdots & \kappa_{4,i,i,j,j,s,v} \end{pmatrix} + \sigma_{i,j,t} \begin{pmatrix} \sigma_{i,j,t} & \cdots & \sigma_{i,j,t} \\ \sigma_{i,j,t} & \cdots & \sigma_{i,j,t} \\ \vdots & \cdots & \vdots \\ \sigma_{i,j,t} & \cdots & \sigma_{i,j,t} \end{pmatrix} + \begin{pmatrix} \sigma_{i,j,t} \cdots \sigma_{i,j,t} \end{pmatrix}.
\]

Therefore,

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{Cov}(e, e, e', e) = \frac{1}{N} \sum_{i=1}^{N} \left( \kappa_{4,i,i,i,i,t,t,t,t} + \sigma_{i,i,t} \begin{pmatrix} \sigma_{i,i,t} & \cdots & \sigma_{i,i,t} \\ \sigma_{i,i,t} & \cdots & \sigma_{i,i,t} \\ \vdots & \cdots & \vdots \\ \sigma_{i,i,t} & \cdots & \sigma_{i,i,t} \end{pmatrix} + \sigma_{i,i,t} \sigma_{i,i,t} \right) + o_p(1) \rightarrow (\kappa_4 + \sigma_4) I_t t' + \sigma_4 I_T.
\]

For the second term of (52):

\[
\sqrt{N} \tilde{U}^{-1} \frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s \eta_{st} = \tilde{U}^{-1} \frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s \tilde{F}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{i,t} = \tilde{U}^{-1} \tilde{F}' \tilde{F} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{i,t} = \tilde{U}^{-1} \tilde{U}^{-1} \tilde{U}^{-1} N(0_r, \sigma^2 U^{-1} Q \Sigma \Lambda Q^T U^{-1}).
\]
where \( Q \Sigma \lambda' = U \). For the third term of (52):

\[
\sqrt{N} \hat{U}^{-1} \sum_{s=1}^{T} \tilde{F}_s \eta_{ts} = \sqrt{N} \hat{U}^{-1} \sum_{s=1}^{T} \tilde{F}_s \eta_{ts} \sum_{i=1}^{N} \lambda_i e_{is} = \hat{U}^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \tilde{F}_s e_{is} \lambda_i F_t
\]

\[
= \hat{U}^{-1} \sum_{s=1}^{T} \tilde{F}'_s \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i \lambda_j \right) F_t = \hat{U}^{-1} \sum_{s=1}^{T} (F_t \otimes \tilde{F}') \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\lambda_i \otimes e_i)
\]

\[
\rightarrow_d N(0_r, \sigma^2 (\frac{F'_t \Sigma \Lambda F_t}{T}) U^{-2}),
\]

where the asymptotic covariance matrix of \( T^{-1} (F'_t \otimes \tilde{F}') \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\lambda_i \otimes e_i) \) is obtained noticing that

\[
T^{-1} (F'_t \otimes \tilde{F}') (\Sigma \Lambda \otimes \sigma^2 I_T) (F_t \otimes FH) = \sigma^2 (F'_t \Sigma \Lambda F_t) I_T.
\]

Given that the three terms are potentially correlated, we need the joint convergence of Assumption 7, as it emerges by re-writing the right hand side of (52) in matrix form:

\[
\tilde{F}_t - \hat{H}' F_t = T^{-1} \hat{U}^{-1} \left[ \tilde{F}'_t, (F'_t \otimes \tilde{F}') \right] \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} (e_i e_{it} - \mathcal{E}(e_i e_{it})) \\ \lambda_i e_{it} \\ (\lambda_i \otimes e_i) \end{bmatrix}
\]

\[(55) \quad \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} (e_i e_{it} - \mathcal{E}(e_i e_{it})) \\ (\lambda_i \otimes e_i) \end{bmatrix}, \quad (56)
\]

where \( \Lambda_t \) is defined in (26). By Assumption 5 the \( e_{it} \) have zero third-moments, implying that the covariance matrix of \( ((e_i e_{it}) - E(e_i e_{it}))', (\lambda_i \otimes e_i)' \) is block-diagonal. The diagonal elements have been established above. When considering (55), one needs the asymptotic (cross) covariance matrix between \( \lambda_i e_{it} = (\lambda_i \otimes e_{it}) \) and \( (\lambda'_j \otimes e'_{ij}) \), which automatically emerges by pre- and post-multiplying the asymptotic covariance matrix in (25) by \( \Lambda_t \) and \( \Lambda'_t \), respectively, as indicated by (56). However, it can also be derived by direct calculations as follows: by our assumptions

\[
\frac{1}{N} \sum_{i,j=1}^{N} (\lambda_i \lambda_j' \otimes \mathcal{E}(e_{it} e'_{ij})) \rightarrow_p \sigma^2 (\Sigma \Lambda \otimes \Lambda'_t, T),
\]

implying

\[
\frac{1}{T} \hat{U}^{-1} (\tilde{F}' F') \frac{1}{N} \sum_{i,j=1}^{N} (\lambda_i \lambda_j' \otimes \mathcal{E}(e_{it} e'_{ij})) (F_t \otimes \tilde{F}) \hat{U}^{-1} \rightarrow_p \frac{\sigma^2}{T} U^{-1} Q(\Sigma \Lambda \otimes \Lambda'_t) (F_t \otimes FH) U^{-1} = \frac{\sigma^2}{T} U^{-1} \Sigma \Lambda F_t F'_t H U^{-1}.
\]

Let us now extend the proof to the time-varying case. We first show that

\[
\frac{1}{N} \sum_{s=1}^{T_0} \left[ (x_s - \Lambda_t F_s)' (x_s - \Lambda_t F_s) - (x_s - \Lambda_t F_s)' (x_s - \Lambda_t F_s) \right] w_s(t) = o_p(1).
\]

This implies that in terms of loadings we are in fact estimating \( \Lambda_t \), thus constant within the interval \( T_t \). In fact,

\[
\left[ (x_s - \Lambda_t F_s)' (x_s - \Lambda_t F_s) - (x_s - \Lambda_t F_s)' (x_s - \Lambda_t F_s) \right]
\]

\[
= \left[ -F'_s \Lambda'_t x_s + F'_s \Lambda'_t x_s - x'_s \Lambda'_t F_s + x'_s \Lambda'_t F_s + F'_s \Lambda'_t F_s - F'_s \Lambda'_t F_s \right]
\]

\[
= \left[ -F'_s (\Lambda_t - \Lambda_s)' x_s - x'_s (\Lambda_t - \Lambda_s) F_s + F'_s (\Lambda_t - \Lambda_s)' \Lambda_s F_s + F'_s (\Lambda_t - \Lambda_s)' (\Lambda_t - \Lambda_s) F_s + F'_s (\Lambda_t - \Lambda_s)' (\Lambda_t - \Lambda_s) F_s \right].
\]
By the mean value theorem \((\Lambda_t - \Lambda_s) = \Lambda_{s^*}^{(1)}(t - s)\) and, by summing across \(i\), one gets the \(o_p(N)\) rate.

We now show how, within the interval \(T_t\), we can replace \(\sigma^2, \sigma_4, \kappa_4\) by \(\sigma^2_{t-1}, \sigma_{4t-1}, \kappa_{4t-1}\) throughout our results. To simplify the exposition, it is convenient to assume that the set \(T_t\) equals \(\{1, \cdots, T\}\) (instead of \(\{t - T + 1, \cdots, t\}\)) without loss of generality. Starting with Lemma \(\mathbf{3}\) we just need to consider \(ee'/N - \sigma^2_{t-1}I_T\) where now

\[
\frac{ee'}{N} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \sigma_{ii,11} & 0 & \cdots & 0 \\ 0 & \sigma_{ii,22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \sigma_{ii,TT} \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} \sigma_{ii,11} & 0 & \cdots & 0 \\ 0 & \sigma_{ii,22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \sigma_{ii,TT} \end{bmatrix}^{\frac{1}{2}} \eta_i \eta_i' + o_p(1)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \sigma_{ii,tt} + \sigma_{ii,1^*}(1 - t) & 0 & \cdots & 0 \\ 0 & \sigma_{ii,tt} + \sigma_{ii,2^*}(2 - t) & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \sigma_{ii,tt} + \sigma_{ii,T^*}(T - t) \end{bmatrix}^{\frac{1}{2}} \eta_i \eta_i' + o_p(1)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \sigma_{ii,tt} + \sigma_{ii,1^*}(1 - t) & 0 & \cdots & 0 \\ 0 & \sigma_{ii,tt} + \sigma_{ii,2^*}(2 - t) & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \sigma_{ii,tt} + \sigma_{ii,T^*}(T - t) \end{bmatrix}^{\frac{1}{2}} + o_p(1)
\]

Let us first generalize Theorem \(\mathbf{[3]}\) Regarding the first term of the corresponding asymptotic covariance matrix:

