Identifying and exploiting alpha in linear asset pricing models with strong, semi-strong, and latent factors*

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

Ron P. Smith†
Birkbeck, University of London

April 17, 2024

Abstract

The risk premia of traded factors are the sum of factor means and a parameter vector we denote by $\boldsymbol{\phi}$ which is identified from the cross-section regression of α_i on the vector of factor loadings, $\boldsymbol{\beta}_i$. If $\boldsymbol{\phi}$ is non-zero one can construct "phi-portfolios" which exploit the systematic components of non-zero alpha. We show that for known values of $\boldsymbol{\beta}_i$ and when $\boldsymbol{\phi}$ is non-zero there exist phi-portfolios that dominate mean-variance portfolios. The paper then proposes a two-step bias corrected estimator of $\boldsymbol{\phi}$ and derives its asymptotic distribution allowing for idiosyncratic pricing errors, weak missing factors, and weak error cross-sectional dependence. Small sample results from extensive Monte Carlo experiments show that the proposed estimator has the correct size with good power properties. The paper also provides an empirical application to a large number of U.S. securities with risk factors selected from a large number of potential risk factors according to their strength and constructs phi-portfolios and compares their Sharpe ratios to mean variance and S&P portfolios.

JEL Classifications: C38, G10

Key Words: Factor strength, pricing errors, risk premia, missing factors, pooled Lasso, mean-variance and phi-portfolios.

*We are grateful to Juan Espinosa Torres and Jing Kong for carrying out the computations and to Hayun Song for excellent research assistance, all are PhD students at USC. We are also grateful for helpful comments from the editor, Fabio Trojani, two anonymous referees, Natalia Bailey, Alessio Sancetta, Takashi Yamagata and participants at the 2023 Kansas Econometrics Institute, the 2023 NBER Summer Insitute and the 2023 IAAE Annual Conference in Oslo.

†Corresponding author, email r.smith@bbk.ac.uk.

1 Introduction

There are two approaches to the estimation of risk premia and testing of market efficiency, often referred as the beta and the SDF (stochastic discount factor) methods, see Jagannathan and Wang (2002). This paper adopts the beta method, and following the literature uses a linear factor pricing model (LFPM) to explain the time series of excess returns on individual securities, $r_{it} = R_{it} - r_t^f$, where R_{it} is the return and r_t^f the risk free rate for $i = 1, 2, ..., n$; $t = 1, 2, ..., T$, by a set of observed tradable risk factors. We use individual securities rather than portfolios since, as we will show, if risk factors are not strong, large n is required for accurate estimation. Ang, Liu, and Schwarz (2020) discuss the general issues in the choice between portfolios and individual stocks. Pesaran and Smith (2023) discuss both the use of portfolios and the relationship between the SDF and LFPM approaches.

The LFPM explains the excess return on each security, r_{it}, by an intercept, α_i, labelled alpha, and a $K \times 1$ vector of traded risk factors, $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{Kt})'$ with loadings, $\boldsymbol{\beta}_i$:

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it}. \tag{1}$$

Under the Arbitrage Pricing Theory, APT, due to Ross (1976) we have

$$E\left(r_{it}\right) = c + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \eta_i, \tag{2}$$

Taking unconditional expectations of (1) and using (2) we have

$$E(r_{it}) = \alpha_i + \boldsymbol{\beta}_i'\mu = c + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \eta_i.$$

where $E(\mathbf{f}_t) = \boldsymbol{\mu}$ which in turn yields

$$\alpha_i = c + \boldsymbol{\beta}_i'\boldsymbol{\phi} + \eta_i, \tag{3}$$

where $\boldsymbol{\phi} = \boldsymbol{\lambda} - \boldsymbol{\mu}$. If $\boldsymbol{\phi} \neq 0$ one can construct what we call "phi-portfolios" that exploit the systematic components of α_i, as captured by the non-zero elements of $\boldsymbol{\phi}$.[1] The idiosyncratic pricing errors, η_i, cannot be exploited, and our proposed phi-portfolios exist only if $\boldsymbol{\phi} \neq \mathbf{0}$. The focus of much of the literature has been on testing $\alpha_i = 0$, or on estimating the risk premium $\boldsymbol{\lambda}$ and mean-variance (MV) portfolios. But given that estimating a non-zero $\boldsymbol{\phi}$ provides a way to identify and exploit the alpha in a linear factor pricing model for large n, $\boldsymbol{\phi}$ is an interesting object in itself, which will be the focus of this paper.

First, we introduce our proposed phi-portfolio in terms of the factor loadings $\mathbf{B}_n = (\boldsymbol{\beta}_1, \beta_2, ..., \beta_n)'$ and $\boldsymbol{\phi}$, and compare its limiting properties (as $n \to \infty$) with the standard MV portfolio. We show that for known factor loadings and non-zero $\boldsymbol{\phi}$ there exist phi-portfolios with strictly positive returns, given by $\boldsymbol{\phi}'\boldsymbol{\phi} > 0$, that are fully diversified (their variance tends to zero with n). The rate at which the variance of phi-portfolio returns tends to zero will depend on the strength of the traded risk factors compared to the strength of the missing (latent) factors, highlighting the importance of factor strengths in portfolio analysis. Since the phi-portfolios have Sharpe ratios that tend to infinity with

[1] We are grateful to one of the anonymous reviewers for suggesting the idea of phi-portfolios, as a way of motivating and interpreting the role of $\boldsymbol{\phi}$ in asset pricing models.

n, they dominate MV portfolios whose Sharpe ratio is bounded in n. Note that unlike MV portfolios, the phi-portfolios do not require knowledge of the inverse of the covariance matrix of returns, which is particularly difficult to estimate accurately when n is large.

Next we consider estimation of and inference about $\boldsymbol{\phi}$ using a large number of individual securities, taking account of firm-specific pricing errors, traded factors that are not strong, missing latent factors, and panel data sets where the time dimension, T, is small relative to n. Factor strength plays a central role in our analysis of $\boldsymbol{\phi}$. We use a measure of factor strength, α_k, developed in Bailey, Kapetanios, and Pesaran (2016, 2021), which is defined in terms of the proportion of non-zero factor loadings, β_{ik}.[2] A factor is strong if this proportion is very close to unity, it is semi-strong if $1 > \alpha_k > 1/2$, and it is weak if $\alpha_k < 1/2$. Use of this measure allows us to be precise about the degree of pervasiveness and show how the strengths of the observed factors, the missing factors and the pricing errors, each influence estimation and inference about $\boldsymbol{\phi}$.

In practice, exploitation of a non-zero $\boldsymbol{\phi}$ requires n to be large and rebalancing such long-short portfolios for so many securities may incur high transactions costs or not be feasible. In addition, model uncertainty, estimation uncertainty, time variation in both $\boldsymbol{\beta}_i$ and in conditional volatility pose additional difficulties in implementing a strategy to exploit the potential returns revealed by $\boldsymbol{\phi}$. In developing the theory we will abstract from such practical difficulties, but in the empirical section we illustrate some of these issues with a comparison of the performance of $\boldsymbol{\phi}$ based portfolios relative to MV portfolios which would face similar difficulties.

We estimate $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_K)'$, using a two-step estimator. In the first step we estimate the intercepts, $\hat{\alpha}_i$, and the factor loadings, $\hat{\boldsymbol{\beta}}_i$, from least squares regressions of excess returns on an intercept and risk factors. In the second step, $\boldsymbol{\phi}$ is estimated from the cross section regression of $\hat{\alpha}_i$ on $\hat{\boldsymbol{\beta}}_i$. As with the two-step estimator of $\boldsymbol{\lambda}$, such a two-step estimator of ϕ_k will also be biased, and requires bias-correction. Following Shanken (1992), we consider a bias-corrected version of the two-step estimator of $\boldsymbol{\phi}$, which we denote by $\tilde{\boldsymbol{\phi}}_{nT}$. We develop the asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT}$ under quite general set of assumptions regarding the idiosyncratic pricing errors, error cross-sectional dependence, and the presence of missing (latent) factors. The paper also investigates the implications of factor strengths for the precision with which $\boldsymbol{\phi}$ can be estimated. The LFPM, following Chamberlain and Rothschild (1983), assumes that all the observed factors are strong and the eigenvalues of the covariance matrix of the errors are bounded.

In developing the arbitrage pricing theory, APT, Ross (1976), whose concerns were primarily theoretical, assumed the factors had mean zero: $\boldsymbol{\mu} = 0$, so $\boldsymbol{\phi} = \boldsymbol{\lambda}$, is the risk premium. For traded factors under market efficiency, where $\boldsymbol{\phi} = \mathbf{0}$, the risk premium is the factor mean $\boldsymbol{\mu} = \boldsymbol{\lambda}$. Were one interested in estimating $\boldsymbol{\lambda}$ there may be statistical

[2] Alpha is used both for the LFPM intercepts and the measure of factor strength, because these are established usages, but it will be clear from context and subscripts which is being referred to.

reasons to estimate ϕ_k and μ_k separately, then summing them to obtain an estimate of λ_k. The factor mean, $\mu_k = E(f_{kt})$, can be estimated consistently at the regular $\sqrt{T}$ rate directly using time series data on the risk factors, f_{kt}, for $t = 1, 2, ..., T$, and does not require knowledge of the factor loadings or n. In contrast, estimation of ϕ_k requires panel data to estimate the factor loadings and hence both n and T dimensions are important. In some cases it may be beneficial to use different time series dimensions, T_μ and T_ϕ, to estimate μ_k and ϕ_k, respectively. For example, when factor loadings are subject to breaks it is advisable to use a relatively short sample, and when some factors are not sufficiently strong a large n is required. Thus one could use large T to estimate μ_k and small T to estimate ϕ_k and λ_k can be simply estimated by adding the estimates of μ_k and ϕ_k. We do not pursue the use of different T, but if the same T is used to estimate the mean of factor k, say $\hat{\mu}_{k,T}$, and the bias-corrected $\tilde{\phi}_{k,nT}$ then their sum is the same as the Shanken (1992) bias-corrected risk premium for factor k, $\tilde{\lambda}_{k,nT}$. This decomposition can be used to obtain the rate at which $\tilde{\lambda}_{k,nT}$ converges to its true value, $\lambda_{0,k} = \phi_{0,k} + \mu_{0,k}$.

The main theoretical results of the paper are set out around five theorems under a number of key assumptions, with proofs provided in the mathematical appendix. The small sample properties of $\tilde{\boldsymbol{\phi}}_{nT}$ are investigated using extensive Monte Carlo experiments, allowing for a mixture of strong and semi-strong observed factors, latent factors, pricing errors, GARCH effects and non-Gaussian errors. Small sample results are in line with our theoretical findings, and confirm that the bias-corrected estimator, $\tilde{\boldsymbol{\phi}}_{nT}$, has the correct size and good power properties for samples with time series dimensions of $T = 120$ and $T = 240$. They also show, in accord with the theory, that the precision with which ϕ_k is estimated falls with α_k, the strength of the k^{th} factor.

Our theoretical derivations and Monte Carlo simulations assume that the list of relevant observed factors is known. But in practice the relevant factors must be selected. Extending the theory to the high dimensional case where factors are selected, rather than given, is beyond the scope of the present paper. Since the rate of convergence of $\tilde{\boldsymbol{\phi}}_{nT}$ to its true value is given by $\sqrt{T}n^{(\alpha_k+\alpha_{\min}-1)/2}$, and the Monte Carlo confirms the crucial role of factor strengths in estimation and inference on $\boldsymbol{\phi}$, it seems sensible to select factors on the basis of their factor strength. Weak factors whose strength is around $1/2$ can be ignored and absorbed into the error term.

The above selection procedure is applied in a high dimensional setting with both a large number of securities (n from $1,090$ to $1,175$) and a large number of potential risk factors (m from 177 to 189), taken from the Chen and Zimmermann (2022), which can all be traded. We used monthly data over the period $1996m1 - 2022m12$ and considered balanced panels obtained by including *all* existing stocks in a given month for which there are T observations. We considered $T = 120$ and 240 months, and focus on the latter which we found to be more reliable, given the large number of securities being considered. Various procedures could be used to select risk factors for a given security,

i. We used Lasso for this purpose and then selected a subset of these factors that were chosen by a sufficiently large number of securities in the sample, and whose estimated strength were above the given threshold value of $\alpha_k > 0.75$. We refer to this selection procedure as pooled Lasso. Using this procedure with $T = 240$ we ended up with 7 risk factors for the sample ending in 2015, declining to 4 in 2021. Interestingly, the three Fama-French factors were always included in the set of factors selected by pooled Lasso.

Accordingly, we considered three linear asset pricing models for our portfolio analysis: the pooled Lasso selected at the start of our evaluation sample, denoted by PL7, the Fama-French three factors model, FF3, and the Fama-French five factors model, FF5, which includes two factors not in PL7. The three models are estimated using rolling samples of size $T = 240$, starting with a sample ending in $2015m12$ and finishing with a sample ending in $2022m11$. The hypothesis that $\boldsymbol{\phi} = \mathbf{0}$ was rejected for all 84 rolling samples and all three models, albeit less strongly over the post Covid-19 period. The test results suggested possible unexploited return opportunities, and to investigate this possibility further, we constructed phi-portfolios and compared their Sharpe ratios with the ones based on standard MV portfolios over the full sample evaluation sample, $2016m1 - 2022m12$, and sample ending $2019m12$, that excludes the Covid-19 period. We find that in five out of the six cases (3 models 2 samples) the phi-portfolio has a higher SR than the corresponding MV portfolio. The exception is the FF5 pre Covid-19. This illustrates that if $\boldsymbol{\phi} \neq 0$, it is possible to construct a portfolio that outperforms the mean variance portfolio. In both the pre Covid-19 sample and the full sample the highest SR was obtained by the PL7 phi-portfolios, which also outperformed the S&P500. The SRs for the sample ending in 2022 were substantially lower than the sample ending in 2019, consistent with a falling value of the probability that $\boldsymbol{\phi}$ was non-zero.

Related literature: On estimation of risk premia, following Fama and MacBeth and Shanken (1992), estimation of risk premia is further examined by Shanken and Zhou (2007), Kan, Robotti, and Shanken (2013), and Bai and Zhou (2015). The survey paper by Jagannathan, Skoulakis, and Wang (2010) provide further references.

Testing for market efficiency dates back to Jensen (1968) who proposes testing $\mathrm{a}_i = 0$ for each i separately. Gibbons, Ross, and Shanken (1989) provide a joint test for the case where the errors are Gaussian and $n < T$. Gagliardini, Ossola, and Scaillet (2016) develop two-pass regressions of individual stock returns, allowing time-varying risk premia, and propose a standardised Wald test. Raponi, Robotti, and Zaffaroni (2019) propose a test of pricing error in cross section regression for fixed number of time series observations. They use a bias-corrected estimator of Shanken (1992) to standardise their test statistic. Ma, Lan, Su, and Tsai (2020) employ polynomial spline techniques to allow for time variations in factor loadings when testing for alphas. Feng, Lan, Liu, and Ma (2022) propose a max-of-square type test of the intercepts instead of the average used in the literature, and recommend using a combination of the two testing procedures. He, He, Huang, and Zhou

(2022) propose two statistics, a Wald type statistic which require n and T to be of the same order of magnitude and a standardised t-ratio. Kleibergen (2009) considers testing in the case where the loadings are small. Pesaran and Yamagata (2012, 2024) consider testing that the intercepts in the LFPM are zero when n is large relative to T and there may be non Gaussian errors and weakly cross-correlated errors.

A large number of risk factors have been considered in the empirical literature. We use the Fama and French (1993) three factors in our Monte Carlo design. In our empirical application we use the five factors proposed by Fama and French (2015) and the large set of factors provided by Chen and Zimmermann (2022). Harvey and Liu (2019) document over 400 factors published in top finance journals. Dello Preite, Uppal, Zaffaroni, and Zviadadze (2022) argue that despite the hundreds of systematic risk factors considered in the literature, there is still a sizable pricing error and that this can be explained by asset specific risk that reflects market frictions and behavioral biases. There is a large Bayesian literature, including Chib, Zeng, and Zhao (2020), and Hwang and Rubesam (2022) on selecting factor models. The issue of factor selection is also addressed by Fama and French (2018).

Strong and weak factors in asset returns are considered by Anatolyev and Mikusheva (2022), Connor and Korajczyk (2022), and Giglio, Xiu, and Zhang (2023). Beaulieu, Dufour, and Khalaf (2020) discuss the lack of identification of risk premia when many of the loadings are zero. There has also been concern about the consequences of omitted factors. Giglio and Xiu (2021) discuss the problem and try to deal with it using a three-pass method which is valid even when not all factors in the model are specified or observed using principal components of the test assets. Onatski (2012) and Lettau and Pelger (2020a,b) provide extensive discussions of weak factor and latent factors, respectively. More recent contributions include Bai and Ng (2023) and Uematsu and Yamagata (2023).

There is also a large literature on portfolio construction. Herskovic, Moreira, and Muir (2019) discuss low cost methods of hedging risk factors. Dello Preite, Uppal, Zaffaroni, and Zviadadze (2024) derive an SDF in which there is compensation for unsystematic risk within the framework of the APT. Korsaye, Quaini, and Trojani (2021) introduce model-free smart SDFs which give rise to non-parametric SDF bounds for testing asset pricing models. Daniel, Mota, Rottke, and Santos (2020) discuss the common practice of creating characteristic portfolios by sorting on characteristics associated with average returns and show that these portfolios capture not only the priced risk associated with the characteristic but also unpriced risk. Quaini, Trojani, and Yuan (2023) propose an estimator of tradable factor risk premia.

Paper's outline: The rest of the paper is organized as follows: Section 2 provides the framework for estimation of $\boldsymbol{\phi}$. Section 3 sets out the assumptions and states the main theorems. Theorem 1 shows that the standard Fama-MacBeth estimator is valid only when there are no pricing errors and $n/T \to 0$. Theorem 2 shows that the Shanken bias-corrected estimator of λ_k continues to be consistent for a fixed T as $n \to \infty$, even in presence of weak pricing errors and weak missing common factors. Theorem 3 provides conditions under which the bias-corrected estimator, $\tilde{\boldsymbol{\phi}}_{nT}$, is consistent for $\boldsymbol{\phi}_0$, and derives the asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT}$, assuming the observed factors are strong ($\alpha_k = 1$ for all k). Theorem 4 extends the results to situations that one or more risk factors are semi-strong, and establishes the rate at which $\tilde{\boldsymbol{\phi}}_{nT}$ converges to its true value. For example, for a factor with strength α_k we show that $\tilde{\phi}_{k,nT} - \phi_{0,k} = O_p\left(T^{-1/2} n^{-(\alpha_k + \alpha_{\min} - 1)/2}\right)$, and as a consequence

$$\tilde{\lambda}_{k,nT} - \lambda_{0,k} = O_p\left(T^{-1/2} n^{-(\alpha_k + \alpha_{\min} - 1)/2}\right) + O_p\left(T^{-1/2}\right),$$

where $\lambda_{0,k}$ is the true value of the risk premia associated to factor f_{kt}, and $\alpha_{\min}$ is the strength of the least strong factor included. This consistency condition is weaker than the one derived by Giglio, Xiu, and Zhang (2023). Finally, Theorem 5 gives conditions for consistent estimation of the asymptotic variance of $\tilde{\boldsymbol{\phi}}_{nT}$, using a suitable threshold estimator of the covariance matrix. Section 4 presents the Monte Carlo (MC) design, its calibration and a summary of the main findings. Section 5 discusses the problem of factor selection from a large number of potential factors. Section 6 gives the empirical application using monthly data on a large number of individual US stocks and risk factors over the period 1996-2021. It selects factors, estimates $\boldsymbol{\phi}$, and compares the performance of MV and phi-portfolios. Section 7 provides some concluding remarks.

Detailed mathematical proofs are provided in a mathematical appendix. Further information on data sources, MC calibration plus some supplementary material for the empirical application are provided in the online supplement A. To save space all MC results are given in the online supplement B.

2 Identification and estimation of $\boldsymbol{\phi}$

Let R_{it} denote the holding period return on traded security i, which can be bought long or short without transaction costs, r_t^f is the risk free rate, and $r_{it} = R_{it} - r_t^f$ is the excess return. We start with the linear factor pricing model (LFPM)

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + u_{it}, \tag{4}$$

for $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, where r_{it} is explained in terms of the $K \times 1$ vector of factors $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{Kt})'$. The intercept α_i and the $K \times 1$ vector of factor loadings, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})'$, are unknown. The idiosyncratic errors, u_{it} have zero means and are assumed to be serially uncorrelated. The factors, $\mathbf{f}_t$, are assumed to be covariance stationary with the constant mean $\boldsymbol{\mu} = E(\mathbf{f}_t)$, and $\boldsymbol{\Sigma}_f = E\left[(\mathbf{f}_t - \boldsymbol{\mu})(\mathbf{f}_t - \boldsymbol{\mu})'\right]$.

