Factor Models for Conditional Asset Pricing *

PAOLO ZAFFARONI

Imperial College Business School

This Draft

March 19, 2025

Abstract

This paper develops a methodology, building on a local-PCA approach, for inference on the pricing ability of conditional asset pricing models designed to mitigate the effect of omitted risk factors and misspecified conditional dynamics. The methodology is designed to exploit the rich information available in large cross-sections of individual stocks. Monte Carlo experiments and an empirical application demonstrate the benefits of this methodology over existing approaches.

Keywords: Local Principal Component Analysis; No-Arbitrage; Latent Factor Model; Risk Premia; Stochastic Discount Factor; Misspecification; Large Cross-Sections.

1 Introduction

Asset pricing theory is intrinsically dynamic with prices and risk premia, i.e., compensation for accepting risk exposure, expressed as conditional moments. However, estimating conditional asset pricing models typically relies on a parsimonious specification of the conditional distribution of asset prices, inevitably guided by computational and statistical considerations. But with "...tractability

* I thank Pasquale Della Corte, Victor DeMiguel, Patrick Gagliardini, Robert Korajczyk, Frank Kleibergen, Markus Pelger, Valentina Raponi, Eric Renault, Olivier Scaillet, Raman Uppal, Dacheng Xiu, Guofu Zhou, Irina Zviadadze, and audiences at seminars in Cambridge, Northwestern, Imperial College London, Luxembourg School of Finance, Yale, Tinbergen Institute (Amsterdam) and Warwick, for suggestions. I thank Massimo Dello Preite for excellent research assistance, Bryan Kelly for sharing their IPCA code, and Ian Martin for sharing his data.

comes misspecification..." (Hansen and Sargent (1999)).[1] Asset pricing factor models are key tools for understanding how different systematic risks affect asset prices, yet finding a complete list of these risk factors has been challenging.[2] Popular models, like those by Fama and French (1993, 2015), have made significant progress by identifying three to five major risk factors, but there is still no agreement on which factors are really "...important and which are...subsumed by others" (Cochrane (2011), Harvey et al. (2016))). Thus, conditional asset pricing factor models are very stylized and likely to be misspecified, adversely affecting their pricing ability.

This paper advances the asset pricing literature with a PCA-based methodology that tackles two critical sources of misspecification in conditional asset pricing factor models: (1) omitted risk factors and (2) incorrect specification of the conditional distribution of asset returns. I consider a *local* version of the conventional Principal Component Analysis (PCA) estimator ("local-PCA") to mitigate these two sources of misspecification. This approach applies PCA to panel data consisting of many assets (N) sampled over a fixed time-series window (T). My methodology leverages the well-known ability of PCA to estimate *latent* risk factors accurately. At the same time, the *local* trait of my local-PCA methodology allows for model-free time variation in certain model parameters.

PCA represents an increasingly popular avenue in empirical asset pricing given that, for portfolio construction and pricing purposes, the risk factors do not need to be observed, as elicited in the APT of Ross (1976), yielding a reduced-form stochastic discount factor (SDF) with substantial pricing ability and robust to omitted priced risk factors.[3] Traditional PCA (hereafter labelled as "constant-PCA") assumes constancy of the model's parameters. Its statistical properties are well-established under the double asymptotic setting, where N and T can diverge.[4] However, when T is fixed, these results are no longer valid. The first contribution of this paper is to derive the asymptotic analysis of the local-PCA.

Several PCA-based procedures have been proposed for conditional asset pricing models, where risk exposures, risk premia, and sometimes alphas, are assumed to be some function of *observed* instruments, such as firms' characteristics.[5]Despite their appeal, these approaches may use the wrong functional form, when parametrically specified, or select the wrong set of instruments, and often lack a formal asymptotic analysis. These sources of misspecification are overcome by using the local-PCA methodology proposed in this paper because it is based on time averages over short time windows of fixed length T, capturing model-free time variation.[6]

[1]Moreover, the typical approach hinges on the unrealistic presumption that investors use the same conditioning information used to construct the conditional asset pricing model.

[2]From a theoretical perspective, factor models have 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).

[3]See Kozak et al. (2018), Lettau and Pelger (2020b), and Giglio and Xiu (2021), among many others.

[4]See Bai and Ng (2002) and Stock and Watson (2002), among others, for constant-PCA, and Lettau and Pelger (2020a) for a penalized version of PCA.

[5]See, among others, Connor et al. (2012), Fan et al. (2016), Kelly et al. (2019), Kim et al. (2019), Pelger and Xiong (2019), Haddad et al. (2020), and Ma and Gagliardini (2020). For more details, see Section OA2.

[6]The use of rolling samples of data is extremely popular in empirical asset pricing since the influential Fama and MacBeth (1973) methodology was coined, further refined in Andersen et al. (2006), Lewellen and Nagel (2006) and

The second contribution of this paper is to show how to rip the benefits of keeping T fixed, but at the same mitigating the estimation noise associated with it, by deriving an asymptotically valid inferential methodology for risk premia, the SDF, and optimal mean-variance portfolios. These represent the primary objects of interest in empirical asset pricing, making my methodology appealing for applications with large cross-sections of individual securities, given the widespread evidence of time variation in their conditional distribution.

My results rely on three insights. First, accurate estimation of risk factors by PCA requires *only* a large N, regardless of T. Second, risk premia and the SDF depend *only* on the risk factors, but not on the risk exposures (i.e., the loadings), whose estimates remain severely biased when T is fixed. However, the moments of the risk exposures (specifically their mean and covariance matrix) do affect risk premia and SDF's standard errors and, hence, need to be estimated consistently. The third insight shows that one can *accurately* estimate the moments of the risk exposures, using bias-adjusted cross-sectional averages of the risk exposures. The same insight applies to mean-variance portfolios; that is, even though the portfolio weights are severely biased as they depend (approximately) linearly on the risk exposures, accurate estimation of mean-variance portfolio returns can still be obtained because these depend on cross-sectional averages of the N risk exposures.

My inferential procedure delivers estimates of risk premia and SDF that are asymptotically *equivalent*, as N diverges, to the estimates based on a rolling sample of dimension T of the true risk factors, which are unfeasible because of their latency.[7] However, when T is finite only a noisy yet unbiased version of the risk factors' true premia, known as the *ex-post risk premia* (see Shanken (1992)), can be identified. Along the same lines, I can only identify the *ex-post SDF*, a noisy version of the true SDF. In other words, although I can estimate the risk factors accurately (because of large N), their first two moments, which contribute to determining the *ex-ante* risk premia and SDF, remain inaccurately estimated (because of fixed T). However, I show that the ex-post risk premia and ex-post SDF are still valuable objects for pricing and portfolio construction purposes.

I aim to evaluate the ability of the conditional factor model to price assets, rather than simply assessing the statistical significance of its parameter estimates. The third contribution of this paper involves developing inferential local approaches to test whether the factors' *ex-ante* risk premia unfold expected returns, and whether the *ex-ante* SDF adequately prices the assets. These procedures are applicable when the number of assets N is large, despite the aforesaid challenge in identifying the ex-ante risk premia and SDF when the time window T is finite.[8]

The asymptotic theory for the local-PCA hinges on a single economic assumption, i.e., the absence

Ang and Kristensen (2012), among others.

[7] Given that, in practice, N is always finite, I provide asymptotically valid standard errors that quantify the uncertainty around these unfeasible estimates of risk premia and SDF. Similarly, I derive asymptotically valid confidence intervals for the most commonly used measures of portfolio performance and risk, such as the Sharpe ratio (SR) and the Value-at-Risk (VaR).

[8] The key insight, that makes my pricing tests asymptotically valid, is that (by exploiting large N) I can disentangle the pricing error, that reflects the genuine pricing performance of the ex-ante risk premia and SDF, from the fixed-T noise affecting the ex-post risk premia and SDF.

of arbitrage, and on several statistical assumptions. No-arbitrage permits to avoid de-meaning the data, in turn, required for the local-PCA to work asymptotically. I also need a certain degree of local smoothness in the risk exposures time-variation over the T periods, and rule out time-variation in the variances and covariances (when averaged across assets) of the idiosyncratic errors, while still allowing cross-sectional dependence in the latter. I also need the number of risk factors to be locally constant. As T can be taken to be almost arbitrarily small, these assumptions rule out abrupt changes in the parameters over the sample period. Moreover, some of these assumptions are akin to the empirical evidence as, for example, stock returns are only mildly autocorrelated across time but strongly so cross-sectionally. Finally, my local-PCA method is robust to mispricing, i.e., it works under the APT, in the sense that it still valid under systematic deviation of expected returns from the expected returns implied by the factor structure. If T also diverges, I obtain the constant-PCA asymptotic results, as one would expect.

My work is very close to Pelger (2019, 2020), Lettau and Pelger (2020a,b), Giglio and Xiu (2021), and Ma and Gagliardini (2020) from an asset pricing perspective. I differ from all these approaches because I develop the inferential theory for conditional asset pricing models using my local-PCA approach, whereas, except for Pelger (2019, 2020), these papers use the constant-PCA and variations of.[9] In terms of econometrics, I extend the consistency result of Connor and Korajczyk (1986) and Bai (2003)[Theorem 4] by providing the inferential theory of the local-PCA, and a consistent estimator of the number of risk factors, in the finite-T setting, allowing for local time variation of the factor model's parameters. A complete literature review is reported in Section OA2 of the Online Appendix.

The paper proceeds as follows. Section 2 presents the conditional factor model and Section 3 provides the intuition and motivation of my methodology. Section 4 develops the analysis of the local-PCA, with the proofs in Section OA9 (except for Theorem 1 whose proof is reported here). Section 5 describes the asset pricing implications of the model, in particular, for inference on the pricing ability of risk premia, the SDF, and mean-variance portfolios. Monte Carlo experiments are in Section 6 and the empirical application is summarized in Section 7, using individual US monthly stock returns (from January 1980 to December 2019 and $N = 19,259$ stocks, sourced CRSP), where I compare the out-of-sample performance of my local-PCA (using $T = 24$) with the one of the constant-PCA (using the full sample). Section 8 concludes. A final appendix lists additional technical assumptions and reports the proof of the asymptotic analysis of my local-PCA. When using the OA prefix, in section and equation numbers, it means that I am referring to the Online Appendix.[10]

[9] My local-PCA can estimate locally-strong risk factors, i.e., factors that live only over a limited period, which will look like weak factors when estimated by constant-PCA using the full sample. Hence, from this angle, my methodology can be seen as complementary to the RP-PCA of Lettau and Pelger (2020a,b).

[10] The OA contains further material, namely additional theoretical results and mathematical proofs, various Monte Carlo exercises, and a detailed account of the empirical analysis.

2 Conditional Factor Model

I assume that x_{is}, the excess return on asset i at time s, satisfies a finite-dimensional conditional factor model, in the sense that each x_{is} is a linear function of a finite number of latent risk factors and an idiosyncratic error, as follows:

$$x_{is} \quad = \alpha_{is-1} + \boldsymbol{\lambda}'_{is-1}\mathbf{F}_s + e_{is} \text{ for every } t \in \mathcal{Z} \text{ and } i \in \mathcal{N}, \tag{1}$$

where α_{is-1} is a time-varying intercept, $\mathbf{F}_s$ defines the $r_{s-1} \times 1$ vector of *latent* risk factors with risk premia $\boldsymbol{\gamma}_{s-1}$, satisfying $\mathcal{E}_{s-1}(\mathbf{F}_s) = \boldsymbol{\gamma}_{s-1}$, where $\mathcal{E}_{s-1}(.)$ denotes the conditional mean, conditioning on all available information up to time $s-1$, $\boldsymbol{\lambda}_{is-1}$ is the $r_{s-1} \times 1$ vector of *latent time-varying* risk exposures (i.e., the loadings), e_{is} is the *idiosyncratic* error, $\mathcal{N} = \{1, \cdots\}$, and $\mathcal{Z} = \{\cdots, -1, 0, 1, \cdots\}$. Finally, I assume here that a risk-free asset, with return r_{fs-1}, is tradable.[11] Formulation (1) together with (2) below represents a conditional version of the Arbitrage Pricing Theory (APT) of Ross (1976).[12]

The following assumption, uncommon in the statistical PCA literature but fundamental in asset pricing, imposes the required structure on $\mathcal{E}_{s-1}(x_{is})$ to develop my methodology. Let $var_{s-1}(\cdot)$ denote the conditional variance, conditioning on all available information up to time $s-1$, and define a portfolio p excess return $r^p_{N,s} \equiv \mathbf{w}^{p\prime}_{Ns-1}\mathbf{x}_{s.}$, with weights $\mathbf{w}^p_{Ns-1} = (w^p_{1ts1}, w^p_{2s-1}, \cdots, w^p_{Ns-1})'$, where $\mathbf{x}_{s.} \equiv (x_{1s}, \cdots, x_{Ns})'$.

Assumption 1 (asymptotic no-arbitrage) *There is no sequence of portfolios* p*, along some subsequence N' for which $var_{s-1}(r^p_{N',s}) \to 0$ almost surely (a.s.), when $N' \to \infty$, and $\mathcal{E}_{s-1}(r^p_{N',s}) \geq C > 0$ a.s. for some $s \in \mathcal{Z}$ and an arbitrary positive scalar C.*

Under the no-arbitrage assumption above, with a limited degree of cross-sectional dependence of the e_{is}, and a sufficient degree of smoothness of the risk exposures $\boldsymbol{\lambda}_{is-1}$ (see Proposition OA.4), the

[11] The analysis of the case in which a risk-free asset is not tradeable is described in Section OA6.