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{Cov}(e_i e_i, e'_j e_j) = \frac{1}{N} \sum_{i=1}^{N} \left( \kappa_{4,iitiisstttt} l_s t'_s + \sigma_{ii,ss} \begin{bmatrix} \sigma_{ii,11} & 0 & \cdots & 0 \\ 0 & \sigma_{ii,22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \sigma_{ii,TT} \end{bmatrix} \right) + \sigma^2_{ii,ss} l_s t'_s + o_p(1)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \left( \left( \kappa_{4,iitiisstttt} + \kappa_{4,iitiisstttt}^{(1)}(s - t) \right) l_s t'_s + \left( \sigma_{ii,tt} + \sigma_{ii,1^*}^{(1)}(t - s) \right) \right) \times \begin{bmatrix} \sigma_{ii,tt} + \sigma_{ii,1^*}(1 - t) & \cdots & 0 \\ 0 & \sigma_{ii,tt} + \sigma_{ii,2^*}(2 - t) & \cdots \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \sigma_{ii,tt} + \sigma_{ii,T^*}(T - t) \end{bmatrix} + \sigma^2_{ii,ss} l_s t'_s + o_p(1)
\]

\[
\rightarrow \left( \kappa_{4t-1} + \kappa_{4t-1} \right) l_s t'_s + \sigma_{4t-1} l_T + o_p(1)
\]
Regarding the second term of the asymptotic covariance matrix:

\[
\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} (\lambda_{it-1} \lambda'_{jt-1} \otimes \mathcal{E}(e_i e'_j)) = \frac{1}{N} \sum_{i=1}^{N} (\lambda_{it-1} \lambda'_{it-1} \otimes \mathcal{E}(e_i e'_i)) + o_p(1)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \left( \lambda_{it-1} \lambda'_{it-1} \otimes \begin{bmatrix}
\sigma_{ii,11} & 0 & \cdots \\
0 & \sigma_{ii,22} & \cdots \\
\cdots & \cdots & \cdots \\
0 & \sigma_{ii,TT} & \cdots \\
\end{bmatrix} \right) + o_p(1)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \left( \lambda_{it-1} \lambda'_{it-1} \otimes \begin{bmatrix}
0 & \sigma_{ii,tt} + \sigma_{ii,1} (1 - t) & \cdots \\
\cdots & \cdots & \cdots \\
0 & \sigma_{ii,tt} + \sigma_{ii,TT} (T - t) & \cdots \\
\end{bmatrix} \right) + o_p(1)
\]

\[
= (\Sigma_{t-1} \otimes \sigma^2_{t-1} I_T) + o_p(1).
\]

Finally, regarding the non-zero covariance term:

\[
\frac{1}{N} \sum_{i,j=1}^{N} (\lambda_{it-1} \lambda'_{jt-1} \otimes \mathcal{E}(e_i e'_j)) = \sigma^2_{t-1} (\Sigma_{t-1} \otimes \ell'_s).
\]

QED

**Proof of Theorem** 2. The result immediately follows from Lemma 4 noting that \(F'_t \Sigma A Q' = F'_t H H^{-1} \Sigma A H^{-1}\) and \(F'_t \Sigma A F'_t = F'_t H H^{-1} \Sigma A H^{-1} H' F_t\).

The time-varying case requires to generalize Lemma 4. Then, part (i) of Lemma 2 follows immediately by the above arguments. For part (ii), one needs to consider terms such as

\[
\frac{1}{NT} \sum_{h=1}^{T} \sum_{i=1}^{N} \mathcal{E} \left( \sum_{s=1}^{T} e_{is} c_{ish} \right)^4 = \frac{1}{NT} \sum_{h=1}^{T} \sum_{i=1}^{N} \sum_{s_1,s_2,s_3,s_4=1}^{T} \kappa_{4,isi,s_1,s_2,s_3,s_4} c_{i1} c_{i2} c_{i3} c_{i4} c_{is_1} c_{is_2} c_{is_3} c_{is_4}
\]

\[
+ \frac{1}{NT} \sum_{h=1}^{T} \sum_{i=1}^{N} \sum_{s_1,s_2,s_3,s_4=1}^{T} (\sigma_{ii,s_1 s_2 s_3 s_4} + \sigma_{ii,s_1 s_3 s_2 s_4} + \sigma_{ii,s_1 s_4 s_2 s_3}) c_{i1} c_{i2} c_{i3} c_{i4}
\]

\[
= \frac{1}{NT} \sum_{h=1}^{T} \sum_{i=1}^{N} \left( \sum_{s_1,s_2=1}^{T} \kappa_{4,isi,s_1,s_2} c_{i1}^4 + \sum_{s_1,s_2=1}^{T} 3\sigma_{ii,s_1 s_2} c_{i1}^2 c_{i2}^2 c_{i3} c_{i4} \right) + o(1)
\]

\[
= \frac{1}{NT} \sum_{h=1}^{T} \sum_{i=1}^{N} \left( \sum_{s_1,s_2=1}^{T} (\kappa_{4,isi,tt} + \kappa_{4,isi,s} (s-t)) c_{i1}^4 + \sum_{s_1,s_2=1}^{T} 3((\sigma_{ii,tt} + (\sigma_{ii,s_1} (s-t)) c_{i1}^2 c_{i2}^2 c_{i3} c_{i4}^2 \right) + o(1).
\]

The remainder four terms of part (ii) follow along the same lines. Finally, by the same arguments, part (iii) easily follow yielding

\[
\frac{1}{N} \sum_{i=1}^{N} \lambda_{it-1} \lambda'_{it-1} \rightarrow_p H_{t-1}^{-1} \Sigma A_{t-1} H_{t-1}^{-1} + \frac{\sigma^2_{t-1}}{T} I_{t-1}.
\]

The proof is concluded setting \(\kappa_{4t-1} = 0\). QED
Proof of Theorem 4. To show that $\text{Prob}(PC(k) < PC(r)) \to 0$ for any $k \neq r$, consider at first case $r < k$, where $PC(k) - PC(r) = V^*(k) - V^*(r) - (r-k)g(N)$. Then

$$\text{Prob}(V^*(k) - V^*(r) < (r-k)g(N)) \to 0,$$

because $g(N) \downarrow 0$ whereas $V^*(k) - V^*(r)$ has a strictly (a.s.) positive limit. Consider now case $k > r$. Then

$$\text{Prob}(PC(k) < PC(r)) = \text{Prob}(V^*(r) - V^*(k) > (k-r)g(N)) \to 0,$$

because $V^*(r) - V^*(k) = O_p(N^{-\frac{1}{2}})$ whereas $N^{-\frac{1}{2}}/g(N) = o(1)$.

In the time-varying case one needs to evaluate the limit of terms such as $N^{-1}\sum_{i=1}^Ne_ie_i$, where, to derive the corresponding asymptotic covariance matrix, one needs to evaluate, for every $1 \leq s,v \leq T$, $\text{Cov}(\tilde{e}_{is},\tilde{e}_{it},\tilde{e}_{js},\tilde{e}_{jt})$. Then

$$\text{Prob}(\tilde{V}^*(t) - \tilde{V}^*(r) < (r-t)g(N)) \to 0,$$

because $\tilde{V}^*(r) - \tilde{V}^*(k) = O_p(N^{-\frac{1}{2}})$ whereas $N^{-\frac{1}{2}}/g(N) = o(1)$.

Proof of Theorem 4. Consider case $a_i = 0$, as the proof follows along the same lines when $a_i \neq 0$ by Lemmas 1 and 2.

(ii) Consistency of $\tilde{\gamma}$ follows by Slutzky theorem and Theorem 1 as:

$$\tilde{\gamma} = \frac{1}{T}\sum_{t=1}^T \tilde{F}_t \to_p \frac{1}{T}\sum_{t=1}^T H'F_t = H'\gamma^P.$$

(iii) To derive the limiting distribution of the risk premia estimator, by the same steps adopted for the proof of Theorem 1.

$$\tilde{\gamma} - H'\gamma^P = \frac{\tilde{F}'1_T}{T} - \frac{H'F'1_T}{T} = \tilde{U}^{-1}\frac{\tilde{F}'(ee' - \sigma^2I_T)}{T} + \tilde{U}^{-1}\frac{\tilde{F}'A'e'1_T}{NT} + \tilde{U}^{-1}\frac{\tilde{F}'eA}{T}. $$