Under the Arbitrage Pricing Theory (APT) due to Ross (1976) the pricing errors, η_i for $i = 1, 2, ..., n$ defined by

$$\eta_i = E(r_{it}) - c - \boldsymbol{\beta}_i' \boldsymbol{\lambda}, \tag{5}$$

are bounded such that

$$\sum_{i=1}^{n} \eta_i^2 < C < \infty, \tag{6}$$

where c is zero-beta expected excess return, $\boldsymbol{\lambda}$ is the $K \times 1$ vector of risk premia. Under APT, we have

$$E\left(r_{it}\right) = c + \boldsymbol{\beta}_i' \boldsymbol{\lambda}_0 + \eta_i, \tag{7}$$

and reduces to the standard beta representation of an unconditional asset pricing model, if $c+\eta_i = 0$. In this case, $\boldsymbol{\beta}_i = -Cov(r_{it}, m_t)/var(m_t)$, where m_t is the stochastic discount factor (SDF), which satisfies the moment condition $E_t(r_{i,t+1}m_{t+1}) = 0$. See Pesaran and Smith (2023) for a discussion of the link. For empirical assessment, we consider the more general unconditional pricing model (7), and relate to the linear factor pricing model. To this end, taking unconditional expectations of (4) and using the APT condition we have

$$E(r_{it}) = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{\mu} = \boldsymbol{\beta}_i' \boldsymbol{\lambda} + c + \eta_i. \tag{8}$$

which in turn yields

$$\alpha_i = c + \boldsymbol{\beta}_i' \boldsymbol{\phi} + \eta_i, \text{ for } i = 1, 2, ..., n. \tag{9}$$

where

$$\boldsymbol{\phi} = \boldsymbol{\lambda} - \boldsymbol{\mu}, \tag{10}$$

The focus of the literature has been on testing for alpha, $\alpha_i = 0$, and the estimation of the risk premia, $\boldsymbol{\lambda}$, using panel data on excess returns, $\{r_{it}, 1, 2, ..., n;\ t = 1, 2, ..., T\}$, and $\mathbf{F}$, the $T \times K$ matrix of observations on the factors. It is clear that $\boldsymbol{\phi}$ plays an important role in tests for alpha in LFPM, and a non-zero $\boldsymbol{\phi}$ implies non-zero alphas, which in turn implies exploitable excess profitable opportunities. More specifically, we show that for known values of $\boldsymbol{\beta}_i$, $i = 1, 2, ..., n$ and when $\boldsymbol{\phi}$ is non-zero there exists phi-based portfolios with non-zero means that are fully diversified (their variance tends to zero with n), namely have Sharpe ratios that tend to infinity with n, and hence dominate mean-variance (MV) portfolios. For the MV portfolios to be efficient it is necessary that $\boldsymbol{\phi} = \mathbf{0}$.

2.1 Why $\boldsymbol{\phi}$ matters: introduction of the phi-portfolios

Substitute the expression for α_i given by (9) in (4) to obtain

$$r_{it} = c + \boldsymbol{\beta}_i' \boldsymbol{\phi} + \eta_i + \boldsymbol{\beta}_i' \mathbf{f}_t + u_{it}, \text{ for } i = 1, 2, ..., n, \tag{11}$$

and write the n return equations more compactly as

$$\mathbf{r}_{\circ t} = c\boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\phi} + \mathbf{B}_n \mathbf{f}_t + \boldsymbol{\eta}_n + \mathbf{u}_{\circ t}, \tag{12}$$

where $\mathbf{r}_{\circ t} = (r_{1t}, r_{2t}, ..., r_{nt})'$, $\boldsymbol{\tau}_n$ is an n-dimensional vector of ones, $\mathbf{B}_n = (\boldsymbol{\beta}_{\circ 1}, \boldsymbol{\beta}_{\circ 2}, ..., \boldsymbol{\beta}_{\circ K})$, $\boldsymbol{\beta}_{\circ k} = (\beta_{1k}, \beta_{2k}, ..., \beta_{nk})'$, $\mathbf{u}_{\circ t} = (u_{1t}, u_{2t}, ..., u_{nt})'$, $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$, and $\mathbf{V}_u = E\left(\mathbf{u}_{\circ t}\mathbf{u}_{\circ t}'\right)$. Suppose that the factors, $\mathbf{f}_t$, are traded, and $\boldsymbol{\phi}'\boldsymbol{\phi} > 0$. Consider the $n \times 1$ vector of phi-portfolio weights, $\mathbf{w}_{\phi}$, given by

$$\mathbf{w}_{\phi} = \mathbf{M}_n \mathbf{B}_n \left(\mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n\right)^{-1} \boldsymbol{\phi}, \tag{13}$$

where $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}_n'$, and Finally, consider the long-short hedged portfolio return

$$\rho_{t,\phi} = \mathbf{w}_\phi'\mathbf{r}_{\circ t} - \boldsymbol{\phi}'\mathbf{f}_t = \boldsymbol{\phi}'\left[\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)^{-1}\mathbf{B}_n\mathbf{M}_n\mathbf{r}_{\circ t} - \mathbf{f}_t\right], \tag{14}$$

and using (12) note that

$$\rho_{t,\phi} = \boldsymbol{\phi}'\boldsymbol{\phi} + \mathbf{w}_\phi'\boldsymbol{\eta}_n + \mathbf{w}_\phi'\mathbf{u}_{\circ t}. \tag{15}$$

For a given vector of pricing errors, $\boldsymbol{\eta}_n$, $E(\rho_{t,\phi}) = \boldsymbol{\phi}'\boldsymbol{\phi} + \mathbf{w}_\phi'\boldsymbol{\eta}_n$, and $Var(\rho_{t,\phi}) = \mathbf{w}_\phi'\mathbf{V}_u\mathbf{w}_\phi$, where $\mathbf{V}_u = E(\mathbf{u}_{\circ t}\mathbf{u}_{\circ t}')$. Using (13) it follows that

$$\left\|\mathbf{w}_\phi'\boldsymbol{\eta}_n\right\|^2 \leq \left\|\boldsymbol{\eta}_n\right\|^2\left\|\mathbf{w}_\phi\right\|^2 \leq \left\|\boldsymbol{\eta}_n\right\|^2\left\|\boldsymbol{\phi}\right\|^2\lambda_{max}\left[\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)^{-1}\right] = \frac{\left\|\boldsymbol{\eta}_n\right\|^2\left\|\boldsymbol{\phi}\right\|^2}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)}.$$

Similarly,

$$Var(\rho_{t,\phi}) \leq \left\|\mathbf{w}_\phi\right\|^2\lambda_{max}(\mathbf{V}_u) = \frac{\lambda_{max}(\mathbf{V}_u)\left\|\boldsymbol{\phi}\right\|^2}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)}. \tag{16}$$

Hence, $E\left(\boldsymbol{\rho}_{t,\phi}\right) \to \boldsymbol{\phi}'\boldsymbol{\phi}$, and $Var(\rho_{t,\phi}) \to 0$, as $n \to \infty$, if

$$\frac{\left\|\boldsymbol{\eta}_n\right\|^2}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)} \to 0, \text{ and } \frac{\lambda_{max}(\mathbf{V}_u)}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)} \to 0. \tag{17}$$

The conditions are met if $\lambda_{min}(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n) \to \infty$, $\lambda_{max}(\mathbf{V}_u) < C$, and the APT condition (6) are met. The first two conditions follow if the LPFM given by (4) is an approximate factor model as assumed by Chamberlain and Rothschild (1983), namely when the factors are strong and the errors, u_{it}, are weakly cross correlated.[3] Therefore, knowledge of $\boldsymbol{\phi}'\boldsymbol{\phi}$ and its statistical significance can play an important role in portfolio analysis.

In the case where $c = 0$, $\boldsymbol{\phi} = \mathbf{0}$, and $\boldsymbol{\eta}_n = \mathbf{0}$, there are no exploitable excess profit opportunities, and portfolios formed using factor loadings reduce to the standard mean-variance (MV) portfolio, given by $\rho_{MV,t} = \boldsymbol{\mu}_R'\mathbf{V}_R^{-1}\mathbf{r}_{\circ t}$, where $\boldsymbol{\mu}_R = \mathbf{B}_n\boldsymbol{\mu}$ and $\mathbf{V}_R = \mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n' + \mathbf{V}_u$. As shown in Lemma A.1, the squared Sharpe ratio of MV portfolio, $SR_{MV}^2 = \boldsymbol{\mu}_R'\mathbf{V}_R^{-1}\boldsymbol{\mu}_R$, is bounded by $\boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}$, and this upper bound is achieved only if $\lambda_{min}(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n) \to \infty$, as $n \to \infty$. Specifically

$$\begin{aligned} SR_{MV}^2 &= \boldsymbol{\mu}_R'\mathbf{V}_R^{-1}\boldsymbol{\mu}_R = \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{B}_n'\left(\mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n' + \mathbf{V}_u\right)^{-1}\mathbf{B}_n\boldsymbol{\mu} \\ &\leq \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu} \end{aligned} \tag{18}$$

In contrast, if it is similarly assumed that $c = 0$ and $\boldsymbol{\eta}_n = \mathbf{0}$, but $\boldsymbol{\phi}'\boldsymbol{\phi} > 0$, then the following hedged portfolio can be formed[4]

$$\begin{aligned} \rho_{t,\phi} &= \boldsymbol{\phi}'\left[\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\mathbf{B}_n\mathbf{V}_u^{-1}\mathbf{r}_{\circ t} - \mathbf{f}_t\right] = \boldsymbol{\phi}'\left[\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\mathbf{B}_n'\mathbf{V}_u^{-1}\left(\mathbf{B}_n\boldsymbol{\phi}_0 + \mathbf{B}_n\mathbf{f}_t + \mathbf{u}_{\circ t}\right) - \mathbf{f}_t\right] \\ &= \boldsymbol{\phi}'\boldsymbol{\phi} + \boldsymbol{\phi}'\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{u}_{\circ t}, \end{aligned}$$

and its squared Sharpe ratio is given by

$$SR_\phi^2 = \frac{\left(\boldsymbol{\phi}'\boldsymbol{\phi}\right)^2}{\boldsymbol{\phi}'\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\boldsymbol{\phi}}. \tag{19}$$

Since,

$$\boldsymbol{\phi}'\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\boldsymbol{\phi} \leq \left(\boldsymbol{\phi}'\boldsymbol{\phi}\right)\lambda_{\max}\left[\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\right] = \frac{\boldsymbol{\phi}'\boldsymbol{\phi}}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)},$$

[3] The diversification conditions in (17) are met more generally, and accommodate the presence of semi-strong traded factors in f_t, and less restrictive conditions on $\mathbf{V}_u$.and the pricing errors η_i. See Remark 1 below.

[4] When $c \neq 0$, it is not possible to eliminate c (which is unpriced), and at the same time exploit the error covariance matrix $\mathbf{V}_u$.

then

$$SR_{\phi}^{2} \geq (\boldsymbol{\phi}'\boldsymbol{\phi})\ \lambda_{min}\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right). \tag{20}$$

As a result, if $\lambda_{min}\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right) \rightarrow \infty$ as $n \rightarrow \infty$, SR_{ϕ}^{2} increases in n without bounds if $\boldsymbol{\phi}'\boldsymbol{\phi} > 0$, and SR_{ϕ}^{2} will eventually dominate SR_{MV}^{2}, since the latter is bounded in n.

It is also worth noting that the Sharpe ratio of the MV portfolio continues to be bounded in n even if $\boldsymbol{\phi} \neq \mathbf{0}$. In this case we have

$$SR_{MV}^{2} = \boldsymbol{\lambda}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda} - \boldsymbol{\lambda}'\mathbf{B}_n'\left(\mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n' + \mathbf{V}_u\right)^{-1}\mathbf{B}_n\boldsymbol{\lambda},$$

and $SR_{MV}^{2} \leq (\boldsymbol{\mu} + \boldsymbol{\phi})'\,\boldsymbol{\Sigma}_f^{-1}(\boldsymbol{\mu} + \boldsymbol{\phi}) < C$. Whether $\boldsymbol{\phi} = \mathbf{0}$ or not affects the magnitude of SR_{MV}^{2} but does not alter the fact that SR_{MV}^{2} will be bounded if $\lambda_{min}\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right) \rightarrow \infty$ as $n \rightarrow \infty$.

To make sure that the mean-variance portfolio is efficient we must have $\boldsymbol{\phi}'\boldsymbol{\phi} = 0$, and it is of special interest to estimate $\boldsymbol{\phi}$ and develop reliable procedures for testing its statistical significance, using a large number of securities. In practice, the number of tradeable securities, n, might not be sufficiently large, and there are important specification and estimation uncertainties, and the Sharpe ratio of the phi-based portfolio, $\boldsymbol{\rho}_{t,\phi}$, is likely to be bounded even if $\boldsymbol{\phi}'\boldsymbol{\phi} > 0$. We turn to these issues in the empirical application provided in Section 6.

2.2 Fama-MacBeth and Shanken estimators of risk premia

It will prove convenient to write (4) in matrix notation by stacking the excess returns by $t = 1, 2, ..., T$, for each security i

$$\mathbf{r}_{i\circ} = \alpha_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ},\ \text{for } i = 1, 2, ..., n, \tag{21}$$

where $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$, and $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones. Similarly, stacking the excess returns by i for each t we have (12) which we rewrite as

$$\mathbf{r}_{\circ t} = \boldsymbol{\alpha}_n + \boldsymbol{B}_n\mathbf{f}_t + \mathbf{u}_{\circ t},\ \text{for } t = 1, 2, ..., T, \tag{22}$$

where $\boldsymbol{\alpha}_n = (\alpha_1, \alpha_2, ..., \alpha_n)' = c\tau_n + \mathbf{B}_n\boldsymbol{\phi} + \eta_n$.

The risk premia are usually estimated using a two-pass procedure suggested by Fama and MacBeth (1973). The first-pass runs time series regressions of excess returns, r_{it}, on the K observed factors to give estimates of the factor loadings, $\boldsymbol{\beta}_i$:

$$\hat{\boldsymbol{\beta}}_{iT} = \left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}. \tag{23}$$

The second-pass runs a cross section regression of average returns, $\bar{r}_{i\circ} = T^{-1}\sum_{t=1}^{T} r_{it}$ on the estimated factor loadings, to obtain the FM estimator of $\boldsymbol{\lambda}$, namely

$$\hat{\boldsymbol{\lambda}}_{nT} = \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}_{n\circ}, \tag{24}$$

where $\hat{\mathbf{B}}_{nT} = (\hat{\boldsymbol{\beta}}_{1T}, \hat{\boldsymbol{\beta}}_{2T}, ..., \hat{\boldsymbol{\beta}}_{nT})'$, $\bar{\mathbf{r}}_{n\circ} = (\bar{r}_{1T}, \bar{r}_{2T}, ..., \bar{r}_{nT})'$, $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}_T'$, $\boldsymbol{\tau}_T$ is a T-dimensional vector of ones, $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}_n'$, and $\boldsymbol{\tau}_n$ is an n-dimensional vector of ones.

As is well known, when T is finite FM's two-pass estimator is biased due the errors in estimation of factor loadings that do not vanish. The small T bias of the two-pass estimator of $\boldsymbol{\lambda}$ has been a source of concern in the empirical literature. Under standard regularity conditions and as $n \to \infty$, we have

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 \to_p \left[\boldsymbol{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Sigma}_{\beta\beta}\mathbf{d}_{fT} - \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_0\right), \tag{25}$$

where $\mathbf{d}_{fT} = \hat{\mu}_T - \mu_0$, $\boldsymbol{\lambda}_0$ and μ_0 are the true values of $\boldsymbol{\lambda}$ and μ, respectively, $\boldsymbol{\Sigma}_{\beta\beta} = \lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)$, and $\bar{\sigma}^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_i^2 > 0$. Following Shanken (1992), $\bar{\sigma}_n^2$ can be consistently estimated (for a fixed T) by

$$\hat{\bar{\sigma}}_{nT}^2 = \frac{\sum_{t=1}^T\sum_{i=1}^n \hat{u}_{it}^2}{n(T-K-1)}, \tag{26}$$

where

$$\hat{u}_{it} = r_{it} - \hat{\alpha}_{iT} - \hat{\boldsymbol{\beta}}_{iT}'\mathbf{f}_t, \tag{27}$$

and as before $\hat{\alpha}_{iT}$ and $\hat{\boldsymbol{\beta}}_{iT}$ are the OLS estimators of a_i and $\boldsymbol{\beta}_i$. Using these results the bias-corrected version of the two-pass estimator is given by[5]

$$\tilde{\boldsymbol{\lambda}}_{nT} = \mathbf{H}_{nT}^{-1}\left(\frac{\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}_{n\circ}}{n}\right), \tag{28}$$

where

$$\mathbf{H}_{nT} = \frac{\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n} - T^{-1}\hat{\bar{\sigma}}_{nT}^2\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}. \tag{29}$$

When all the risk factors are strong, under certain regularity conditions, there exists a fixed T_0 such that for all $T > T_0$, then

$$p\lim_{n\to\infty}\left(\tilde{\boldsymbol{\lambda}}_{nT}\right) = \boldsymbol{\lambda}_T^* = \boldsymbol{\lambda}_0 + \left(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0\right), \tag{30}$$

where $\boldsymbol{\mu}_0$ indicates the true value of the factor mean. Shanken refers to $\boldsymbol{\lambda}_T^*$ as "ex-post" risk premia to be distinguished from $\boldsymbol{\lambda}_0$, referred to as "ex ante" risk premia. See also Section 3.7 of Jagannathan, Skoulakis, and Wang (2010).

In this paper we exploit Shanken's bias correction procedure by applying it to $\boldsymbol{\phi} = \boldsymbol{\lambda} - \boldsymbol{\mu}$ which we identify directly using (9) from the regression of a_i on $\boldsymbol{\beta}_i$ for $i = 1, 2, ..., n$, assuming the idiosyncratic pricing errors, η_i, are sufficiently weak relative to the strengths of the risk factors in a sense which will be made precise below.

2.3 Estimation of $\boldsymbol{\phi}$

In view of (9), the estimation of $\boldsymbol{\phi}$ can be carried out following a two-step procedure whereby in the first step a_i and $\boldsymbol{\beta}_i$ are estimated from the least squares regressions of r_{it} on an intercept and $\mathbf{f}_t$, and these are then used in a second step regression to estimate $\boldsymbol{\phi}$, namely,

$$\hat{\boldsymbol{\phi}}_{nT} = \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\boldsymbol{\alpha}}_{nT}, \tag{31}$$

where $\hat{\boldsymbol{\alpha}}_{nT} = (\hat{\alpha}_{1T}, \hat{\alpha}_{1T}, ..., \hat{\alpha}_{nT})' = \bar{\mathbf{r}}_{nT} - \hat{\mathbf{B}}_{nT}\hat{\boldsymbol{\mu}}_T$, and as before $\hat{\mathbf{B}}_{nT} = (\hat{\boldsymbol{\beta}}_{1T}, \hat{\boldsymbol{\beta}}_{2T}, ..., \hat{\boldsymbol{\beta}}_{nT})'$. This estimator is consistent for $\boldsymbol{\phi}_0$ so long as n and $T \to \infty$, and bias-corrections are

[5] See also Shanken and Zhou (2007), Kan, Robotti, and Shanken (2013), and Bai and Zhou (2015), and the survey paper by Jagannathan, Skoulakis, and Wang (2010) for further references.

necessary to ensure the large n consistency of the estimator when T is fixed. A Shanken type bias-corrected estimator of $\boldsymbol{\phi}_0$ is given by

$$\tilde{\boldsymbol{\phi}}_{nT} = \mathbf{H}_{nT}^{-1} \left[\frac{\hat{\mathbf{B}}_{nT}' \mathbf{M}_n \hat{\mathbf{a}}_{nT}}{n} + T^{-1} \widehat{\sigma}_{nT}^2 \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \hat{\boldsymbol{\mu}}_T \right], \tag{32}$$

where $\mathbf{H}_{nT}$ and $\widehat{\sigma}_{nT}^2$ are given by (29) and (26), respectively. It is also easily established that

$$\tilde{\boldsymbol{\phi}}_{nT} = \tilde{\boldsymbol{\lambda}}_{nT} - \hat{\boldsymbol{\mu}}_T, \tag{33}$$

and for a fixed T and as $n \to \infty$, we have

$$p \lim_{n\to\infty} \tilde{\boldsymbol{\phi}}_{nT} = p \lim_{n\to\infty} \tilde{\boldsymbol{\lambda}}_{nT} - \hat{\boldsymbol{\mu}}_T.$$

Hence, upon using (30)

$$p \lim_{n\to\infty} \tilde{\boldsymbol{\phi}}_{nT} = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) - \hat{\boldsymbol{\mu}}_T = \boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0 = \boldsymbol{\phi}_0, \tag{34}$$

and there exists a fixed T_0 such that for all $T > T_0$, $\tilde{\boldsymbol{\phi}}_{nT}$ converges to $\boldsymbol{\phi}_0$ as $n \to \infty$. Also using (30) and (33), and noting that $\boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0 = \boldsymbol{\phi}_0$, interestingly we have

$$\tilde{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \tilde{\boldsymbol{\phi}}_{nT} + \hat{\boldsymbol{\mu}}_T - \boldsymbol{\lambda}_T^* = \tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0.$$

So inference using the Shanken bias-corrected estimator of $\boldsymbol{\lambda}$ around $\boldsymbol{\lambda}_T^*$, is the same as making inference using $\tilde{\boldsymbol{\phi}}_{nT}$ around $\boldsymbol{\phi}_0$.

The asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT}$ depends on both n and T. Assuming the observed factors are strong and under certain regularity conditions, to be introduced below, we have

$$\sqrt{nT} \left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 \right) \to_d N \left(\mathbf{0}, \boldsymbol{\Sigma}_{\beta\beta}^{-1} \mathbf{V}_\xi \boldsymbol{\Sigma}_{\beta\beta}^{-1} \right), \tag{35}$$

where

$$\mathbf{V}_\xi = \left(1 + \boldsymbol{\lambda}_0' \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0 \right) p \lim_{n\to\infty} \left[n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \right].$$

The variance of $\tilde{\boldsymbol{\phi}}_{nT}$ is consistently estimated by

$$\widehat{Var\left(\tilde{\boldsymbol{\phi}}_{nT} \right)} = T^{-1} n^{-1} \mathbf{H}_{nT}^{-1} \hat{\mathbf{V}}_{\xi,nT} \mathbf{H}_{nT}^{-1}, \tag{36}$$

where $\mathbf{H}_{nT}$ is given by (29),

$$\hat{\mathbf{V}}_{\xi,nT} = (1 + \hat{s}_{nT}) \left(n^{-1} \hat{\mathbf{B}}_n' \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_n \right), \tag{37}$$

and

$$\hat{s}_{nT} = \tilde{\boldsymbol{\lambda}}_{nT}' \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \tilde{\boldsymbol{\lambda}}_{nT}, \tag{38}$$

$\tilde{\boldsymbol{\lambda}}_{nT}$ is defined by (28), and $\tilde{\mathbf{V}}_u$ is a suitable estimator of $\mathbf{V}_u = E\left(\mathbf{u}_{\circ t} \mathbf{u}_{\circ t}' \right)$. How to estimate $\mathbf{V}_u$ and the conditions under which $Var\left(\tilde{\boldsymbol{\phi}}_{nT} \right)$ is consistently estimated by $\widehat{Var\left(\tilde{\boldsymbol{\phi}}_{nT} \right)}$ is discussed in sub-section 3.2.