[12] The APT induces formulation (1) as follows: one postulates $x_{is} = a_{is-1} + \boldsymbol{\lambda}'_{is-1}\mathbf{f}_s + e_{is}$ for some initial factors $\mathbf{f}_s$ (for example, non-tradable risk factors) and a time-varying intercept a_{is-1}. When $\mathcal{E}_{s-1}(\mathbf{f}_s) = \mathbf{0}_{r_{s-1}}$ then $a_{is-1} = \mathcal{E}_{s-1}(x_{is})$. Under no-arbitrage (and the regularity conditions on the e_{is} and $\boldsymbol{\lambda}_{is-1}$ mentioned above), there exists an $r_{s-1} \times 1$ vector $\boldsymbol{\gamma}_{s-1}$ (i.e., the risk premia vector) such that $\mathcal{E}_{s-1}(x_{is}) = \alpha_{is-1} + \boldsymbol{\lambda}'_{is-1}\boldsymbol{\gamma}_{s-1}$ where the *pricing errors* α_{is-1} satisfy (2), yielding (1) by setting $\mathbf{F}_s \equiv \mathbf{f}_s + \boldsymbol{\gamma}_{s-1} - \mathcal{E}_{s-1}(\mathbf{f}_s)$. The risk premia vector satisfies $\boldsymbol{\gamma}_{s-1} = -r_{fs-1}\left(\mathcal{E}_{s-1}(m_{s-1,s}\mathbf{f}_s) - \frac{\mathcal{E}_{s-1}(\mathbf{f}_s)}{r_{fs-1}}\right)$ and $\boldsymbol{\gamma}_{s-1} = \mathcal{E}_{s-1}(\mathbf{F}_s)$, where $m_{s-1,s}$ denotes the SDF of model (1) (see Hansen and Richard (1987)[Theorem 2.1]). I show that the $\mathbf{F}_s$ are asymptotically equivalent to tradable mimicking portfolios when $N \to \infty$ (see Section OA5).

intercept terms α_{is-1} satisfy,[13]

$$\sum_{i=1}^{\infty} \alpha_{is-1}^2 < \infty \ \ a.s. \text{ for every } s \in \mathcal{Z}. \tag{2}$$

The square summability condition (2) on the intercept terms α_{is-1}, known as *pricing errors* in asset pricing, implies that I do not need to estimate them, hence allowing me to avoid de-meaning the x_{is}. Whereas, when using constant-PCA, the observables can be demeaned without any loss of information, this would adversely impact the asymptotic properties of the local-PCA estimator because of the fixed T.[14]

This paper aims to estimate the conditional factor model (1) by local-PCA, and provide a methodology for inference on its pricing ability, as I now explain.

3 Local-PCA: How It Works and Why it Matters

My aim here is to illuminate my approach and to highlight how it works, some of its limitations, and especially why it matters for asset pricing, as $N \to \infty$. I will discuss, which features are of the essence, what is observed and what is not, what varies with time and what does not, what can be estimated consistently and what cannot, and what is the economic and statistical role of no-arbitrage. I will skip technical details: the precise statements are in the next sections. I will write the salient features in *italics*.

Fix t, some possibly small $T \geq r_t$, and the time window $s \in \mathcal{T}_t = \{t - T + 1, \ldots, t\}$. For ease of notation, I will mostly drop the time subscripts in this section, when it depends on t. For the exposition, I assume in this section the risk exposures are constant over time, $\lambda_{i,s-1} \equiv \lambda_i$. Generally, Assumption 3 requires the latter to be true asymptotically, as $N \to \infty$. By Assumption 2, the number of risk factors satisfies $r_s \equiv r$. Write (1) in matrix form as

$$\mathbf{X} = \boldsymbol{\alpha} + \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{e}, \tag{3}$$

where $\mathbf{X}$, $\boldsymbol{\alpha}$ and $\mathbf{e}$ are $T \times N$, $\mathbf{F}$ is $T \times r$ and $\boldsymbol{\Lambda}$ is $N \times r$. As is well known, $\mathbf{F}$ and $\boldsymbol{\Lambda}$ are identified only up to linear transformations, i.e., $\mathbf{FH}$ and $\boldsymbol{\Lambda}\mathbf{H}^{-1\prime}$ also satisfy (3) for any non-singular $r \times r$ matrix $\mathbf{H}$. As usual in factor analysis, all these transformations are equally valid. I seek to estimate $\mathbf{F}$ and $\boldsymbol{\Lambda}$, up to such a linear transformation. The key property of equation (3) is the square-

[13] Condition (2) is the main prediction of the APT, i.e., that $\mathcal{E}_{s-1}(x_{is})$ can differ from $\boldsymbol{\lambda}'_{is-1}\mathcal{E}_{s-1}(\mathbf{F}_s)$ although not too much (in the sense of (2)); for the unconditional setting see Huberman (1982), Chamberlain (1983), and Ingersoll (1984) among others, and for the conditional setting see Stambaugh (1983), here extended by Proposition OA.4. In general, (2) involves the *inverse* of the conditional covariance matrix of the e_{is} with elements $\sigma^{(-1)}_{ij,s-1s-1}$ (see, for example, Ingersoll (1984)[Equation 9]), i.e., $\sum_{i,j=1}^{\infty} \alpha_{is-1}\alpha_{js-1}\sigma^{(-1)}_{ij,s-1s-1} < \infty$ *a.s.*, but the latter simplifies to (2) under our assumptions.

[14] For my local-PCA to work, the data can be only mildly temporally dependent. De-meaning the observables induces the error term $e^*_{is} \equiv e_{is} - T^{-1}\sum_{r=1}^{T} e_{ir}$ and, except for large T, a case I rule out, the e^*_{is} are serially correlated even when the e_{is} are temporally independent.

summability condition (2), i.e., $\sum_{i=1}^{\infty} a_{i,s-1}^2 < \infty \ a.s.$, which arises from the no-arbitrage condition, see Assumption 1 and Proposition OA.4. Below, I will discuss, why this property arises and why it matters. Equation (3) above implies

$$\frac{\mathbf{X}\mathbf{X}'}{NT} = \frac{1}{T}\mathbf{F}\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N}\mathbf{F}' + \frac{1}{T}\hat{\mathbf{\Sigma}}_e, \tag{4}$$

where $\hat{\mathbf{\Sigma}}_e = \left(\mathbf{F}\mathbf{\Lambda}'\left(\boldsymbol{\alpha}' + \mathbf{e}'\right) + \left(\boldsymbol{\alpha} + \mathbf{e}\right)\mathbf{\Lambda}\mathbf{F}'\right)/N + \boldsymbol{\alpha}\boldsymbol{\alpha}'/N + \left(\boldsymbol{\alpha}\mathbf{e}' + \mathbf{e}\boldsymbol{\alpha}'\right)/N + \mathbf{e}\mathbf{e}'/N$. As $N \to \infty$, all terms of $\hat{\mathbf{\Sigma}}_e$ converge to zero, except the last. For the terms involving $\boldsymbol{\alpha}$, this follows from the square summability condition (2) on the entries in $\boldsymbol{\alpha}$ and some mild restrictions on $\mathbf{F}$, $\mathbf{\Lambda}$, and $\mathbf{e}$, and for the $\mathbf{e}\mathbf{e}'/N$ term this follows from the law of large numbers. Therefore $\hat{\mathbf{\Sigma}}_e$ has the same limit (in probability) as

$$\frac{1}{N}\sum_{i=1}^{N} \mathcal{E}(\mathbf{e}_i\mathbf{e}_i') \to \mathbf{\Sigma}_e, \tag{5}$$

where $\mathbf{e}_i \equiv (e_{it-T+1}, \ldots, e_{it})'$ for asset i. In words, the matrix $\mathbf{\Sigma}_e$ is the covariance matrix of the residuals e_{is} across dates $s = t - T + 1, \ldots, t$, averaged across i. Due to Assumption 7, the matrix $\mathbf{\Sigma}_e$ is diagonal. Likewise,

$$\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N} \to \mathbf{\Sigma}_\Lambda \tag{6}$$

is an $r \times r$ matrix, given by the sum $\boldsymbol{\mu}_\Lambda\boldsymbol{\mu}_\Lambda' + cov(\boldsymbol{\lambda})$ of the (limiting) cross-section outer product of the cross-sectional means (i.e., $\sum_{i=1}^{N} \boldsymbol{\lambda}_i/N \to \boldsymbol{\mu}_\Lambda$) and cross-sectional variance-covariance matrix of the $r \times 1$ risk exposure vectors $\boldsymbol{\lambda}_i$. I will call $\mathbf{\Sigma}_\Lambda$ (somewhat imprecisely) the risk exposure variance-covariance matrix.

As $N \to \infty$, taking the limit of (4) as $N \to \infty$,

$$\frac{\mathbf{X}\mathbf{X}'}{NT} \to_p \mathbf{M} \equiv \frac{1}{T}\mathbf{F}\mathbf{\Sigma}_\Lambda\mathbf{F}' + \frac{1}{T}\mathbf{\Sigma}_e. \tag{7}$$

The difference between that limit, i.e., $\mathbf{M}$, and the finite-N version, i.e., $\mathbf{X}\mathbf{X}'/NT$, allows me to provide central limit results and tests, but (7) is useful to gain a key insight. *The risk factors* $\mathbf{F}$*, and their number, can be consistently estimated as* $N \to \infty$*, up to a linear transformation, even for fixed* T*.*

The matrix $\mathbf{M}$ (or, more precisely, its finite-N counterpart) can be calculated from the data, is symmetric and can be decomposed into its eigenvalues and eigenvectors, i.e., $\mathbf{M} = T^{-1}\mathbf{V}\mathbf{U}\mathbf{V}'$, where $\mathbf{U}$ is diagonal. Notice that the eigenvectors of a symmetric matrix are real and orthogonal, and (w.l.o.g.) I have normalized $\mathbf{V}$ so that $\mathbf{V}\mathbf{V}'/T = \mathbf{I}_T$, the latter denoting the $T \times T$ identity matrix, to help the interpretation of the risk factors' estimator. By Assumption 7 the diagonal elements of $\mathbf{\Sigma}_e$ are all equal to some σ^2. Combining (7) with the spectral decomposition of $\mathbf{M}$ (and recalling the

normalization of $\mathbf{V}$) one gets $\mathbf{V}'\mathbf{M}\mathbf{V} = T\mathbf{U} = \mathbf{V}'\left(T^{-1}\mathbf{F}\mathbf{\Sigma}_\Lambda\mathbf{F}' + T^{-1}\mathbf{\Sigma}_e\right)\mathbf{V}$ or

$$\mathbf{U} - \frac{\sigma^2}{T}\mathbf{I}_T = \frac{\mathbf{V}'\mathbf{F}}{T}\mathbf{\Sigma}_\Lambda\frac{\mathbf{F}'\mathbf{V}}{T}. \tag{8}$$

Since the left-hand side is diagonal, it follows that $\mathbf{V}$ is also the eigenvector matrix for $\mathbf{F}\mathbf{\Sigma}_\Lambda\mathbf{F}'/T^2$.[15] The assumption that the cross-section expected variance of the error term e_{is} is constant, i.e., $\mathbf{\Sigma}_e = \sigma^2\mathbf{I}_T$, is essential for this result. With aggregate stochastic volatility (i.e., when averaged across assets) and therefore distinct values for the diagonal elements of $\mathbf{\Sigma}_e$, the eigenvectors of $\mathbf{M}$ are not the eigenvectors of $\mathbf{F}\mathbf{\Sigma}_\Lambda\mathbf{F}'/T^2$ and the factors cannot be consistently estimated if $N \to \infty$ but T stays fixed. In economic terms, one cannot distinguish aggregate volatility of the risk factors from aggregate (i.e., averaged across assets) volatility in the asset-specific error, unless T becomes large.[16]

It turns out that, given (8), there is therefore some $r \times r$ matrix $\mathbf{H}$ such that some r columns of $\mathbf{V}$ coincide with $\mathbf{F}\mathbf{H}$, and assume (w.l.o.g.) these to be the the first r columns of $\mathbf{V}$. As discussed at the beginning, $\mathbf{F}\mathbf{H}$ and $\mathbf{\Lambda}\mathbf{H}^{-1\prime}$ make an equally valid parameterization of (3). Since the column of $\mathbf{V}$ are orthogonal one another, with the first r of these columns satisfying $\mathbf{H}'\mathbf{F}'\mathbf{V}/T = (\mathbf{I}_r, \mathbf{0}_{r\times T-r})$, in view of $\mathbf{H}'\mathbf{F}'\mathbf{F}\mathbf{H}/T = \mathbf{I}_r$, where I re-write $(\mathbf{V}'\mathbf{F}/T)\mathbf{\Sigma}_\Lambda(\mathbf{F}'\mathbf{V}/T) = (\mathbf{V}'\mathbf{F}\mathbf{H}/T)\mathbf{H}^{-1}\mathbf{\Sigma}_\Lambda\mathbf{H}^{-1\prime}(\mathbf{H}'\mathbf{F}'\mathbf{V}/T)$, it follows that $\mathbf{H}^{-1}\mathbf{\Sigma}_\Lambda\mathbf{H}^{-1\prime}$ must be diagonal, and (8) becomes

$$\mathbf{U} - \frac{\sigma^2}{T}\mathbf{I}_T = \begin{pmatrix} \mathbf{I}_r \\ \mathbf{0}_{T-r\times r} \end{pmatrix}\mathbf{H}^{-1}\mathbf{\Sigma}_\Lambda\mathbf{H}^{-1\prime}(\mathbf{I}_r, \mathbf{0}_{r\times T-r}) = \begin{bmatrix} \mathbf{H}^{-1}\mathbf{\Sigma}_\Lambda\mathbf{H}^{-1\prime} & \mathbf{0}_{r\times T-r} \\ \mathbf{0}_{T-r\times r} & \mathbf{0}_{T-r\times T-r} \end{bmatrix}. \tag{9}$$

The first r diagonal entries of $\mathbf{U}$ are strictly larger than the last $T-r$, where the diagonal entries of $\mathbf{H}^{-1}\mathbf{\Sigma}_\Lambda\mathbf{H}^{-1\prime}$ are distinct by Assumption 6. Ordering the diagonal entries of $\mathbf{U}$ from largest to smallest then provides a unique solution $\mathbf{F}\mathbf{H}$ as the r eigenvectors corresponding to the largest element of $\mathbf{U}$, up to signs, i.e., up to multiplying each column with $+1$ or -1. This shows, that the factors are identified (and consistently estimated) for $N \to \infty$, up to some linear transformation $\mathbf{H}$. The magnitude of the diagonal entries if $\mathbf{U}$ is used to calculate the number of factors. In particular, my estimator selects r such that $T-r$ eigenvalues of $\mathbf{U} - \frac{\sigma^2}{T}\mathbf{I}_T$ are small (going to zero when N becomes large) or, more precisely, it selects r such that the difference between the $(r+2)$th and $(r+1)$th eigenvalues is small. Although this works well whenever one accidentally picks an $r' < r$, leading one to consider $r'+1$ and so on, this criterion will be silent when $r' > r$, i.e., would not distinguish r' from r'' when are both bigger than r. To overcome this difficulty, I need to introduce a penalization term (function of N) in my estimation criterion that pushes the estimator to disregard the irrelevant $r'-r$ small eigenvalues, leading to a consistent estimator for r. More in general, the distinct elements of $\mathbf{H}^{-1}\mathbf{\Sigma}_\Lambda\mathbf{H}^{-1\prime}$ should be well-separated and not be too small either, i.e., not too close to zero.

[15] Formal arguments are spelt in the Central Lemma OA.3.

[16] As T can be taken to be (almost) arbitrarily small, i.e., $T \geq r$, the assumption of constant error variance (when averaged across assets) over the sample interval, becomes mild if the stochastic volatility of the e_{is} is very persistent, i.e., smooth, and jumps are ruled out.