Regarding the first term on the right hand side:

$$\sqrt{N}\tilde{U}^{-1}\frac{\tilde{F}'(ee' - \sigma^2I_T)}{T} \to_d N(0, \frac{1}{T^2}\tilde{U}^{-1}H'(\kappa_4 + \sigma_4\tilde{\Sigma}_F + \sigma_4\tilde{F}\tilde{F}'\tilde{F}^T)\tilde{H}U^{-1}),$$

where, to derive the corresponding asymptotic covariance matrix, one needs to evaluate, for every $1 \leq s,v \leq T$,

$$\text{Cov}(\tilde{e}_{is},\tilde{e}_{it},\tilde{e}_{js},\tilde{e}_{jt}) = \kappa_{4,iijj,svtr} + \mathcal{E}(\tilde{e}_{is}\tilde{e}_{ij})\mathcal{E}(\tilde{e}_{it}\tilde{e}_{jr}) + \mathcal{E}(\tilde{e}_{is}\tilde{e}_{jr})\mathcal{E}(\tilde{e}_{it}\tilde{e}_{jr}) = \kappa_{4,iijj,svtr} + \sigma_{ij,sv}\sigma_{ij,tr} + \sigma_{ij,sv}\sigma_{ij,tv},$$

yielding, in matrix form,

$$\text{Cov}(\tilde{e}_{is},\tilde{e}_{it},\tilde{e}_{js},\tilde{e}_{jt}) = \begin{pmatrix}
\kappa_{4,iijj,11T} & \cdots & \kappa_{4,iijj,1TT} \\
\kappa_{4,iijj,21T} & \cdots & \kappa_{4,iijj,2TT} \\
\vdots & \vdots & \vdots \\
\kappa_{4,iijj,T1T} & \cdots & \kappa_{4,iijj,TTT}
\end{pmatrix} + \sigma_{ij,tr} \begin{pmatrix}
\sigma_{ij,11} & \cdots & \sigma_{ij,1T} \\
\sigma_{ij,21} & \cdots & \sigma_{ij,2T} \\
\vdots & \vdots & \vdots \\
\sigma_{ij,T1} & \cdots & \sigma_{ij,TT}
\end{pmatrix} + \begin{pmatrix}
\sigma_{ij,11} \\
\sigma_{ij,21} \\
\vdots \\
\sigma_{ij,T1}
\end{pmatrix} \begin{pmatrix}
\sigma_{ij,1r} & \cdots & \sigma_{ij,Tr}
\end{pmatrix}.
Therefore,
\[
T^{-2} \sum_{t=1}^{T} \sum_{r=1}^{T} \left( \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{Cov}(e_i e_{it}, e_j e_{jr}) \right) = T^{-2} \sum_{t=1}^{T} \sum_{r=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \left( \kappa_i \sum_{i=1}^{T} t^t e_{it} + \sigma_{ii,tt} \right) + \sigma_{ii,tt} t^t e_{it} + o_p(1)
\]

\[
\rightarrow \frac{(\kappa_i + \sigma_i)}{T} I_T + \sigma_i \frac{T}{T^2}. \]

For the second term:
\[
\sqrt{N} \hat{U}^{-1} T \hat{F}^T \hat{A} \hat{e}^T 1_T \rightarrow d N(0_r, \frac{\sigma^2}{T} U^{-1} Q \Sigma_A Q^T U^{-1}),
\]

as
\[
\frac{1}{NT^2} \sum_{i=1}^{T} \sum_{j=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \text{Cov}(1_T e_i, e_j 1_T) = \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{r=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \text{Cov}(e_{it}, e_{jr}) \]
\[
= \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{r=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \sigma_{ii,tt} + o_p(1) \rightarrow \frac{\sigma^2}{T} \Sigma_A,
\]

and for the third term:
\[
\sqrt{N} \hat{U}^{-1} T \hat{F}^T \frac{e \Lambda}{N} \hat{F} = \hat{U}^{-1}(\hat{F}^T \otimes \frac{e \Lambda}{T}) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\lambda_i \otimes e_i).
\]

It remains to obtain the covariance terms, with the only non-zero term being
\[
\frac{1}{T} \sum_{t=1}^{T} \hat{U}^{-1} T \hat{F}^T \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} (\lambda_i \lambda_j \otimes \mathcal{E}(e_{it} e_{jt}))(\hat{F}^T \otimes \frac{e \Lambda}{T}) \hat{U}^{-1} \rightarrow \frac{\sigma^2}{T} U^{-1} Q \Sigma_A \hat{F}^T \hat{U}^{-1}.
\]

Finally, the expression for the asymptotic covariance matrix of \( \hat{\gamma} \) simplifies recalling the identities \( Q \Sigma_A Q^T = U \) and \( H^T \Sigma_F H' = I_r \). The above four quantities make the matrix \( C_T \).

(iv) Consistency of \( \hat{C}_T \) easily follows along the lines of the proof to Theorem 2, recalling the identity \( QH = I_r \), replacing \( H' \hat{F} \) by \( \hat{F} \) and recalling that \( \Sigma_A \rightarrow_h H^{-1} \Sigma_A H^{-1} \). QED

**Proof of Theorem 5** Define the following function of an arbitrary, full column rank, \( T \times r \) matrix \( G \)
\[
m(G; t) = \frac{1}{rf} - \frac{1}{rf} \left( \frac{1}{T} G \left( \frac{G' M_1 G}{T} \right)^{-1} (G' (t - \frac{1}{T}) \right).
\]

Then \( m_{t-1, t} = m(F; t) \) and \( m_{r-1, t} = m(F; t) \). Notice that \( m_{t-1, t} = m(FC; t) \) for any non-singular \( r \times r \) matrix \( C \), including the special case \( C = H \).
8.4 Appendix D: Proofs of Theorems

Given that \( T \) is fixed, we can view \( \text{vec}(F') \) as a vector of parameters, and use the delta method, in particular expanding \( m(\tilde{F}; t) \) around \( m(\tilde{H}'F'; t) \), to derive the limiting distribution of \( \tilde{m}_t \). Considering the first-order differential of \( m(\tilde{G}; t) \) with respect to the arbitrary argument (matrix) \( G \):

\[
dm(\tilde{G}; t) = -\frac{1}{r_f} \left( \frac{1}{T} \right) dG \left( \frac{G'M_{1T}G}{T} \right)^{-1} (G'(t_t - \frac{1}{T})) + \frac{1}{T} \left( \frac{1}{T} \right) G \left( \frac{G'M_{1T}G}{T} \right)^{-1} (\frac{1}{T} G'(t_t - \frac{1}{T}))
\]

implies

\[
\frac{\partial m(\tilde{F}; t)}{\partial \text{vec}G'} = \frac{1}{r_f} \left( -((\frac{1}{T}) \otimes I_r) \tilde{G}'^{-1} \tilde{F}'(t_t - \frac{1}{T})) - ((t_t - \frac{1}{T}) \otimes I_r) \tilde{G}'^{-1} \tilde{F}' \left( \frac{1}{T} \right) \right) + (M_{1T} \tilde{F} \otimes I_r) (\tilde{G}'^{-1} \tilde{F}'(t_t - \frac{1}{T}) \otimes \tilde{G}'^{-1} \tilde{F}' \left( \frac{1}{T} \right) \right)
\]

satisfying

\[
\frac{\partial m(\tilde{F}; t)}{\partial \text{vec}G'} \rightarrow_p (I_T \otimes H^{-1}) \frac{\partial m(\tilde{F}; t)}{\partial \text{vec}G'}.
\]

We now evaluate the asymptotic covariance matrix of \( \tilde{F}' - \tilde{H}'F' \), where

\[
\tilde{F}' - \tilde{H}'F' = \tilde{U}^{-1} \tilde{F}' \left( \frac{e'e'}{N} - \sigma^2 I_T \right) + \tilde{U}^{-1} \tilde{F}' F' \Lambda e' \left( \frac{1}{N} \right) + \tilde{U}^{-1} \tilde{F}' eA \left( \frac{1}{N} \right) F'.
\]

implying

\[
\text{vec}(\tilde{F}' - \tilde{H}'F') = \text{vec}(\tilde{U}^{-1} \tilde{F}' \left( \frac{e'e'}{N} - \sigma^2 I_T \right) + \tilde{U}^{-1} \tilde{F}' F' \Lambda e' \left( \frac{1}{N} \right) + \tilde{U}^{-1} \tilde{F}' eA \left( \frac{1}{N} \right) F')
\]

\[
= (I_T \otimes \tilde{U}^{-1} \tilde{F}' \left( \frac{1}{N} \right) \sum_{i=1}^{N} ((e_i \otimes e_i) - \sigma^2 \text{vec}(I_T)) + (I_T \otimes \tilde{U}^{-1} \tilde{F}' \left( \frac{1}{N} \right) \sum_{i=1}^{N} (e_i \otimes \lambda_i) + (F \otimes \tilde{U}^{-1} \tilde{F}' \left( \frac{1}{N} \right) \sum_{i=1}^{N} (\lambda_i \otimes e_i).}
\]
Then

\[
\text{Var}(\sqrt{N} \text{vec}(\bar{F}' - \bar{H}'F')) = \\
+ (I_T \otimes U^{-1} \frac{F'F}{T}) \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}[(e_i \otimes \lambda_i)(e_i' \otimes \lambda_i')((I_T \otimes \frac{F'F}{T}) \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}[(\lambda_i \otimes e_i)(\lambda_i' \otimes e_i')])((I_T \otimes \frac{F'F}{T}) \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}[(\lambda_i \otimes e_i)(\lambda_i' \otimes e_i')])]
\]

Re-arranging terms, one obtains

\[
\text{Var}(\sqrt{N} \text{vec}(\bar{F}' - \bar{H}'F')) \rightarrow \\
= (I_T \otimes U^{-1} \frac{H'H'}{T}) \mu_{\alpha}(I_T \otimes \frac{F'H'}{T} U^{-1}) + (\sigma^2 \mu_{\alpha} \otimes U^{-1}) + (F \Sigma_{\alpha} F' \otimes \frac{\sigma^2}{T} U^{-2})
\]

using the property \( \kappa_{\alpha\alpha}(A \otimes B) = (B \otimes A) \kappa_{\alpha\alpha} \) for every \( m \times n \) matrix \( A \) and \( p \times q \) matrix \( B \), and \( \kappa'_{ab} = \kappa^{-1}_{ba} = \kappa_{ba} \) (see [Magnus & Neudecker 2007][Chapter 3, Section 7, Theorem 9]), and using the identities