Note that even when $\mathbf{V}_u = \sigma^2 \mathbf{I}_T$ the variance of $\tilde{\boldsymbol{\phi}}_{nT}$ does not reduce to $\sigma^2 \boldsymbol{\Sigma}_{\beta\beta}^{-1}$, the standard least squares formula used for the case of known factor loadings. When the loadings are estimated the scaling term $\left(1 + \boldsymbol{\lambda}_0' \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0 \right)$ is required and its neglect can lead to serious over-rejection even if $n/T \to 0$ as n and $T \to \infty$.

2.4 Factor strength

In this paper we deviate from the standard literature and allow the observed and latent factors to have different degrees of strength, depending on how pervasively they impact the security returns. Bailey, Kapetanios, and Pesaran (2021) define the strength of factor, f_{kt}, in terms of the number of its non-zero factor loadings. For a factor to be strong almost all of its n loadings must differ from zero. Given our focus on estimation of risk premia, we adopt the following definition which directly relates to the covariance of $\boldsymbol{\beta}_i$. See also Chudik, Pesaran, and Tosetti (2011).

Definition 1 *(Factor strengths) The strength of factor f_{kt} is measured by its degree of pervasiveness as defined by the exponent α_k in*

$$\sum_{i=1}^{n} \left(\beta_{ik} - \bar{\beta}_k\right)^2 = \ominus(n^{\alpha_k}), \tag{39}$$

and $0 < \alpha_k \leq 1$. We refer to $\{\alpha_k,\ k = 1, 2, ..., K\}$ as factor strengths. Factor f_{kt} is said to be strong if $\alpha_k = 1$, semi-strong if $1 > \alpha_k > 1/2$, and weak if $0 \leq \alpha_k \leq 1/2$. Condition (39) applies irrespective of whether the loadings, β_{ik}, are viewed as deterministic or stochastic.

In the above definition $\ominus_p(n^{\alpha_k})$ denotes the rate at which additional securities add to the factor's strength and α_k can be viewed as a logarithmic expansion rate in terms of n and relates to the proportion of non-zero factor loadings. In the literature it is commonly assumed that the covariance matrix of factor loadings defined by

$$\boldsymbol{\Sigma}_{\beta\beta} = p \lim_{n\to\infty} \left[n^{-1} \sum_{i=1}^{n} \left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_n\right)\left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}_n\right)' \right], \tag{40}$$

is positive definite, where $\bar{\boldsymbol{\beta}}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{\beta}_i = (\bar{\beta}_1, \bar{\beta}_2, ..., \bar{\beta}_k)'$. For $\boldsymbol{\Sigma}_{\beta\beta}$ to be positive definite matrix it is *necessary* that all the K risk factors under consideration are strong in the sense that

$$p \lim_{n\to\infty} \left[n^{-1} \sum_{i=1}^{n} \left(\beta_{ik} - \bar{\beta}_k\right)^2 \right] > 0, \text{ for } k = 1, 2, ..., K. \tag{41}$$

In terms of our definition of factor strength, $\boldsymbol{\Sigma}_{\beta\beta}$ will be positive definite if all the observed factors are strong, namely if $\alpha_k = 1$ for $k = 1, 2, ..., K$. However, such an assumption is quite restrictive and is unlikely to be satisfied for many risk factors being considered in the literature. Bailey, Kapetanios, and Pesaran (2021) show that, apart from the market factor, only a handful of 144 factors in the literature considered by Feng, Giglio, and Xiu (2020) come close to being strong. Giglio, Xiu, and Zhang (2023) consider the estimation of PCA-based risk premia in presence of weak factors. However, their definition of factor strength involves both n and T, and is best viewed as a consistency condition rather than factor strength as such. See the discussion following Theorem 4. Our notion of factor strength, α_k, is in line with the recent literature. See, for example, Bai and Ng (2023) and Uematsu and Yamagata (2023).

2.5 Missing factor

We now turn to the structure of the errors, u_{it}, in the returns equations, and consider two possible sources of error cross-sectional dependence: a missing or latent factor and production networks. The issue of missing factors has been investigated in the recent literature by Giglio and Xiu (2021) and Anatolyev and Mikusheva (2022). The issue of production networks has been investigated in the recent literature by Herskovic (2018), who derives two risk factors based on the changes in network concentration and network sparsity, and Gofman, Segal, and Wu (2020), who focus on the vertical dimension of production by modeling a supply chain, in terms of supplier-customer links. They find that the further away a firm is from final consumers the higher its return. They use this to create a factor TMB (top minus bottom). Both sources of cross-sectional error dependence could be important, since network dependence cannot be represented using latent factor models. See Section 3 of Chudik, Pesaran, and Tosetti (2011).

To allow for both forms of error cross-sectional dependence we consider the following decomposition of u_{it}

$$u_{it} = \gamma_i g_t + v_{it}, \tag{42}$$

where g_t is the missing (latent) factor and v_{it} is weakly cross-correlated in the sense of approximate factor models due to Chamberlain (1983) and Chamberlain and Rothschild (1983). Here we allow for a single missing factor to simplify the exposition, but note that increasing the number of missing factors has little impact on our analysis, so long as the number of missing factors is fixed. Using the normalization $E(g_t^2) = 1$, and assuming that $\gamma_i g_t$ and v_{it} are independently distributed then $E(u_{it}u_{jt}) = \sigma_{ij} = \gamma_i\gamma_j + \sigma_{v,ij}$, and as shown in Lemma A.2, $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{ij}| = O(1)$ so long as the strength of g_t, $\alpha_\gamma < 1/2$ and $\lambda_{\max}(\mathbf{V}_v) < \infty$. This is despite the fact that $\lambda_{\max}(\mathbf{V}_u) = O(n^{\alpha_\gamma})$, where $\mathbf{V}_v = E(\mathbf{v}_i\mathbf{v}_i')$ and $\mathbf{V}_u = E(\mathbf{u}_i\mathbf{u}_i')$.[6] In the Monte Carlo experiments, we consider the possibility of missing factors, as well as weak spatial and network cross-dependence that satisfy conditions of approximate factor models.

2.6 Pricing errors

The APT condition (6), given by (18) in Theorem II of Ross (1976), ensures that under APT the (idiosyncratic) pricing errors are sparse. In this paper we relax the Ross's condition to

$$\sum_{i=1}^{n} \eta_i^2 = O(n^{\alpha_\eta}), \tag{43}$$

where the exponent α_η measures the degrees of pervasiveness of pricing errors. Deviations from APT are measured in terms of α_η ($0 \leq \alpha_\eta < 1$). We investigate the robustness of our proposed estimator of $\boldsymbol{\phi}$ to α_η. This extension is important for tests of market efficiency

[6]Note that Chamberlain's approximate factor model specification requires $\lambda_{\max}(\mathbf{V}_u) = O(1)$ and is violated if $\alpha_\gamma > 0$.

where the null of interest is $H_0 : \alpha_i = c$ for all i in (9). We note that under the alternative hypothesis $H_1 : \alpha_i = c + \boldsymbol{\beta}_i'\boldsymbol{\phi} + \eta_i$, therefore it is desirable to develop a test of $\boldsymbol{\phi} = \mathbf{0}$ which is robust to a wider class of pricing errors than those entertained originally by Ross, where $\alpha_\eta = 0$.

Remark 1 *Having formalized the concepts of factor strength, missing factors, and the less restrictive APT condition given by (43), it is now of interest to revisit the conditions under which the phi-portfolio fully diversifies. Consider (17) and note that*

$$\frac{\|\boldsymbol{\eta}_n\|^2}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)} = O\left[n^{-(\alpha_{min}-\alpha_\eta)}\right], \text{ and } \frac{\lambda_{max}\left(\mathbf{V}_u\right)}{\lambda_{min}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)} = O\left[n^{-(\alpha_{min}-\alpha_\gamma)}\right].$$

Using these results in (15) and (16) and we have

$$E\left(\rho_{t,\phi}\right) = \boldsymbol{\phi}'\boldsymbol{\phi} + O\left[n^{-(\alpha_{min}-\alpha_\eta)}\right], \text{ and } Var\left(\rho_{t,\phi}\right) = O\left[n^{-(\alpha_{min}-\alpha_\gamma)}\right]$$

Therefore, for phi-portfolio to dominate the MV portfolio in addition to $\boldsymbol{\phi}'\boldsymbol{\phi} \neq 0$, it is also required that the strengths of the traded factors α_k, $k = 1, 2, ..., K$ are strictly larger than the strength of the missing factor, α_γ, as well as the strength of the idiosyncratic pricing errors, α_η.

3 Assumptions and theorems

We make the following standard assumptions about $\mathbf{f}_t, g_t$, v_{it}, $\boldsymbol{\beta}_i$, η_i, and γ_i (the drivers of asset returns):

Assumption 1 *(Observed common factors) (a) The $K \times 1$ vector of observed risk factors, $\mathbf{f}_t$, follows the general linear process*

$$\mathbf{f}_t = \boldsymbol{\mu} + \sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_\ell \zeta_{t-\ell}, \tag{44}$$

where $\|\boldsymbol{\mu}\| < C$, $\zeta_t \sim IID(\mathbf{0}, \mathbf{I}_K)$, and $\boldsymbol{\Psi}_\ell$ are $K \times K$ exponentially decaying matrices such that $\|\boldsymbol{\Psi}_\ell\| < C\rho^\ell$ for some $C > 0$ and $0 < \rho < 1$. (b) The $T \times K$ data matrix $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$ is full column rank and there exists T_0 such that for all $T > T_0$, $\hat{\boldsymbol{\Sigma}}_f = T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ is a positive definite matrix, $\lambda_{\max}\left[(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\right] < C$, $\hat{\boldsymbol{\Sigma}}_f \to_p \boldsymbol{\Sigma}_f = E\left(\mathbf{f}_t - \boldsymbol{\mu}_0\right)\left(\mathbf{f}_t - \boldsymbol{\mu}_0\right)' > \mathbf{0}$, where $\boldsymbol{\mu}_0$ is the true value of $\boldsymbol{\mu}$.

Assumption 2 *(Observed factor loadings) (a) The factor loadings β_{ik} for $i = 1, 2, ..., n$ and $k = 1, 2, ..., K$ are stochastically bounded such that $sup_{ik}E\left(\beta_{ik}^2\right) < C$,*

$$\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)^2 = \ominus_p(n^{\alpha_k}), \text{ for } k = 1, 2, ..., K. \tag{45}$$

(b)The $n \times K$ matrix of factor loadings, $\mathbf{B}_n = (\boldsymbol{\beta}_{\circ 1}, \boldsymbol{\beta}_{\circ 2}, ..., \boldsymbol{\beta}_{\circ K})$, where $\boldsymbol{\beta}_{\circ k} = (\beta_{1k}, \beta_{2k}, ..., \beta_{nk})'$ satisfy

$$0 < c < \lambda_{min}\left(\mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1}\right) < \lambda_{max}\left(\mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1}\right) < C < \infty, \tag{46}$$

for some small and large positive constants, c and C, where $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}_n'$, $\boldsymbol{\tau}_n = (1, 1, ..., 1)'$, and $\mathbf{D}_\alpha$ is the $n \times n$ diagonal matrix

$$\mathbf{D}_\alpha = Diag(n^{\alpha_1/2}, n^{\alpha_2/2}, ..., n^{\alpha_K/2}). \tag{47}$$

Assumption 3 *(latent factor) (a) The latent factor, g_t, in (42) is distributed independently of $\mathbf{f}_{t'}$, for all t and t', g_t is serially independent with mean zero, $E(g_t) = 0$, $E(g_t^2) = 1$, and a finite fourth order moment, $sup_t E(g_t^4) < C$. (b) The loadings γ_i are such that $sup_i |\gamma_i| < C$ and*

$$\sum_{i=1}^{n} |\gamma_i| = O(n^{\alpha_\gamma}). \tag{48}$$

Assumption 4 *(idiosyncratic errors) (a) The errors $\{v_{it},\ i = 1, 2, ..., n;\ t = 1, 2, ..., T\}$ are distributed independently of the factors $f_{k,t'}$, and g_t, for all i, t, t' and $k = 1, 2, ..., K$, and their associated loadings β_{ik}, and γ_i. They are serially independent with $E(v_{it}) = 0$ and finite fourth order moments $E(v_{it}^4) < \infty$, and covariances $E(v_{it}v_{jt}) = \sigma_{v,ij}$, such that*

$$\sup_i \sum_{j=1}^{n} |\sigma_{v,ij}| < \infty, \text{ and } \sup_i \sum_{j=1}^{n} Cov(v_{it}^2, v_{jt}^2) < \infty, \tag{49}$$

with $\lambda_{\min}(\mathbf{V}_v) > 0$, where $\mathbf{V}_v = (\sigma_{v,ij})$. (b) The degree of cross-sectional dependence of v_{it} is sufficiently weak so that

$$T^{-1/2} n^{-1/2} \sum_{t=1}^{T} \sum_{i=1}^{n} (\beta_{ik} - \bar{\beta}_k) v_{it} \rightarrow_d N(0, \omega_k^2), \text{ for } k = 1, 2, ..., K, \tag{50}$$

where

$$\omega_k^2 = p \lim_{n\to\infty} n^{-\alpha_k} \sum_{i=1}^{n} \sum_{j=1}^{n} (\beta_{ik} - \bar{\beta}_k)(\beta_{jk} - \bar{\beta}_k)\sigma_{v,ij}. \tag{51}$$

Assumption 5 *(Pricing errors) The pricing errors, η_i, for $i = 1, 2, ..., n$ are individually bounded, $sup_j |\eta_j| < C$, and are distributed independently of the factor loadings, β_{jk}, and γ_j for all i, j and $k = 1, 2, ..., K$, as well as satisfying the condition*

$$\sum_{i=1}^{n} |\eta_i| = O\left(n^{\alpha_\eta}\right), \tag{52}$$

with $\alpha_\eta < 1/2$.

Remark 2 *Under Assumption 1 $E(\mathbf{f}_t) = \mu$, and $Var(\mathbf{f}_t) = \mathbf{\Sigma}_f = \sum_{\ell=0}^{\infty} \mathbf{\Psi}_\ell \mathbf{\Psi}_\ell'$. Also since $\|\mathbf{\Sigma}_f\| \leq \sum_{\ell=0}^{\infty} \|\mathbf{\Psi}_\ell\|^2$ it then follows from part (a) of Assumption 1 that $\|\mathbf{\Sigma}_f\| < C$.*

Remark 3 *Under Assumption 2*

$$\mathbf{D}_\alpha^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} \rightarrow_p \mathbf{\Sigma}_{\beta\beta}(\boldsymbol{\alpha}) > 0, \tag{53}$$

where $\mathbf{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$ is a $k \times k$ symmetric positive definite matrix which is a function of $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_K)'$. This follows from (46) since for any non-zero $n \times 1$ vector $\boldsymbol{\kappa}$,

$$\boldsymbol{\kappa}' \mathbf{D}_\alpha^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} \boldsymbol{\kappa} \geq (\kappa'\kappa)\, \lambda_{min}\left(\mathbf{D}_\alpha^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1}\right) > 0.$$

In the standard case where the factors are all strong ($\alpha_k = 1$ for all k), the above limit reduces to $n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \rightarrow_p \mathbf{\Sigma}_{\beta\beta}(\boldsymbol{\tau}_K) = \mathbf{\Sigma}_{\beta\beta} > 0$.

Remark 4 *The high level condition (50) in Assumption 4 is required for establishing the asymptotic normality of the estimator of $\boldsymbol{\phi}_0$, and is clearly met when v_{it} and/or β_{ik} are independently distributed. It is also possible to establish (50) under weaker conditions assuming that v_{it} and/or β_{ik} satisfy some time-series type mixing conditions applied to cross section.*

Remark 5 *The exponent parameter, α_η, of the pricing condition in (52), can be viewed as the degree to which pricing errors are pervasive in large economies (as $n \to \infty$). Letting $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$ we have*

$$\sum_{i=1}^{n} \eta_i^2 = \|\boldsymbol{\eta}_n\|^2 \leq \|\boldsymbol{\eta}_n\|_\infty \|\boldsymbol{\eta}_n\|_1 = sup_j |\eta_j| \left(\sum_{i=1}^{n} |\eta_i| \right), \tag{54}$$

and under Assumption (5) it also follows that

$$\sum_{i=1}^{n} \eta_i^2 = O\left(n^{\alpha_\eta}\right). \tag{55}$$

Similarly

$$\sum_{i=1}^{n} \gamma_i^2 = O(n^{\alpha_\gamma}). \tag{56}$$

Remark 6 *Whilst (52) implies (55), the reverse does not follow. By allowing for $\alpha_\eta > 0$ we are relaxing the Ross's boundedness condition that requires setting $\alpha_\eta = 0$.*

Remark 7 *The assumption that the observed and missing factors, $\mathbf{f}_t$ and $g_{t'}$, are distributed independently is not restrictive and can be relaxed. For example, suppose that*

$$g_t = \mu_g + \boldsymbol{\theta}'\mathbf{f}_t + v_{gt},$$

where $\boldsymbol{f}_t$ and v_{gt} are independently distributed. Then using (42) we have

$$u_{it} = \gamma_i \mu_g + \gamma_i \left(\boldsymbol{\theta}'\mathbf{f}_t\right) + \gamma_i v_{gt} + v_{it},$$

and the return equation (4) can be written as

$$r_{it} = (\alpha_i + \gamma_i \mu_g) + (\boldsymbol{\beta}_i + \gamma_i \boldsymbol{\theta})' \mathbf{f}_t + \gamma_i v_{gt} + v_{it},$$

with v_{gt} now acting as the missing common factor, which, by construction, is distributed independently of $\mathbf{f}_t$.

Remark 8 *Assumptions 3 and 4 allow u_{it} to be cross-sectionally weakly correlated, but require the errors to be serially uncorrelated. This requirement is not strong for asset pricing models, since realized returns are only mildly serially correlated and most likely such serial dependence will be captured by the serial correlation in the observed factors.*

As we shall see, to estimate and conduct inference on the risk premia associated with the observed factors, f_{kt}, we require $\alpha_k > \alpha_\gamma < 1/2$, where α_γ denotes the strength of the latent factor, g_t, and similarly defined by $\sum_{i=1}^{n} \gamma_i^2 = \ominus(n^{\alpha_\gamma})$. Namely, the latent factor must be sufficiently weak so that ignoring it will be inconsequential, and observed factors sufficiently strong so that they can be distinguished from the weak latent factor.

The main theoretical results of the paper are set out around five theorems. Theorem 1 considers the Fama-MacBeth two-step estimator and derives its limiting property as n and $T \to \infty$. To eliminate the bias of Fama-MacBeth estimator we require $n/T \to 0$, and to eliminate the effects of pricing errors we need $Tn^{\alpha_\eta}/n \to 0$, which results in a contradiction. Thus the Fama-MacBeth estimator is valid only when there are no pricing errors ($\eta_i = 0$ for all i) and when $n/T \to 0$. Theorem 2 provides a proof that the estimator of $\bar{\sigma}_n^2$ (denoted by $\widehat{\bar{\sigma}}_{nT}^2$) proposed by Shanken (1992) continues to be unbiased for a fixed T as

$n \to \infty$, even under the general setting of the current paper that allows for missing factors as well as pricing errors. Theorem 2 also establishes that $\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2 \to O_p(n^{-1/2}T^{-1/2})$, which is essential for establishing the results for the bias-corrected estimator of $\boldsymbol{\phi}_0$, namely $\tilde{\boldsymbol{\phi}}_{nT}$ given by (32), summarized in Theorem 3. This theorem provides conditions under which $\tilde{\boldsymbol{\phi}}_{nT}$ is a consistent estimator of $\boldsymbol{\phi}_0$, and derives its asymptotic distribution assuming the observed factors are strong, again allowing for pricing errors, a missing factor, and other forms of weak error cross-sectional dependence. Theorem 4 extends the results of Theorem 3 to the case where one or more of the observed risk factors are semi-strong and shows how factor strength impacts the precision with which the elements of $\boldsymbol{\phi}_0$ are estimated. Finally, Theorem 5 presents the conditions under which the asymptotic variance of $\tilde{\boldsymbol{\phi}}_{nT}$ can be consistently estimated.

Theorem 1 *(Small T bias of Fama-MacBeth estimator of $\boldsymbol{\lambda}$) Consider the multi-factor linear return model (22) with the missing factor g_t in u_{it} as defined by (42) and the associated risk premia, $\boldsymbol{\lambda}$, defined by (5). Suppose that Assumptions 1, 2, 4, 3 and 5 hold and all observed factors are strong. Suppose further that the true value of the risk premia, $\boldsymbol{\lambda}_0$, is estimated by Fama-MacBeth two-pass estimator, $\hat{\boldsymbol{\lambda}}_{nT}$, defined by (24). Then for any fixed $T > T_0$ such that $\lambda_{\min}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}) > 0$, we have (as $n \to \infty$)*

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) - \frac{\bar{\sigma}^2}{T}\left[\boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2 \frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* + o_p(1), \quad (57)$$

where $\hat{\boldsymbol{\mu}}_T = T^{-1}\sum_{t=1}^T \mathbf{f}_t$,

$$\boldsymbol{\Sigma}_{\beta\beta} = \lim_{n\to\infty}\left(\frac{\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n}{n}\right), \text{ and } \bar{\sigma}^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \sigma_i^2 > 0.$$

The proof is provided in Section B.1 of the mathematical appendix.