For ease of exposition here, assume $\boldsymbol{\alpha} = 0$ in (3). Armed with $\mathbf{FH}$, the OLS estimator $\hat{\boldsymbol{\Lambda}} \equiv \mathbf{X}'\mathbf{FH}(\mathbf{H}'\mathbf{F}'\mathbf{FH})^{-1} = \mathbf{X}'\mathbf{FH}/T$ provides the best fit and is an "estimator" of $\boldsymbol{\Lambda}\mathbf{H}^{-1\prime}$, with residuals $\hat{\mathbf{e}} = (\mathbf{I}_T - \mathbf{FHH}'\mathbf{F}'/T)\,\mathbf{X} = (\mathbf{I}_T - \mathbf{FHH}'\mathbf{F}'/T)\,\mathbf{e}$ and $\hat{\boldsymbol{\Sigma}}_\Lambda \equiv \hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}}/N \to_p \mathbf{H}^{-1}\boldsymbol{\Sigma}_\Lambda\mathbf{H}^{-1\prime} + (\sigma^2/T)\mathbf{I}_r \geq \mathbf{H}^{-1}\boldsymbol{\Sigma}_\Lambda\mathbf{H}^{-1\prime}$. If $T > r$, then σ^2 is consistently estimated by $\hat{\sigma}^2 \equiv tr(\hat{\mathbf{e}}\hat{\mathbf{e}}')/N(T-r) = tr[(\mathbf{I}_T - \mathbf{FHH}'\mathbf{F}'/T)\,\mathbf{ee}'/N]/(T-r) \to_p \sigma^2\,(T - tr(\mathbf{H}'\mathbf{F}'\mathbf{FH}/T))\,/(T-r) = \sigma^2$, with $tr(\cdot)$ denoting the trace, and therefore the risk exposures covariance matrix (up to some linear transformation) can be accurately estimated as $\hat{\boldsymbol{\Sigma}}_\Lambda - (\hat{\sigma}^2/T)\mathbf{I}_r$. Since T is fixed, residual uncertainty remains, even as $N \to \infty$. For a particularly simple case, suppose that $T = r = 1$. In that case, $\mathbf{FH} = 1$ and $\hat{\boldsymbol{\Lambda}} = \mathbf{X}'$ implying $\hat{\mathbf{e}} = \mathbf{0}_N'$, i.e., one obtains a perfect fit, and σ^2 can no longer be estimated. *Although the risk exposures cannot be accurately estimated, I can consistently estimate their moments, by suitable bias adjustments, up to a linear transformation, leveraging on a consistent estimator of σ^2 whenever $T > r$.*[17]

Consider now well-diversified portfolios with weights $\mathbf{w} \in \mathbb{R}^N$, i.e., $\mathbf{ew} \to_p \mathbf{0}_T$ as $N \to \infty$. I show how $\mathbf{x}^w \equiv \mathbf{Xw}$, i.e., the portfolio excess returns of investing w_i in asset i, $i = 1, \ldots, N$, pays off some weighted average $\mathbf{Fw}^F$ with $\mathbf{w}^F \in \mathbb{R}^r$, with the additional uncertainty vanishing as $N \to \infty$ and without arbitrage opportunities. In fact,

$$\mathbf{x}^w = \boldsymbol{\alpha}\omega + \mathbf{F}\,(\boldsymbol{\Lambda}'\mathbf{w}) + \mathbf{ew} \to_p \mathbf{F}\boldsymbol{\mu}_\Lambda^w,$$

the desired result, under mild conditions assuring the convergence, with the factors' weights $\mathbf{w}^F$ being given by the weighted mean (limit of) of the risk exposures $\boldsymbol{\mu}_\Lambda^w$. In particular, I need that $\boldsymbol{\alpha}\mathbf{w} \to_p \mathbf{0}_T$, which is assured via the square-summability condition (2), otherwise arbitrage opportunities arise. A portfolio excess returns with an identical performance (with large N) can be constructed using the local-PCA by replacing $\mathbf{X}$ with its estimated common component, i.e., $\hat{\mathbf{x}}^w \equiv \mathbf{FH}\hat{\Lambda}'\mathbf{w} \to_p \mathbf{Fw}^F$. To better understand the role of the no-arbitrage condition (2), assume $T = r = H = 1$, implying $\hat{\lambda}_i = \alpha_i + e_{i1} + \lambda_i$, assuming that the distribution of $e_{i1} + \lambda_i$ is symmetric around zero.[18] Set $w_i = 1/N$ for all $i = 1, \ldots, N$, i.e., consider the equally-weighted portfolio. As $N \to \infty$, the uncertainty in the payoff vanishes, given $\boldsymbol{\mu}_\Lambda^{1/N} = 0$, and the portfolio return $\mathbf{x}^{1/N}$ delivers the certain payoff $N^{-1}\sum_{i=1}^N \alpha_i \to \mu_\alpha$. Unless $\mu_\alpha = 0$, this violates no-arbitrage, and (2) ensures that this does not happen.

The no-arbitrage condition (2) plays a crucial role in my statistical analysis: it implies that I do not need to de-mean the data to account for $\boldsymbol{\alpha}$, which represents an intercept term for (3). De-meaning the data would lead to the error term $e_{is} - T^{-1}\sum_{v=1}^T e_{iv}$ in the estimated model, ruling out $\boldsymbol{\Sigma}_e = \sigma^2\mathbf{I}_T$, and thus invalidating the consistency property of the local-PCA estimator of $\mathbf{F}$, as

[17] In practice, as elicited in the next section, the above arguments continue to hold replacing $\mathbf{FH}$ by the PCA $\tilde{\mathbf{F}}$, namely the $\tilde{r}$ (the estimate of r) eigenvectors corresponding to the largest $\tilde{r}$ eigenvalues of $\mathbf{XX}'/NT$ (multiplied by $\sqrt{T}$), and $\hat{\boldsymbol{\Lambda}}$ is replaced by $\tilde{\boldsymbol{\Lambda}} = \mathbf{X}'\tilde{\mathbf{F}}/T$. Likewise, $\hat{\boldsymbol{\Sigma}}_\Lambda$ and $\hat{\sigma}^2$ are replaced by the corresponding PCA estimators.

[18] To be more specific, suppose that e_{i1} is normally distributed with mean zero and variance σ^2 and that $\lambda_i \in \{-1, 1\}$, drawn iid with equal probability. The density for $e_{i1} + \lambda_i$ is then the equal average of two normal densities with means -1 and 1.

explained above. *The no-arbitrage condition* (2)*, plays a crucial role in determining the large-N statistical properties of my local-PCA estimator.*

An important class of well-diversified portfolios are mean-variance portfolios, whose weights are linear in the risk exposures $\mathbf{\Lambda}$, which however remain un-precisely estimated by $\hat{\mathbf{\Lambda}}$ even for $N \to \infty$, Thus, the excess returns $\mathbf{X}$ and the estimates $\hat{\mathbf{\Lambda}}$ are both affected by $\mathbf{e}$. This induces a bias in the mean-variance portfolio returns (because it involves the product of $\mathbf{X}$ and $\hat{\mathbf{\Lambda}}$ inducing averages in the squared $\mathbf{e}$), which can nevertheless be estimated and removed, just like I did for correcting the estimated moments of the risk exposures. *I can construct mean-variance portfolio returns that are immune to the estimation noise affecting their weights, i.e., affecting* $\hat{\mathbf{\Lambda}}$*, achieving the same performance as if I was able to estimate the* $\mathbf{\Lambda}$ *accurately.*

In empirical asset pricing, we are interested in estimating risk premia and the SDF. *As my local-PCA can estimate the risk factors precisely even when* T *is fixed, I can construct the risk premia and the SDF as if one would observe the true* $\mathbf{F}$ *because these objects are only functions of the risk factors.* However, fixing T raises an identification issue, namely one can only identify the ex-post risk premia $\bar{\mathbf{F}} \in \mathbb{R}^r$ (see Shanken (1992)), i.e., the cross-time average of the risk factors which is a rather inaccurate measure of the ex-ante risk premia $\boldsymbol{\gamma}_{s-1} = \mathcal{E}_{s-1}(\mathbf{F}_s)$ for small T, given by (see footnote 12)

$$\bar{\mathbf{F}} \equiv \bar{\boldsymbol{\gamma}} + \bar{\mathbf{f}} - T^{-1} \sum_{s=t-T+1}^{t} \mathcal{E}_{s-1}(\mathbf{f}_s), \tag{10}$$

where the barred quantities indicate time averages. In fact, by averaging (1) across the T observations one obtains $\bar{x}_i \approx \bar{\alpha}_i + \boldsymbol{\lambda}_i' \bar{\mathbf{F}} + \bar{e}_i$, i,.e., the time-average of excess returns $\bar{x}_i$ is *linear* in $\bar{\mathbf{F}}$ with (locally) constant risk exposures $\boldsymbol{\lambda}_i$, that is $\bar{\mathbf{F}}$ *prices* the cross-section of average returns. For all practical purposes, i.e., estimation and use of the asset pricing model, the time-variation of the parameters can be ignored under my assumptions without affecting my asymptotics, allowing one to work with a model with (locally) constant parameters over the T observations.[19]

Likewise, I can identify (net of rotation) the ex-post sample cross-time variance-covariance matrix $\hat{\mathbf{\Omega}} \equiv \sum_{s=t-T+1}^{t} (\mathbf{F}_s - \bar{\mathbf{F}})(\mathbf{F}_s - \bar{\mathbf{F}})'/T$ of $\mathbf{F}$, which is also an inaccurate measure of the true (ex-ante) factor covariance matrix $\mathbf{\Omega}_{s-1} \equiv var_{s-1}(\mathbf{F}_s)$. With this and the risk-free return r_f, one can identify the *ex-post* SDF

$$\mathcal{M}_{s-1,s} \equiv \frac{1}{r_f} - \frac{1}{r_f} \bar{\mathbf{F}} \hat{\mathbf{\Omega}}^{-1} (\mathbf{F}_s - \bar{\mathbf{F}}) \text{ for every } s \in \mathcal{T}_t, \tag{11}$$

a noisy proxy of the true (ex-ante) SDF $m_{s-1,s} \equiv 1/r_f - (1/r_f)\mathcal{E}_{s-1}(\mathbf{F}_s')\mathbf{\Omega}_{s-1}^{-1}(\mathbf{F}_s - \mathcal{E}_{s-1}(\mathbf{F}_s))$ (see, for example, Cochrane (2005)[Chapters 6 and 8]). Note that the SDF is immune, i.e., is not influenced by, the rotation matrix $\mathbf{H}$, unlike the risk premia. The cross-time average of the ex-post SDF satisfies $\overline{\mathcal{M}} = 1/r_f$, and $T^{-1} \sum_{s=t-T+1}^{t} \mathcal{M}_{s-1,s} \hat{x}_s^w = 0$ for the estimated portfolio returns $\hat{\mathbf{x}}^w$, the counterparts

[19] The strength of my methodology, not requiring the specification of parameters' time-variation, implies however that the analysis of how parameters change from one time-window to the other is necessarily excluded, as it would require a fully-parametric approach and thus prone to misspecification.

to $1 = \mathcal{E}(M_{s-1,s} r_f)$ and $0 = \mathcal{E}(M_{s-1,s} \mathbf{F}'_s \mathbf{w}^F)$.[20] When using the ex-post risk premia and the ex-post SDF for pricing, one commits pricing errors that are a mixture of the true pricing performance of the model, i.e., a function of the ex-ante risk prema $\mathcal{E}_{s-1}(\mathbf{F}_s)$ and SDF $m_{s-1,s}$, and of the small-T estimation noise, which can be substantial for small T. *I am able to separate the estimation noise and the pricing errors, leading to formal tests of pricing ability for the ex-ante risk premia and SDF despite not being identified.*

4 Estimation: Local-PCA with Large N

In this section, I present the limiting distribution theory for the local-PCA estimated risk factors, including consistent estimation of their asymptotic covariance matrix and of the time-varying number of risk factors, the proofs of which are relegated to the final appendix and Section OA9. The asymptotic analysis of the local-PCA represents my first contribution.

Hereafter, I assume that data and estimators always refer to the time interval $\mathcal{T}_t \equiv (t-T+1, \cdots, t)$ of length T for a *given* t, although for simplicity I will not make this always explicit, in contrast to the population quantities which are always indexed by a time indicator.[21] I will *always* refer to population quantities, and seldom to sample quantities, indexed by the letter t when these are constant within the interval $\mathcal{T}_t$, but can change from interval to interval.

My local-PCA relies on several technical assumptions listed in the final Appendix (Section 10) but I introduce some of them now because these are pivotal to the *local* nature of my analysis, allowing me to express the conditional factor model (1) as a *locally*-constant-parameters model. While my derivations rely on these assumptions, there are several other technical assumptions inherent to local-PCA, which I list in the Appendix to maintain the ease of the discussion.

Specifically, I require the risk exposures and the number of risk factors to satisfy the following smoothing conditions. Hereafter $\| \mathbf{A} \| = [tr(\mathbf{A}'\mathbf{A})]^{\frac{1}{2}}$ denotes the Frobenius norm of a generic matrix $\mathbf{A}$, and $a_N = O(b_N), a_N = o(b_N)$ mean that $|a_N|/b_N < C$, some $C > 0$ and that $|a_N|/b_N \to 0$, as $N \to \infty$, respectively, for deterministic sequences $\{a_N\}$ and $\{b_N\}$, writing $o_p(\cdot), O_p(\cdot)$ when these statements hold in probability.

Assumption 2 (number of factors - smoothness) *For every $s \in \mathcal{T}_t$, the number of risk factors is locally constant:*

$$r_s = r_{t-1}.$$

Assumption 3 (risk exposures - smoothness) *For $\mathbf{\Lambda}_s \equiv (\boldsymbol{\lambda}_{1s} \cdots \boldsymbol{\lambda}_{Ns})'$,*

$$sup_{s \in \mathcal{T}_t} \| \mathbf{\Lambda}_s - \mathbf{\Lambda}_{t-1} \|^2 = o(N^{\frac{1}{2}}).$$

[20] The ex-post risk premia $\bar{\mathbf{F}}$ and ex-post SDF $\mathcal{M}_{s-1,s}$, although different from their ex-ante counterparts especially for small T, are valuable objects for pricing purposes. For instance, $\mathcal{E}(\mathcal{M}_{s-1,s} x_{is}) = O(T^{-1})$ despite $\mathcal{M}_{s-1,s}$ and $m_{s-1,s}$ differ by an $O(T^{-1/2})$ term (see Proposition OA.2 for the proof and discussion) leading to robust pricing performance, corroborated by the empirical application.

[21] For example, if $T = 24$, then $\mathcal{T}_t$ is comprised by the observations from $t - 23$ to t.