\( H' \Sigma_{\alpha} H = I_r \), \( U^{-1} H^{-1} \Sigma_{\alpha} H^{-1} U^{-1} = U^{-1} \), in particular in second variance term above, and \( U^{-1} H' \Sigma_{\alpha} H = U^{-1} H' \Sigma_{\alpha} H H^{-1} \Sigma_{\alpha} = U^{-1} H' \Sigma_{\alpha} H H^{-1} \Sigma_{\alpha} = U^{-1} H' \Sigma_{\alpha} H H^{-1} \Sigma_{\alpha} = H' \) in last two covariance terms. QED

**Proof of Theorem 6** For part (i), it is convenient to re-write \( F_{mp}^{\alpha} \) as

\[
F_{mp}^{\alpha} = (\gamma, \Omega) \left( \begin{array}{c} a' + \gamma \Lambda' \\ \Lambda' \end{array} \right) \left\{ (a + \Lambda \gamma), \Lambda \right\} \left( \begin{array}{cc} 1 & 0' \\ 0 & \Omega \end{array} \right) \left[ a' + \gamma \Lambda' \right] + \Sigma_{\rho}^{-1} x_t.
\]

\[
=(\gamma, \Omega) \Lambda' (\Lambda' + \Sigma_{\rho})^{-1} x_t.
\]

\[
=(\gamma, \Omega) \Lambda' (\Lambda' + \Sigma_{\rho})^{-1} \left( \Lambda' (f_t - \mathcal{E}(f_t)) + e_t \right),
\]

setting

\[
\Lambda_t \equiv (a + \Lambda \gamma, \Lambda), \quad \Omega_t \equiv \left( \begin{array}{cc} 1 & 0' \\ 0 & \Omega \end{array} \right).
\]
Then
\[ F'^{mp}_t = (\gamma, \Omega) \left( \mathbf{I}_{t+1} - \Lambda'_e \Sigma^{-1}_e \Lambda'_{\mathbf{I}_{t+1}} (\Omega^{-1}_t + \Lambda'_e \Sigma^{-1}_e \Lambda'_{\mathbf{I}_{t+1}})^{-1} \right) \Lambda'_e \Sigma^{-1}_e \Lambda'_{\mathbf{I}_{t+1}} \left( \mathbf{f}_t - \mathbf{f}(\mathbf{f}_t) \right) + \mathbf{e}_t \]

\[ = (\gamma, \Omega) \Omega^{-1}_t \left( 1 + \frac{a' + \gamma' \Lambda'}{\Lambda'_e} (\mathbf{a} + \Lambda \gamma) \right) \frac{(a' + \gamma' \Lambda') \Sigma^{-1}_e \Lambda'}{\Omega^{-1} + \Lambda'_e \Sigma^{-1}_e \Lambda'}^{-1} \Lambda'_e \Sigma^{-1}_e \Lambda'_{\mathbf{I}_{t+1}} \left( \mathbf{f}_t - \mathbf{E}(\mathbf{f}_t) \right) + \mathbf{e}_t \]

\[ = I + II. \]

Consider first term. Setting for simplicity \( a \equiv 1 + (a' + \gamma' \Lambda') \Sigma^{-1}_e (\mathbf{a} + \Lambda \gamma) \), \( b \equiv \Lambda' \Sigma^{-1}_e (\mathbf{a} + \Lambda \gamma) \) and \( B \equiv \Omega^{-1} + \Lambda' \Sigma^{-1}_e \Lambda' \), and using the block-wise formula for the inverse of a matrix, gives:

\[ \left( \frac{1}{\Lambda'_e} (\mathbf{a} + \Lambda \gamma) \right)^{-1} = \left( \begin{array}{cc} a^{-1} + a^{-2} b' (B - \frac{bb'}{a})^{-1} b & \frac{b'}{a} (B - \frac{bb'}{a})^{-1} \\ (-B + \frac{bb'}{a})^{-1} b & (B - \frac{bb'}{a})^{-1} \end{array} \right) \]

Post multiplying the latter expression by \( \Lambda'_e \Sigma^{-1}_e \Lambda' = \left( \begin{array}{c} a^{-1} \ b' \\ b \ B - \Omega^{-1} \end{array} \right) \) yields

\[ \left( \begin{array}{cc} a^{-1} + a^{-2} b' (B - \frac{bb'}{a})^{-1} b & \frac{b'}{a} (B - \frac{bb'}{a})^{-1} \\ (-B + \frac{bb'}{a})^{-1} b & (B - \frac{bb'}{a})^{-1} \end{array} \right) \left( \begin{array}{cc} a^{-1} & b' \\ b & n^{-1} \end{array} \right) = \]

\[ \left( \begin{array}{cc} a^{-1} - \frac{a^{-2} b' (B - \frac{bb'}{a})^{-1} b}{(B - \frac{bb'}{a})^{-1}} & b' \\ (-B + \frac{bb'}{a})^{-1} b & n^{-1} \end{array} \right) \left( \begin{array}{cc} a^{-1} & \frac{b'}{a} (B - \frac{bb'}{a})^{-1} \\ b & (B - \frac{bb'}{a})^{-1} \end{array} \right) \]

We now evaluate the behaviour of \( (B - \frac{bb'}{a})^{-1} \) and \( b/a \) as \( N \to \infty \). One needs first to express \( B - \frac{bb'}{a} \) suitably, as:

\[ B - \frac{bb'}{a} = \Omega^{-1} + \Lambda' \Sigma^{-1}_e \Lambda' \frac{(a + \Lambda \gamma) (a + \Lambda \gamma)' \Sigma^{-1}_e \Lambda'}{(1 + (a + \Lambda \gamma)' \Sigma^{-1}_e (a + \Lambda \gamma))} \]

\[ = \Omega^{-1} + \Lambda' \Sigma^{-1}_e \Lambda' [I_N - \frac{\Sigma^{-1/2}_e (a + \Lambda \gamma) (a + \Lambda \gamma)' \Sigma^{-1}_e \Lambda'}{(1 + (a + \Lambda \gamma)' \Sigma^{-1}_e (a + \Lambda \gamma))}] \]

\[ = \Omega^{-1} + \Lambda' \Sigma^{-1}_e \Lambda' [I_N + \Sigma^{-1/2}_e (a + \Lambda \gamma) (a + \Lambda \gamma)' \Sigma^{-1}_e \Lambda'] \]

given the identity

\[ [I_N - \frac{\Sigma^{-1/2}_e (a + \Lambda \gamma) (a + \Lambda \gamma)' \Sigma^{-1}/2_e}{(1 + (a + \Lambda \gamma)' \Sigma^{-1}_e (a + \Lambda \gamma))}] = [I_N + \Sigma^{-1/2}_e (a + \Lambda \gamma) (a + \Lambda \gamma)' \Sigma^{-1}_e]. \]

Setting the non-singular matrix, where non-singularity holds for \( N \) large enough as the last two terms of \( C \) only affect its maximum eigenvalue,

\[ C \equiv \Sigma^{-1}_e + aa' + \Lambda \gamma a' + a \gamma' \Lambda'. \]
and using it within the inverse of the last expression above, gives

\[
(B - \frac{bb'}{a})^{-1} = \\
\Omega - \Omega \Lambda' \Sigma_e^{-1/2} (I_N + \Sigma_e^{-1/2} (a + \Lambda \gamma)(a + \Lambda \gamma)') \Sigma_e^{-1/2} + \Sigma_e^{-1/2} \Lambda \Omega \Lambda' \Sigma_e^{-1/2} \Sigma_e^{-1/2} \Lambda \Omega
\]

\[
= \Omega - \Omega \Lambda' (\Sigma_e^{-1} + (a + \Lambda \gamma)(a + \Lambda \gamma)' + \Lambda \Omega \Lambda')^{-1} \Lambda \Omega
\]

\[
= \Omega - \Omega (\Lambda' C + \Lambda (\gamma \gamma' + \Omega) \Lambda')^{-1} \Lambda \Omega
\]

\[
= \Omega - \Omega (\Lambda' C^{-1} \Lambda)((\gamma \gamma' + \Omega)^{-1} + \Lambda' C^{-1} \Lambda)^{-1} (\gamma \gamma' + \Omega)^{-1} \Omega
\]

implying, as \(N \to \infty\), in particular by the divergence of \(\Lambda' C^{-1} \Lambda\),

\[
(B - \frac{bb'}{a})^{-1} \to_p \Omega - \Omega (\gamma \gamma' + \Omega)^{-1} \Omega = \frac{\gamma \gamma'}{(1 + \gamma \Omega^{-1} \gamma)}.
\]