To derive the asymptotic distribution of $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0$ it is required that both n and $T \to \infty$, jointly. Also, noting that

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) + \left(\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right),$$

it is clear that increasing n is not relevant for the distribution of $\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0$, but joint n and T asymptotics are required when investigating the distribution of $\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$. Focussing on the latter, and using result (B.12) in the Appendix, we have

$$\left(n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)\sqrt{nT}\left(\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) = n^{-1/2}T^{1/2}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1/2}T^{1/2}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n$$
$$+ n^{-1/2}T^{1/2}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - n^{-1/2}T^{1/2}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*.$$

Where $\boldsymbol{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$, $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$, $\bar{\mathbf{u}}_{n\circ} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, and $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^T u_{it}$. Consider first the terms that include the pricing errors, $\boldsymbol{\eta}_n$, and using the results in Lemma A.3 note that

$$n^{-1/2}T^{1/2}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(T^{1/2}n^{-1/2+\alpha_\eta}\right), \quad n^{-1/2}T^{1/2}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(n^{-1/2+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right).$$

It is clear that the effects of pricing errors on the distribution of $\hat{\boldsymbol{\phi}}_{nT}$ vanish only if $T^{1/2}n^{-1/2+\alpha_\eta} \to 0$, and $\alpha_\eta + \alpha_\gamma < 1$. Also

$$n^{-1/2}T^{1/2}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} = O_p\left(T^{-1/2}\right),$$

$$n^{-1/2}T^{1/2}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^* = \sqrt{\frac{n}{T}}\bar{\sigma}_n^2\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* + O_p\left(T^{-1/2}\right).$$

Finally, for the first two terms involving $\mathbf{B}_n$ and $\mathbf{U}_{nT}$ we have

$$n^{-1/2}T^{1/2}\mathbf{B}_n'\mathbf{M}_n\left(\bar{\mathbf{u}}_{n\circ} - \mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right) = O_p\left(1\right). \tag{58}$$

It is clear that the small T bias of the asymptotic distribution of the two-step estimator, given by $\sqrt{\frac{n}{T}}\bar{\sigma}_n^2\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^*$, does not vanish unless, $n/T \to 0$. At the same time for the pricing errors to have no impact on the distribution of the two-step estimator we must have $Tn^{a_\eta}/n \to 0$. Both conditions cannot be met simultaneously. It is possible to derive the asymptotic distribution of $\hat{\boldsymbol{\phi}}_{nT}$, and hence that of $\hat{\boldsymbol{\lambda}}_{nT}$, when $n/T \to 0$ and $\eta = \mathbf{0}$, but these are quite restrictive conditions, and to avoid them we follow Shanken (1992) and instead consider a bias-corrected version of $\hat{\boldsymbol{\phi}}_{nT}$, namely $\tilde{\boldsymbol{\phi}}_{nT}$ given by (32). As noted earlier $\tilde{\boldsymbol{\phi}}_{nT} = \tilde{\boldsymbol{\lambda}}_{nT} - \hat{\boldsymbol{\mu}}_T$, where $\tilde{\boldsymbol{\lambda}}_{nT}$ is the bias-corrected version of $\hat{\boldsymbol{\lambda}}_{nT}$ originally proposed by Shanken.

To investigate the asymptotic properties of $\tilde{\boldsymbol{\phi}}_{nT}$ we first need to establish conditions under which $\widehat{\bar{\sigma}}_{nT}^2$, defined by (26), is a consistent estimator of $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \sigma_i^2$, which enters the bias-corrected estimator. The proof of consistency in the literature does not allow for missing factors or pricing errors and only considers the case where T is fixed as $n \to \infty$. For derivation of asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT}$ we also need to consider the limiting properties of $\widehat{\bar{\sigma}}_{nT}^2$ under joint n and T asymptotics. The following theorem provides the required results for $\widehat{\bar{\sigma}}_{nT}^2$ as an estimator of $\bar{\sigma}_n^2$.

Theorem 2 *Consider $\widehat{\bar{\sigma}}_{nT}^2$, the estimator of $\bar{\sigma}_n^2$ given by (see (26)),*

$$\widehat{\bar{\sigma}}_{nT}^2 = \frac{\sum_{t=1}^T\sum_{i=1}^n \hat{u}_{it}^2}{n(T-K-1)}, \tag{59}$$

and suppose that Assumptions 1, 3, and 4, are satisfied. Then for a fixed T

$$\lim_{n\to\infty} E\left(\widehat{\bar{\sigma}}_{nT}^2\right) = \bar{\sigma}^2, \tag{60}$$

where $\bar{\sigma}^2 = \lim_{n\to\infty}\bar{\sigma}_n^2$, and $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \sigma_i^2$. Furthermore

$$\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2 = O_p\left(T^{-1/2}n^{-1/2}\right). \tag{61}$$

For a proof see sub-section B.2 in the Appendix.

Result (61) shows that $\widehat{\bar{\sigma}}_{nT}^2$ continues to be a consistent estimator of $\bar{\sigma}^2 = \lim_{n\to\infty}\bar{\sigma}_n^2$ for a fixed T as $n \to \infty$, even in the presence of pricing errors and a missing common factor. This result also holds when one or more of the factors are semi-strong.

Equipped with the above result we are now in a position to present the theorem that sets out the asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT}$.

Theorem 3 *Consider, $\tilde{\boldsymbol{\phi}}_{nT}$, the bias-corrected estimators of $\boldsymbol{\phi}_0$ given by (32). Suppose Assumptions 1, 2, 4, 3 and 5 hold, all the observed factors are strong, ($\alpha_k = 1$, for*

$k = 1, 2, ..., K$), and the strength of the missing factor, α_γ defined by (48), satisfies α_γ $< 1/2$.

$$\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 = O_p\left(T^{-1/2}n^{-1/2}\right) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(n^{-1+\alpha_\eta}\right) + O_p\left(T^{-1}n^{-1/2}\right), \tag{62}$$

where α_η denotes the degree of pervasiveness of the pricing errors defined by (43). (a) When T is fixed, $\alpha_\gamma < 1/2$ and $\alpha_\eta < 1$, then there exists T_0 such that for all $T > T_0$

$$p\lim_{n\to\infty}\left(\tilde{\boldsymbol{\phi}}_{nT}\right) = \boldsymbol{\phi}_0. \tag{63}$$

Also

$$\sqrt{nT}\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) = \boldsymbol{\Sigma}_{\beta\beta}^{-1}\boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2}+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{1/2}n^{-1/2+\alpha_\eta}\right) + O_p\left(T^{-1/2}\right), \tag{64}$$

where $\boldsymbol{\Sigma}_{\beta\beta} = p\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)$,

$$\boldsymbol{\xi}_{nT} = n^{-1/2}T^{-1/2}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{a}_T, \tag{65}$$

and $\mathbf{a}_T = \tau_T - \mathbf{M}_T\mathbf{F}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\boldsymbol{\lambda}_T^*$. (b) If $\alpha_\gamma < 1/2$, $\alpha_\eta < 1/2$, and $\sqrt{\frac{T}{n}}n^{\alpha_\eta} \to 0$, as n and $T \to \infty$ jointly, then

$$\sqrt{nT}\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) \to_d N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\beta\beta}^{-1}\mathbf{V}_\xi\boldsymbol{\Sigma}_{\beta\beta}^{-1}\right), \tag{66}$$

where

$$\mathbf{V}_\xi = \left(1 + \boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0\right)\ p\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\right). \tag{67}$$

For a proof see sub-section B.3 in the Appendix.

Result (62) establishes the finite T consistency of $\tilde{\boldsymbol{\phi}}_{nT}$ for $\boldsymbol{\phi}_0$ so long as $\alpha_\gamma < 1/2$ and $\alpha_\eta < 1$, thus extending the Shanken result to a much more general setting. To the best of our knowledge the asymptotic distribution in (66) is new and shows that the asymptotic covariance matrix of $\tilde{\boldsymbol{\phi}}_{nT}$ includes the term $\boldsymbol{\lambda}_T^{*\prime}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\boldsymbol{\lambda}_T^*$, that arises from the first stage estimation of the factor loadings, and must be included in the analysis for valid inference. It is also clear that this additional term does not vanish with $T \to \infty$, and tends to $\boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0 \geq (\boldsymbol{\lambda}_0'\boldsymbol{\lambda}_0)\lambda_{\max}\left(\boldsymbol{\Sigma}_f^{-1}\right) = (\boldsymbol{\lambda}_0'\boldsymbol{\lambda}_0)\lambda_{\min}\left(\boldsymbol{\Sigma}_f\right) > 0$, which is strictly non-zero unless $\boldsymbol{\lambda}_0 = \mathbf{0}$. Shanken type bias correction addresses the mean of the asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT}$, but not its covariance.

The $O_p\left(T^{-1/2}\right)$ term in (64) arises from the sampling errors involved in the estimation of the factor loadings and $\bar{\sigma}_n^2$, and tends to zero at the regular $\sqrt{T}$ rate. But n has to be sufficiently large to eliminate the effects of pricing errors on identification of $\boldsymbol{\phi}_0$, as dictated by condition $\sqrt{\frac{T}{n}}n^{\alpha_\eta} \to 0$, as n and $T \to \infty$.[7] The requirement that T need not be too large relative to n for estimation of $\boldsymbol{\phi}_0$ is consistent with separating the estimation of $\boldsymbol{\phi}_0$ from that of $\boldsymbol{\mu}_0$, allowing the use a relatively small T and a large n to estimate $\boldsymbol{\phi}_0$ and a relatively large T when estimating $\boldsymbol{\mu}_0$.

[7]The condition $\sqrt{\frac{T}{n}}n^{\alpha_\eta} \to 0$ can be weakened somewhat to $\sqrt{\frac{T}{n}}n^{\alpha_\eta/2} \to 0$ if we also assume that $\beta_{ik} - \bar{\beta}_k$ are independently distributed over i, but will still require n to be larger than T.

3.1 What if one or more of the risk factors are semi-strong?

We now turn to an intermediate case where one or more of the observed factors are semi-strong, in the sense that their factor strength, α_k lies between 1/2 and 1. The case of weak risk factors is already covered in the proceeding analysis, and such factors can be included in the error term, u_{it}, with little consequence for the estimation of risk premia of the remaining factors that are strong or semi-strong. Weak factors do not have any explanatory power and can be dropped from the analysis.

When one or more of the observed factors is semi-strong $\mathbf{\Sigma}_{\beta\beta}$ is no longer positive definite and Theorem 3 does not apply, but it is possible to adapt the proofs to establish the limiting properties of $\tilde{\phi}_{k,nT}$ (the k^{th} element of $\tilde{\boldsymbol{\phi}}_{nT}$) for different values of α_k .

To this end, analogously to $\tilde{\boldsymbol{\phi}}_{nT}$, we introduce the following estimator of $\boldsymbol{\phi}_0$

$$\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) = \mathbf{H}_{nT}^{-1}(\boldsymbol{\alpha}) \left[\mathbf{D}_{\alpha}^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{a}}_{nT} + \frac{n}{T}\hat{\bar{\sigma}}_{nT}^2\mathbf{D}_{\alpha}^{-1}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\hat{\boldsymbol{\mu}}\right], \tag{68}$$

where

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_{\alpha}^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\mathbf{D}_{\alpha}^{-1} - \frac{n}{T}\hat{\bar{\sigma}}_{nT}^2\mathbf{D}_{\alpha}^{-1}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\mathbf{D}_{\alpha}^{-1}. \tag{69}$$

It is now easily seen that

$$\mathbf{D}_{\alpha}\left(\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0\right) = \mathbf{H}_{nT}^{-1}(\boldsymbol{\alpha})\,\mathbf{q}_{nT}(\boldsymbol{\alpha}), \tag{70}$$

where

$$\mathbf{q}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_{\alpha}^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{a}}_{nT} + \frac{n}{T}\hat{\bar{\sigma}}_{nT}^2\mathbf{D}_{\alpha}^{-1}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\hat{\boldsymbol{\mu}}_T - \mathbf{H}_{nT}(\boldsymbol{\alpha})\mathbf{D}_{\alpha}\boldsymbol{\phi}_0. \tag{71}$$

$\mathbf{D}_{\alpha}$ is defined by (47), and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_K)'$. It is easily established that numerically $\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha})$ is identical to $\tilde{\boldsymbol{\phi}}_{nT}$, and its introduction is primarily for the purpose of establishing the limiting properties of $\tilde{\phi}_{k,nT} - \phi_{0,k}$ that do depend on α_k. Note that

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = n\mathbf{D}_{\alpha}^{-1}\mathbf{H}_{nT}\mathbf{D}_{\alpha}^{-1}, \text{ and } \mathbf{q}_{nT}(\boldsymbol{\alpha}) = n\mathbf{D}_{\alpha}^{-1}\mathbf{s}_{nT}$$

where $\mathbf{s}_{nT}$ and $\mathbf{H}_{nT}$ are already defined by (B.23) and (B.24). Using these in (70) we have

$$\mathbf{D}_{\alpha}\left(\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0\right) = \left(n\mathbf{D}_{\alpha}^{-1}\mathbf{H}_{nT}\mathbf{D}_{\alpha}^{-1}\right)^{-1} n\mathbf{D}_{\alpha}^{-1}\mathbf{s}_{nT} = \mathbf{D}_{\alpha}\mathbf{H}_{nT}^{-1}\mathbf{s}_{nT},$$

and it follows that $\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0 = \mathbf{H}_{nT}^{-1}\mathbf{s}_{nT} = \tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\tau}_K) = \tilde{\boldsymbol{\phi}}_{nT}$. See (B.22).

The convergence results for $\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha})$ are set out in the following theorem.

Theorem 4 *Consider, $\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha})$, the bias-corrected estimators of $\boldsymbol{\phi}_0$ given by (68), and suppose Assumptions 1, 2, 4, 3 and 5 hold, the strength of observed factors, $\boldsymbol{f}_t = (f_{1t}, f_{2t}, ..., f_{Kt})'$, is given by $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_K)'$, and the strength of the missing factor, g_t, defined by (48) is α_{γ}. Let $\alpha_{\min} = \min_k(\alpha_k)$ and suppose that $\alpha_{\gamma} < 1/2$. Then*

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_{\alpha}^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_{\alpha}^{-1} + O_p\left(T^{-1}n^{-\alpha_{\min}+1/2}\right),$$

where $\mathbf{H}_{nT}(\boldsymbol{\alpha})$ is given by (69), and by part (b) of Assumption 2, $\mathbf{H}_{nT}(\boldsymbol{\alpha}) \to_p \mathbf{\Sigma}_{\beta\beta}(\boldsymbol{\alpha}) > 0$, for any fixed $T > T_0$ such that $\lambda_{\max}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} < C$ and $\alpha_{\min} > 1/2 > \alpha_{\gamma}$. Also

$$\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} = O_p\left(n^{-(\alpha_k+\alpha_{\min})/2+1/2}T^{-1/2}\right) + O_p\left(n^{\frac{-(\alpha_k+\alpha_{\min})+\left(\alpha_{\eta}+\alpha_{\gamma}\right)}{2}}T^{-1/2}\right) \tag{72}$$

$$+ O_p\left(n^{-(\alpha_k+\alpha_{\min})/2+\alpha_{\eta}}\right) + O_p\left(n^{-(\alpha_k+\alpha_{\min})/2+1/2}T^{-1}\right).$$

See sub-section B.4 of the Appendix for a proof.

The result in (72) establishes the consistency of $\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) = \tilde{\phi}_{k,nT}$ even if f_{kt} is semi-strong so long as $n \to \infty$, and $\alpha_{\min} > 1/2$, $\alpha_k + \alpha_{\min} > \alpha_\eta + \alpha_\gamma$, and $\alpha_k + \alpha_{\min} > 2\alpha_\eta$. Clearly, these results reduce to the case of strong factors where $\alpha_{\min} = \alpha_k = 1$. Turning to the asymptotic distribution of $\tilde{\phi}_{k,nT}(\boldsymbol{\alpha})$, again only convergence rates are affected, and instead of the regular rate of $\sqrt{nT}$, we have $\sqrt{T}n^{(\alpha_k+\alpha_{\min}-1)/2}$, and using (72) we have

$$\sqrt{T}n^{(\alpha_k+\alpha_{\min}-1)/2}\left(\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k}\right) \tag{73}$$

$$= O_p(1) + O_p\left(n^{-1/2+\frac{(\alpha_\eta+\alpha_\gamma)}{2}}\right) + O_p\left(\sqrt{T}n^{-1/2+\alpha_\eta}\right) + O_p\left(T^{-1/2}\right),$$

and The conditions needed for eliminating the effects of the pricing errors are the same as before and are given by $\alpha_\eta + \alpha_\gamma < 1$ and $\sqrt{T}n^{-1/2+\alpha_\eta} \to 0$. The asymptotic distribution is unaffected except for the slower rate of convergence alluded to above. It is also of interest to note that adding semi-strong factors can adversely affect the convergence rate of the strong factor with $\alpha_k = 1$. As an example suppose the asset pricing model contains two factors, one strong, $\alpha_1 = 1$ and one semi strong with $\alpha_2 < 1$. Then the convergence rate of $\tilde{\phi}_{1,nT}(\boldsymbol{\alpha}) - \phi_{0,1}$ is given by$\sqrt{T}\ n^{(1+\alpha_2-1)/2}$ which is slower than the rate we would have obtained for $\tilde{\phi}_{1,nT}(\boldsymbol{\alpha}) - \phi_{0,1}$ if both factors were strong ($\alpha_{min} = \alpha_k = 1$), namely the regular rate of $\sqrt{nT}$.

Furthermore, when conditions $\alpha_\eta + \alpha_\gamma < 1$ and $\sqrt{T}n^{-1/2+\alpha_\eta} \to 0$ are met we have

$$\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} = O_p\left(T^{-1/2}n^{-(\alpha_k+\alpha_{\min}-1)/2}\right), \tag{74}$$

and $\phi_{0,k}$ is consistently estimated if $T^{-1/2}n^{-(\alpha_k+\alpha_{\min}-1)/2} \to 0$. Also, using (33), an estimator of the risk premia, λ_k, is given by $\tilde{\lambda}_{k,nT}(\boldsymbol{\alpha}) = \tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) + \hat{\mu}_{k,T}$, where $\hat{\mu}_{k,T} = T^{-1}\sum_{t=1}^{T} f_{kt}$. Hence

$$\tilde{\lambda}_{k,nT}(\boldsymbol{\alpha}) - \lambda_{0,k} = \left[\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k}\right] + \left(\hat{\mu}_{k,T} - \mu_{0,k}\right).$$

Under Assumption 1 $\hat{\mu}_{k,T} - \mu_{0,k} = O_p\left(T^{-1/2}\right)$, and using (74) it then follows that

$$\tilde{\lambda}_{k,nT}(\boldsymbol{\alpha}) - \lambda_{0,k} = O_p\left(T^{-1/2}n^{-(\alpha_k+\alpha_{\min}-1)/2}\right) + O_p\left(T^{-1/2}\right), \tag{75}$$

and $\tilde{\lambda}_{k,nT}(\boldsymbol{\alpha})$ is a consistent estimator of $\lambda_{0,k}$ if $T \to \infty$ as well as $T^{-1/2}n^{-(\alpha_k+\alpha_{\min}-1)/2} \to 0$. More specifically, suppose $T = \ominus\left(n^d\right)$ for some $d > 0$, where $\ominus(\cdot)$ denotes T and n^d are of the same order of magnitude. Then for any $d > 0$, the condition for consistency of $\tilde{\lambda}_{k,nT}(\boldsymbol{\alpha})$ is given by $\alpha_k + \alpha_{min} + d > 1$ and $d > 0$. In the case where all risk factors have the same strength, α, the consistency condition reduces to $\alpha > (1-d)/2$, which is weaker than the one derived by Giglio, Xiu, and Zhang (2023), namely $n/(\|\beta\|^2 T) \to 0$, where, in terms of our notation, $\|\beta\|^2 = \ominus(n^\alpha)$. This latter condition will be met if $\alpha > 1-d$. In practice where T is small relative to n, the accuracy of $\tilde{\lambda}_{k,nT}(\boldsymbol{\alpha})$ as an estimator $\lambda_{0,k}$ does depend on α, and our weaker condition on $\alpha > (1-d)/2$ is advantageous.