The number of risk factors is allowed to change over time but I rule out that such change occurs within the time interval of size T. In terms of risk exposures, several interesting examples of time-variation are comprised by Assumption 3.[22] For example, one can have $\mathbf{\Lambda}_{s-1} = \mathbf{\Lambda}_0 + \mathbf{\Lambda}_1 \mathbf{z}_{s-1}$, for some matrices of parameters $\mathbf{\Lambda}_0, \mathbf{\Lambda}_1$, with the condition $\mathbf{\Lambda}_1' \mathbf{\Lambda}_1 = o(N^{\frac{1}{2}})$, where $\mathbf{z}_{s-1}$ is a vector of observed variables. The time-variation permitted by my approach includes, for example, the case when $\mathbf{z}_{s-1}$ behaves like a (multidimensional) random walk. A further smoothing assumption is required in terms of the time behaviour of the variances and covariances of the idiosyncratic error e_{is}, implying them to be only mildly temporally dependent with the time interval of size T, building on Connor and Korajczyk (1986) and Bai (2003), ruling out aggregate (i.e., when averaged across assets) time-heteroskedasticity of the e_{is}.[23] In particular, I need $N^{-1} \sum_{i=1}^{N} var(e_{is})$ to converge the locally-constant σ_{t-1}^2, at some rate for every $s \in \mathcal{T}_t$ (see Assumption 7). This is laid out in the final Appendix, containing the additional regularity conditions (Assumptions 4 to 9) required by my local-PCA. As the time window of size T can be taken almost arbitrarily small, these smoothness assumptions appear relatively mild.

For estimation, I replace model (1) with a *locally*-constant-parameters model (by Assumptions 2 and 3) without intercept (by Assumption 1), leading to the local-PCA estimator for the risk exposures and the risk factors, obtained by solving the following well-known least-squares problem:

$$\tilde{\mathbf{F}}, \tilde{\mathbf{\Lambda}} \quad = argmin_{\mathbf{F},\mathbf{\Lambda}} \frac{1}{NT} \sum_{s \in \mathcal{T}_t} \sum_{i=1}^{N} (x_{is} - \mathbf{F}_s' \boldsymbol{\lambda}_i)^2, \tag{12}$$

where $\mathbf{X} \equiv (\mathbf{x}_{\cdot 1} \cdots \mathbf{x}_{\cdot i} \cdots \mathbf{x}_{\cdot N}) \equiv (\mathbf{x}_{t-T+1.} \cdots \mathbf{x}_{s.} \cdots \mathbf{x}_{t.})'$, setting $\mathbf{x}_{\cdot i} \equiv (x_{it-T+1} \cdots x_{is} \cdots x_{it})'$ and $\mathbf{x}_{s.} \equiv (x_{1s} \cdots x_{is} \cdots x_{Ns})'$, with $\tilde{\mathbf{F}}$ being equal to the set of r_{t-1} eigenvectors, multiplied by $\sqrt{T}$, corresponding to the largest r_{t-1} eigenvalues of $\mathbf{X}\mathbf{X}'/NT$, hence satisfying $\tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T = \mathbf{I}_{r_{t-1}}$ (the identity matrix of size $r_{t-1} \times r_{t-1}$) and $\tilde{\mathbf{\Lambda}} \equiv \mathbf{X}'\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1} = \mathbf{X}'\tilde{\mathbf{F}}/T$.[24]

The asymptotic properties, as $N \to \infty$, of the local-PCA estimator are established below. Hereafter, $\to_p$ and $\to_d$ denote convergence in probability and distribution, respectively, and $\mathbf{0}_{a \times b}$ denotes

[22] Consistency of the local-PCA requires the slightly weaker condition $sup_{s \in \mathcal{T}_t} \parallel \mathbf{\Lambda}_s - \mathbf{\Lambda}_{t-1} \parallel^2 = o(N)$, whereas the limiting distribution result demands the $o(N^{\frac{1}{2}})$-rate. I do not distinguish between these two cases to simplify the exposition. For a similar smoothness condition see Bates et al. (2013), although there the rate of $sup_{s \in \mathcal{T}_t} \parallel \mathbf{\Lambda}_s - \mathbf{\Lambda}_{t-1} \parallel^2$ involves both N and T because they study the constant-PCA.

[23] My assumption permits time-heteroskedasticity of the e_{is} that evolves independently across stocks (or with some mild deviation from it) but rules out a common component in the e_{is} second-moment dynamics, except when the latter changes very slowly, e.g., when extremely persistent (see also footnote 16).

[24] See Magnus and Neudecker (2007), Theorem 7 of Chapter 17, for an elegant proof. An asymptotically equivalent PCA estimator can be obtained by setting $\hat{\mathbf{\Lambda}}$ equal to the eigenvectors of $\mathbf{X}'\mathbf{X}/NT$, multiplied by $\sqrt{N}$, corresponding to the r_{t-1} largest eigenvalues, and $\hat{\mathbf{F}} = \mathbf{X}\hat{\mathbf{\Lambda}}/N$ (see Bai and Ng (2002)). As N is much larger than T, the latter approach is computationally less convenient than the one described above. Without loss of generality for N, assume that $N > T \geq r_{r-1}$, as I will only allow N to diverge, where $T \geq r_{t-1}$ avoids the existence of the zero eigenvalues of certain relevant matrixes.

the $a \times b$ matrix of zeros.[25]

Theorem 1 (rate of convergence and limiting distribution of the risk factors estimator) *Under Assumptions 1-9, as $N \to \infty$ and for every $s \in \mathcal{T}_t$, $\tilde{\mathbf{F}} \equiv (\tilde{\mathbf{F}}_{t-T+1}, \cdots, \tilde{\mathbf{F}}_s, \cdots, \tilde{\mathbf{F}}_t)'$ satisfy*

(i) $\| \tilde{\mathbf{F}}_s - \tilde{\mathbf{H}}'_{t-1}\mathbf{F}_s \| = O_p(N^{-\frac{1}{2}})$.

(ii) $\sqrt{N}\left(\tilde{\mathbf{F}}_s - \tilde{\mathbf{H}}'_{t-1}\mathbf{F}_s\right) \to_d N(\mathbf{0}_{r_{t-1}}, \mathbf{A}'_{t-1}\mathbf{B}_{s,t-1}\mathbf{A}_{t-1})$, *setting* $\mathbf{A}_t \equiv (\mathbf{I}_{r_t}, \mathbf{I}_{r_t}, \mathbf{I}_{r_t})'$ *and*

$$\mathbf{B}_{s,t} \equiv \begin{bmatrix} \frac{\sigma_{4t}}{T}\mathbf{U}_t^{-2} + \frac{(\kappa_{4t}+\sigma_{4t})}{T^2}\mathbf{U}_t^{-1}\mathbf{H}'_t\mathbf{F}_s\mathbf{F}'_s\mathbf{H}_t\mathbf{U}_t^{-1} & \mathbf{0}_{r_t\times r_t} & \mathbf{0}_{r_t\times r_t} \\ \mathbf{0}_{r_t\times r_t} & \sigma_t^2\mathbf{U}_t^{-1} & \frac{\sigma_t^2}{T}\mathbf{U}_t^{-1}\mathbf{Q}_t\mathbf{\Sigma}_{\mathbf{\Lambda}_t}\mathbf{F}_s\mathbf{F}'_s\mathbf{H}_t\mathbf{U}_t^{-1} \\ \mathbf{0}_{r_t\times r_t} & \frac{\sigma_t^2}{T}\mathbf{U}_t^{-1}\mathbf{H}'_t\mathbf{F}_s\mathbf{F}'_s\mathbf{\Sigma}_{\mathbf{\Lambda}_t}\mathbf{Q}'_t\mathbf{U}_t^{-1} & \frac{\sigma_t^2}{T}(\mathbf{F}'_s\mathbf{\Sigma}_{\mathbf{\Lambda}_t}\mathbf{F}_s)\mathbf{U}_t^{-2} \end{bmatrix},$$

where $\mathbf{F} \equiv (\mathbf{F}_{t-T+1}\cdots\mathbf{F}_s\cdots\mathbf{F}_t)'$, $\hat{\mathbf{\Sigma}}_{\mathbf{F}}$, $\mathbf{\Sigma}_{\mathbf{\Lambda}_t}$, σ_t^2, σ_{4t}, and κ_{4t}, are defined in Assumptions 4, 5, and 7, respectively, $\mathbf{U}_t$ denotes the diagonal $r_t \times r_t$ matrix of distinct eigenvalues of $\mathbf{\Sigma}_{\mathbf{\Lambda}_t}^{\frac{1}{2}}\hat{\mathbf{\Sigma}}_{\mathbf{F}}\mathbf{\Sigma}_{\mathbf{\Lambda}_t}^{\frac{1}{2}}$ with corresponding eigenvectors $\mathbf{\Upsilon}_t$, $\mathbf{Q}_t \equiv \mathbf{U}_t^{\frac{1}{2}}\mathbf{\Upsilon}'_t\mathbf{\Sigma}_{\mathbf{\Lambda}_t}^{-\frac{1}{2}}$, for a.s.-nonsingular matrices

$$\tilde{\mathbf{H}}_t \equiv \left(\frac{\mathbf{\Lambda}'_t\mathbf{\Lambda}_t}{N}\right)\left(\frac{\mathbf{F}'\tilde{\mathbf{F}}}{T}\right)\breve{\mathbf{U}}_t^{-1}, \quad \mathbf{H}_t \equiv \mathbf{\Sigma}_{\mathbf{\Lambda}_t}^{\frac{1}{2}}\mathbf{\Upsilon}_t\mathbf{U}_t^{-\frac{1}{2}} = \mathbf{Q}_t^{-1} = \mathbf{\Sigma}_{\mathbf{\Lambda}_t}\mathbf{Q}'_t\mathbf{U}_t^{-1},$$

with $\breve{\mathbf{U}}_t$ denoting the diagonal $r_t \times r_t$ matrix of the largest r_t eigenvalues of $\left(\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{T}\sigma_t^2\mathbf{I}_T\right)$.

The proof builds on the fact that the PCA-estimated risk factors can be interpreted as portfolio returns (i.e., weighted cross-sectional averages) in the estimated risk exposures or, alternatively, as the estimated coefficients when regressing returns on the (estimated) risk exposures. As said, the latter remain very noisy with fixed T but their estimation noise dissipates as $N \to \infty$ when the risk exposures are used as weights in cross-averages. Factors are estimated up to an (unknown) rotation, induced by the matrix $\tilde{\mathbf{H}}_{t-1}$, which can be written as $\tilde{\mathbf{H}}^{\dagger}_{t-1}\tilde{\mathbf{J}}_{t-1}$, where the (non-diagonal) matrix $\tilde{\mathbf{H}}^{\dagger}_{t-1}$ normalizes the *magnitude* of the estimated risk factors, whereas the diagonal matrix $\tilde{\mathbf{J}}_{t-1}$ is made by the numbers $+1$ and -1 (see Section OA4 for details).[26]

To construct confidence intervals and perform hypothesis testing on the estimated factors, one needs to estimate consistently the asymptotic covariance matrix reported in Theorem 1-(ii). This involves the estimation of $\mathbf{\Sigma}_{\mathbf{\Lambda}_{t-1}}$, which appears particularly problematic because the $\boldsymbol{\lambda}_{it-1}$ are not consistently estimated when T is fixed. In particular, using Theorem 1-(i) and setting $\mathbf{e}_{.i} \equiv (e_{it-T+1}\cdots e_{it})'$ and $\tilde{\mathbf{\Lambda}} \equiv (\tilde{\boldsymbol{\lambda}}_1, \cdots, \tilde{\boldsymbol{\lambda}}_N)'$, one gets, for some $\mathbf{l}_{it-1}$ satisfying $N^{-1}\sum_{i=1}^N \mathbf{l}_{it-1}\mathbf{l}'_{it-1} =$

[25] My local-PCA estimator (12) can be interpreted as Nadaraya-Watson nonparametric estimator with a rectangular window and a bandwidth $h \equiv T/T_0$, with $T_0 > T$ denoting the overall time-series length. I do not pursue this interpretation any further because my fixed-T setting rules out any large-T asymptotic analysis, including the optimal selection of the bandwidth h.

[26] For some empirical analyses, one needs to reconstruct the *glued* time series of the estimated risk factors across a sequence of periods longer than T, which involves mapping the sequence of rotation matrices $\tilde{\mathbf{H}}_{t-1}$ (one for each sub-sample of length T) into a single rotation matrix (see Proposition OA.5 for details).

$o(1)$,[27]

$$\tilde{\boldsymbol{\lambda}}_i \to_p \mathbf{H}_{t-1}^{-1}\boldsymbol{\lambda}_{it-1} + \frac{\mathbf{H}_{t-1}'\mathbf{F}'\mathbf{e}_{.i}}{T} + \mathbf{l}_{it-1} \text{ as } N \to \infty. \tag{13}$$

However, despite (13), I can still estimate consistently the first-moment vector and the second-moment matrix of the $\boldsymbol{\lambda}_{it-1}$ for T fixed.[28] In fact, averaging both sides of (13) and letting $N \to \infty$, one obtains $N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{\lambda}}_i \to_p \mathbf{H}_{t-1}^{-1}\boldsymbol{\mu}_{\Lambda_{t-1}}$ as the (cross-sectional) average of the bias in (13) is washed out, with $\boldsymbol{\mu}_{\Lambda_{t-1}}$ denoting the limit of the average of the true risk exposures. Along the same lines, $N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{\lambda}}_i\tilde{\boldsymbol{\lambda}}_i' \to_p \mathbf{H}_{t-1}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_{t-1}}\mathbf{H}_{t-1}^{-1\prime} + (\sigma_{t-1}^2/T)\mathbf{I}_{r_{t-1}}$, where now a bias arises due to the averaging of the squared e_{is}. A suitable bias adjustment, based on a consistent estimator for σ_{t-1}^2 is now required, given by $\tilde{\sigma}^2 \equiv T/(T-r_t)V(r_t)$ with $V(r_t) \equiv (1/NT)tr\left(\tilde{\mathbf{e}}\tilde{\mathbf{e}}'\right)$, using the local-PCA residuals $\tilde{\mathbf{e}} \equiv (\mathbf{X} - \tilde{\mathbf{F}}\tilde{\boldsymbol{\Lambda}}')$. In particular, Theorem OA.3 establishes $\| \tilde{\mathbf{B}}_{s,t} - \mathbf{B}_{s,t} \| \to_p \mathbf{0}$ for every $s \in \mathcal{T}_t$, as $N \to \infty$, under Assumptions 1-9, for a feasible estimator $\tilde{\mathbf{B}}_{s,t}$. For this result, I need to set κ_{4t-1}, the cross-sectional average fourth-order cumulant of the e_{it}, equal to zero for identification purposes (see Lemma OA.4).[29]

4.1 Number of Factors

The above arguments implicitly assume that the number of factors r_{t-1} is known. In practice, one replaces r_{t-1} with a consistent estimator, which I now describe, asymptotically valid when $N \to \infty$ and T remains fixed. For some function $g(\cdot)$ of N, let

$$PC_{\mathcal{T}_t}(k) \equiv (\frac{T}{T-k})V(k) + kg(N),$$

with $V(k) \equiv (NT)^{-1}tr\left((\mathbf{X} - \tilde{\mathbf{F}}^k\tilde{\boldsymbol{\Lambda}}^{k\prime})(\mathbf{X} - \tilde{\mathbf{F}}^k\tilde{\boldsymbol{\Lambda}}^{k\prime})'\right)$, with $\tilde{\mathbf{F}}^k$ denoting the $T \times k$ matrix of the (normalized) k eigenvectors associated with the k largest eigenvalues of $\mathbf{X}\mathbf{X}'/NT$. Similarly, $\tilde{\boldsymbol{\Lambda}}^k = \mathbf{X}'\tilde{\mathbf{F}}^k(\tilde{\mathbf{F}}^{k\prime}\tilde{\mathbf{F}}^k)^{-1} = T^{-1}\mathbf{X}'\tilde{\mathbf{F}}^k$ denotes the $N \times k$ matrix of estimated risk exposures.