In fact, divergence of \(\Lambda' C^{-1} \Lambda\) follows as it is bounded from below by \(\Lambda' A g_N (C^{-1})\). In turn \(g_N (C^{-1}) = g_1^{-1}(C)\) and thus one needs the maximum eigenvalue of \(C\) to not diverge too fast. However, the maximum eigenvalue of \(C\) diverges at most as \(a' \Lambda \gamma = O_p (N^{-\frac{1}{2}})\), given that \(a' a = O(1)\) by no arbitrage and \(\Sigma_e\) has minimum eigenvalue bounded away from zero, whereas \(\Lambda' \Lambda = O_p (N)\). It follows that

\[
b' - \frac{\Lambda' \Sigma_e^{-1}(a + \Lambda \gamma)}{\Omega^{-1} + \Lambda' \Sigma_e^{-1} \Lambda} \to_p D \gamma, \gamma / D \gamma.
\]

setting \(D\) to be probability limit of \(\Lambda' \Sigma_e^{-1} A / N\). Therefore, putting terms together,

\[
1 + (a' + \gamma \Lambda') \Sigma_e^{-1}(a + \Lambda \gamma) / \Omega^{-1} + (a' + \gamma \Lambda') \Sigma_e^{-1} \Lambda = \left( \begin{array}{cc} (a-1) & -\frac{1}{a} b' (B - bb')^{-1} b \\ bb' (B - bb')^{-1} b - \frac{1}{a} & I_r - (B - bb')^{-1} \Omega^{-1} \end{array} \right)
\]

\[
\to_p \left( \begin{array}{cc} \frac{\gamma \Omega^{-1} \gamma}{(1 + \gamma \Omega^{-1} \gamma)} & \frac{\gamma \Omega^{-1} \gamma}{(1 + \gamma \Omega^{-1} \gamma)} \\ \frac{\gamma \Omega^{-1} \gamma}{(1 + \gamma \Omega^{-1} \gamma)} & I_r - \frac{\gamma \Omega^{-1} \gamma}{(1 + \gamma \Omega^{-1} \gamma)} \end{array} \right)
\]

It follows that, as \(N\) diverges, term \(I\) satisfies:

\[
I \to_p (\gamma, I_r) \left( \begin{array}{cc} 1 & 1 \\ \frac{1}{(1 + \gamma \Omega^{-1} \gamma)} & I_r - \frac{1}{1 + \gamma \Omega^{-1} \gamma} \end{array} \right) (f_t - \mathcal{E}(f_t)) = (\gamma, I_r) \left( \begin{array}{c} 1 \\ f_t - \mathcal{E}(f_t) \end{array} \right) = F_t.
\]

Next, considering the variance of term II (conditioning on the \(\Lambda\)), we will show that it converges to zero as \(N\) diverges. In fact,

\[
\text{var} \left( \left( \gamma, \Omega \right) \left( I_r + 1 - (\Lambda' \Sigma_e^{-1} \Lambda_r) (\Omega^{-1} + (\Lambda' \Sigma_e^{-1} \Lambda_r)^{-1}) (\Lambda' \Sigma_e^{-1} \eta_t) \right) \right)
\]

\[
= (\gamma, I_r) \left( \begin{array}{cc} a & b' \\ b & B \end{array} \right)^{-1} (a - 1) \left( \begin{array}{cc} a & b' \\ b & B - \Omega^{-1} \right) (\gamma, I_r)' \left( \begin{array}{cc} a & b' \\ b & B \end{array} \right)^{-1}
\]

\[
\to_p (\gamma, I_r) \left( \begin{array}{cc} \frac{1}{(1 + \gamma \Omega^{-1} \gamma)} & \frac{1}{(1 + \gamma \Omega^{-1} \gamma)} \\ \frac{1}{(1 + \gamma \Omega^{-1} \gamma)} & I_r - \frac{1}{(1 + \gamma \Omega^{-1} \gamma)} \end{array} \right) (\gamma, I_r)' = 0_{r \times r},
\]

using

\[
\left( \begin{array}{cc} a & b' \\ b & B \end{array} \right)^{-1} \to_p \frac{1}{(1 + \gamma \Omega^{-1} \gamma)} \left( \begin{array}{cc} 1 & -\gamma' \\ -\gamma & \gamma \gamma' \end{array} \right).
\]

Therefore \(II \to_p 0_r\) establishing part (i).
Part (ii) follows from the identity
\[ \tilde{F}^{mp} - \tilde{F} = -\sigma^2(\tilde{a}^2 I_r + \tilde{A}'\tilde{A})^{-1}\tilde{F}, \]
and recalling that, under our assumptions, \( \tilde{A}'\tilde{A} = O_p(N^{-1}) \). QED

**Proof of Theorem** 
Consider at first cases (ii), (iii) and (iv) for the static model, where \( \tilde{a}_T = \tilde{a}1_T \) as \( a_{it} = a_i \) with \( \tilde{a} \equiv a'1_T/N \). By simple algebraic steps:

\[
(\frac{\tilde{\gamma}_0}{\tilde{\mu}_\Lambda}) - (\frac{\gamma_0 + \tilde{a}}{H^{-1}\tilde{\mu}_\Lambda}) = (\tilde{D}'\tilde{D})^{-1}\tilde{D}'\tilde{x}_T - (\frac{\gamma_0 + \tilde{a}}{H^{-1}\tilde{\mu}_\Lambda})
\]

\[
= (\tilde{D}'\tilde{D})^{-1}\tilde{D}'(\frac{\gamma_0 + \tilde{a}}{\mu_\Lambda} + \tilde{e}_T) - (\frac{\gamma_0 + \tilde{a}}{H^{-1}\tilde{\mu}_\Lambda})
\]

\[
= (\tilde{D}'\tilde{D})^{-1}\tilde{D}'(\frac{\gamma_0 + \tilde{a}}{\mu_\Lambda}) - (\frac{\gamma_0 + \tilde{a}}{H^{-1}\tilde{\mu}_\Lambda})
\]

\[
= (\tilde{D}'\tilde{D})^{-1}\tilde{D}'(\frac{1}{0_r} - \tilde{H}) - (\frac{\gamma_0 + \tilde{a}}{H^{-1}\tilde{\mu}_\Lambda})
\]

\[
= (\tilde{D}'\tilde{D})^{-1}\tilde{D}' + (\tilde{D}'\tilde{D})^{-1}(\tilde{D}'\tilde{D})^{-1}\tilde{D}'(\frac{1}{0_r} - \tilde{H}) - (\frac{\gamma_0 + \tilde{a}}{H^{-1}\tilde{\mu}_\Lambda}) = I + II.
\]

Consider now

\[
\tilde{D}'\tilde{D}^{-1}\tilde{D}' = (0, 0) - (0, 0) = 0,
\]

implying that term \( II \) satisfies

\[
(\tilde{D}'\tilde{D})^{-1}\tilde{D}'(\frac{\gamma_0 + \tilde{a}}{\mu_\Lambda}) = (\tilde{D}'\tilde{D})^{-1}\tilde{D}'(\tilde{F} - \tilde{F}^*)\tilde{H}^{-1}\tilde{\mu}_\Lambda
\]

Asymptotic normality of \( \sqrt{N}\text{vec}(\tilde{F}' - \tilde{H}'\tilde{F}') \), and the corresponding asymptotic covariance matrix, follows from the proof to Theorem 5 whereas asymptotic normality of \( \sqrt{N}\tilde{e}_T \) follows from our assumptions. Thus, to derive the asymptotic distribution of \( (\gamma_0, \mu_\Lambda)' \), we need to further derive the asymptotic covariance between \( \sqrt{N}\text{vec}(\tilde{F}' - \tilde{H}'\tilde{F}') \) and \( \sqrt{N}\tilde{e}_T \), given by:

\[
N\text{Cov}(\text{vec}(\tilde{F}' - \tilde{H}'\tilde{F}'), \tilde{e}_T) = N\text{Cov}(\text{vec}(\tilde{I}_T \otimes \tilde{U}^{-1}\tilde{F}' - \tilde{H}'\tilde{F}'\tilde{U}^{-1})/\tilde{T}_N, \tilde{e}_T)
\]

\[
+ N\text{Cov}((\tilde{I}_T \otimes \tilde{U}^{-1}\tilde{F}')/\tilde{T}_N, \tilde{e}_T) + N\text{Cov}(\text{vec}(\tilde{F}' - \tilde{H}'\tilde{F}')/\tilde{T}_N, \tilde{e}_T)
\]

\[
= N(\tilde{I}_T \otimes \tilde{U}^{-1}Q) \frac{1}{N^2} \sum_{i=1}^{N^2} (\varepsilon(e_i, e_i') \otimes \lambda_i) + N(\tilde{U}^{-1}H'\tilde{F}') \frac{1}{N^2} \sum_{i=1}^{N^2} (\lambda_i \otimes \varepsilon(e_i, e_i')) + o(1)
\]

\[
\rightarrow \sigma^2(\tilde{I}_T \otimes \tilde{U}^{-1}Q)(\tilde{I}_T \otimes \mu_\Lambda) + \sigma^2(\tilde{F}' - \tilde{H}'\tilde{F}')/(\mu_\Lambda \otimes \tilde{I}_T) = \sigma^2(\tilde{I}_T \otimes \tilde{U}^{-1}H'\tilde{F}')/(\mu_\Lambda \otimes \tilde{I}_T) + \sigma^2(\tilde{F}' - \tilde{H}'\tilde{F}')/(\mu_\Lambda \otimes \tilde{I}_T),
\]
recollecting that $Q = \mathbf{H}^{-1}$. Putting terms together, parts (i),(ii) and (iv) follow, where the matrix $G$ is obtained using the property $K_{a1} = I_a$. Part (iii) follows precisely along the same steps of Theorem 2.