3.2 Consistent estimation of the variance of $\tilde{\boldsymbol{\phi}}_{nT}$

To carry out inference on $\boldsymbol{\phi}_0$, or any of its elements individually, we require a consistent estimator of $Var\left(\tilde{\boldsymbol{\phi}}_{nT}\right)$. Using (66) and (67) we first note that $\boldsymbol{\Sigma}_{\beta\beta}$ is consistently estimated by $\mathbf{H}_{nT}$ given by (29). Therefore, it is sufficient to find a suitable estimator of $\mathbf{V}_u = (\sigma_{ij})$ such that $\mathbf{V}_\xi$ given by (67) is consistently estimated. Under suitable sparsity restrictions $\mathbf{V}_u$ can be consistently estimated using the various thresholding procedures advanced in the statistical literature by Bickel and Levina (2008a,b), Cai and Liu (2011), and Bailey, Pesaran, and Smith (2019, BPS). Fan, Liao, and Mincheva (2011, 2013) also show that the adaptive threshold technique of Cai and Liu applies equally to the residuals from an approximate factor model. Here we consider the threshold estimator proposed by BPS which does not require cross-validation and is shown to have desirable small sample properties. It is given by $\tilde{\mathbf{V}}_u = (\tilde{\sigma}_{ij})$

$$\begin{aligned} &\tilde{\sigma}_{ii} = \hat{\sigma}_{ii} \\ &\tilde{\sigma}_{ij} = \hat{\sigma}_{ij}\mathbf{1}\left[|\hat{\rho}_{ij}| > T^{-1/2}c_\alpha(n,\delta)\right],\ i = 1, 2, \ldots, n-1,\ j = i+1, \ldots, n, \end{aligned} \tag{76}$$

where

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt},\ \hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}},\quad \hat{u}_{it} = r_{it} - \hat{\alpha}_{i,T} - \hat{\boldsymbol{\beta}}'_{i,T}\mathbf{f}_t, \tag{77}$$

and $c_p(n,d) = \Phi^{-1}\left(1 - \frac{p}{2n^d}\right)$, is a normal critical value function, p is the the nominal size of testing of $\sigma_{ij} = 0$, $(i \neq j)$ and d is chosen to take account of the $n(n-1)/2$ multiple tests being carried out. Monte Carlo experiments carried out by BPS suggest setting $d = 2$. The variance estimator given by (76) does not require a knowledge of the factor strength and applies to risk factors of differing degrees.

Under Assumptions 1, 3, and 4, $\|\mathbf{V}_u\| = O\left(n^{\alpha_\gamma}\right)$, and using results in Fan, Liao, and Mincheva (2011, 2013) we have

$$\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\| = O_p\left(n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right). \tag{78}$$

Consider the following estimator of $\mathbf{V}_\xi$

$$\hat{\mathbf{V}}_{\xi,nT} = (1 + \hat{s}_{nT})\left(n^{-1}\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right).$$

where $\hat{s}_{nT} = \tilde{\boldsymbol{\lambda}}'_{nT}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\tilde{\boldsymbol{\lambda}}_{nT}$. Under Assumption 1, $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F} \to_p \boldsymbol{\Sigma}_f$ and using the results above we have $\tilde{\boldsymbol{\lambda}}_{nT} = \tilde{\boldsymbol{\phi}}_{nT} + \hat{\boldsymbol{\mu}}_T \to_p \boldsymbol{\phi}_0 + \boldsymbol{\mu}_0 = \boldsymbol{\lambda}_0$. Hence, $\hat{s}_{nT} \to_p \boldsymbol{\lambda}'_0\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0$ as $n, T \to \infty$, jointly, and it is sufficient to show that

$$n^{-1}\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT} - n^{-1}\mathbf{B}'_n\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n \to_p \mathbf{0}. \tag{79}$$

The following theorem provides a formal statement of the conditions under which $\hat{\mathbf{V}}_{\xi,nT}$ is a consistent estimator of $\mathbf{V}_\xi$.

Theorem 5 *Suppose Assumptions 1, 2, 4, 3 and 5 hold, and all the observed factors are strong, ($\alpha_k = 1$, for $k = 1, 2, ..., K$), and the strength of the missing factor, g_t, defined by (48), $\alpha_\gamma < 1/2$. Then*

$$\left\|\hat{\mathbf{V}}_{\xi,nT} - \mathbf{V}_\xi\right\| = O_p\left(n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right), \tag{80}$$

where

$$\hat{\mathbf{V}}_{\xi,nT} = (1+\hat{s}_{nT})\left(n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right), \tag{81}$$

$\tilde{\mathbf{V}}_u = (\tilde{\sigma}_{ij})$, $\tilde{\sigma}_{ij}$ *is the threshold estimator of* σ_{ij} *given by (76), and*

$$\mathbf{V}_\xi = \left(1+\boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0\right)\ p\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\ \right), \tag{82}$$

For a proof see sub-section B.5 in the Appendix.

This theorem shows that consistent estimation of $Var\left(\tilde{\boldsymbol{\phi}}_{nT}\right)$ can be achieved by using a suitable threshold estimator of $\mathbf{V}_u$, so long as the strength of the missing factor, α_γ, is sufficiently weak in the sense that $n^{\alpha_\gamma}\sqrt{\ln(n)/T} \to 0$ as $n, T \to \infty$.

4 Small sample properties of the estimators and tests for $\boldsymbol{\phi}$

4.1 Monte Carlo Design

This section presents Monte Carlo simulations to investigate the small sample properties of estimators and tests for $\boldsymbol{\phi}_0$. In the empirical application of the next section the factors are selected from a large list. But here we assume $K = 3$ and mimic the 3 Fama-French factors, namely the market return minus the risk free rate, MKT, the value factor (high minus low book to market portfolios, HML) and the size factor (small minus big portfolios, SMB). These are denoted by f_{kt}, $k = M, H, S$.[8] For further details see Section C of the online supplement A.

4.1.1 Loadings and factor strengths

To calibrate the loadings, β_{ik}, we used excess returns on a large number securities observed over the shorter sample covering the 20 years $2002m1$ $-2021m12$ ($T = 240$). Monthly returns for NYSE and NASDAQ stocks code 10 and 11 from CRSP were downloaded from Wharton Research Data Services and converted to excess returns over the risk free rate, taken from Kenneth French's webpages, in percent per month. Only stocks with available data for the full sample were included, yielding a balanced panel, and to avoid outliers influencing the results, stocks with a kurtosis greater than 16 were excluded. There were 1289 stocks before exclusion on the basis of kurtosis and 1175 after. The summary statistics giving mean, median, standard deviation of the estimates of β_{ik} and their histograms are provided in the online supplement A.

For factor strength, we considered a range of DGPs. Given the evidence that most factors, other than the market factor, are not strong, we focus on the case where there is one strong factor, namely the market factor with $\alpha_M = 1$, plus two semi-strong factors,

[8] Data on factors and the risk free rate are downloaded from Kenneth French's data library: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

with the value factor, HML, being quite strong with $\alpha_H = 0.85$, and the size factor, SML, being only moderately strong with $\alpha_S = 0.65$. These estimates are also informed by the results provided in Bailey, Kapetanios, and Pesaran (2021) who propose methods for estimation of factor strength. For a given factor strength, α_k, the associated loadings, β_{ik}, are generated as $\boldsymbol{\beta}_k = (\beta_{1k}, \beta_{2k}, , ..., \beta_{\alpha_k}, 0, 0, ...0)$ where $n_{\alpha_k} = \lfloor n^{\alpha_k} \rfloor$ the integer part of n^{α_k}, with non-zero and zero values of $\boldsymbol{\beta}_k$ given by

$$\beta_{ik} \sim IIDN(\mu_{\beta_k}, \sigma^2_{\beta_k}), \text{ for } i = 1, 2, ..., \lfloor n^{\alpha_k} \rfloor,$$

$$\beta_{ik} = 0 \text{ for } i = \lfloor n^{\alpha_k} \rfloor + 1, \lfloor n^{\alpha_k} \rfloor + 2, ..., n,$$

where $\lfloor n^{\alpha_k} \rfloor$ denotes the integer part of n^{α_k}. Since the security returns are randomly generated, it does not matter how zero and non-zero values of β_{ik} are distributed across i. Also, the zero loadings can also be replaced by an exponentially decaying sequence without any implications for the simulation results.[9] We also set

$$\mu_{\beta_M} = 1, \sigma_{\beta_M} = 0.4; \ \mu_{\beta_H} = 0.2, \sigma_{\beta_H} = 0.5$$

$$\mu_{\beta_S} = 0.6, \sigma_{\beta_S} = 0.5,$$

which match the mean and standard deviation of the estimates of β_{ik}. See above.

4.1.2 Generation of pricing errors

The pricing errors in (6) can be considered as firm-specific characteristics and are set as $\boldsymbol{\eta}_n = \left(\eta_1, \eta_2, ..., \eta_{n_\eta}, 0, 0, ..., 0\right)'$. The non-zero loadings of $\boldsymbol{\eta}_n$ for $i \leq n_\eta = \lfloor n^{\alpha_\eta} \rfloor$ are drawn from $IIDU(0.7, 0.9)$, and $\eta_i = 0$ for $i = n_\eta + 1, n_\eta + 2, ..., n$. We consider $\alpha_\eta = (0, 0.3)$. When $\alpha_\eta = 0$ we have $\eta_i = 0$, for all i. As in the case of factor loadings the non-zero values of $\boldsymbol{\eta}_n$ must be randomly allocated to different groups.

4.1.3 Generation of return equation errors

The return equation errors, u_{it}, are generated following (42) as a combination of a missing factor, $g_t \sim IIDN(0, 1)$ plus an idiosyncratic error, v_{jt}. The loadings $\boldsymbol{\gamma} = \left(\gamma_1, \gamma_2, ..., \gamma_{n_\gamma}, 0, 0, ..., 0\right)'$ of the missing factor are set as

$$\gamma_i \sim IIDU(0.7, 0.9), \text{ for } i = 1, 2, ..., \lfloor n^{\alpha_\gamma} \rfloor,$$

$$\gamma_i = 0, \text{ for } i = \lfloor n^{\alpha_\gamma} \rfloor + 1, \lfloor n^{\alpha_\gamma} \rfloor + 2, ..., n,$$

where α_γ is the strength of the missing factor g_t. We consider $\alpha_\gamma = 1/4$ and $1/2$.

For the idiosyncratic errors, v_{it}, we consider spatial as well as a block diagonal specification, with the spatial specification including a diagonal specification as the special case. Under the spatial specification the idiosyncratic errors are generated as the first order spatial autoregressive model $v_{it} = \rho_\varepsilon \sum_{j=1}^n w_{ij} v_{jt} + \kappa \varepsilon_{it}$, which can be written in matrix notation as $\mathbf{v}_t = \rho_\varepsilon \mathbf{W} \mathbf{v}_t + \kappa \boldsymbol{\varepsilon}_t$, and solved for as $\mathbf{v}_t = \kappa \left(\mathbf{I}_n - \rho_\varepsilon \mathbf{W}\right)^{-1} \boldsymbol{\varepsilon}_t$. Adding the missing factor now yields

$$\mathbf{u}_t = \boldsymbol{\gamma}\ g_t + \kappa \left(\mathbf{I}_n - \rho_\varepsilon \mathbf{W}\right)^{-1} \boldsymbol{\varepsilon}_t. \tag{83}$$

[9] See also footnote 5 of Bailey, Kapetanios, and Pesaran (2016, p.942).

The spatial coefficient ρ_ε is such that $|\rho_\varepsilon| < 1$, $\mathbf{W} = (w_{ij})$ with $w_{ii} = 0$, and $\sum_{j=1}^{n} w_{ij} = 1$. The diagonal case is obtained by setting $\rho_\varepsilon = 0$, with $\rho_\varepsilon = 0.5$ characterizing the SAR specification. The weight matrix $\mathbf{W} = (w_{ij})$ is set to follow the familiar rook pattern where all its elements are set to zero except for $w_{i+1,i} = w_{j-1,j} = 0.5$ for $i = 1, 2, ..., n-2$ and $j = 3, 4..., n$, with $w_{1,2} = w_{n,n-1} = 1$.

Under the block error covariance specification, $\mathbf{v}_t$ is generated as $\mathbf{v}_t = \kappa \hat{\mathbf{S}} \boldsymbol{\varepsilon}_t$, where $\hat{\mathbf{S}}$ is a block diagonal matrix with its b^{th} block given by $\hat{\mathbf{S}}_b$ for $b = 1, 2, ..., B$, and $\boldsymbol{\varepsilon}_t = (\varepsilon'_{1t}, \varepsilon'_{2t}, ..., \varepsilon'_{Bt})'$, and $\boldsymbol{\varepsilon}_{bt} = (\varepsilon_{b,1t}, \varepsilon_{b,2t}, ..., \varepsilon_{b,n_b,t})'$. $\hat{\mathbf{S}}$ is set as a Cholesky factor of the correlation matrix of $\mathbf{u}_t$. Denoting this correlation matrix by $\hat{\mathbf{R}}_u$,

$$\hat{\mathbf{R}}_u = \left[Diag(\hat{\mathbf{V}}_{Bu})\right]^{-1/2} \hat{\mathbf{V}}_{Bu} \left[Diag(\hat{\mathbf{V}}_{Bu})\right]^{-1/2} = Diag(\hat{\mathbf{R}}_{bu},\ b = 1, 2, ..., B),$$

where $\hat{\mathbf{V}}_{Bu}$ is the threshold estimator of $\mathbf{V}_u$ subject to the additional restriction that $\mathbf{V}_u$ is block diagonal. For each block $\hat{\mathbf{R}}_{bu}$ we set the number of distinct non-zero elements of this block equal to the integer part of $[n_b(n_b - 1)/2] \times q_b$ where q_b is the proportion of non-zero distinct elements in block b of our calibrated sample and computed by the calibration over the sample $2001m10 - 2021m9$. The non-zero elements are drawn randomly from $IIDU(0, 0.5)$. Similarly, adding the missing factor, we have

$$\mathbf{u}_t = \boldsymbol{\gamma} g_t + \kappa \hat{\mathbf{S}} \boldsymbol{\varepsilon}_t. \tag{84}$$

The block diagonal structure is intended to capture possible within industry correlations not picked up by observed or weak missing factors, with each block representing an industry or sector. To calibrate the block structure estimates of the pair-wise correlations between the residuals of the return regressions using the Fama-French three factors of the $T = 240$ sample ending in 2021 were obtained. Then all the statistically insignificant correlations were set to zero, allowing for the multiple testing nature of the tests. For the majority of securities (668 out of the 1168), the pair-wise return correlations were not statistically significant. The securities with a relatively large number of non-zero correlations were either in the banking or energy related industries. Considering stocks by 2-digit SIC classifications, a division into $B = 14$ contiguous groups ranging in size from 33 to 145 stocks, seemed sensible. More detail on the process is given in Section D of the online supplement A.

The primitive errors, ε_{it} for $i = 1, 2, ..., n$ in (83) and (84) are generated as $\varepsilon_{it} = \sqrt{\sigma_{ii}} \varpi_{it}$, where $\varpi_{it} \sim IIDN(0, 1)$, and $\varepsilon_{it} = \sqrt{\sigma_{ii}} \left[\sqrt{\frac{v-2}{v}} \varpi_{it}\right]$, where $\varpi_{it} \sim IID\ t(v)$, with $t(v)$ denotes a standard t distributed variate with $v = 5$ degrees of freedom. Also $\sigma_{ii} \sim IID\ 0.5(1+\chi_1^2)$, $for\ i = 1, 2, ..., n_b$ and $b = 1, 2, ..., B$. In this way, it is ensured that $Var(\varepsilon_{b,it}) = \sigma_{b,ii}$, and on average $E\left[Var(\varepsilon_{b,it})\right] = E(\sigma_{ii}) = 1$, under both Gaussian and t-distributed errors. Note that $Var(\nu_{b,it}) = v/(v-2)$. All the experiments are designed to give an R^2 of about 0.3, similar to that obtained in the empirical applications. For further details see sub-section C.5 of the online supplement A.

4.1.4 Experiments

In total, we consider 12 experimental designs: six designs with Gaussian errors and six with $t(5)$ distributed errors. We considered designs with GARCH effects, with and without pricing errors, η_i, and with and without the missing factor, g_t. We also considered designs with spatial patterns in the idiosyncratic errors, v_{it}. All experiments are implemented using $R = 2,000$ replications. Details of of the 12 experiments are summarized in Table S-1 of the online supplement B (MC results).

4.1.5 Alternative estimators of $\mathbf{V}_u$

Subsection 3.2 considered consistent estimation of the variance of $\tilde{\phi}_{nT}$ using $\tilde{\mathbf{V}}_u$ a threshold estimator for $\mathbf{V}_u$, given by equation (76). For comparison purposes we also considered two other estimators of $\mathbf{V}_u$. These were the sample covariance matrix $\hat{\mathbf{V}}_u = \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' / T$ and a diagonal covariance matrix, where the off-diagonal elements of $\hat{\mathbf{V}}_u$, $\hat{\sigma}_{ij}$, are set to zero. Thus we have three designs for the return error covariance matrix, $\mathbf{V}_u$, and three different estimators of it. A comparison of the results for the different covariance matrices is available on request. The diagonal estimator, as to be expected, performed poorly when the true covariance matrix was not diagonal, particularly for the spatial error covariance matrix and when the strength of the missing factor was close to $1/2$. For these designs the sample and threshold estimators of the covariance matrix generally performed similarly and given that there is a theoretical justification for the threshold estimator and there are structures of the error covariance matrix for which the sample estimator is unlikely to perform well we report the results using the threshold estimator in the simulations below.

4.2 Monte Carlo results

We focus on a comparison of two-step (defined by (31) and the bias-corrected (BC) estimator (defined by (32)), and report bias, root mean square error (RMSE) and size for testing $H_{0j} : \phi_{0k} = 0$, $k = M, H, S$ at the five per cent nominal level, for all $n = 100, 500, 1,000, 3,000$ and $T = 60$, 120, 240 combinations. The results for all 12 experiments are summarized in Tables S-A-E1-S-A-E12 in the online supplement B. In terms of bias and RMSE the two-step estimator does much better than the bias-corrected (BC) estimator when $T = 60$ and $n = 100$, but this gap closes quickly as n is increased. In fact for $T = 60$ and $n = 3,000$, the bias and RMSE of the BC estimator (at 0.0010 and 0.0607) are much less than those of the two-step estimator (at -0.0080 and 0.1489). This pattern continues to hold when $T = 120$ and 240. Bias correction can cause the RMSE to "blow up" for small samples, such as $n = 100$ and $T = 60$, but for $n = 500$ and above the bias-corrected estimator always has a smaller RMSE than the two-step estimator. As discussed in the theoretical section, having a large n is important for the properties of the estimators.

But most importantly, the two-step estimator is subject to substantial size distortions, particularly when T is small relative to n. As predicted by the theory, the degree of over-rejection of the tests based on FM estimator falls with T, but increases with n. For example, the two-step test sizes rise from 11.1% when $T = 60$ and $n = 100$ to 60.9% when $T = 60$ and $n = 3,000$. Increasing T reduces the size distortion of the two-step estimator but test sizes are still substantially above the 5% nominal value when n is large. The strong tendency of the tests based on the two-step estimator to over-reject could be an important contributory factor leading to false discovery of a large number of apparently significant factors in the literature. In contrast, sizes of the tests based on the BC estimator, using the variance estimator given by (36), are all close to its nominal value, irrespective of the factor strength or sample size combinations. We only note some elevated test sizes in the case of the experimental design 12, and when we consider the semi-strong factors. The highest test size of 7.85 per cent is obtained for the least strong factor, f_{st}, when $n = 3000$ and $T = 60$. See Table S-A-E12 of the online supplement B.

We also experimented with raising α_η from 0.3 to 0.5, making the pricing errors much more pervasive and ρ_ε from 0.5 to 0.85, introducing more spatial correlation. This increased the rejection rate in experiments 9 and 10.

4.2.1 Empirical power functions

Plots of the empirical power functions for testing the null hypothesis $H_{0j} : \phi_{0k} = 0$, $k = M, H, S$, are also provided in the online supplement B (Figures S-A-E1 to S-A-E12). All power curves have the familiar bell curve shape and tend to unity as n and/or T are increased, showing the test has satisfactory power, particularly for n sufficiently large even when $T = 60$. Again the power functions are quite similar across the 12 different experiments and show similar patterns for strong and semi-strong factors. However, this similarity hides the fact that the test of $\phi_k = 0$ for the strong factor is much more powerful than corresponding tests for the semi-strong factors, with the test power declining as factor strength is reduced.

4.2.2 Differences in performance of strong and semi-strong factors

These differences in the effects of factor strength on the power of the test of $\phi_k = 0$ are in line with our theoretical results, and are also reflected in the rate at which the RMSE of the estimators of ϕ_M, ϕ_H and ϕ_S fall with n. For example, using results in Table S-B-E12, for the bias-corrected estimator and design 12 with $T = 240$, the ratio of RMSE of $n = 3,000$ to $n = 100$ is 17% for the strong factor ($\alpha_M = 1$), 23% for the first semi-strong factor with $\alpha_H = 0.85$, and 36% for the second semi-strong factor with $\alpha_S = 0.65$. As strength falls one needs larger cross section samples of securities to attain the same level of precision. In the case of the two-step estimator there was the same pattern, but the fall in the RMSE with n was much slower. The ratio for $T = 240$ of RMSE of $n = 3,000$

to $n = 100$ is 25% for ϕ_M, rather than 17%, which it was for the bias-corrected estimator; 48% rather than 23% for ϕ_H, and the coefficient of the third least strong factor the RMSE fell then increased with n.

4.2.3 Semi-strong versus weak factors

So far, we have assumed that the DGP is correctly specified with two semi-strong and no observed weak factors. Here we consider the implications of incorrectly excluding semi-strong factors or correctly including weak factors on the small sample properties of the BC estimator of ϕ_M the coefficient of the strong factor. Using the same DGP (which includes one strong factor and two semi-strong factors), we carried out additional MC experiments (designs 1-12) where we also estimated ϕ_M without the semi-strong factors being included in the regression. Comparative results with and without the semi-strong factors. The results are summarized in Tables S-C-E1 to S-C-E12 of the online supplement B. We find that incorrectly excluding semi-strong factors can be quite costly, both in terms of bias and RMSE as well as size distortions. In terms of RMSE it was almost always better to estimate the model with the semi strong factors included. The exception was for the case of $T = 60$, $n = 100$, where including the semi-strong factors with a very small sample size caused the RMSE to blow up. Size distortions resulting from the exclusion of the semi-strong factors tended to be more pronounced for large n and T samples. These conclusions were not sensitive to the choice of the experimental design. For these experiments the lesson seems to be that it is important to have n large and include relevant semi-strong factors providing that they are sufficiently strong.