Theorem 2 (consistent estimation of the number of factors) *Set* $\tilde{r} \equiv \text{argmin}_{0\le k\le T} PC_{\mathcal{T}_t}(k)$. *Under Assumptions 1-9, and for* $g(N) \to 0$ *together with* $\sqrt{N}g(N) \to \infty$,

$$Prob(\tilde{r} = r_{t-1}) \to 1 \text{ as } N \to \infty.$$

Although the existing literature offers several methods for estimating the number of latent risk factors (see the literature review in Section OA2), these methods are only applicable when both N and T

[27] Set $\mathbf{l}_{it-1} \equiv T^{-1}\mathbf{H}_{t-1}'\sum_{s\in\mathcal{T}_t}\mathbf{F}_s\mathbf{F}_s'(\boldsymbol{\lambda}_{is-1} - \boldsymbol{\lambda}_{it-1}) + T^{-1}\mathbf{H}_{t-1}'\sum_{s\in\mathcal{T}_t}\mathbf{F}_s\alpha_{is-1}$.

[28] The lack of consistency of the risk exposures is *unrelated* to the latency of the factors (except for the presence of the rotation matrix): It will still hold when the $\mathbf{F}_t$ are observed risk factors and T is finite.

[29] Higher-order cumulants of the e_{it} are not constrained to be zero, implying that $\kappa_{4t-1} = 0$ is not equivalent to Gaussianity. Moreover, fat-tails of the e_{is} can still be manifested employing a mixed-distribution argument, whereby σ_{4t-1} (the cross-sectional average of the squared variance of the e_{is}) is much larger than the square of σ_{t-1}^2 (the square of the cross-sectional average variance of the e_{is}). In the data, the estimated σ_{4t-1} is more than ten times larger than the estimated squared σ_{t-1}^2.

tend to infinity, ruling them out in my local fixed-T setting. This is the gap filled by Theorem 2, whose proof spells out the properties required by the function $g(.)$ to be a suitable penalization function when T remains fixed (i.e., $g(N) + 1/(\sqrt{N}g(N)) \to 0$ as $N \uparrow \infty$). As one can replace r_{t-1} with $\tilde{r}$ without affecting the asymptotic analysis (by standard asymptotic arguments), I will assume hereafter that r_{t-1} is known. As a by-product of Theorem 2, I show that when an erroneously larger number of risk factors is considered, that is $k > r_{t-1}$, the risk premia of the $\tilde{\mathbf{F}}^k$ consist of a linear combination of the risk premia associated with the correct number (viz. r_{t-1}) of risk factors.[30]

Finally, note that a model with a time-varying number of risk factors and one with a constant number of factors, together with risk exposures of varying magnitude or varying cross-sectional dependence, are observationally equivalent. This un-determinacy can only be avoided by a fully specified parameterization of the time-variation, say employing observed instruments, although this could induce misspecification when the wrong set of instruments is pre-selected. My local-PCA methodology departs from this approach.

5 Asset Pricing Implications

Local Estimation with Large N. Given my local-PCA methodology, I construct and analyze *local* estimators for risk premia, the SDF, and mean-variance portfolios. This represents my second contribution, the proofs of which are presented in Sections OA9 and OA12. A natural estimator for $\bar{\mathbf{F}}$ is $\bar{\tilde{\mathbf{F}}} \equiv T^{-1}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s$, as the vector of risk premia $\bar{\mathbf{F}}$ is the sample mean of the true risk factors. My estimator is also economically justified as $\tilde{\mathbf{F}}_s$ is (asymptotically) equivalent to the vector of (estimated) mimicking portfolio returns (see Section OA5).[31] Interestingly, $\bar{\tilde{\mathbf{F}}}$ turns out to be numerically identical to the two-pass estimator (see Proposition OA.3), that is $\bar{\tilde{\mathbf{F}}} = (\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Lambda}})^{-1}\tilde{\mathbf{\Lambda}}'\bar{\mathbf{x}}_N$, setting $\bar{\mathbf{x}}_N \equiv (\bar{x}_1, \cdots, \bar{x}_i, \cdots, \bar{x}_N)'$, providing further support to my risk premia estimator.[32]

For the risk premia estimator, by Theorem OA.4,

$$\| \bar{\tilde{\mathbf{F}}} - \tilde{\mathbf{H}}'_{t-1}\bar{\mathbf{F}} \| = O_p(N^{-\frac{1}{2}}) \text{ and } \sqrt{N}(\bar{\tilde{\mathbf{F}}} - \tilde{\mathbf{H}}'_{t-1}\bar{\mathbf{F}}) \to_d N(\mathbf{0}_{r_{t-1}}, \mathbf{A}'_{t-1}\mathbf{C}_{t-1}\mathbf{A}_{t-1}),$$

with $\| \tilde{\mathbf{C}}_t - \mathbf{C}_t \| \to_p 0$, for some estimator $\tilde{\mathbf{C}}_t$ of $\mathbf{C}_t$, and where $\mathbf{A}_{t-1}$ is defined in Theorem 1.

For the SDF, I construct the local-PCA estimator as

$$\widetilde{\mathcal{M}}_{s-1,s} \equiv \frac{1}{r_{fs-1}} - \frac{1}{r_{fs-1}}\bar{\tilde{\mathbf{F}}}\tilde{\mathbf{\Omega}}^{-1}(\tilde{\mathbf{F}}_s - \bar{\tilde{\mathbf{F}}}) \text{ for every } s \in \mathcal{T}_t, \tag{14}$$

[30] This suggests that the use of the local-PCA avoids the problems arising from erroneously including spurious (observed) factors, i.e., factors that are uncorrelated with the test assets (see Kan and Zhang (1999)) or, more generally, weak factors (see Lettau and Pelger (2020b)). See Section OA15.5 for further arguments.

[31] A mimicking portfolio for $\mathbf{F}_t$ is given by $\mathbf{F}_t^{mp} \equiv \mathcal{E}_{t-1}(\mathbf{F}_t\mathbf{x}'_{t.})(\mathcal{E}_{t-1}(\mathbf{x}_{t.}\mathbf{x}_{t.}))^{-1}\mathbf{x}_{t.} = \mathbf{w}_{Nt-1}^{mp\prime}\mathbf{x}_{t.}$ (see Huberman et al. (1987), with its local-PCA estimator being $\tilde{\mathbf{F}}_t^{mp} \equiv \left(\bar{\tilde{\mathbf{F}}}\bar{\tilde{\mathbf{F}}}' + \tilde{\mathbf{\Omega}}\right)\tilde{\mathbf{\Lambda}}'\left((\tilde{\mathbf{\Lambda}}\bar{\tilde{\mathbf{F}}})(\tilde{\mathbf{\Lambda}}\bar{\tilde{\mathbf{F}}})' + \tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Omega}}\tilde{\mathbf{\Lambda}}' + \tilde{\sigma}^2\mathbf{I}_N\right)^{-1}\mathbf{x}_{t.}$.

[32] The two-pass estimator of Fama and MacBeth (1973) is the most popular risk premia estimator when considering observable non-tradable risk factors. Shanken (1992) and Jagannathan and Wang (1998) provide the large-T asymptotics, Bai and Zhou (2015) and Gagliardini et al. (2016) study the large-(N,T) case, and Raponi et al. (2020) consider the case when N diverges and T is fixed.

setting $\tilde{\mathbf{\Omega}} \equiv T^{-1}\sum_{s\in\mathcal{T}_t}(\tilde{\mathbf{F}}_s - \bar{\tilde{\mathbf{F}}})(\tilde{\mathbf{F}}_s - \bar{\tilde{\mathbf{F}}})' = \mathbf{I}_{r_{t-1}} - \bar{\tilde{\mathbf{F}}}\bar{\tilde{\mathbf{F}}}'$, which, by Theorem OA.5, satisfies

$$\| \widetilde{\mathcal{M}}_{s,s+1} - \mathcal{M}_{s,s+1} \| = O_p(N^{-\frac{1}{2}}) \text{ and } \sqrt{N}\Big(\widetilde{\mathcal{M}}_{s,s+1} - \mathcal{M}_{s,s+1}\Big) \to_d N(0, \mathbf{D}'_{s,t-1}\mathbf{E}_{t-1}\mathbf{D}_{s,t-1}) \text{ for every } s \in \mathcal{T}_t,$$

with $\| \tilde{\mathbf{D}}_{s,t} - \mathbf{D}_{s,t} \| \to_p 0$, $\| \tilde{\mathbf{E}}_t - \mathbf{E}_t \| \to_p 0$, for some estimators $\tilde{\mathbf{D}}_{s,t}$ and $\tilde{\mathbf{E}}_{s,t}$ of $\mathbf{D}_{s,t}$ and $\mathbf{E}_{s,t}$, respectively, defined in the statement of the theorem.

Assessing the economic significance of using my methodology can be based on evaluating the performance of mean-variance portfolio strategies constructed using the local-PCA estimated asset pricing model, such as for the tangency portfolio excess returns $r_t^{tg} \equiv \mathbf{w}_{t-1}^{tg\prime}\mathbf{x}_{t.}$, namely the mean-variance portfolio with the largest Sharpe ratio, setting $\mathbf{w}_{t-1}^{tg} \equiv \mathbf{\Sigma}_{xt-1}^{-1}\mathbf{\Lambda}_{t-1}\boldsymbol{\gamma}_{t-1}/(\mathbf{1}'_N\mathbf{\Sigma}_{xt-1}^{-1}\mathbf{\Lambda}_{t-1}\boldsymbol{\gamma}_{t-1})$, with $\mathbf{\Sigma}_{xt-1} \equiv \big(\mathbf{\Lambda}_{t-1}\mathbf{\Omega}_{t-1}\mathbf{\Lambda}'_{t-1} + \sigma^2_{t-1}\mathbf{I}_N\big)$.[33] As the local-PCA risk exposures' estimates remain biased for finite T, a plug-in estimator for the $\mathbf{w}_{t-1}^{tg}$ is inaccurate. Despite this, I can estimate the tangency portfolio *return* r_t^{tg} accurately by exploiting the averaging induced by $\mathbf{w}_{t-1}^{tg\prime}\mathbf{x}_{t.}$, as $N \to \infty$.[34] See Section OA12 for details and extension to the other types of mean-variance portfolios.

Testing Pricing Performance with Large N. In my third contribution, I derive formal testing procedures to assess the pricing ability of my conditional asset pricing model (1), hence evaluating the effect of the model-free time-variation for my asset pricing model (1) in terms of *economic* criteria. When the asset pricing model (1) is correctly specified, it follows that the pricing errors $\delta_{is-1} \equiv \mathcal{E}_{s-1}(x_{is}) - \boldsymbol{\lambda}'_{is-1}\mathcal{E}_{s-1}(\mathbf{F}_s)$, satisfy $\delta_{is-1} = \alpha_{is-1}$ for every i, and $\boldsymbol{\gamma}_{s-1} = \mathcal{E}_{s-1}(\mathbf{F}_s)$ is pricing correctly the cross-section of expected excess returns $\mathcal{E}_{s-1}(x_{is})$.[35]Alternatively, when considering the pricing ability through the lenses of the SDF one examines $h_{is-1} \equiv \mathcal{E}_{s-1}(m_{s-1,s}x_{is}) = \delta_{is-1}/r_{fs-1}$, yielding $h_{is-1} = \alpha_{is-1}/r_{fs-1}$ for every i under correct model specification because $m_{s-1,s}$ is pricing correctly the cross-section of payoffs x_{is}. In this case δ_{is-1} and h_{is-1} are negligible by (2), given that $N^{-1}\sum_{i=1}^N h^2_{is-1}$ and $N^{-1}\sum_{i=1}^N \delta^2_{is-1}$ are both $O(N^{-1})$, whereas $\sum_{i=1}^N h^2_{is-1}$ and $\sum_{i=1}^N \delta^2_{is-1}$ diverge when the model (1) is misspecified.

Having estimated the model by local-PCA, one obtains $\tilde{\delta}_i \equiv \bar{x}_i - \tilde{\boldsymbol{\lambda}}'_i\bar{\tilde{\mathbf{F}}}$ and $\tilde{h}_i \equiv T^{-1}\sum_{s\in\mathcal{T}_t}\widetilde{\mathcal{M}}_{s-1,s}x_{is}$, the empirical counterparts to $T^{-1}\sum_{s\in\mathcal{T}_t}\delta_{is-1}$ and $T^{-1}\sum_{s\in\mathcal{T}_t}h_{is-1}$. However, whereas the latter depends on the ex-ante risk premia and SDF, respectively, $\tilde{\delta}_i$ and $\tilde{h}_i$ depend on the estimated ex-post quantities. It follows that $\tilde{\delta}_i \to_p \bar{\alpha}_i + \bar{e}_i - (\mathbf{e}'_{.i}\mathbf{F}/T)\mathbf{H}_{t-1}\mathbf{H}'_{t-1}\bar{\mathbf{F}} + \bar{\Delta}_{it-1} + \mathbf{l}'_{it-1}\mathbf{H}'_{t-1}\bar{\mathbf{F}}$ and $\tilde{h}_i \to_p T^{-1}\sum_{s\in\mathcal{T}_t}\mathcal{M}_{s-1,s}(\alpha_{is-1} + e_{is} + \Delta_{ist-1})$ as $N \to \infty$ when (1) holds, that is their limits are non-zero even under correct model specification, due to the finite-T bias.[36] Despite this problem, I

[33]As it is well-known, there is a one-to-one correspondence between the SDF and mean-variance portfolios: the minimum-variance SDF has a variance that coincides (bar the risk-free rate term) with the squared Sharpe ratio of the tangency portfolio, as elucidated in Hansen and Jagannathan (1991).

[34]Along the same lines, I can construct asymptotically valid confidence intervals for the Sharpe ratio, the value-at-risk (VaR), and the expected shortfall (ES) associated with the mean-variance portfolio return r_t^{tg}.