Consider now case (i), which is only relevant in the dynamic case. Following the previous steps, but recognizing time-variation of the $a_{lt-1}, \gamma_{0t-1}$ and $\lambda_{lt-1}$, as $\bar{a} \equiv a_T^t 1_T / T = O_p(N^{-\frac{1}{2}})$, $e_T = O_p(N^{-\frac{1}{2}})$ and $\| F^* - F_H^{t-1} \| = O_p(N^{-\frac{1}{2}})$,

\[
\begin{pmatrix}
\hat{\gamma}_0 \\
\hat{\mu}_\Lambda
\end{pmatrix} = (\hat{\mathbf{D}}'\hat{\mathbf{D}})^{-1} \hat{\mathbf{D}}' \left( \hat{\mathbf{a}}_T + \gamma_0 + \hat{\mathbf{F}}_\mathbf{H}_t \hat{\mu}_{\Lambda_{t-1}} + \hat{\mathbf{e}}_T \right) + O_p(N^{-\frac{1}{2}})
\]

\[
= \left( \frac{\hat{\mathbf{D}}'\hat{\mathbf{D}}}{T} \right)^{-1} \hat{\mathbf{D}}' \left( \hat{\mathbf{a}}_T + \gamma_0 + \hat{\mathbf{F}}_\mathbf{H}^{t-1} \hat{\mu}_{\Lambda_{t-1}} \right) + O_p(N^{-\frac{1}{2}})
\]

\[
= 1 + \hat{\mathbf{F}}^{\ast} \hat{\Omega}^{t-1} - \hat{\mathbf{F}}^{\ast} \hat{\Omega}^{t-1} \left( \hat{\mathbf{F}}_\mathbf{H}^{t-1} \hat{\mu}_{\Lambda_{t-1}} + \hat{\mathbf{e}}_T \right) + O_p(N^{-\frac{1}{2}})
\]

\[
= \left( \begin{pmatrix}
\hat{\gamma}_0 \\
\hat{\mu}_\Lambda
\end{pmatrix} - \left( \frac{\hat{\mathbf{F}}^{\ast} \hat{\Omega}^{t-1} \text{Cov}(\hat{\mathbf{F}}^{\ast}, \gamma_0)}{T} + O_p(N^{-\frac{1}{2}}) \right) \right)
\]

\[\text{QED}\]

### 8.5 Appendix E: The matrixes $U_e, U_{lt-1}$ and $\hat{U}_e, \hat{U}_{lt-1}$

We now present the closed-form expression for $U_e$, which is defined as the limit:

$$
\frac{1}{N} \sum_{i=1}^{N} E \left( ((e_{i} \otimes e_{i}) - \sigma^2 vec(I_T))((e_{i}^t \otimes e_{i}^t) - \sigma^2 vec'(I_T)) \right) \rightarrow U_e.
$$

In particular, Raponi et al. (2018) established that the $T^2 \times T^2$ matrix $U_e$ has the form:

\[
U_e = \begin{bmatrix}
U_{11} & \cdots & U_H & \cdots & U_{1T} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
U_{H1} & \cdots & U_H & \cdots & U_{HT} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
U_{T1} & \cdots & U_{TT} \
\end{bmatrix}. \quad (56)
\]

Each block of $U_e$ is a $T \times T$ matrix. The blocks along the main diagonal, denoted by $U_{tt}$, $t = 1, 2, \ldots, T$, are themselves diagonal matrixes with $(\kappa_4 + 2\sigma_4)$ in the $(t,t)$-th position and $\sigma_4$ in the $(s,s)$ position for every
s \neq t$, that is,

\[
U_{tt} = \begin{bmatrix}
\sigma_4 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & \cdots & \sigma_4 & 0 & \cdots & \cdots & 0 \\
0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & 0 & \cdots & 0 \\
0 & \cdots & \cdots & 0 & \sigma_4 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix}.
\]

(57)

The blocks outside the main diagonal, denoted by $U_{ts}$, $s, t = 1, 2, \ldots, T$ with $s \neq t$, are all made of zeros except for the $(s, t)$-th position that contains $\sigma_4$, that is,

\[
U_{ts} = \begin{bmatrix}
0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
0 & \cdots & 0 & \sigma_4 & 0 & \cdots & 0 \\
0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix}.
\]

(58)

The time-varying matrix $U_{et}$ is defined by replacing $\kappa_4, \sigma_4$ with $\kappa_{4t}, \sigma_{4t}$ in the expressions above.

Finally, noticing that $U_e = U_e(\kappa_4, \sigma_4)$ and $U_{et-1} = U_e(\kappa_{4t-1}, \sigma_{4t-1})$, consistent plug-in estimators can be easily obtained as $\hat{U}_e = U_e(0, \hat{\sigma}_4)$ and $\hat{U}_{et-1} = U_e(0, \hat{\sigma}_{4t-1})$ when $\kappa_4 = \kappa_{4t-1} = 0$ and $\hat{\sigma}_4, \hat{\sigma}_{4t-1}$ denote the consistent estimators adopted in the paper.

8.6 Tables
<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$PC_{p_1}$</th>
<th>$IC_{p_1}$</th>
<th>$AIC_1$</th>
<th>$BIC_1$</th>
<th>$PC(0.10)$</th>
<th>$PC(0.25)$</th>
<th>$PC(0.45)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.28</td>
<td>1.132</td>
<td>1.052</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.316</td>
<td>1.08</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.256</td>
<td>1.024</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.248</td>
<td>1.008</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.196</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.184</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.136</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1.072</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table I: Estimated number of factors, $\hat{k}$ (average across Monte Carlo iterations). The true number of factors is $r = 1$ and we search for $0 \leq k \leq T = 2, 3$. The $e_{it}$ are assumed iid $N(0, 1)$ across $i$ and $t$. Columns 3 to 6 correspond to the competing criteria $IP_{p_1}, IC_{p_1}, AIC_1, BIC_1$ (see Bai & Ng (2002) [Section 5] for details). Columns 7 to 9 correspond to our large-$N$ criterion $PC(\epsilon_2) = (\frac{T}{T-k})V(k) + kg(N)$ with $g(N) = \frac{(\log(N))^{1}N^{1/2}}{\sqrt{N}}$ and $\epsilon_1 = 0, \epsilon_2 = 0.1, 0.25, 0.45$. 
<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$PC_{p1}$</th>
<th>$IC_{p1}$</th>
<th>$AIC_1$</th>
<th>$BIC_1$</th>
<th>$PC(0.10)$</th>
<th>$PC(0.25)$</th>
<th>$PC(0.45)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>5.16</td>
<td>5.11</td>
<td>5.13</td>
<td>5.19</td>
<td>1.11</td>
<td>1.01</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>5.03</td>
<td>5.03</td>
<td>5.03</td>
<td>5.04</td>
<td>1.06</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>5.05</td>
<td>5.03</td>
<td>5.03</td>
<td>5.05</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>5.07</td>
<td>5.06</td>
<td>5.06</td>
<td>5.07</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>5.07</td>
<td>5.07</td>
<td>5.07</td>
<td>5.07</td>
<td>6</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>5</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>5</td>
<td>5.01</td>
<td>5.01</td>
<td>5.01</td>
<td>5.01</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1.02</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>8</td>
<td>7.51</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>10</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>1.5</td>
<td>1</td>
<td>8</td>
<td>3.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>30</td>
<td>1</td>
<td>4.84</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>30</td>
<td>1</td>
<td>1.29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>3.54</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>2.18</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>60</td>
<td>1</td>
<td>4.78</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>120</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>120</td>
<td>1.13</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>120</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>4.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>120</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>120</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>120</td>
<td>1</td>
<td>1.43</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3000</td>
<td>120</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table II: Estimated number of factors, $\hat{k}$ (average across Monte Carlo iterations). The true number of factors is $r = 1$ and we search for $0 \leq k \leq T$. The $e_{it}$ are assumed iid $N(0, 1)$ across $i$ and $t$. Columns 3 to 6 correspond to the competing criteria $IP_{p1}, IC_{p1}, AIC_1, BIC_1$ (see Bai & Ng (2002) [Section 5] for details). Columns 7 to 9 correspond to our large-$N$ criterion $PC(\epsilon_2) = (\frac{T}{T-k})V(k) + kg(N)$ with $g(N) = \frac{(\log(N))^{e_2}N^2}{\sqrt{N}}$ and $e_1 = 0, e_2 = 0.1, 0.25, 0.45$. 
<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$PC_{p1}$</th>
<th>$IC_{p1}$</th>
<th>$AIC_1$</th>
<th>$BIC_1$</th>
<th>$PC(0.10)$</th>
<th>$PC(0.25)$</th>
<th>$PC(0.45)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
<td>4.38</td>
<td>3.57</td>
<td>2.23</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>5.01</td>
<td>5</td>
<td>5</td>
<td>5.01</td>
<td>6</td>
<td>5.98</td>
<td>2.43</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>5.01</td>
<td>5</td>
<td>5.01</td>
<td>6</td>
<td>6</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
<td>6</td>
<td>6</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>5.05</td>
<td>5.02</td>
<td>5.01</td>
<td>6</td>
<td>6</td>
<td>5.11</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>5</td>
<td>5.03</td>
<td>5.02</td>
<td>5</td>
<td>5.03</td>
<td>6</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td>5</td>
<td>5.08</td>
<td>5.07</td>
<td>5.06</td>
<td>5.08</td>
<td>6</td>
<td>5.85</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4.4</td>
<td>3.39</td>
<td>2.04</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>6.19</td>
<td>3.23</td>
<td>1.8</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>6.76</td>
<td>2.87</td>
<td>1.76</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>6.5</td>
<td>2.9</td>
<td>1.87</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5.77</td>
<td>2.91</td>
<td>1.86</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5.29</td>
<td>2.95</td>
<td>1.95</td>
</tr>
<tr>
<td>3000</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4.39</td>
<td>2.99</td>
<td>1.94</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4.22</td>
<td>2.88</td>
<td>1.88</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>6.34</td>
<td>2.93</td>
<td>8</td>
<td>8</td>
<td>3.61</td>
<td>2.91</td>
<td>1.94</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>3.09</td>
<td>3</td>
<td>8</td>
<td>4.46</td>
<td>3.04</td>
<td>3</td>
<td>2.07</td>
</tr>
<tr>
<td>200</td>
<td>30</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.15</td>
</tr>
<tr>
<td>500</td>
<td>30</td>
<td>3</td>
<td>3</td>
<td>5.52</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.26</td>
</tr>
<tr>
<td>1000</td>
<td>30</td>
<td>3</td>
<td>3</td>
<td>3.03</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.38</td>
</tr>
<tr>
<td>3000</td>
<td>30</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.55</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4.15</td>
<td>2.79</td>
<td>1.95</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>4.18</td>
<td>2.98</td>
<td>8</td>
<td>8</td>
<td>3.02</td>
<td>2.95</td>
<td>2.03</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>3.41</td>
<td>3</td>
<td>3</td>
<td>2.23</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.32</td>
</tr>
<tr>
<td>500</td>
<td>60</td>
<td>3</td>
<td>3</td>
<td>5.49</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.62</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
</tr>
<tr>
<td>3000</td>
<td>60</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.76</td>
</tr>
<tr>
<td>10</td>
<td>120</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>3.53</td>
<td>2.78</td>
<td>1.98</td>
</tr>
<tr>
<td>30</td>
<td>120</td>
<td>3.01</td>
<td>2.99</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>2.97</td>
<td>2.07</td>
</tr>
<tr>
<td>100</td>
<td>120</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>4.91</td>
<td>3</td>
<td>3</td>
<td>2.37</td>
</tr>
<tr>
<td>200</td>
<td>120</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.45</td>
</tr>
<tr>
<td>500</td>
<td>120</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.81</td>
</tr>
<tr>
<td>1000</td>
<td>120</td>
<td>3</td>
<td>3</td>
<td>3.12</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.9</td>
</tr>
<tr>
<td>3000</td>
<td>120</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.96</td>
</tr>
</tbody>
</table>