When the DGP includes one strong factor ($\alpha_M = 1$) and two weak factors ($\alpha_H = \alpha_S = 0.5$), in terms of the bias and RMSE for ϕ_M it is unambiguously better to exclude the weak factors from the regression, even though they are in the DGP. Weak factors are best treated as missing and absorbed in the error term.

4.3 Main conclusions from MC experiments

The conclusions from the Monte Carlo simulations are that the bias corrected estimator of $\boldsymbol{\phi}_0$, generally works well. Although, it can generate a large RMSE for small n and T, this can be solved by increasing n. This performance is robust to non-Gaussian errors, GARCH effects, missing weak factors, pricing errors and cross-sectional dependence of the types considered in these experiments. The size is generally correct and the power good. The rate at which RMSEs decline with n depends on the strength of the underlying factors. Semi-strong factors need much larger values of n for precise estimation. Tests of the joint significance of $\phi_M = \phi_H = \phi_S = 0$, not reported here, also performed well, as might be expected given the good power performance of the separate induced tests which are reported. Including weak factors could be harmful, but there are potential advantages of adding semi-strong factors, although issue of how best to select such factors is an open question to which we now turn.

5 Factor selection when the number of securities and the number of factors are both large

Our theoretical derivations and Monte Carlo simulations both assume that the number of risk factors included in the return regressions is fixed and the factors are known. For the empirical application we face the additional challenge of selecting a small number of relevant risk factors from a possibly large number of potential factors, m. This problem has been the subject of a number of recent studies. Harvey, Liu, and Zhu (2016) propose a multiple testing approach aimed at controlling the false discovery rate in the process of factor selection, and in a more recent paper Harvey and Liu (2020) suggest using a double-bootstrap method to calibrate the t-statistic used for controlling the desired level of the false discovery rate. Giglio and Xiu (2021) suggest applying the double-selection Lasso procedure by Belloni, Chernozhukov, and Hansen (2014) to second pass regressions. None of these methods distinguish between strong, semi-strong or weak factors in their selection process, whereas the theory and simulations presented above indicate the importance of factor strength for estimation and inference.

Here we propose an alternative selection procedure where we first estimate the strength of all the m factors under consideration, and then select factors with strength above a given threshold, the value of which is informed by the convergence results of Theorem 4. This theorem showed that if factor f_{kt} has strength α_k, then for a given T the BC estimator of ϕ_k converges to its true value, ϕ_{0k}, at the rate of $n^{(\alpha_k+\alpha_{\min}-1)/2}$, where $\alpha_{\min} = \min_i(\alpha_i)$. As is recognized in non-parametric estimation literature, if the rate of convergence is less than 1/3, the gain in precision with n is so slow that the estimator may not be that useful.[10] To achieve rate of $n^{1/3}$ we need to set the threshold value of α_k, denoted by $\underline{\alpha}$, such that $\underline{\alpha} + \alpha_{\min} > 1 + 2/3$. The smallest value of such a threshold is obtained when $\alpha_{\min} = \underline{\alpha}$ or if $\underline{\alpha} > 1/2 + 1/3$. Given the threshold the main issue is how to estimate factor strength. This problem is already addressed in Bailey, Kapetanios, and Pesaran (2021, BKP) when m is fixed. In this setting they base their estimation on the statistical significance of f_{kt} in the first stage time-series regressions of excess returns on all the factors under consideration, whilst allowing for the n multiple testing problem which their approach entails. When m (the number of factors) is also large the first stage regressions will also be subject to the multiple testing problem and penalized regression techniques such as Lasso or the one covariate at a time (OCMT) selection procedure technique proposed by

[10] The Manski (1985) maximum score estimator for a binary response model has $n^{1/3}$ convergence and this is regarded as very slow and there are suggested modifications such as Horowitz (1992) to increase the rate of convergence to $n^{2/5}$.

Chudik, Kapetanios, and Pesaran (2018) could be used. Irrespective of selection technique used at the level of individual security returns, we end up with n different subsets of the m factors under consideration. Factor strengths can then be estimated similarly to BKP from their selection frequencies across the n securities.

To be more specific, denote the set of m factors under consideration by $\mathcal{S}$ and denote the set of selected factors for security i by $\hat{S}_i$ and their numbers by $\hat{m}_i = |\hat{S}_i|$. Clearly $\hat{S}_i \subseteq \mathcal{S}$, and $\hat{m}_i \leq m$, for $i = 1, 2, ..., n$. Then compute the proportion of stocks in which the k^{th} factor is selected, $\hat{\pi}_k$, for $k \in \{1, 2, ..., m\}$ based on $\hat{S}_i$, $i = 1, 2, ..., n$ by $\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}\{k \in \hat{S}_i\}$. Then the strength of the k^{th} factor is measured by

$$\hat{\alpha}_k = \begin{cases} 1 + \frac{\ln \hat{\pi}_k}{\ln n}, & \text{if } \hat{\pi}_k > 0, \\ 0, & \text{if } \hat{\pi}_k = 0. \end{cases} \tag{85}$$

The transformation from $\hat{\pi}_k$ to $\hat{\alpha}_k$ is explained and justified in BKP, where it is shown that considering strength aids interpretation because it is not dependent on n. It is beyond the scope of the present paper to provide theoretical justification for the proposed factor selection procedure, but using extensive Monte Carlo experiments Yoo (2022) has shown that the proposed method has desirable small sample properties whether Lasso or OCMT is used for factor selection at the level of individual security returns.

6 An empirical application using a large number of U.S. securities and a large number of risk factors

This section uses the results above in the explanation of monthly returns for a large number, n, of U.S. securities, by a large active set of m potential risk factors. We first briefly describe the sources and characteristics of the data for the stock returns and factors, which cover different sub-samples over the period $1996m1 - 2022m12$. We then consider the selection of a subset of K factors from the active set. Finally we test $\boldsymbol{\phi}_0 = 0$, and construct and evaluate phi-portfolios and corresponding mean-variance portfolios for alternative models.

Monthly returns (inclusive of dividends) for NYSE and NASDAQ stocks from CRSP with codes 10 and 11 were downloaded from Wharton Research Data Services. They were converted to excess returns by subtracting the risk free rate, which was taken from Kenneth French's data base. To obtain balanced panels of stock returns and factors, only variables for which there was data for the full sample under consideration were used. Excess returns are measured in percent per month. To avoid outliers influencing the results, stocks with a kurtosis greater than 16 were excluded. To examine factor selection, four samples were considered, each had 20 years of data, $T = 240$, ending in $2015m12$, $2017m12$, $2019m12$, $2021m12$. Filtering out the stocks with kurtosis larger than 16 removed about 100 of the roughly 1200 stocks. The number of stocks (n) considered for each of the four $T = 240$ samples are given in panel A of Table 1. Further detail is given in Section E of the online supplement A.[11] For analysis of the phi-portfolios, the factors selected in the sample ending in $2015m12$ were used to construct portfolios up to $2022m12$.

6.1 Factor selection

For factor selection we used a sample of $T = 240$ observations[12]. The set of factors considered combine the 5 Fama-French factors with the 207 factors from the Chen and Zimmermann (2022), Open Source Asset Pricing webpages, both downloaded July 6 2022. Only factors with data for the full sample were considered so the return regressions constitute a balanced panel. The number of factors in each of the four 20 year samples ending in the years 2015, 2017, 2019, 2021 is also given in panel A of Table 1, and range between 187 to 199. Summary statistics for the factors in the active set are given in E.2 of the online supplement A.

To implement the factor selection procedure set out in Section 5, Lasso is used to carry out selection in the return regressions for individual securities and we refer to the factor selection procedure as pooled Lasso (PL). As is well known, Lasso does not work well with too many highly correlated regressors, therefore, factors with an absolute correlation with the market factor greater than 0.70 were dropped. This still left between 177 and 190 risk factors in the active set $\mathcal{S}$, depending on the sample period (see panel A of Table 1).

Specifically, Lasso was applied n times to the regressions of excess returns, $r_{it} = R_{it} - r_t^f$, for $i = 1, 2, ..., n$, on the 177 to 190 factors in the active set, $\mathcal{S}$, to select the sub-set $\hat{S}_i$ for each i over the four 20-year samples, separately. Following the literature, the tuning parameters in the Lasso algorithm were set by ten-fold cross-validation.[13] Interestingly, the market factor was selected by Lasso for almost all the securities, thus confirming the pervasive nature of the market factor. No other factor came close to being selected for all the securities. Lasso tended to choose a lot of non-market factors and every factor got chosen in at least one return regression. The mean number of non-market factors chosen by Lasso fell from 11.6 in the 2015 sample to 9.9 in the 2021 sample. The median was lower, falling from 10 to 8. There was a long right tail because Lasso tended to choose a very large number of non-market factors for some securities, ranging from a maximum of 48 in the 2021 sample to 54 in the 2017 sample.

Apart from the market factor, there are no systematic patterns for the rest of selected

[11] Summary statistics for the excess returns across the different samples are given in Table 7 of the online supplement A.

[12] Some results for $T = 120$ are included in the online supplement.

[13] The post-Lasso and one covariate multiple testing (OCMT) approach of Chudik, Kapetanios, and Pesaran (2018, CKP). were also investigated, but Lasso seemed to work reasonably well. The details of the Lasso procedure used are given in Section 2.2 of the online supplement of CKP paper.

factors in the return equations for the individual securities. Following the theory set out in Section 5, the K factors used to estimate $\boldsymbol{\phi}$ are chosen on the basis of their factor strength.[14] A minimum threshold of 0.7 was used. This is below the threshold value of $\underline{\alpha} = 1/2 + 1/3$ required to achieve the convergence rate of $n^{1/3}$, and is intended to capture borderline semi-strong factors. We also consider the values of 0.75,and 0.80 that are quite close to $\underline{\alpha}$. The number of selected factors for different choices of factor strength threshold is given in panel B of Table 1, for the four different samples.

Using the threshold of 0.70, 17 factors (inclusive of the market factor) were selected for the sample ending in 2015, with the number of selected factors declining to 15, 13 and 11, for the samples ending 2017, 2019 and 2021, respectively. At the other extreme, setting the threshold at 0.80, the number of selected factors dropped to 4 for the samples ending in 2015, 2017 and 2021, and 2 for the sample ending in 2021. Since 17 factors seemed too large and 2 factors too small, the threshold value was set at the intermediate value of 0.75. We considered always conditioning on the market factor, but since Lasso almost always selected it, this was unnecessary.

Table 1: Summary statistics for the number of stocks and number of selected factors using the factor strength threshold of 0.75, for four twenty years ($T = 240$) samples ending in 2021, 2019, 2017 and 2015

$T = 240$ with end dates	2021	2019	2017	2015
Panel A: Number of stocks and factors under consideration				
Number of stocks	1289	1276	1243	1181
Number of stocks with kurtosis < 16	1175	1143	1132	1090
Number of non-market factors	187	198	199	197
Number of non-market factors with $r < 0.70$	177	189	190	189
Panel B: Number of selected factors by strength threshold				
Number with strength > 0.80	2	4	4	4
Number with strength > 0.75	4	6	7	7
Number with strength > 0.70	11	13	15	17

Note: Panel A shows the number of stocks and risk factors used before and after filtering by the specified criterion. Panel B shows the number of risk factors selected with strength greater than the specified threshold level using Lasso to select factors at the level of individual securities.

The list of factors selected by pooled Lasso for the four samples are given in Table 2. The three Fama-French factors, Market, HML and SMB, are all selected in all four periods.[15] Of the Fama-French three, only the market factor is strong, with estimated

[14] The idea of using factor strength could also be viewed as a kind of averaging of the factors selected in individual return regressions.

[15] We also considered selecting the risk factors using the generalized one covariate at a time (OCMT) method proposed by Sharifvaghefi (2023). Using GOCMT the Fama-French three factors were again amongst the five strongest factors selected. The use of GOCMT for factor selection is also investigated

strength in excess of 0.98 across the four periods.[16] The other factor which is selected across all the four periods is "short selling". This is proposed by Dechow, Hutton, Meulbroek, and Sloan (2001) who argue that short-sellers target firms that are priced high relative to fundamentals. It measures the extent to which investors are shorting the market as reflected in Compustat data. Two additional factors are selected in periods ending in 2019 and earlier. One is "Beta Tail Risk" proposed by Kelly and Jiang (2014) which estimates a time-varying tail exponent from the cross section of returns. The other is "Cash Based Operating Profitability" (CBOP) suggested by Ball, Gerakos, Linnainmaa, and Nikolaev (2016). This is operating profit less accruals, with working capital and R&D adjustments. For periods ending in 2017 and 2015 the "Sin Stock" indicator proposed by Hong and Kacperczyk (2009) is also selected. It takes the value of unity if the stock in question is involved in producing alcohol, tobacco, and gaming. They find that such stocks are held less by norm-constrained institutions such as pension plans.

The factor strengths are relatively stable across the periods, with many of the estimates close to the threshold value of 0.75. Apart from the market factor only SMB, Short Selling and Beta Tail Risk factors have strengths in excess of 0.85 when averaged across the four periods. From the large number of factors in the active set we have ended up with relatively few factors that are reasonably strong and for which $\boldsymbol{\phi}_0$ can be estimated reasonably accurately.[17]

Table 2: Selected factors with estimated strength in excess of the threshold 0.75 for the samples of size T=240 ending in 2021, 2019, 2017 and 2015

End date	2021	2019	2017	2015
Selected Factors	Estimated strength (α)			
Mkt.	0.99	0.98	0.98	0.98
SMB	0.90	0.84	0.86	0.86
Short Selling	0.77	0.85	0.85	0.83
HML	0.76	0.77	0.76	0.75
BetaTailRisk		0.87	0.86	0.86
CBOP		0.76	0.77	0.77
Sin Stock		.	0.76	0.76

Note: The risk factors listed are the market factor (Mkt.), size (SMB), Short Selling that measures the extent of short sales in the market, the value factor (HML), the cash-based operating profitability factor (CBOP), Beta Tail Risk, and Sin Stock which is a binary indicator taking the value of unity if the stock in question is involved in so called "Sin" industries producing alcohol, tobacco, and gaming. Further details on these risk factors are provided by Chen and Zimmermann (2022).

by Yoo (2022), using Monte Carlo and empirical applications.

[16] These results also support the choice of 3 Fama-French factors and their strength used in our Monte Carlo simulations.

[17] We do not report estimates of individual ϕ_k. Because of correlations between the loadings, the sign, size and significance of the coefficients are difficult to interpret and for phi-portfolio construction, discussed below, what matters is $\boldsymbol{\phi}'\boldsymbol{\phi}$ which determines the return on the portfolio.

The strengths of the selected factors are also closely related to the average measures of fit often used in the literature. Here we consider both $Ave\bar{R}^2 = n^{-1}\sum_{i=1}^{n}\bar{R}_i^2$, a simple average of the fit of the individual return regressions adjusted for degrees of freedom, $\bar{R}_i^2 = 1-(T-K-1)^{-1}\sum_{t=1}^{T}\hat{u}_{it}^2/T^{-1}\sum_{t=1}^{T}(r_{it}-\bar{r}_{i\circ})^2$, and the adjusted pooled R^2 defined by $\overline{PR}^2 = 1-\widehat{\bar{\sigma}}_{nT}^2/s_{r,nT}^2$, where $\widehat{\bar{\sigma}}_{nT}^2$ is the bias-corrected estimator of $\bar{\sigma}_n^2$ defined by (59) and $= (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(r_{it}-\bar{r}_{i\circ})^2$. Both of these measures behave very similarly, but the pooled version is less sensitive to outliers. As shown in Appendix F, for sufficiently large n and T, $\overline{PR}^2$ is dominated by the contribution of the most strong factor(s). Since the only strong factor selected is the market factor in Table 3 we report the $Ave\bar{R}^2$ and $\overline{PR}^2$ in the case of return regressions which just include the market factor and those which include all other factors with strength in excess of 0.75. First, we note that the $\overline{PR}^2$ values are generally lower than the $Ave\bar{R}^2$. Second, the additional factors do add to the fit, but their relative contributions vary considerably across sample sizes and periods. In general, the marginal contribution of non-market factors tend to be smaller when T is larger, which is consistent with the theory for adjusted pooled R^2 set out in Section F of the online supplement A.

Table 3: Average and pooled R squared for the return regressions when using market factor alone or factors chosen by Pooled Lasso plus 0.75 threshold

End Year		2021		2019		2017		2015	
No. of stocks, n		1175		1143		1132		1090	
No. of selected factors		4		6		7		7	
		Mkt.	Selected	Mkt.	Selected	Mkt.	selected	Mkt.	selected
	T								
$Ave\bar{R}^2$	240	0.22	0.28	0.18	0.26	0.16	0.25	0.16	0.24
	120	0.23	0.30	0.24	0.31	0.25	0.33	0.25	0.32
$\overline{PR}^2$	240	0.20	0.24	0.17	0.24	0.15	0.23	0.15	0.23
	120	0.18	0.24	0.19	0.24	0.22	0.29	0.22	0.29

Note: This table shows for each of the four end years and the two sample sizes the adjusted average and pooled $R^2(Ave\bar{R}^2$ and $\overline{PR}^2)$ for the return regressions when using market factor alone or factors selected with strength higher than 0.75. The list of selected factors are given in Table 2.

6.2 Testing for non-zero ϕ

In principle, if $\phi \neq 0$ there are potentially exploitable excess returns. In practice, to construct an effective phi-portfolio a large number of securities is required and rebalancing such long-short portfolios for so many securities may not be feasible or may incur high transactions costs. In addition, model uncertainty, estimation uncertainty, time variation in both $\boldsymbol{\beta}_i$ and in conditional volatility pose additional difficulties in implementing a strategy to exploit the potential returns revealed by $\boldsymbol{\phi}$. We will abstract from such practical difficulties to provide some indication of the performance of phi-portfolios relative to alternatives which would face similar difficulties.

As our preferred asset pricing model, we consider the seven factors selected by pooled Lasso using the sample ending in $2015m12$, which we label as PL7.[18] Recall that the PL7 includes the 3 Fama-French factors (Mkt., SMB, and HML) plus the four risk factors, Short Selling, CBOP, Beta Tail Risk, and Sin Stocks. But given uncertainties that surround the problem of model selection we also considered the two popular FF factor models, namely FF3 and FF5. The latter augments FF3 with RMW (robust minus weak operating profitability) and CMA (conservative minus aggressive investment portfolios). The three models are estimated using twenty-year rolling windows covering the 84 months from $2015m12$ to $2022m11$, so that we can generate out of sample return forecasts for the months $2016m1 - 2022m12$.[19] For all 3×84 model-sample interactions we computed the following Wald test statistics

$$W^2_{t|T} = \tilde{\boldsymbol{\phi}}'_{t|T} \left[\widehat{Var\left(\tilde{\boldsymbol{\phi}}_{t|T}\right)} \right]^{-1} \tilde{\boldsymbol{\phi}}_{t|T},$$

for testing the null hypothesis $\boldsymbol{\phi} = 0$, where $\tilde{\boldsymbol{\phi}}_{t|T}$ and $\widehat{Var\left(\tilde{\boldsymbol{\phi}}_{t|T}\right)}$ denote the rolling versions of (32) and (36).[20] For the PL7 the range of the test statistic was from 141.8 to 25.8, as compared to the 5 per cent $\chi^2(7)$ critical value of 14.07. Thus the hypothesis that $\boldsymbol{\phi} = 0$ is strongly rejected in all the 84 rolling sample for PL7. This is also true for the FF5 and FF3 models where the test statistic ranged from 162.7 to 25.0, and from 67.9 to 25.8, compared to the 5 per cent critical values of 11.07 and 7.8, respectively.

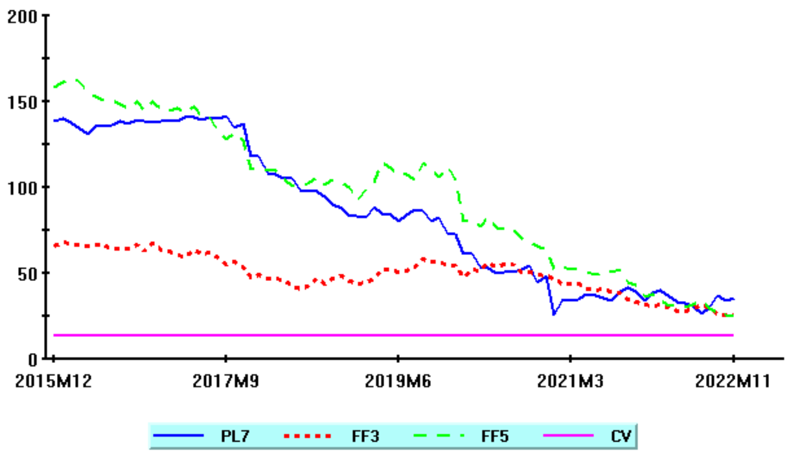
The rolling values of the Wald statistics for testing $\boldsymbol{\phi} = 0$ for the three models are shown in Figure 1. The horizontal line (pink) represents the critical value of the χ^2_7 distribution at the 5 per cent level. The time profiles of these test statistics clearly show that $\boldsymbol{\phi} = 0$ is rejected for all rolling samples and for all three models. But there is also a clear downward trend showing that the evidence against $\boldsymbol{\phi} = 0$ has been getting weaker over time, irrespective of the choice of asset pricing model.

[18]The selection of the PL7 model was reported in the earlier version of the paper submitted for publication and was not informed by the performance the phi-portfolio that we report in this version of the paper.

[19]Due to entry and exit of securities the number of securities included in our analysis varied across the rolling sample periods. We started with $n = 1,090$ securities for the first rolling sample ending in $2015m12$, with the number of available securities with 240 months of data falling to 953 by 2017, 838 by 2019, 767 by 2021, and 736 by 2022.

[20]The formulae for the rolling estimates are provided in the sub-section C.8 of the online appendix.

Figure 1: Rolling chi-squared statistics for testing $\boldsymbol{\phi} = 0$ using a window of size 240 for FF3, FF5, and PL7 models.