[35]If a risk-free asset is not tradeable, denoting by γ_{0s-1} the zero-beta rate and by x_{is} the *gross* returns, one obtains $\delta_{is-1} = \mathcal{E}_{s-1}(x_{is}) - \gamma_{0s-1} - \boldsymbol{\lambda}'_{it-1}\mathcal{E}_{s-1}(\mathbf{F}_s)$, with the arguments of this section extending to this case (see Section OA6).

[36]The terms $(\bar{\Delta}_{it-1} + \mathbf{l}'_{it-1}\mathbf{H}'_{t-1}\bar{\mathbf{F}})$ and $\sum_{s\in\mathcal{T}_t}\mathcal{M}_{s-1,s}\Delta_{is,t-1}/T$ in $\tilde{\delta}_i$ and $\tilde{h}_i$, respectively, arise from the constant-parameters local approximation that I use, namely $\bar{x}_i = \bar{\alpha}_i + \boldsymbol{\lambda}'_{it-1}\bar{\mathbf{F}} + \bar{e}_i + \bar{\Delta}_{it-1}$, with $\bar{\Delta}_{it-1} \equiv T^{-1}\sum_{s\in\mathcal{T}_t}\Delta_{is,t-1} =$

can conduct asymptotically correct inference on the pricing ability of the (ex-ante) risk premia $\boldsymbol{\gamma}_{t-1}$ and SDF $m_{s-1,s}$ by studying the behaviour of $N^{-1}\sum_{i=1}^{N}\tilde{\delta}_i^2$ and $N^{-1}\sum_{i=1}^{N}\tilde{h}_i^2$.

Theorem 3 (limiting distribution of tests for pricing performance) *Under Assumptions 1-9, as $N\to\infty$, for some $c>0$,*

(i) Setting $\mathcal{S}_{\delta,t-1}\equiv\sqrt{N}(\tilde{\mathbf{G}}'_{\delta,t-1}\tilde{\mathcal{U}}_{\epsilon}\tilde{\mathbf{G}}_{\delta,t-1})^{-\frac{1}{2}}\left(\sum_{i=1}^{N}\tilde{\delta}_i^2/N-\tilde{\sigma}^2(1-\bar{\tilde{\mathbf{F}}}'\bar{\tilde{\mathbf{F}}})/T\right)$, *with* $\tilde{\mathbf{G}}_{\delta,t-1}\equiv(T^{-1}(\mathbf{1}_T-\tilde{\mathbf{F}}\bar{\tilde{\mathbf{F}}})\otimes T^{-1}(\mathbf{1}_T-\tilde{\mathbf{F}}\bar{\tilde{\mathbf{F}}}))-\frac{1}{T(T-r_{t-1})}vec(\mathbf{I}_T-\tilde{\mathbf{F}}\tilde{\mathbf{F}}'/T)$,

$$\mathcal{S}_{\delta,t-1}\to_d N(0,1)\ \textit{when}\ \delta_{is-1}=\alpha_{is-1}\ \textit{for every}\ i\ \textit{and}\ s\in\mathcal{T}_t,$$

and $|\mathcal{S}_{\delta,t-1}|\to_p\infty$ *under misspecification, i.e.,* $\sum_{i=1}^{N}\delta_{is-1}^2\geq C(N^{\frac{1}{2}+c})$, *some* $C>0$.

(ii) Setting $\mathcal{S}_{h,t-1}\equiv\sqrt{N}(\tilde{\mathbf{G}}'_{h,t-1}\tilde{\mathcal{U}}_{\epsilon}\tilde{\mathbf{G}}_{h,t-1})^{\frac{1}{2}}(\sum_{i=1}^{N}\tilde{h}_i^2/N-\tilde{\sigma}^2\sum_{s\in\mathcal{T}_t}\widetilde{\mathcal{M}}_{s-1,s}^2/T^2)$, *with* $\tilde{\mathbf{G}}_{h,t-1}\equiv(T^{-1}\sum_{s\in\mathcal{T}_t}\widetilde{\mathcal{M}}_{s-1,s}\boldsymbol{\iota}_s-T^{-1}\sum_{s\in\mathcal{T}_t}\widetilde{\mathcal{M}}_{s-1,s}\boldsymbol{\iota}_s)-\sum_{s\in\mathcal{T}_t}(\widetilde{\mathcal{M}}_{s-1,s}^2/(T^2(T-r_{t-1})))vec(\mathbf{I}_T-\tilde{\mathbf{F}}\tilde{\mathbf{F}}'/T)$,

$$\mathcal{S}_{h,t-1}\to_d N(0,1)\ \textit{when}\ \delta_{is-1}=\alpha_{is-1}\ \textit{for every}\ i\ \textit{and}\ s\in\mathcal{T}_t,$$

and $|\mathcal{S}_{h,t-1}|\to_p\infty$ *under misspecification, i.e.,* $\sum_{i=1}^{N}\delta_{is-1}^2\geq C(N^{\frac{1}{2}+c})$, *some* $C>0$.

The proof is based on the insight that I can accurately estimate the finite-T bias arising from averaging the squared $\tilde{\delta}_i$ and $\tilde{h}_i$, which can then be netted out to measure the genuine pricing performance of the (ex-ante) risk premia and SDF.[37] It is straightforward to show that the distribution of my test statistics $\mathcal{S}_{\delta,t-1}$, and $\mathcal{S}_{h,t-1}$ is invariant to asset repackaging, under the null hypothesis of correct model specification.[38]

6 Monte Carlo Experiments

Comparing Pricing Performance. I simulate the conditional factor model of Kelly et al. (2019)[Eq. (3)], without intercept,

$$x_{is}=\boldsymbol{\lambda}_{is-1}\boldsymbol{f}_s+e_{is}\ \text{where}\ \boldsymbol{\lambda}_{is-1}=\mathbf{z}'_{is-1}\boldsymbol{\Gamma}+\boldsymbol{\nu}_i,\ \text{for}\ i=1,\cdots,N\ \text{and}\ s=1,\cdots,T_0,\qquad(15)$$

with $r=3$ latent risk factors $\mathbf{f}_s$, with $\mathbf{z}_{is}$ denoting a vector of $L=10$ observables (so-called the instruments) with coefficients $\boldsymbol{\Gamma}$ and with $\boldsymbol{\nu}_i$ representing the unobserved portion of the risk exposures orthogonal to the instruments. I estimate (15) using both my local-PCA and Kelly et al. (2019)'s

$T^{-1}\sum_{s\in\mathcal{T}_t}(\boldsymbol{\lambda}_{is}-\boldsymbol{\lambda}_{it-1})'\mathbf{F}_s$, which can be ignored by Assumptions 2 and 3 (when taking cross-sectional averages with large N) without any consequence on the asymptotic analysis of my estimators and test statistics.

[37] A tight analogy exists conceptually between my local-PCA-based tests for *conditional* pricing ability under fixed T and Giacomini and White (2006), who derive the asymptotics for a test of *conditional* predictive ability with a *fixed* estimation window.

[38] Similar specification tests are studied in Gagliardini et al. (2016) and Raponi et al. (2020), when N and T diverge and when only N diverges, respectively, for the case of observed risk factors.

IPCA methodology, specifically designed for (15).[39]

I consider three variations of this Monte Carlo exercise. Experiment (A) considers the case when $\mathbf{z}_{is}$ is an L-dimensional stationary vector autoregression, setting $\boldsymbol{\nu}_i = \mathbf{0}_{r\times 1}$ for every asset i, and assuming that the whole vector $\mathbf{z}_{is}$ is *known*. Experiment (B) is identical to (A) except that now one implements IPCA assuming that only the *first* element of $\mathbf{z}_{is}$ is observed. My local-PCA estimator is identical in these two experiments as it does not rely on the knowledge of the $\mathbf{z}_{is}$ whereas IPCA will estimate the wrong model under (B). Experiment (C) is identical to experiment (A) except that now $\mathbf{z}_{is-1}$ is an L-dimensional random walk. As a comparison metric between my local-PCA and the IPCA estimator, I adopt my criterion of pricing ability $\sum_{i=1}^{N}\tilde{\delta}_i^2/N$ (both in- and out-of-sample versions).[40] I consider $1,000$ Monte Carlo iterations and set the pair (N,T) equal to $(15,5),(15,25),(100,5),(100,25),(1,000,5),(1,000,25)$, respectively. The local-PCA is estimated over rolling windows of size $T < T_0$ and the IPCA uses the total time series length of $T_0 = 100$ for the in-sample exercise. For the out-of-sample exercise, I construct the model forecasts for time t using the estimates corresponding to the time window $[t-T, t-1]$ (rolling window of size T) and $[1, t-1]$ (expanding window) for local-PCA and IPCA, respectively.

TABLE (I) HERE

Panel (A) of Table I presents the most favourable scenario to IPCA as it assumes perfect knowledge of the instruments and sets $\boldsymbol{\nu}_i = \mathbf{0}_{r\times 1}$. I find little difference between my local-PCA and the IPCA estimators out-of-sample, although in-sample my local-PCA appears superior.[41] Panel (B) demonstrates the negative effect of misspecification for the IPCA methodology, with a substantial deterioration of the out-of-sample performance, whereas this does not affect my local-PCA. Panel (C) exemplifies the effect of substantial time-variation in the risk exposures with random-walk instruments: my local-PCA appears to handle this case well, whereas the IPCA's performance deteriorates. Finally, across the three experiments, I found only a little difference between the in-sample and out-of-sample performance for IPCA whereas the difference is more evident for my local-PCA, as expected given the local trait of my methodology. Overall, my local-PCA fares particularly well with large N.

Locally-Strong Factors.[42] My local-PCA approach permits the identification of *locally-strong* factors, that is factors that are present only during certain periods. A risk factor may be deemed as a weak factor unconditionally, that is when its risk exposures are small, in the sense that are square

[39] I consider the special case of the IPCA estimator with the exact analytical solution provided in Kelly et al. (2019)[Internet Appendix B].

[40] In Section OA16.3, where I provide further details of the data-generating processes for experiments (A), (B), and (C), together with time-series plots of $\sum_{i=1}^{N}\tilde{\delta}_i^2/N$ that complement Table (I), I report further results for two other metrics of comparison between local-PCA and IPCA: a statistical measure (the correlations between estimated and true risk factors and risk exposures) and another economic measure (the total-R^2 of Kelly et al. (2019), both with its in- and out-of-sample versions).

[41] When one considers the more general case $\boldsymbol{\nu}_i \neq \mathbf{0}_{r\times 1}$ in (15) but retaining the same design of Panel (A), IPCA performance deteriorates substantially because $\boldsymbol{\nu}_i$ cannot be estimated by IPCA, whereas the performance remains unchanged for my local-PCA. Details are available upon request.

[42] I am grateful to an anonymous referee for highlighting this further insight into my local-PCA methodology.

summable, when estimated across a long time series but are locally big when estimated over short periods.[43] Specifically, I define a risk factor with risk exposures λ_{is-1}^{lst} to be *locally-strong* over the period $\mathcal{T}_t$ if, as $N \to \infty$,

$$N^{-1}\sum_{i=1}^{N}(\lambda_{is-1}^{lst})^2 = \begin{cases} O(1) & \text{for any } s \in \mathcal{T}_t, \\ O(N^{-1}) & \text{for any } s \notin \mathcal{T}_t. \end{cases} \tag{16}$$

My inferential procedure, when applied to the interval $\mathcal{T}_t$ of length T remains valid even when (16) holds, without any additional modification. Moreover, definition (16) includes the case $\sum_{i=1}^{N}(\lambda_{is-1}^{lst})^2 = 0$ for $s \notin \mathcal{T}_t$, that is when the risk exposures are equal to zero.[44] Therefore, my local-PCA methodology is potentially able to identify risk factors that are useless in some periods but strong in others, including the case of new risk factors that arise (permanently) starting from a given period onwards (e.g., the ESG risk factors).[45]

To demonstrate the ability of local-PCA to estimate locally-strong factors, and the risk of using constant-PCA, I simulate model (1) with $r_{t-1} = r = 2$ across $T + T^0$ periods with $N = 1,000, T = 24, T^0 = 240$, where the second factor is only locally-strong (i.e., strong when $1 \leq s \leq T$ and weak when $T + 1 \leq s \leq T + T^0$).[46] The local-PCA is estimated using the first $T = 24$ observations and the constant-PCA (denoted by $\tilde{\mathbf{F}}^0$) using the whole $240 + T = 264$ observations.

TABLE (II) HERE

Table II demonstrates the accuracy of the local-PCA, in terms of both the risk premia and the risk factors themselves, and the danger arising from using the constant-PCA for the case of locally-strong factors. The last column demonstrates that the second factor's (i.e., the locally-strong one) strength for the constant-PCA estimates is one order of magnitude smaller than for the local-PCA, suggesting that the estimated risk factors would be wrongly deemed as (unconditionally) weak when using constant-PCA.[47]

7 Empirical Application

Table III compares the pricing performance of my local-PCA with the constant-PCA, both in terms of SDF (top two panels) and risk premia (bottom two panels), and both in-sample and out-

[43] Weak factors, which can arise if either their risk exposures are small, or sparse, across all assets, pose a great challenge to empirical asset pricing (see Lettau and Pelger (2020b)). Recent advances help mitigate these challenges (see Anatolyev and Mikusheva (2002) and Giglio et al. (2021)).

[44] This is the case of *spurious* or *useless* factors (see Kan and Zhang (1999)), the most extreme case of weak factors (see Kleibergen (2009), and Bryzgalova (2016) among others).

[45] When identifying locally-strong risk factors, it is unnecessary to glue the local-PCA estimates using the approach of Section OA4, as the aim here is to identify factors that are strong relative to a given period, rather than reconstructing them across the full length of the available data.

[46] I set $\boldsymbol{\lambda}_{is} = \boldsymbol{\lambda}_i = (\lambda_{1i}, \lambda_{2i})' \sim N(\mathbf{0}_2, \mathbf{I}_2)$ for $1 \leq s \leq T$, with $\lambda_{1is} = \lambda_{1i}^0 \sim N(0,1)$ and $\lambda_{2is} = \lambda_{2i}^0 \sim N(0, 1/N)$ for $T + 1 \leq s \leq T^0$, with $\lambda_{1i}, \lambda_{2i}, \lambda_{1i}^0, \sqrt{N} \times \lambda_{2i}^0$ iid across i and mutually independent (see Section OA16.3 for details).