Table III: Estimated number of factors, $\hat{k}$ (average across Monte Carlo iterations). The true number of factors is $r = 3$ and we search for $0 \leq k \leq T$. The $e_{it}$ are assumed iid $N(0, 1)$ across $i$ and $t$. Columns 3 to 6 correspond to the competing criteria $IP_{p1}, IC_{p1}, AIC_1, BIC_1$ (see Bai & Ng (2002) [Section 5] for details). Columns 7 to 9 correspond to our large-$N$ criterion $PC(\epsilon_2) = \left(\frac{T}{T-k}\right)V(k) + kg(N)$ with $g(N) = \frac{(\log(N))^{1.5}}{\sqrt{N}}$ and $\epsilon_1 = 0$, $\epsilon_2 = 0.1, 0.25, 0.45$. 
### Table IV: Sample correlations between $\tilde{F}_t$ and $F_t$ for $1 \leq t \leq T$ (average across Monte Carlo iterations).

<table>
<thead>
<tr>
<th></th>
<th>N=10</th>
<th>N=100</th>
<th>N=1000</th>
<th>N=3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T=2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$T=5$</td>
<td>0.942</td>
<td>0.996</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>$T=50$</td>
<td>0.955</td>
<td>0.996</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

### Table V: Sample correlations between $\tilde{\lambda}_i$ and $\lambda_i$ for $1 \leq i \leq N$ (average across Monte Carlo iterations).

<table>
<thead>
<tr>
<th></th>
<th>N=10</th>
<th>N=100</th>
<th>N=1000</th>
<th>N=3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T=2$</td>
<td>0.610</td>
<td>0.667</td>
<td>0.674</td>
<td>0.675</td>
</tr>
<tr>
<td>$T=5$</td>
<td>0.875</td>
<td>0.905</td>
<td>0.907</td>
<td>0.907</td>
</tr>
<tr>
<td>$T=50$</td>
<td>0.989</td>
<td>0.991</td>
<td>0.992</td>
<td>0.992</td>
</tr>
</tbody>
</table>
Table VI: Identifying risk factors: entire period (Dec 1960 to Dec 2013). We report the correlation between $mkt_t$ (market excess return), $smb_t$ (small-minus-big portfolio return), $hml_t$ (high-minus-low portfolio return), $rmw_t$ (profitability portfolio return), $cmi_t$ (investment portfolio return) and the first five estimated risk factors. Rolling windows of $T = 12$ observations from Dec 1960 to Dec 2013. Data source: Kenneth French’s website.
Table VII: Identifying risk factors: from the burst of the sub-prime bubble to recovery (Aug 2007 to Feb 2009). We report the correlation between $mkt_t$ (market excess return), $smb_t$ (small-minus-big portfolio return), $hml_t$ (high-minus-low portfolio return), $rmw_t$ (profitability portfolio return), $cma_t$ (investment portfolio return) and the first five estimated risk factors. Rolling windows of $T = 12$ observations from Dec 1960 to Dec 2013. Data source: Kenneth French’s website.

<table>
<thead>
<tr>
<th></th>
<th>mkt</th>
<th>smb</th>
<th>hml</th>
<th>rmw</th>
<th>cma</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor 1</td>
<td>0.367</td>
<td>0.083</td>
<td>0.378</td>
<td>0.173</td>
<td>-0.064</td>
</tr>
<tr>
<td>factor 2</td>
<td>-0.194</td>
<td>-0.037</td>
<td>-0.311</td>
<td>-0.160</td>
<td>0.109</td>
</tr>
<tr>
<td>factor 3</td>
<td>0.161</td>
<td>0.433</td>
<td>0.439</td>
<td>-0.125</td>
<td>-0.046</td>
</tr>
<tr>
<td>factor 4</td>
<td>0.071</td>
<td>0.085</td>
<td>-0.342</td>
<td>-0.361</td>
<td>-0.462</td>
</tr>
<tr>
<td>factor 5</td>
<td>0.007</td>
<td>0.191</td>
<td>-0.232</td>
<td>-0.155</td>
<td>-0.412</td>
</tr>
</tbody>
</table>
Table VIII: Column one reports the out-of-sample Sharpe ratio (annualized) corresponding to equally-weighted portfolio excess return. Columns two to five report the out-of-sample Sharpe ratios (annualized) corresponding to the tangency portfolio excess returns:

\[
\tilde{r}_{t+1}^{mv,k} = \tilde{w}_t^{mv'}x_{t+1},
\]

where \( \tilde{\Sigma}_t = (\tilde{\Lambda}_t\tilde{\Omega}\tilde{\Lambda}_t' + \tilde{\sigma}_t^2I_N) \).