6.3 Comparative performance of phi and MV portfolios

Having established that most likely $\boldsymbol{\phi} \neq \mathbf{0}$ for the asset pricing models we have considered, we now turn to the performance of phi-portfolios based on these models. Using the recursive version of the phi-portfolio given by (14), we consider the following phi-portfolio returns

$$\hat{\rho}_{t+1,\phi} = \tilde{\boldsymbol{\phi}}_{t|T} \left[\left(\hat{\mathbf{B}}'_{t|T} \mathbf{M}_n \hat{\mathbf{B}}_{t|T} \right)^{-1} \hat{\mathbf{B}}_{t|T} \mathbf{M}_n \mathbf{r}_{\circ,t+1} - \mathbf{f}_t \right],$$

for $t = 2016m1, 2016m2, ..., 2022m12$, using the rolling estimates $\hat{\mathbf{B}}_{t|T} = \left(\hat{\beta}_{1t|T}, \hat{\beta}_{2t|T} ...., \hat{\beta}_{nt|T} \right)'$ with $T = 240$, for each of the three factor models, FF3, FF5 and PL7. We compare the annualized Sharpe ratios of phi-portfolios with the ones based on associated MV portfolios, given by $\rho_{t+1,MV} = \boldsymbol{\mu}'_R \mathbf{V}_R^{-1} \mathbf{r}_{\circ,t+1}$.[21] Although in principle, MV portfolios can be constructed without a reference to a particular factor model, reliable estimation of $\boldsymbol{\mu}_R$ and $\mathbf{V}_R^{-1}$ are challenging when n is relatively large. For example, the rolling sample covariance matrix estimator of $\mathbf{V}_R$, given by $\mathring{\mathbf{V}}_{R,t|T} = T^{-1} \sum_{\tau=t-T+1}^{t} \left(\mathbf{r}_{\circ,\tau} - \bar{\mathbf{r}}_{\circ,t|T} \right) \left(\mathbf{r}_{\circ,\tau} - \bar{\mathbf{r}}_{\circ,t|T} \right)'$, with $\bar{\mathbf{r}}_{\circ,t|T} = T^{-1} \sum_{\tau=t-T+1}^{t} \mathbf{r}_{\circ,\tau}$, will be singular when $n > T$, and can be very poorly estimated if T is not sufficiently large relative to n. There is a vast literature on consistent estimation of high dimensional covariance matrices like $\mathbf{V}_R$. Fan, Liao, and Mincheva (2011) use observed factors while Fan, Liao, and Mincheva (2013) use principal components to filter out the effects of strong factors, in both cases assuming $\mathbf{V}_u$ is sparse, and then using a threshold method to estimate it. Shrinkage estimators of $\mathbf{V}_R$ are also proposed in the literature with a recent survey provided by Ledoit and Wolf (2022). However, the shrinkage estimators require n and T to be of the same order of magnitude and do

[21] Given our focus on the Sharpe ratios, we have set the scaling of the MV portfolio to unity.

not work well when n is much larger than T, as in the present application. We follow Fan, Liao, and Mincheva (2011) and base our estimation of $\boldsymbol{\mu}_R$ and $\mathbf{V}_R$ on the same factor model used to construct the phi-portfolio. For a given factor model, characterized by $c, \mathbf{B}$, and $\mathbf{F}$, we compute the MV portfolio returns as

$$\hat{\rho}_{t+1.MV} = \hat{\mu}'_{R,t|T} \hat{\mathbf{V}}^{-1}_{R,t|T} \mathbf{r}_{\circ,t+1},$$

where $\hat{\mu}_{R,t|T} = \hat{c}_{t|T} \tau_n + \hat{\mathbf{B}}_{t|T} \tilde{\boldsymbol{\lambda}}_{t|T}$, $\tilde{\boldsymbol{\lambda}}_{t|T} = \tilde{\boldsymbol{\phi}}_{t|T} + \hat{\boldsymbol{\mu}}_{t|T}$,

$$\hat{\mathbf{V}}_{R,t|T} = \hat{\mathbf{B}}'_{t|T} \left(\frac{\mathbf{F}'_{t|T} \mathbf{M}_T \mathbf{F}_{t|T}}{T} \right)^{-1} \hat{\mathbf{B}}_{t|T} + \tilde{\mathbf{V}}_{u,t|T},$$

$\hat{\boldsymbol{\mu}}_{t|T} = \bar{\boldsymbol{f}}_{t|T} = T^{-1} \sum_{\tau=t-T+1}^{t} \mathbf{f}_\tau$, and $\tilde{\mathbf{V}}_{u,t|T}$ is the rolling estimate of $\mathbf{V}_u$. The algorithms used to compute the recursive estimates for the MV portfolio can be found in the subsection C.8 in the online supplement A.

Table 4 presents annualised SR of the phi-portfolios, for the FF3, FF5, and PL7 models, and their corresponding MV portfolios for two samples, both beginning in 2016m1, one ending in 2019m12, pre Covid-19, and one ending in 2022m12. In five of the six SR ratios reported in this table, the phi-portfolio has a higher SR than the corresponding MV portfolio. The exception is the SR associated to the FF5 model for the pre Covid-19 sample. This illustrates that if $\boldsymbol{\phi} \neq \mathbf{0}$, it is possible to construct portfolios that outperform the mean-variance portfolio. Amongst the 3 models considered, the phi-portfolio based on the PL7 model performed best during the pre Covid-19 and the full sample, even beating the S&P 500. The Sharpe ratio of the phi-portfolio based on the Pl7 model was 1.95 compared to 0.94 for the S&P 500 during the pre Covid-19 period, and fell sharply to 0.65 for the full sample as compared to 0.58 for the S&P 500. The SR for the MV portfolio using the same model were 0.87 and 0.42 for the two samples, respectively.

The sharp decline in the SRs as we add the post Covid-19 years is in line with the strong downward trend in the Wald statistics for the test of $\boldsymbol{\phi} = \mathbf{0}$ shown in Figure 1. As is well known SRs have large standard errors, and in the case of our application that are around 1, so none of the SRs are significantly different from zero, with the possible except of the largest SR of 1.95 for phi-portfolio based on PL7 model for the pre Covid-19 period. We also note that the reported Sharpe ratios do not allow for transaction costs and the fact that shorting might not be feasible for all the securities.

Table 4: Annualised Sharpe ratios of realized monthly returns for alternative portfolios based on 240 month rolling window estimates.

	Mean-variance portfolios		Phi-portfolios	
	$2016m1$ - $2019m12$	$2016m1$ - $2022m12$	$2016m1$-$2019m12$	$2016m1$-$2022m12$
Models	Pre Covid-19	Full sample	Pre Covid-19	Full sample
FF3	0.57	0.35	0.86	0.51
FF5	0.49	0.43	0.35	0.57
Lasso 7	0.87	0.42	1.95	0.65

Note: The annualized Sharpe ratio (SR) is computed as $\sqrt{12}\bar{\rho}/s_{\rho}$, where $\bar{\rho}$ is the mean of monthly returns, and s_{ρ} is the standard deviation of monthly returns. For comparison the SR of the monthly returns on S&P 500 were 0.94 and 0.58 over the periods 2016m1-2019m12, and 2016m1-2022m12, respectively.

7 Concluding remarks

In this paper we have highlighted the importance of decomposing the risk premia, $\boldsymbol{\lambda}$, into the the factor mean, $\boldsymbol{\mu}$, and $\boldsymbol{\phi}$, and writing the alpha of security i, α_i, in terms of $\boldsymbol{\phi}$ and the idiosyncratic pricing errors. We have shown that when $\boldsymbol{\phi} \neq \mathbf{0}$, it is possible to construct a portfolio, denoted as phi-portfolio, that dominates the associated mean-variance portfolio for the number securities, n, is sufficiently large and the risk factors are sufficiently strong. Given the pivotal role played by $\boldsymbol{\phi}$ both for estimating the risk premia and for formation of large portfolios, and tests of market efficiency, we have exclusively focussed on estimation of $\boldsymbol{\phi}$, and its asymptotic distribution under quite a general setting that allows for missing factors and idiosyncratic pricing errors. Since factor means, $\boldsymbol{\mu}$, can be estimated at the regular rate of $T^{-1/2}$ from time series data, it is relatively straightforward to develop a mixed strategy for estimation of $\boldsymbol{\lambda}$ by adding a time series estimate of $\boldsymbol{\mu}$ to the bias-corrected estimator of $\boldsymbol{\phi}$. If we use the same time series sample, such an estimator reduces to the Shanken bias-corrected estimator of $\boldsymbol{\lambda}$. But in practice, given the concern over the instability of the factor loadings β_{ik} over time, one could use relatively long time series, say T_μ, when estimating $\boldsymbol{\mu}$, and a shorter time series, say $T_\phi < T_\mu$, when estimating $\boldsymbol{\phi}$. The distributional and small sample properties of such a mixed estimator of risk premia is a topic for further research.

Our theoretical and Monte Carlo results further highlight the important role played by factor strengths in estimation and inference on $\boldsymbol{\phi}$, and hence on $\boldsymbol{\lambda}$. For a fixed T_ϕ, factors with strength below 2/3 lead to estimates of $\boldsymbol{\phi}$ with convergence rate of $n^{-1/3}$ or worse, and their use in asset pricing models can be justified only when n is very large. We have also shown that weak factors, with strength below 1/2 are best treated as missing and absorbed in the error term. We have shown that estimation of $\boldsymbol{\phi}$ for strong or semi-strong factors is robust to weak missing factors, and the explicit inclusion of weak factors in the empirical analysis is likely to have adverse spill over effects on the estimates of $\boldsymbol{\phi}$

for strong and semi-strong factors. In view of these results we have proposed a factor selection procedure where only factors with strength above $1/2 + 1/3$ are included in the asset pricing model. Developing a formal statistical theory for the proposed selection is another topic for future research.

The paper also provides an empirical application to a large number of U.S. securities with risk factors selected from a large number of potential risk factors according to their strength, and use a pooled Lasso approach to select 7 risk factors out of over 180. We find strong statistical evidence against $\boldsymbol{\phi} = \mathbf{0}$ for the selected model as well as for the two popular Fama-French models (FF3 and FF5). Using rolling estimates of $\boldsymbol{\phi}$ we also construct phi portfolios with better Sharpe ratios as compared to associated mean-variance portfolios. But we also warn that these portfolio comparisons are preliminary and need to be further investigated by allowing for transaction costs, and the feasibility of the long-short trading strategies that are involved.

References

ANATOLYEV, S., AND A. MIKUSHEVA (2022): "Factor models with many assets: strong factors, weak factors, and the two-pass procedure," *Journal of Econometrics*, 229, 103–126.

ANG, A., J. LIU, AND K. SCHWARZ (2020): "Using Stocks or Portfolios in Tests of Factor Models," *Journal of Financial and Quantitative Analysis*, 55(3), 709 750.

BAI, J., AND S. NG (2023): "Approximate factor models with weaker loadings," *Journal of Econometrics*, 235, 1893–1916.

BAI, J., AND G. ZHOU (2015): "Fama-MacBeth two-pass regressions: improving risk premia estimates," *Finance Research Letters*, 15, 31–40.

BAILEY, N., G. KAPETANIOS, AND M. H. PESARAN (2016): "Exponent of cross-sectional dependence: Estimation and inference," *Journal of Applied Econometrics*, 31, 929–960.

——— (2021): "Measurement of factor strength: theory and practice," *Journal of Applied Econometrics*, 36, 587–613.

BAILEY, N., M. H. PESARAN, AND L. V. SMITH (2019): "A multiple testing approach to the regularisation of large sample correlation matrices," *Journal of Econometrics*, 208, 507–534.

BALL, R., J. GERAKOS, J. T. LINNAINMAA, AND V. NIKOLAEV (2016): "Accruals, cash flows, and operating profitability in the cross section of stock returns," *Journal of Financial Economics*, 121, 28–45.

BEAULIEU, M.-C., J.-M. DUFOUR, AND L. KHALAF (2020): "Arbitrage pricing, weak beta, strong beta: identification-robust and simultaneous inference," *Cireq Working paper 15-2020.*

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81, 608–650.

BICKEL, P. J., AND E. LEVINA (2008a): "Covariance regularization by thresholding," *The Annals of Statistics*, 36, 2577–2604.

——— (2008b): "Regularized estimation of large covariance matrices," *The Annals of Statistics*, 36, 199–227.

CAI, T., AND W. LIU (2011): "Adaptive thresholding for sparse covariance matrix estimation," *Journal of the American Statistical Association*, 106, 672–684.

CHAMBERLAIN, G. (1983): "Funds, factors, and diversification in arbitrage pricing models," *Econometrica*, 51, 1305–1323.

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

CHEN, A. Y., AND T. ZIMMERMANN (2022): "Open source cross-sectional asset pricing," *Critical Finance Review*, 27, 207–264.

CHIB, S., X. ZENG, AND L. ZHAO (2020): "On comparing asset pricing models," *The Journal of Finance*, 75, 551–577.

CHUDIK, A., G. KAPETANIOS, AND M. H. PESARAN (2018): "A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models," *Econometrica*, 86, 1479–1512.

CHUDIK, A., M. H. PESARAN, AND E. TOSETTI (2011): "Weak and strong cross-section dependence and estimation of large panels," *The Econometrics Journal*, 14, C45–C90.

CONNOR, G., AND R. A. KORAJCZYK (2022): "Semi-strong factors in asset returns," *Journal of Financial Econometrics*, 22(1).

DANIEL, K., L. MOTA, S. ROTTKE, AND T. SANTOS (2020): "The Cross-Section of Risk and Returns," *The Review of Financial Studies*, 33(5), 1927–1979.

DECHOW, P. M., A. P. HUTTON, L. MEULBROEK, AND R. G. SLOAN (2001): "Short-sellers, fundamental analysis, and stock returns," *Journal of Financial Economics*, 61, 77–106.

DELLO PREITE, M., R. UPPAL, P. ZAFFARONI, AND I. ZVIADADZE (2022): "What is Missing in Asset-Pricing Factor Models?," *Available at SSRN 4135146.*

——— (2024): "Cross-Sectional Asset Pricing with Unsystematic Risk," *SSRN:*

https://ssrn.com/abstract=4135146.

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.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

——— (2018): "Choosing factors," *Journal of Financial Economics*, 128, 234–252.

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 M. Mincheva (2011): "High dimensional covariance matrix estimation in approximate factor models," *Annals of Statistics*, 39, 3320–3356.

——— (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 603–680.

Feng, G., S. Giglio, and D. Xiu (2020): "Taming the factor zoo: a test of new factors," *The Journal of Finance*, 75, 1327–1370.

Feng, L., W. Lan, B. Liu, and Y. Ma (2022): "High-dimensional test for alpha in linear factor pricing models with sparse alternatives," *Journal of Econometrics*, 229, 152–175.

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

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A test of the efficiency of a given portfolio," *Econometrica*, 57, 1121–1152.

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 (2023): "Test assets and weak factors," *Chicago Booth Research Paper 21-04, Available at SSRN: https://ssrn.com/abstract=3768081 or http://dx.doi.org/10.2139/ssrn.3768081*.

Gofman, M., G. Segal, and Y. Wu (2020): "Production networks and stock returns: The role of vertical creative destruction," *The Review of Financial Studies*, 33, 5856–5905.

Harvey, C. R., and Y. Liu (2019): "A census of the factor zoo," *Available at SSRN 3341728*.

——— (2020): "False (and missed) discoveries in financial economics," *The Journal of Finance*, 75, 2503–2553.

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, A., S. He, D. Huang, and G. Zhou (2022): "Testing Asset Pricing Models Using Pricing Error Information," *Available at SSRN 3143752*.

Herskovic, B. (2018): "Networks in production: asset pricing implications," *The Journal of Finance*, 73, 1785–1818.

Herskovic, B., A. Moreira, and T. Muir (2019): "Hedging risk factors," *SSRN: https://ssrn.com/abstract=3148693*.

Hong, H., and M. Kacperczyk (2009): "The price of sin: the effects of social norms on markets," *Journal of Financial Economics*, 93, 15–36.

Horowitz, J. L. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, 60, 505–531.

Hwang, S., and A. Rubesam (2022): "Bayesian selection of asset pricing factors using individual stocks," *Journal of Financial Econometrics*, 20, 716–761.

Jagannathan, R., G. Skoulakis, and Z. Wang (2010): "The analysis of the cross-section of security returns," in *Handbook of Financial Econometrics: Applications*, ed. by Y. Ait-Sahalia, and L. P. Hansen, vol. 2, chap. 14, pp. 73–134. Elsevier, Amsterdam.

Jagannathan, R., and Z. Wang (2002): "Empirical Evaluation of Asset-Pricing Models: A Comparison of the SDF and Beta Methods," *The Journal of Finance*, 57(5), 2337–2367.

Jensen, M. C. (1968): "The performance of mutual funds in the period 1945-1964," *The Journal of Finance*, 23, 389–416.

Kan, R., C. Robotti, and J. Shanken (2013): "Pricing model performance and the two-pass cross-sectional regression methodology," *The Journal of Finance*, 68, 2617–2649.

Kelly, B., and H. Jiang (2014): "Tail risk and asset prices," *The Review of Financial Studies*, 27, 2841–2871.

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

Korsaye, S. A., A. Quaini, and F. Trojani (2021): "Smart Stochastic Discount Factors," *Swiss Finance Institute Research Paper No. 21-51, Available at SSRN: https://ssrn.com/abstract=3878300.*

Ledoit, O., and M. Wolf (2022): "The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation," *Journal of Financial Econometrics*, 20, 187–218.

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

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

Ma, S., W. Lan, L. Su, and C.-L. Tsai (2020): "Testing alphas in conditional time-varying factor models with high-dimensional assets," *Journal of Business & Economic Statistics*, 38, 214–227.

Manski, C. F. (1985): "Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333.

Onatski, A. (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors," *Journal of Econometrics*, 168, 244–258.

Pesaran, M. H., and R. P. Smith (2023): "Arbitrage pricing theory, the stochastic discount factor and estimation of risk premia from portfolios," *Econometrics and Statistics*, 26, 17–30.

Pesaran, M. H., and T. Yamagata (2012): "Testing CAPM with a large number of assets," *IZA Discussion Papers No. 6469.*

——— (2024): "Testing for alpha in linear factor pricing models with a large number of securities," *Journal of Financial Econometrics*, 22, 407–460.

Quaini, A., F. Trojani, and M. Yuan (2023): "Tradable Factor Risk Premia and Oracle Tests of Asset Pricing Models," *Swiss Finance Institute Research Paper No. 23-81, Available at SSRN: https://ssrn.com/abstract=4574683.*

Raponi, V., C. Robotti, and P. Zaffaroni (2019): "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.

Shanken, J., and G. Zhou (2007): "Estimating and testing beta pricing models: alternative methods and their performance in simulations," *Journal of Financial Economics*, 84, 40–86.

Sharifvaghefi, M. (2023): "Variable selection in linear regressions with many highly correlated covariates," *Available at SSRN 4159979.*

Uematsu, Y., and T. Yamagata (2023): "Inference in sparsity-induced weak factor models," *Journal of Business & Economic Statistics*, 41(1), 126–139.

Yoo, J. (2022): "Factor strengths in high-dimensional setting with applications to finance and macroeconomics," *PhD Job Market Paper*, Department of Economics, University of Southern California.

A Mathematical Appendix

A.1 Introduction

In this mathematical appendix we first introduce the notations used in our mathematical treatment, and state and establish a number of lemmas. We then provide detailed proofs of Theorems 1, 2, 3, 4 and 5 presented in the paper.

Notations: Generic positive finite constants are denoted by C when large, and c when small. They can take different values at different instances. $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the maximum and minimum eigenvalues of $\boldsymbol{A}$. $\boldsymbol{A} > 0$ denotes that $\boldsymbol{A}$ is a positive definite matrix. $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_F = [Tr(\mathbf{A}'\mathbf{A})]^{1/2}$, $\|\boldsymbol{A}\|_p = (E\|\boldsymbol{A}\|^p)^{1/p}$, for $p \geq 2$ denote spectral, Frobenius, and ℓ_p norms of matrix $\boldsymbol{A}$, respectively. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequences of positive real numbers, then $f_n = O(g_n)$, if there exists C such that $|f_n|/g_n \leq C$ for all n. $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. Similarly, $f_n = O_p(g_n)$ if f_n/g_n is stochastically bounded, and $f_n = o_p(g_n)$, if $f_n/g_n \to_p 0$, where $\to_p$denotes convergence in probability. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \ominus(g_n)$ if there exists $n_0 \geq 1$ such that $\inf_{n\geq n_0}(f_n/g_n) \geq C$, and $\sup_{n\geq n_0}(f_n/g_n) \leq C$.