[47] An empirical example is presented Section OA15.5, where I study the extent to which my local-PCA risk factors are (locally) related to the intermediary-capital-risk factor of He et al. (2017).

of-sample.[48]

TABLE (III) HERE

The local-PCA is evaluated using both the balanced and unbalanced panels; the balanced panel case is directly comparable to the constant-PCA.[49] Each entry in the table represents the ratio (in absolute value) between the bias-adjusted pricing statistics of the constant-PCA (numerator) and the local-PCA (denominator). The SDF pricing statistics $N^{-1}\sum_{i=1}^{N}\tilde{h}_i^2$ (top two panels) and risk premia pricing statistics $N^{-1}\sum_{i=1}^{N}\tilde{\delta}_i^2$ (bottom two panels), described in Theorem 3, have been averaged across all the available rolling windows before evaluating the ratios. Panels two and four report the in-sample results and panels one and three report the out-of-sample results. They correspond to the (time-varying) estimated number of factors $\tilde{r}$ (column two), as described in Theorem 2, and a number of risk factors ranging from one to 10 (columns three-12), respectively. The local-PCA is evaluated with $T = 24$ for both the balanced and unbalanced panels.[50]

8 Conclusions

This paper develops a local-PCA-based methodology for inference on conditional asset pricing models linear in latent risk factors, particularly suited when using large cross-sections of individual stock returns. I focus on factors' risk premia and the SDF spanned by latent risk factors. The strength of my setup is that it naturally handles asset pricing models where every feature is allowed to be time-varying, including risk premia and their exposures, idiosyncratic risk, and the number of risk factors, providing an inferential theory for their pricing ability. I demonstrate that allowing for time variation in the asset pricing model's parameters carries significant economic value, both in terms of small pricing errors and large Sharpe ratios.

9 Data Availability

Code replicating the tables and figures in this article, and information about the proprietary data used can be found in Zaffaroni (2024).

References

Anatolyev, S., and A. Mikusheva, 2002, Factor models with many assets: Strong factors, weak factors, and the two-pass procedure, *Journal of Econometrics* 229, 103–126.

Andersen, T., T. Bollerselv, F. Diebold, and J. Wu, 2006, Realized beta: Persistence and predictability, in T. Fomby, and D. Terrell, eds., *Advances in Econometrics: Econometric Analysis*

[48] See Section OA15 for full details on the design of the empirical exercise with several extensions.

[49] I am grateful to an anonymous referee for raising this issue. For the out-of-sample exercises, the constant-PCA estimates are based on the first half of the data whereas the local-PCA estimates use the T observations before the forecasting period. The balanced panel used by the constant-PCA and the local-PCA (balanced case) is made by $N = 283$ stocks whereas the unbalanced panel used a number of assets in the range $1,820 \leq N \leq 2,360$ when $T = 24$.

[50] Table OA2 in Section OA15.3 replicates this exercise when $T = 120$.

of Economic and Financial Time Series in Honor of R.F. Engle and C.W.J. Granger, vol. B, 1–40 (Elsevier, Amsterdam, Netherlands).

Ang, A., and D. Kristensen, 2012, Testing conditional factor models, *Journal of Financial Economics* 106, 132–156.

Bai, J., 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135–171.

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

Bai, J., and G. Zhou, 2015, Fama-MacBeth two-pass regressions: Improving risk premia estimates, *Finance Research Letters* 15, 31–40.

Bates, J. B., M. Plagborg-Møller, J.H. Stock, and M.W. Watson, 2013, Consistent factor estimation in dynamic factor models with structural instability, *Journal of Econometrics* 177, 289–304.

Bryzgalova, S., 2016, Spurious factors in linear asset pricing models, *Working Paper, LBS* .

Chamberlain, G., 1983, Funds, factors, and diversification in arbitrage pricing models, *Econometrica* 51, 1305–1323.

Cochrane, J. H., 2005, *Asset Pricing* (Princeton University Press, NJ).

Cochrane, J. H., 2011, Presidential address: Discount rates, *The Journal of Finance* 66, 1047–1108.

Connor, G., and R. A. Korajczyk, 1986, Performance measurement with the Arbitrage Pricing Theory: A new framework for analysis, *Journal of Financial Economics* 15, 373–394.

Connor, G., Hagmann. M., and O. Linton, 2012, Efficient estimation of a semiparametric characteristic-based factor model of security returns, *Econometrica* 80, 713–754.

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

Fama, E. F., and K. R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

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

Fan, J., Y. Liao, and W. Wang, 2016, Projected principal component analysis in factor models, *The Annals of Statistics* 44, 219–254.

Gagliardini, P., E. Ossola, and O. Scaillet, 2016, Time-varying risk premium in large cross-sectional equity data sets, *Econometrica* 84, 985–1046.

Giacomini, R., and H. White, 2006, Tests of conditional predictive ability, *Econometrica* 74, 1545–1578.

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

Giglio, S., D. Xiu, and D. Zhang, 2021, Test assets and weak factors, *Chicago Booth Research Paper 21-04* .

Haddad, V., S. Kozak, and S. Santosh, 2020, Factor timing, *The Review of Financial Studies* 33, 1980–2018.

Hansen, L. P., and S. F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, L. P., and T. J. Sargent, 1999, Wanting robustness in macroeconomics, *Working Paper, University of Chicago* .

Hansen, L.P., and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Harvey, C. R, Y. Liu, and H. Zhu, 2016, ... and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.

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

Huberman, G., 1982, A simple approach to Arbitrage Pricing Theory, *Journal of Financial Economics* 28, 183–191.

Huberman, G., S. Kandel, and R. F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *The Journal of Finance* 42, 1–9.

Ingersoll, J. E., 1984, Some results in the theory of arbitrage pricing, *The Journal of Finance* 39, 1021–1039.

Jagannathan, R., and Z. Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *The Journal of Finance* 53, 1285–1309.

Kan, R., and C. Zhang, 1999, Two-pass tests of asset pricing models, *Journal of Finance* 54, 203–235.

Kelly, B. T., S. Pruitt, and Y. Su, 2019, Characteristics are covariances: A unified model of risk and return., *Journal of Financial Economics* 134, 501–524.

Kim, K., R. A. Korajczyk, and A. Neuhierl, 2019, Arbitrage portfolios, *Working Paper, Kellogg School of Management* .

Kleibergen, F., 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149–173.

Kozak, S., S. Nagel, and S. Santosh, 2018, Interpreting factor models, *The Journal of Finance* 73, 1183–1223.

Lettau, M., and M. Pelger, 2020a, Estimating latent asset-pricing factors, *Journal of Econometrics* 218, 1–31.

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

Lewellen, Jonathan, and Stefan Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289–314.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics* 47, 13–37.

Ma, H., and P. Gagliardini, 2020, Extracting statistical factors when betas are time-varying, *Working Paper, Swiss Finance Institute* .

Magnus, Jan, and R. H. Neudecker, 2007, *Matrix Differential Calculus with Applications in Statistics and Econometrics.* (J. Wiley & Sons, Revised Edition, Chichester).

Merton, R., 1973, An Intertemporal Capital Asset Pricing Model, *Econometrica* 41, 867–87.

Pelger, M., 2019, Large-dimensional factor modeling based on high-frequency observations, *Journal of Econometrics* 208, 23–42.

Pelger, M., 2020, Understanding systematic risk: A high-frequency approach, *Journal of Finance* 75, 2179–2220.

Pelger, M., and R. Xiong, 2019, State-varying factor models of large dimensions, *Journal of Business & Economic Statistics* 40, 1315–1333.

Raponi, V., C. Robotti, and P. Zaffaroni, 2020, Testing beta-pricing models using large cross-sections, *The Review of Financial Studies* 33, 2796–2842.

Ross, S. A., 1976, The Arbitrage Theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Shanken, J., 1992, On the estimation of beta-pricing models, *The Review of Financial Studies* 5, 1–33.

Sharpe, W. F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425–442.

Stambaugh, R. F., 1983, Arbitrage pricing with information, *Journal of Financial Economics* 12, 357–369.

Stock, J. H., and M. W. Watson, 2002, Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* 97, 1167–1179.

Zaffaroni, P., 2024, Replication data for: "Factor models for conditional asset pricing", https://doi.org/10.7910/DVN/PS7K7Z/.

Table (I) **Pricing Ability: local-PCA vs IPCA**

(N,T) :	(15,5)	(15,25)	(100,5)	(100,25)	(1,000,5)	(1,000,25)
Experiment (A): Correct Model - Stationary Instruments						
	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)
In-Sample:	5.851	2.036	5.213	1.479	5.304	2.150
Out-of-Sample:	1.137	1.009	1.454	1.487	1.223	2.130
Experiment (B): Misspecified Model - Stationary Instruments						
	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)
In-Sample:	11.668	3.891	6.997	2.764	10.023	3.204
Out-of-Sample:	2.145	3.401	1.967	2.002	2.574	3.012
Experiment (C): Correct Model - Random Walk Instruments						
	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)	(IPCA/local-PCA)
In-Sample:	9.980	4.568	9.025	3.356	11.655	5.924
Out-of-Sample:	1.681	3.494	1.942	2.849	2.077	4.436

This table presents ratios of the pricing ability criterion $N^{-1}\sum_{i=1}^{N}\tilde{\delta}_i^2$ corresponding to the IPCA (numerator) and local-PCA (denominator), averaged across $1,000$ Monte Carlo iterations, and evaluated in-sample and out-of-sample, for various combinations of (N,T). The data generating process is (15) where the vector $\mathbf{z}_{it-1}$ is a 10-sized stationary VAR (Experiment (A)) and a multivariate random walk (Experiment (C)), respectively. Experiment (B) differs from (A) because the IPCA-estimated model is misspecified (one instrument instead of the 10 required instruments).

Table (II) **Weak Factors: local-PCA vs constant-PCA** ($T = 24, r_{t-1} = 2$)

$\#\, factor$	$cor(\tilde{\mathbf{F}}_s, \tilde{\mathbf{H}}'_{T-1}\mathbf{F}_s)$	$cor(\tilde{\mathbf{F}}^0_s, \tilde{\mathbf{H}}^{0\prime}\mathbf{F}_s)$	$(\frac{\tilde{\mathbf{H}}'_{T-1}\bar{\mathbf{F}}}{\bar{\tilde{\mathbf{F}}}} - 1)\%$	$(\frac{\tilde{\mathbf{H}}^{0\prime}\bar{\mathbf{F}}}{\bar{\tilde{\mathbf{F}}}^{\mathbf{0}}} - 1)\%$	$\left(\frac{\tilde{\mathbf{H}}^{-1}_{T-1}\tilde{\mathbf{\Sigma}}_\Lambda\tilde{\mathbf{H}}^{-1\prime}_{T-1}}{(\tilde{\mathbf{H}}^0)^{-1}\tilde{\mathbf{\Sigma}}^0_\Lambda(\tilde{\mathbf{H}}^0)^{-1\prime}}\right)$
#1	0.99	0.02	0.03	-86.65	3.54
#2	0.99	0.11	-0.51	-212.79	26.82

Each row reports in columns two, three, four, five, and six, respectively, the correlation between the local-PCA risk factor ($\tilde{\mathbf{F}}_s$) and the true (rotated) risk factor, and between the constant-PCA risk factor ($\tilde{\mathbf{F}}^0_s$) and the true (rotated) risk factor (averaged across T and $T + T^0$ periods, respectively), the bias (in per cent) of the local-PCA risk premia and the constant-PCA risk premia, and the ratio between local-PCA and constant-PCA cross-averaged squared (rotated) risk exposures (Monte Carlo averages, $M = 1,000$ replications). Row one (two) row corresponds to the first (second) risk factor, respectively.

Table (III) **Pricing Ability: local-PCA vs constant-PCA - SDF** ($T = 24$)

Pricing Ability: SDF

Out-of-Sample	$\tilde{r}$	1	2	3	4	5	6	7	8	9	10
constant-PCA/local-PCA (bal)	3.262	27.791	9.503	8.116	4.875	3.106	2.302	1.871	1.323	0.915	0.712
constant-PCA/local-PCA (unb)	1.096	5.416	3.003	1.776	3.329	4.38	1.745	1.468	0.581	0.864	0.721
In-Sample	$\tilde{r}$	1	2	3	4	5	6	7	8	9	10
constant-PCA/local-PCA (bal)	1.870	12.83	4.475	3.213	2.198	1.683	1.323	1.089	0.889	0.740	0.613
constant-PCA/local-PCA (unb)	0.849	3.510	2.079	1.533	1.082	0.804	0.633	0.513	0.429	0.365	0.298

Pricing Ability: Risk Premia

Out-of-Sample	$\tilde{r}$	1	2	3	4	5	6	7	8	9	10
constant-PCA/local-PCA (bal)	0.845	10.470	6.567	4.784	3.674	2.927	2.369	1.955	1.631	1.375	1.169
constant-PCA/local-PCA (unb)	1.492	5.534	3.407	2.461	1.880	1.493	1.205	0.992	0.827	0.697	0.592
In-Sample	$\tilde{r}$	1	2	3	4	5	6	7	8	9	10
constant-PCA/local-PCA (bal)	4.634	23.767	8.151	6.337	5.0468	3.550	3.720	2.465	2.660	2.420	1.738
constant-PCA/local-PCA (unb)	2.908	161.306	10.825	5.052	3.246	3.148	3.210	2.094	1.455	1.280	1.330

Each entry represents the ratio between the pricing statistics for the constant-PCA over the local-PCA for both the balanced (bal) and unbalanced (unb) panel case: ratios (in absolute value) of the time-averages (across $480 - 2T$ windows) of the bias-adjusted SDF pricing statistics $N^{-1}\sum_{i=1}^{N}\tilde{h}_i^2$ (top two panels) and of the bias-adjusted risk premia pricing statistics $N^{-1}\sum_{i=1}^{N}\tilde{\delta}_i^2$ (bottom two panels), both described in Theorem 3. Panels two and four report the in-sample results and panels one and three report the out-of-sample results, correspondingly to the estimated $\tilde{r}$ (column two), described in Section OA15.1, and fixed risk factors ranging from one to 10 (columns three-12), respectively (monthly data Jan-1980 to Dec-2019, $T = 24$, $N = 19,259$, sourced from CRSP).

10 Appendix

Assumptions.

Assumption 4 (factors) $\hat{\mathbf{\Sigma}}_{\mathbf{F}} = \mathbf{F}'\mathbf{F}/T > 0$ *a.s. for any* T *such that* $r_{t-1} \leq T < \infty$.

Assumption 5 (risk exposures) $\mathbf{\Lambda}'_{t-1}\mathbf{\Lambda}_{t-1}/N \to \mathbf{\Sigma}_{\Lambda_{t-1}} > 0$ *for every* $t \in \mathcal{Z}$.