Rolling windows of \( T = 36 \) observations from Dec 1960 to Dec 2013. Data source: individual equity returns from CRSP. The second row report the \( t \)-ratios for testing equality between the Sharpe ratios of the portfolios associated with the factor model and the Sharpe ratio of the equally weighted portfolio (see Lo (2003) for details).
Figure I: $T = 10, N = 100$. The figure presents the histograms (across Monte Carlo iterations) of the sample correlations between $\mathbf{F}_t$ and $\mathbf{F}_t$ (dark blue) and of the sample correlations between $\lambda_i$ and $\lambda_i$ (red). The correlations have been taken in absolute value.
Figure II: $T = 100, N = 100$. The figure presents the histograms (across Monte Carlo iterations) of the sample correlations between $\tilde{F}_t$ and $F_t$ (dark blue) and of the sample correlations between $\tilde{\lambda}_i$ and $\lambda_i$ (red). The correlations have been taken in absolute value.
Figure III: \( T = 2, N = 10 \). The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely \( f_{s\text{small-}} = (A'\Bar{B}_s A)^{-\frac{1}{2}}N\frac{1}{2}(\Bar{F}_s - \Bar{H}F_s) \) (green area) and \( f_{s\text{large-}} = (\Hat{\sigma}^2\Hat{U}^{s-1})^{-\frac{1}{2}}N\frac{1}{2}(\Bar{F}_s - \Bar{H}F_s) \) (blue area) for \( s = 1 \). The standard normal density function is superimposed on the histograms (dark blue line).
Figure IV: $T = 2, N = 1000$. The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely $f_{s_{\text{small}}} = (\mathbf{A}'\mathbf{B}_s\mathbf{A})^{-\frac{1}{2}} N\frac{1}{2}(\mathbf{F}_s - \mathbf{H}\mathbf{F}_s)$ (green area) and $f_{s_{\text{large}}} = (\hat{\sigma}^2\hat{U}^{-1})^{-\frac{1}{2}} N\frac{1}{2}(\tilde{\mathbf{F}}_s - \tilde{\mathbf{H}}\mathbf{F}_s)$ (blue area) for $s = 1$. The standard normal density function is superimposed on the histograms (dark blue line).
Figure V: $T = 5, N = 10$. The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely $f_{s}^{\text{small-}T} = (A'\tilde{B}_sA)^{-\frac{1}{2}}N^\frac{1}{2}(\tilde{F}_s - \tilde{HF}_s)$ (green area) and $f_{s}^{\text{large-}T} = (\hat{\sigma}^2\hat{U}_s^{-1})^{-\frac{1}{2}}N^\frac{1}{2}(\tilde{F}_s - \tilde{HF}_s)$ (blue area) for $s = 1$. The standard normal density function is superimposed on the histograms (dark blue line).
Figure VI: $T = 5, N = 1000$. The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely $f_{s, \text{small}}^{T} = (A'\hat{B}_s A)^{-\frac{1}{2}} N_{\frac{1}{2}}^{\frac{1}{2}}(\hat{F}_s - \hat{H}_s F_s)$ (green area) and $f_{s, \text{large}}^{T} = (\hat{\sigma}^2 \hat{U}^{-1})^{-\frac{1}{2}} N_{\frac{1}{2}}^{\frac{1}{2}}(\hat{F}_s - \hat{H}_s F_s)$ (blue area) for $s = 1$. The standard normal density function is superimposed on the histograms (dark blue line).
Figure VII: $T = 50, N = 10$. The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely $f_{s_{\text{small}}} = (A'\tilde{B}_sA)^{-\frac{1}{2}}N^{\frac{1}{2}}(\tilde{F}_s - \tilde{H}F_s)$ (green area) and $f_{s_{\text{large}}} = (\tilde{\sigma}^2\tilde{U}_s^{-1})^{-\frac{1}{2}}N^{\frac{1}{2}}(\tilde{F}_s - \tilde{H}F_s)$ (blue area) for $s = 1$. The standard normal density function is superimposed on the histograms (dark blue line).
Figure VIII: $T = 50, N = 1000$. The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely $f^\text{small}\sim T_s = (A'\tilde{B}_sA)^{-\frac{1}{2}}N^{\frac{1}{2}}(\tilde{F}_s - \tilde{H}F_s)$ (green area) and $f^\text{large}\sim T_s = (\tilde{\sigma}^2\tilde{U}^{s-1})^{-\frac{1}{2}}N^{\frac{1}{2}}(\tilde{F}_s - \tilde{H}F_s)$ (blue area) for $s = 1$. The standard normal density function is superimposed on the histograms (dark blue line).
Figure IX: \( T = 50, N = 1000 \). The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely \( f_s^{small-T} = (A' \tilde{B}_s A)^{-\frac{1}{2}} N^\frac{1}{2} (\tilde{F}_s - \tilde{H} F_s) \) (green area) and \( f_s^{large-T} = (\tilde{\sigma}^2 \tilde{U}^{-1})^{-\frac{1}{2}} N^\frac{1}{2} (\tilde{F}_s - \tilde{H} F_s) \) (blue area) for \( s = 1 \). We also report (red area) the histogram of (studentized) estimated factors using Bai (2003) robust standard errors. The standard normal density function is superimposed on the histograms (dark blue line).
Figure X: $T = 100, N = 1000$. The figure presents the histogram of (studentized) estimated factors (across Monte Carlo iterations), namely $f^{\text{small-}T}_s = (A'\tilde{B}_s A)^{-\frac{1}{2}} N^{\frac{1}{2}} (\tilde{F}_s - \tilde{H} F_s)$ (green area) and $f^{\text{large-}T}_s = (\tilde{\sigma}^2 \tilde{U}^{*-1})^{-\frac{1}{2}} N^{\frac{1}{2}} (\tilde{F}_s - \tilde{H} F_s)$ (blue area) for $s = 1$. We also report (red area) the histogram of (studentized) estimated factors using Bai (2003) robust standard errors. The standard normal density function is superimposed on the histograms (dark blue line).
Figure XI: $T = 5, N = 10$. The figure presents the time series (average across Monte Carlo iterations) of the 95% confidence intervals (red lines) for the true factor (black line):

$$
\left( \hat{\beta} \bar{F}_t - 1.96 \frac{\hat{\beta}}{N^{\frac{1}{2}}(A'B_t A)^{\frac{1}{2}}} , \hat{\beta} \bar{F}_t + 1.96 \frac{\hat{\beta}}{N^{\frac{1}{2}}(A'B_t A)^{\frac{1}{2}}} \right), \ t = 1, \cdots, T,
$$

where $\hat{\beta}$ is the OLS estimator from projecting $\bar{F}$ on $F$ (without intercept).
Figure XII: $T = 5, N = 1000$. The figure presents the time series (average across Monte Carlo iterations) of the 95% confidence intervals (red lines) for the true factor (black line):

$$
\left( \tilde{\beta} \hat{F}_t - 1.96 \frac{\tilde{\beta}}{N^{1/2}(A'\tilde{B}_t A)^{1/2}}, \tilde{\beta} \hat{F}_t + 1.96 \frac{\tilde{\beta}}{N^{1/2}(A'\tilde{B}_t A)^{1/2}} \right), \ t = 1, \cdots, T,
$$

where $\tilde{\beta}$ is the OLS estimator from projecting $\tilde{F}$ on $F$ (without intercept).

Data are monthly from Dec 1960 until Dec 2013.
8.7 Figures


Data are monthly from Dec 1960 until Dec 2013.
Figure XVII: The figure presents the time series of the estimated correlation, over rolling windows of 60 months, of the rmw and cma returns with the fourth PCA, where the PCA is implemented over rolling windows of $T = 12$ observations. The number of stocks, corresponding to each rolling window, varies from 800 to 5,000 circa. The grey bands indicating the NBER recession indicators and various economic and financial crises, and are numbered as follows:

The figure presents the time series of the first five estimated risk factors, based on rolling windows of $T = 12$ observations. The number of stocks, corresponding to each rolling window, varies from 800 to 5,000 circa. The grey bands indicating the NBER recession indicators and various economic and financial crises, and are numbered as follows:

Data are monthly from Dec 1960 until Dec 2013.

Data are monthly from Dec 1960 until Dec 2013.
Figure XX: The figure presents the time series of the adjusted-$R^2$ of the regression, over rolling windows of 60 observations, of $\hat{\mu}_{t} = w_N^{\text{ew}} \hat{\Lambda}_t \hat{\gamma}$, where $w_N^{\text{ew}} = N^{-1} 1_N$, based on the estimated number of factors reported in Figure XIV, over four state variables, namely the market dividend yield (ratio of aggregate dividends to the S&P 500 index; source Robert Shiller’s website), the default spread (Moody’s Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity, Not Seasonally Adjusted; source FRED), the term spread (10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity, Not Seasonally Adjusted; source FRED) and the CAPE ratio (Cyclically Adjusted Price Earnings Ratio P/E10; source Robert Shiller website).


Data are monthly from Dec 1960 until Dec 2013.

Data are monthly from Dec 1960 until Dec 2013.
Data are monthly from Dec 1960 until Dec 2013.
Figure XXIV: The figure presents the time series (black line), estimated over rolling windows of $T = 12$ observations, of the pricing performance statistic:

$$
\tilde{\Delta}_{t-1} = \frac{1}{N} \sum_{i=1}^{N} \tilde{\delta}_{it-1}^2 - \frac{\tilde{\sigma}_{t-1}^2}{T} (1 - \tilde{\gamma}^\prime \tilde{\gamma}),
$$

where $\tilde{\delta}_{it-1} = \bar{x}_i - \tilde{\lambda}_t^\prime \tilde{\gamma}$. The black line corresponds to the estimated number of factors $\tilde{k}_{t-1}$ with our criterion. We report cases $k = 1$ (red line), $k = 2$ (green line), $k = 3$ (blue line) and $k = 5$ (light blue).


Data are monthly from Dec 1960 until Dec 2013.
Figure XXV: The figure presents the time series (black line) of the estimated SDF:

$$\tilde{m}_{t,t+1} \equiv \frac{1}{r_{ft}} - \frac{1}{r_{ft}} \tilde{\gamma}^t \tilde{\Omega}^{-1} (\tilde{F}_{t+1} - \tilde{\gamma}).$$


Data are monthly from Dec 1960 until Dec 2013.
Figure XXVI: We estimate the three linear regressions, using rolling windows of 60 months,

\[ \tilde{m}_{t-1,t} = \alpha + \beta_1 mkt_t + u_t, \]
\[ \tilde{m}_{t-1,t} = \alpha + \beta_1 mkt_t + \beta_2 smb_t + \beta_3 hml_t + u_t, \]
\[ \tilde{m}_{t-1,t} = \alpha + \beta_1 mkt_t + \beta_2 smb_t + \beta_3 hml_t + \beta_4 rmw_t + \beta_5 cma_t + u_t, \]


Data are monthly from Dec 1960 until Dec 2013.