Assumption 6 (eigenvalues) *For every* $t \in \mathcal{Z}$, *the eigenvalues of the* $r_{t-1} \times r_{t-1}$ *matrix* $(\mathbf{\Sigma}_{\Lambda_{t-1}}\hat{\mathbf{\Sigma}}_{\mathbf{F}})$ *are distinct a.s.*

Assumption 7 (idiosyncratic error - temporal and cross-sectional dependence) *Set* $\sigma_{ij,uv} = \mathcal{E}(e_{iu}e_{jv})$, *with* $\mathbf{\Sigma}_{et} \equiv [\sigma_{ij,tt}]_{i,j=1}^{N}$, *and* $\kappa_{h,i_1\ldots i_h,t_1\ldots t_h} = cumulant(e_{i_1t_1}, e_{i_2t_2}, \ldots, e_{i_ht_h})$. *For for every* $1 \leq i, j \leq N$, *every* $s, v \in \mathcal{T}_t$ *and* $t \in \mathcal{Z}$, *and some positive constant* C, *where* $g_1(\cdot)$ *denotes the maximum eigenvalue operator and* $\mathbb{1}_{(\cdot)}$ *is the indicator function,*

$$\mathcal{E}(e_{is}) = 0, \mathcal{E}(e_{it}e_{is}e_{jv}) = 0, |\sigma_{ij,sv}| \leq C, |\kappa_{iijj,ttvs}| \leq C$$
$$\frac{1}{N}\sum_{i=1}^{N}(e_{is}e_{iv} - \sigma_{ii,sv}) \to_p 0 \text{ and } \frac{1}{N}\sum_{i=1}^{N}(e_{is}^4 - \kappa_{4,iiii,ssss} - 3\sigma_{ii,ss}^2) \to_p 0, \text{ with}$$
$$\frac{1}{N}\sum_{i=1}^{N}(\sigma_{ii,sv} - \mathbb{1}_{(s=v)}\sigma_{t-1}^2) = o(\frac{1}{\sqrt{N}}), \frac{1}{N}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}|\sigma_{ij,sv}| \to 0 \text{ and } \sup_{t\in\mathcal{Z}} g_1(\mathbf{\Sigma}_{et-1}) = o(N^{\frac{1}{2}}),$$
$$\frac{1}{N}\sum_{i=1}^{N}(\sigma_{ii,ss}^2 - \sigma_{4t-1}) = o(\frac{1}{\sqrt{N}}), \frac{1}{N}\sum_{i=1}^{N}(\kappa_{4,iiii,ttsv} - \mathbb{1}_{(t=v=s)}\kappa_{4t-1}) = o(\frac{1}{\sqrt{N}}),$$

and $N^{-1}\sup_{i_1}\sum_{i_2,\ldots,i_h=1}^{N}|\kappa_{h,i_1\ldots i_h,t_1\ldots t_h}| \to 0$, *for every* $3 \leq h \leq 8$, *every* $t_j \in \mathcal{T}_t$ $(1 \leq j \leq h)$ *and for at least one* i_j $(2 \leq j \leq h)$ *different from* i_1, *for some* $\sigma_t^2, \sigma_{4t}, \kappa_{4t}$ *satisfying*

$$0 < \inf_{t\in\mathcal{Z}}\sigma_t^2 \leq \sup_{t\in\mathcal{Z}}\sigma_t^2 < \infty, 0 < \inf_{t\in\mathcal{Z}}\sigma_{4t} \leq \sup_{t\in\mathcal{Z}}\sigma_{4t} < \infty, 0 \leq \inf_{t\in\mathcal{Z}}|\kappa_{4t}| \leq \sup_{t\in\mathcal{Z}}|\kappa_{4t}| < \infty.$$

Assumption 8 (mixed moments) *For every* $t \in \mathcal{Z}$,

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_{it-1}\mathbf{e}'_{\cdot i} \to_p \mathbf{0}_{r_{t-1}\times T} \text{ and } \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\lambda}_{it-1}\boldsymbol{\lambda}'_{it-1} \otimes \mathcal{E}(\mathbf{e}_{\cdot i}\mathbf{e}'_{\cdot i})) \to (\mathbf{\Sigma}_{\mathbf{\Lambda}_{t-1}} \otimes \sigma_{t-1}^2\mathbf{I}_T).$$

Assumption 9 (joint convergence in distribution) *For every* $s \in \mathcal{T}_t$ *and* $t \in \mathcal{Z}$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{bmatrix}(\mathbf{e}_{\cdot i}e_{is} - \mathcal{E}(\mathbf{e}_{\cdot i}e_{is}))\\(\boldsymbol{\lambda}_{it-1}\otimes\mathbf{e}_{\cdot i})\end{bmatrix} \to_d N\Big(\mathbf{0}_{T(r_{t-1}+1)}, \begin{bmatrix}(\kappa_{4t-1}+\sigma_{4t-1})\iota_s\iota'_s + \sigma_{4t-1}\mathbf{I}_T & \mathbf{0}_{T\times Tr_{t-1}}\\ \mathbf{0}_{Tr_{t-1}\times T} & (\mathbf{\Sigma}_{\mathbf{\Lambda}_{t-1}}\otimes\sigma_{t-1}^2\mathbf{I}_T)\end{bmatrix}\Big).$$

Proof of Theorem 1. Consider case $\alpha_{is-1}=0, \Delta_{is,t-1}\equiv \mathbf{F}_s'(\boldsymbol{\lambda}_{is}-\boldsymbol{\lambda}_{it-1})=0$, for all $s\in\mathcal{T}_t$, as the general case follows by Lemmas OA.1 and OA.2. The following identity holds for every $v\in\mathcal{T}_t$ (see Bai (2003), equation (D.1)):

$$\tilde{\mathbf{F}}_v-\tilde{\mathbf{H}}_{t-1}'\mathbf{F}_v=\breve{\mathbf{U}}_{t-1}^{-1}T^{-1}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\zeta_{sv}+\breve{\mathbf{U}}_{t-1}^{-1}T^{-1}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\eta_{sv}+\breve{\mathbf{U}}_{t-1}^{-1}T^{-1}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\eta_{vs}, \tag{17}$$

where, by Assumptions 6 and 7, for every given $s,v\in\mathcal{T}_t$,

$$\zeta_{sv}\equiv\frac{\mathbf{e}_{s.}'\mathbf{e}_{v.}}{N}-\mathcal{E}(\frac{\mathbf{e}_{s.}'\mathbf{e}_{v.}}{N})=O_p(N^{-\frac{1}{2}}),\ \ \eta_{sv}\equiv\frac{\mathbf{F}_s'\boldsymbol{\Lambda}_{t-1}'\mathbf{e}_{v.}}{N}=O_p(N^{-\frac{1}{2}}). \tag{18}$$

The result follows, noting that, by Lemma OA.3, $\parallel\breve{\mathbf{U}}_{t-1}^{-1}\parallel=O_p(1)$, $\parallel\tilde{\mathbf{F}}'\mathbf{F}/T\parallel=O_p(1)$ for N large enough, and $\parallel\tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T\parallel=O(1)$ by construction.

(ii) By part (i), the first term on the right-hand side of (17) satisfies

$$\breve{\mathbf{U}}_{t-1}^{-1}T^{-1}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\zeta_{sv}=\breve{\mathbf{U}}_{t-1}^{-1}T^{-1}\tilde{\mathbf{F}}'(\frac{\mathbf{e}\mathbf{e}_{v.}}{N}-\mathcal{E}(\frac{\mathbf{e}\mathbf{e}_{v.}}{N}))=\mathbf{U}_{t-1}^{-1}T^{-1}\mathbf{H}_{t-1}'\mathbf{F}'(\frac{\mathbf{e}\mathbf{e}_{v.}}{N}-\mathcal{E}(\frac{\mathbf{e}\mathbf{e}_{\mathbf{v}.}}{N}))+o_p(N^{-\frac{1}{2}}),$$

where, by Lemma OA.3, $\tilde{\mathbf{H}}_{t-1}\to_p\boldsymbol{\Sigma}_{\Lambda_{t-1}}\mathbf{Q}_{t-1}'\mathbf{U}_{t-1}^{-1}=\mathbf{H}_{t-1}=\boldsymbol{\Sigma}_{\Lambda_{t-1}}^{\frac{1}{2}}\boldsymbol{\Upsilon}_{t-1}\mathbf{U}_{t-1}^{-\frac{1}{2}}=\mathbf{Q}_{t-1}^{-1}$. Recalling that $\mathbf{H}_{t-1}'\hat{\boldsymbol{\Sigma}}_F\mathbf{H}_{t-1}=\mathbf{I}_{r_{t-1}}$, one obtains

$$\sqrt{N}\breve{\mathbf{U}}_{t-1}^{-1}\frac{1}{T}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\zeta_{sv}\to_d N(\mathbf{0}_{r_{t-1}},\sigma_{4t-1}\frac{\mathbf{U}_{t-1}^{-2}}{T}+\frac{(\mu_{4t-1}-2\sigma_{4t-1})}{T^2}\mathbf{U}_{t-1}^{-1}\mathbf{H}_{t-1}'\mathbf{F}_v\mathbf{F}_v'\mathbf{H}_{t-1}\mathbf{U}_{t-1}^{-1}).$$

The same reasoning applies to all other limiting distributions that follow. In particular, to derive the asymptotic covariance matrix of $\sqrt{N}\breve{\mathbf{U}}_{t-1}^{-1}\frac{1}{T}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\zeta_{sv}$, consider that he $(s,v)th$ element of $\mathcal{E}(\mathbf{e}_{.i}e_{iw}\mathbf{e}_{.j}'e_{jw})-\mathcal{E}(\mathbf{e}_{.i}e_{iw})\mathcal{E}(\mathbf{e}_{.j}'e_{jw})$, for every $s,v,w\in\mathcal{T}_t$, satisfies $cov(e_{is}e_{iw},e_{jv}e_{jw})=\kappa_{4,iijj,swvw}+\mathcal{E}(e_{is}e_{jv})\mathcal{E}(e_{iw}e_{jw})+\mathcal{E}(e_{is}e_{jw})\mathcal{E}(e_{iw}e_{jv})=\kappa_{4,iijj,swvw}+\sigma_{ij,sv}\sigma_{ij,ww}+\sigma_{ij,sw}\sigma_{ij,wv}$. Therefore, $N^{-1}\sum_{i=1}^N\sum_{j=1}^N cov(\mathbf{e}$ $(\kappa_{4t-1}+\sigma_{4t-1})\iota_w\iota_w'+\sigma_{4t-1}\mathbf{I}_T$. For the second term of (17),

$$\sqrt{N}\breve{\mathbf{U}}_{t-1}^{-1}\frac{1}{T}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\eta_{sv}=\breve{\mathbf{U}}_{t-1}^{-1}\frac{\tilde{\mathbf{F}}'\mathbf{F}}{T}\frac{1}{\sqrt{N}}\sum_{i=1}^N\boldsymbol{\lambda}_{it-1}e_{iv}\to_d N(\mathbf{0}_{r_{t-1}},\sigma_{t-1}^2\mathbf{U}_{t-1}^{-1}\mathbf{Q}_{t-1}\boldsymbol{\Sigma}_{\Lambda_{t-1}}\mathbf{Q}_{t-1}'\mathbf{U}_{t-1}^{-1}),$$

where $\mathbf{Q}_{t-1}\boldsymbol{\Sigma}_{\Lambda_{t-1}}\mathbf{Q}_{t-1}'=\mathbf{U}_{t-1}$. For the third term of (17),

$$\sqrt{N}\breve{\mathbf{U}}_{t-1}^{-1}\frac{1}{T}\sum_{s\in\mathcal{T}_t}\tilde{\mathbf{F}}_s\eta_{vs}=\breve{\mathbf{U}}_{t-1}^{-1}\frac{1}{T}(\mathbf{F}_v'\otimes\tilde{\mathbf{F}}')\frac{1}{\sqrt{N}}\sum_{i=1}^N(\boldsymbol{\lambda}_{it-1}\otimes\mathbf{e}_{.i})\to_d N(\mathbf{0}_{r_{t-1}},\sigma_{t-1}^2(\frac{\mathbf{F}_v'\boldsymbol{\Sigma}_{\Lambda_{t-1}}\mathbf{F}_v}{T})\mathbf{U}_{t-1}^{-2}),$$

where the asymptotic covariance matrix of $T^{-1}(\mathbf{F}_v'\otimes\tilde{\mathbf{F}}')\frac{1}{\sqrt{N}}\sum_{i=1}^N(\boldsymbol{\lambda}_{it-1}\otimes\mathbf{e}_{.i})$ is obtained, noticing that $T^{-1}(\mathbf{F}_v'\otimes\mathbf{H}_{t-1}'\mathbf{F}')(\boldsymbol{\Sigma}_{\Lambda_{t-1}}\otimes\sigma_{t-1}^2\mathbf{I}_T)(\mathbf{F}_v\otimes\mathbf{F}\mathbf{H}_{t-1})=\sigma_{t-1}^2(\mathbf{F}_v'\boldsymbol{\Sigma}_{\Lambda_{t-1}}\mathbf{F}_v)\mathbf{I}_T$. Given that the three terms are potentially correlated, I need the joint convergence of Assumption 9, as it emerges by

rewriting the right-hand side of (17) in matrix form,

$$\tilde{\mathbf{F}}_v - \tilde{\mathbf{H}}'_{t-1}\mathbf{F}_v \quad = T^{-1}\breve{\mathbf{U}}_{t-1}^{-1}\left[\tilde{\mathbf{F}}', (\tilde{\mathbf{F}}'\mathbf{F}), (\mathbf{F}'_v \otimes \tilde{\mathbf{F}}')\right]\mathbf{A}_v \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}(\mathbf{e}_{\cdot i}e_{iv} - \mathcal{E}(\mathbf{e}_{\cdot i}e_{iv}))\\ (\boldsymbol{\lambda}_{it-1} \otimes \mathbf{e}_{\cdot i})\end{bmatrix}, \tag{19}$$

where $\mathbf{A}_v$ is the $(T + r_v + r_vT) \times (T + r_vT)$ matrix $\mathbf{A}_v \equiv \begin{pmatrix}\mathbf{I}_T & \mathbf{0}_{T\times r_vT}\\ \mathbf{0}_{r_v\times T} & (\mathbf{I}_{r_v} \otimes \boldsymbol{\iota}'_v)\\ \mathbf{0}_{r_vT\times T} & \mathbf{I}_{r_vT\times r_vT}\end{pmatrix}$ for every $v \in \mathcal{T}_t$. By Assumption 7, the e_{iv} have zero third-moments, implying that the covariance matrix of $((\mathbf{e}_{\cdot i}e_{iv} - E(\mathbf{e}_{\cdot i}e_{iv}))', (\boldsymbol{\lambda}_{it-1} \otimes \mathbf{e}_{\cdot i})')'$ is block-diagonal. The diagonal elements have been established above. When considering (19), one needs the asymptotic covariance matrix between $\boldsymbol{\lambda}_{it-1}e_{iv} = (\boldsymbol{\lambda}_{it-1} \otimes e_{iv})$ and $(\boldsymbol{\lambda}'_{jt-1} \otimes \mathbf{e}'_{\cdot j})$, which emerges by pre- and post-multiplying the asymptotic covariance matrix in Assumption (9) by $\mathbf{A}_v$ and $\mathbf{A}'_v$, respectively. QED