A.2 Statement of lemmas and their proofs

Lemma A.1 *Suppose the linear factor pricing model (12) holds subject to the restrictions* $c = 0$, $\boldsymbol{\phi} = \mathbf{0}$, $\eta = \mathbf{0}$, $\|\mu\| < C$, *and* $\lambda_{min}\left(\boldsymbol{\Sigma}_f^{-1}\right) > 0$. *Consider the* $K \times 1$ *beta-based portfolio* $\rho_{B,t} = \mathbf{W}'\mathbf{r}_{\circ t}$ *formed from the factor loadings,* $\mathbf{B}_n$, *where* $\mathbf{W}' = \left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\mathbf{B}_n'\mathbf{V}_u^{-1}$. *Then the mean-variance portfolio return, given by* $\rho_{MV,t} = \boldsymbol{\mu}_R'\mathbf{V}_R^{-1}\mathbf{r}_{\circ t}$, *where* $\boldsymbol{\mu}_R = \mathbf{B}_n\boldsymbol{\mu}$ *and* $\mathbf{V}_R = \mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n' + \mathbf{V}_u$, *and the optimal portfolio return formed using the* K *underlying beta portfolios,* $\boldsymbol{\rho}_{B,t}$, *have the same Sharpe ratio:*

$$SR_B^2 = SR_{MV}^2 \leq \mu'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}. \tag{A.1}$$

Further if the factors are sufficiently strong such that $\lambda_{min}\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right) \to \infty$, *then*

$$SR_B^2 = SR_{MV}^2 \to \mu'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}, \text{ as } n \to \infty. \tag{A.2}$$

Proof. Under $c = 0$ and $\boldsymbol{\phi} = \mathbf{0}$, $\mathbf{r}_{\circ t} = \mathbf{B}\mathbf{f}_t + \mathbf{u}_{\circ t}$, $\rho_{B,t} = \mathbf{f}_t + \mathbf{W}'\mathbf{u}_{\circ t}$, and

$$E(\rho_{B,t}) = \boldsymbol{\mu}, \text{ and } Var(\rho_{B,t}) = \boldsymbol{\Sigma}_f + \mathbf{W}'\mathbf{V}_u\mathbf{W} = \boldsymbol{\Sigma}_f + \left(\mathbf{B}'\mathbf{V}_u^{-1}\mathbf{B}\right)^{-1}.$$

The best linear combination of these K portfolios is obtained by finding the $K\times 1$ vector weights, $\mathbf{w}_f$, that minimizes the $Var\left(\mathbf{w}_f'\rho_t\right)$ for a given mean, $\mathbf{w}_f'\boldsymbol{\mu}$. The solution to this optimization problem is given by ■

$$\mathbf{w}_f = \kappa^{-1}\left(\boldsymbol{\Sigma}_f + \left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\right)^{-1}\boldsymbol{\mu},$$

where κ is a risk aversion coefficient. The squared Sharpe ratio of $\rho_{B,t} = \mathbf{w}_f'\rho_{B,t}$ is given by

$$SR_B^2 = \boldsymbol{\mu}'\left(\boldsymbol{\Sigma}_f + \left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1}\right)^{-1}\boldsymbol{\mu}, \tag{A.3}$$

which can be written equivalently as

$$SR_B^2 = \boldsymbol{\mu}' \left[\left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right) \boldsymbol{\Sigma}_f \right]^{-1} \boldsymbol{\mu} \qquad \text{(A.4)}$$
$$= \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right) \boldsymbol{\mu}$$

The squared Sharpe ratio of the mean-variance efficient portfolio is given by

$$SR_{MV}^2 = \boldsymbol{\mu}'\mathbf{B}_n' \left(\mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n' + \mathbf{V}_u\right)^{-1} \mathbf{B}_n\boldsymbol{\mu}.$$

However, since $\mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n'$ is rank deficient then

$$\left(\mathbf{B}_n\boldsymbol{\Sigma}_f\mathbf{B}_n' + \mathbf{V}_u\right)^{-1} = \mathbf{V}_u^{-1} - \mathbf{V}_u^{-1}\mathbf{B} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1},$$

and

$$SR_{MV}^2 = \boldsymbol{\mu}'\mathbf{B}_n' \left[\mathbf{V}_u^{-1} - \mathbf{V}_u^{-1}\mathbf{B} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1} \right] \mathbf{B}_n\boldsymbol{\mu}$$
$$= \boldsymbol{\mu}'\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu}$$

Furthermore,

$$\boldsymbol{\mu}'\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu}$$
$$= \boldsymbol{\mu}' \left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n + \boldsymbol{\Sigma}_f^{-1} - \boldsymbol{\Sigma}_f^{-1}\right) \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu}$$
$$= \boldsymbol{\mu}'\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu}.$$

Hence

$$SR_{MV}^2 = \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\boldsymbol{\mu}. \qquad \text{(A.5)}$$

The first part of result (A.1) now follows from a direct comparison of (A.5) and (A.4). To establish the second part note that

$$SR_{MV}^2 - \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu} = \boldsymbol{\mu}' \left[\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n - \boldsymbol{\Sigma}_f^{-1} \right] \boldsymbol{\mu}$$
$$= \boldsymbol{\mu}' \left[\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \left(\mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n + \boldsymbol{\Sigma}_f^{-1} - \boldsymbol{\Sigma}_f^{-1}\right) - \boldsymbol{\Sigma}_f^{-1} \right] \boldsymbol{\mu}$$
$$= \boldsymbol{\mu}' \left[\boldsymbol{\Sigma}_f^{-1} \left(\mathbf{I}_k - \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \boldsymbol{\Sigma}_f^{-1} \right) - \boldsymbol{\Sigma}_f^{-1} \right] \boldsymbol{\mu}$$
$$= -\boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}.$$

Since $\boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu} \geq 0$, it then follows that $SR_{MV}^2 \leq \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}$. To establish result (A.2), we first note that

$$\left| SR_{MV}^2 - \boldsymbol{\mu}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu} \right| \leq \|\boldsymbol{\mu}\|^2 \left\| \boldsymbol{\Sigma}_f^{-1} \right\|^2 \left\| \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n'\mathbf{V}_u^{-1}\mathbf{B}_n\right)^{-1} \right\|.$$

where $\|\boldsymbol{\mu}\| < C$, and $\left\|\boldsymbol{\Sigma}_f^{-1}\right\|^2 = \lambda_{max}\left(\boldsymbol{\Sigma}_f^{-1}\right) = 1/\lambda_{min}\left(\boldsymbol{\Sigma}_f\right) < C$, since by Assumption

$\lambda_{min}\left(\boldsymbol{\Sigma}_f\right) > 0$. Also

$$\left\| \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right)^{-1} \right\| = \lambda_{\max} \left[\left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right)^{-1} \right] = \frac{1}{\lambda_{\min} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right)}.$$

Since $\boldsymbol{\Sigma}_f^{-1}$ and $\mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n$ are both symmetric matrices, then (see Section 5.3.2 in Lutkepohl, 1996)

$$\lambda_{\min} \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right) \geq \lambda_{\min} \left(\boldsymbol{\Sigma}_f^{-1} \right) + \lambda_{min} \left(\mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right),$$

and

$$\left\| \left(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right)^{-1} \right\| \leq \frac{1}{\lambda_{\min} \left(\boldsymbol{\Sigma}_f^{-1} \right) + \lambda_{min} \left(\mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right)}.$$

Hence

$$\left\| SR_{MV}^2 - \boldsymbol{\mu}' \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\mu} \right\| \leq \frac{\left\| \mu_0 \right\|^2 \left\| \boldsymbol{\Sigma}_f^{-1} \right\|^2}{\lambda_{\min} \left(\boldsymbol{\Sigma}_f^{-1} \right) + \lambda_{min} \left(\mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right)},$$

$\left\| SR_{MV}^2 - \boldsymbol{\mu}' \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\mu} \right\| \to 0$, if $\lambda_{min} \left(\mathbf{B}_n' \mathbf{V}_u^{-1} \mathbf{B}_n \right) \to \infty$, and result (A.2) follows.

Lemma A.2 *Consider the errors* $\{u_{it}, i = 1, 2, ..., n; t = 1, 2, ..., T\}$ *defined by (42) and suppose that Assumptions 3 and 4 hold with* $\alpha_\gamma < 1/2$, *and* $\mathbf{V}_u = (\sigma_{ij})$. *Set* $\overline{u}_{i\circ} = T^{-1} \sum_{t=1}^{T} u_{it}$, *and* $E(u_{it}^2) = \sigma_i^2$. *Then*

$$\left\| \mathbf{V}_u \right\| = \lambda_{\max} \left(\mathbf{V}_u \right) \leq \sup_i \sum_{j=1}^{n} \left| \sigma_{ij} \right| = O \left(n^{\alpha_\gamma} \right), \tag{A.6}$$

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \sigma_{ij} \right| = O \left(1 \right), \text{ and } n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}^2 = O(1). \tag{A.7}$$

Also, for any t *and* t'

$$a_{n,tt} = \frac{1}{n} \sum_{i=1}^{n} \left(u_{it}^2 - \sigma_i^2 \right) = O_p \left(n^{-1/2} \right), \tag{A.8}$$

$$Var(a_{n,tt}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(u_{it}^2, u_{jt}^2) = O(n^{-1}), \tag{A.9}$$

$$a_{n,tt'} = n^{-1} \sum_{i=1}^{n} u_{it} u_{it'} = O_p \left(n^{-1/2} \right), \text{ for } t \neq t', \tag{A.10}$$

$$Var \left(a_{n,tt'} \right) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}^2 = O \left(n^{-1} \right), \text{ for } t \neq t', \tag{A.11}$$

$$b_{n,t} = \frac{1}{n} \sum_{i=1}^{n} (u_{it} \overline{u}_{i\circ} - T^{-1} \sigma_i^2) = O_p \left(T^{-1/2} n^{-1/2} \right), \tag{A.12}$$

and

$$Var(b_{n,t}) = O \left(T^{-1} n^{-1} \right). \tag{A.13}$$

Proof. Result (A.6) follows noting that under Assumptions 3 and 4, $\sigma_{ij} = \gamma_i \gamma_j + \sigma_{v,ij}$, $\sum_{j=1}^{n} \left| \sigma_{ij} \right| \leq \sup_i \left| \gamma_i \right| \sum_{j=1}^{n} \left| \gamma_j \right| + \sum_{j=1}^{n} \left| \sigma_{v,ij} \right|$, and by assumption $\sup_i \left| \gamma_i \right| < C$, $sup_i \sum_{j=1}^{n} \left| \sigma_{v,ij} \right| <$

C, and $\sum_{j=1}^{n}|\gamma_j| = O\left(n^{\alpha_\gamma}\right)$. To prove (A.7)

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{ij}| \leq n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\gamma_i|\,|\gamma_j| + n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{v,ij}|$$
$$= n^{-1}\left(\sum_{i=1}^{n}|\gamma_i|\right)^2 + n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{v,ij}|.$$

By assumption $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{v,ij}| = O(1)$, and $\sum_{i=1}^{n}|\gamma_i| = O(n^{\alpha_\gamma})$, then, in view of (48) and (49) $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{ij}| = O\left(n^{-1+2\alpha_\gamma}\right) + O(1) = O(1)$, since $\alpha_\gamma < 1/2$. Similarly $\sigma_{ij}^2 = \gamma_i^2\gamma_j^2 + \sigma_{v,ij}^2 + 2\gamma_i\gamma_j\sigma_{v,ij}$, and

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2 = n^{-1}Tr(\mathbf{V}_u^2) = n^{-1}\left(\sum_{i=1}^{n}\gamma_i^2\right)^2 + n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{v,ij}^2 + 2n^{-1}\gamma'\mathbf{V}_v\gamma.$$

But

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{v,ij}^2 = n^{-1}Tr\left(\mathbf{V}_v^2\right) \leq \lambda_{\max}^2\left(\mathbf{V}_v\right) = O(1), \tag{A.14}$$
$$\boldsymbol{\gamma}'\mathbf{V}_v\boldsymbol{\gamma} \leq \left(\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)\lambda_{\max}\left(\mathbf{V}_u\right) = O\left(n^{\alpha_\gamma}\right).$$

Overall

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2 = O\left(n^{-1+2\alpha_\gamma}\right) + O(1) + O\left(n^{-1+\alpha_\gamma}\right),$$

and since $\alpha_\gamma < 1/2$, then it follows that $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2 = O(1)$. To prove (A.8) we first note that (using the normalization $E(g_t^2) = 1$)

$$u_{it}^2 - E(u_{it}^2) = (g_t^2 - 1)\gamma_i^2 + \left(v_{it}^2 - \sigma_{v,ii}\right) + 2g_t\gamma_i v_{it}, \tag{A.15}$$

and

$$a_{n,tt} = (g_t^2 - 1)\left(n^{-1}\sum_{i=1}^{n}\gamma_i^2\right) + n^{-1}\sum_{i=1}^{n}\left(v_{it}^2 - \sigma_{v,ii}\right) + 2g_t\left(n^{-1}\sum_{i=1}^{n}\gamma_i v_{it}\right)$$
$$= O\left(n^{-1+\alpha_\gamma}\right) + O_p\left(n^{-1/2}\right) + O\left(n^{-1+\alpha_\gamma/2}\right) = O_p(n^{-1/2}), \text{ since } \alpha_\gamma < 1/2.$$

To establish (A.9), using (A.15) we first note that (for given loadings γ_i)

$$Cov\left(u_{it}^2, u_{jt}^2\right) = \gamma_i^2\gamma_j^2\left[E\left(g_t^2 - 1\right)^2\right] + Cov(v_{it}^2, v_{jt}^2) + 4E\left(g_t^2\right)\gamma_i\gamma_j\sigma_{v,ij},$$

$$Var(a_{n,tt}) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(u_{it}^2, u_{jt}^2) = Var(g_t^2)\left(n^{-1}\sum_{i=1}^{n}\gamma_i^2\right)^2$$
$$+ n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(v_{it}^2, v_{jt}^2) + 4E\left(g_t^2\right)n^{-2}\gamma'\mathbf{V}_v\gamma = O\left(n^{-2+2\alpha_\gamma}\right) + O(n^{-1}) + O\left(n^{-2+\alpha_\gamma}\right),$$

and since $\alpha_\gamma < 1/2$, then $Var(a_{n,tt}) = O\left(n^{-1}\right)$, as required (which also corroborate (A.8)). Consider now (A.10) and since u_{it} is serially independent then $E\left(a_{n,tt'}\right) = 0$ for $t \neq t'$, and we have

$$E\left(u_{it}u_{it'}u_{jt}u_{jt'}\right) = E\left(u_{it}u_{jt}u_{it'}u_{jt'}\right) = E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right) = \sigma_{ij}^2 \text{ for } t \neq t',$$

and

$$\begin{aligned} Var\left(a_{n,tt'}\right) = E\left(a_{n,tt'}^2\right) &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(u_{it}u_{it'}u_{jt}u_{jt'}\right), \text{ for } t \neq t' \\ &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2 = O\left(n^{-1}\right), \text{ using (A.7)}, \end{aligned}$$

which establishes (A.11) and (A.10). To prove (A.12) set $z_{it} = u_{it}\overline{u}_{i\circ} - T^{-1}\sigma_i^2$, and write $b_{n,t} = \frac{1}{n}\sum_{i=1}^{n} z_{it}$. Also note that $u_{it}\overline{u}_{i\circ} = T^{-1}\sum_{s=1}^{T} u_{it}u_{is}$, and given that $\{u_{it}\}$ is serially independent then $E(z_{it}) = 0$ and $E(b_{n,t}) = 0$. and

$$\begin{aligned} Var(b_{n,t}) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(z_{it}z_{jt}\right) &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E(u_{it}u_{jt}\overline{u}_{i\circ}\overline{u}_{j\circ}) - T^{-2}\bar{\sigma}_n^4 \\ &= n^{-2}T^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{s'=1}^{T}E\left(u_{it}u_{jt}u_{is}u_{js'}\right) - T^{-2}\bar{\sigma}_n^4. \end{aligned} \tag{A.16}$$

Also $E\left(u_{it}u_{jt}u_{is}u_{js'}\right) = 0$ for all t if $s \neq s'$. We are left with one case where $s = s' = t$, and one case where $s = s' \neq t$. Hence

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{s'=1}^{T}E\left(u_{it}u_{jt}u_{is}u_{js'}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}E\left(u_{it}u_{jt}u_{is}u_{js}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1,s=t}^{T}E\left(u_{it}u_{jt}u_{is}u_{js}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1,s\neq t}^{T}E\left(u_{it}u_{jt}u_{is}u_{js}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}E\left(u_{it}^2u_{jt}^2\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1,s\neq t}^{T}E\left(u_{it}u_{jt}\right)E\left(u_{is}u_{js}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[Cov\left(u_{it}^2, u_{jt}^2\right) + \sigma_i^2\sigma_j^2\right] + (T-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2, \end{aligned}$$

and hence

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{s'=1}^{T}E\left(u_{it}u_{jt}u_{is}u_{js'}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(u_{it}^2, u_{jt}^2\right) + \left(\sum_{i=1}^{n}\sigma_i^2\right)^2 + (T-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2,$$

$$\begin{aligned} Var(b_{n,t}) =& n^{-2}T^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{s'=1}^{T}E\left(u_{it}u_{jt}u_{is}u_{js'}\right) - T^{-2}\bar{\sigma}_n^4 \\ &= T^{-2}n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(u_{it}^2, u_{jt}^2\right) + T^{-2}\bar{\sigma}_n^4 + (T-1)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2 - T^{-2}\bar{\sigma}_n^4. \end{aligned}$$

Using this result in (A.16) we have

$$Var(b_{n,t}) = n^{-2}T^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(u_{it}^2, u_{jt}^2\right) + \frac{T-1}{T^2}\left(n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2\right).$$

Now using (A.9) and (A.7) we have $Var(b_{n,t}) = O\left(T^{-1}n^{-1}\right)$, which establishes (A.13), and result (A.12) follows by Markov inequality. ■

Lemma A.3 *Consider the* $n \times T$ *error matrix* $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, *where* $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$, *the* $n \times k$ *matrix of factor loadings,* $\mathbf{B}_n = (\boldsymbol{\beta}_{\circ 1}, \boldsymbol{\beta}_{\circ 2}, ..., \boldsymbol{\beta}_{\circ K})$, *where* $\boldsymbol{\beta}_{\circ k} = (\beta_{1k}, \beta_{2k}, ..., \beta_{nk})'$, *the* $n \times 1$ *vector of pricing errors* $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$, *with the pervasiveness coefficient,* α_η, *the observed factors,* $\mathbf{f}_k = (f_{k1}, f_{k2}, ..., f_{kT})'$ *are strong (with* $\alpha_k = 1$, *for* $k = 1, 2, ..., K$*), the missing factor* g_t, *has strength* $\alpha_\gamma < 1/2$, $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_K)$, $\mathbf{M}_n = \mathbf{I}_n - \frac{1}{n}\boldsymbol{\tau}_n\boldsymbol{\tau}_n'$, $\overline{\mathbf{u}}_{n\circ} = (\overline{u}_{1\circ}, \overline{u}_{2\circ}, ..., \overline{u}_{n\circ})'$, $\overline{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$, $\overline{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n}\sigma_i^2$, *and* $\boldsymbol{\tau}_n$ *and* $\boldsymbol{\tau}_T$ *are, respectively,* $n \times 1$ *and* $T \times 1$ *vectors of ones. Suppose that Assumptions 1, 3, 4, 2 and 5 hold. Then*

$$n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(n^{-1+\alpha_\eta}\right), \tag{A.17}$$

$$n^{-1}\mathbf{B}_n'\mathbf{M}_n\overline{\mathbf{u}}_{n\circ} = O_p\left(T^{-1/2}n^{-1/2}\right), \tag{A.18}$$

$$n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T = O_p\left(T^{-1/2}n^{-1/2}\right), \tag{A.19}$$

$$n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right), \tag{A.20}$$

$$n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\overline{\mathbf{u}}_{n\circ} = O_p\left(T^{-1}n^{-1/2}\right), \tag{A.21}$$

$$\mathbf{G}_T'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT} - \bar{\sigma}_n^2\mathbf{I}_T\right)\mathbf{G}_T = O_p\left(T^{-1}n^{-1/2}\right). \tag{A.22}$$

Proof. To establish (A.17) we first note that the k^{th} element of $n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n$ can be written as

$$\pi_{k,n} = n^{-1}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)\eta_i, \tag{A.23}$$

where $\bar{\beta}_k = n^{-1}\boldsymbol{\tau}_n'\boldsymbol{\beta}_{\circ k}$. Since η_j and β_{ik} are distributed independently for all i and j, then $E\left(\pi_{k,n}\right) = 0$, [22]

$$E\left(\left|\pi_{k,n}\right|\left|\boldsymbol{\eta}_n\right.\right) = n^{-1}\sum_{i=1}^{n}E\left|\beta_{ik} - \bar{\beta}_k\right|\left|\eta_i\right| \leq \left[\sup_{i,k}E\left|\beta_{ik} - \bar{\beta}_k\right|\right]\left(n^{-1}\sum_{i=1}^{n}\left|\eta_i\right|\right) = O\left(n^{-1+\alpha_\eta}\right),$$

for $k = 1, 2, ..., K$, and by Markov inequality we have $n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(n^{-1+\alpha_\eta}\right)$, as required. To establish (A.18), noting that

$$n^{-1}\overline{\mathbf{u}}_{n\circ}'\mathbf{M}_n\mathbf{B}_n = \left[n^{-1}\overline{\mathbf{u}}_{n\circ}'\left(\boldsymbol{\beta}_{\circ 1} - \tau_n\bar{\beta}_1\right), n^{-1}\overline{\mathbf{u}}_{n\circ}'\left(\boldsymbol{\beta}_{\circ 2} - \tau_n\bar{\beta}_2\right), ..., n^{-1}\overline{\mathbf{u}}_{n\circ}'\left(\boldsymbol{\beta}_{\circ K} - \tau_n\bar{\beta}_K\right)\right],$$

and the k^{th} element of $n^{-1}\mathbf{B}_n'\mathbf{M}_n\overline{\mathbf{u}}_{n\circ}$ is given by $c_{k,nT} = n^{-1}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)\bar{u}_{i\circ}$. We have $E\left(c_{k,nT}\right) = 0$, and $Var\left(c_{k,nT}\left|\boldsymbol{\beta}_{\circ k}\right.\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)\left(\beta_{jk} - \bar{\beta}_k\right)E\left(\overline{u}_{i\circ}\overline{u}_{j\circ}\right)$, with $E\left(\overline{u}_{i\circ}\overline{u}_{j\circ}\right) = T^{-1}\sigma_{ij}$. Hence (recalling that $\mathbf{V}_u = (\sigma_{ij})$)

$$Var\left(c_{nT,k}\right) = T^{-1}n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}E\left[\left(\beta_{ik} - \bar{\beta}_k\right)\left(\beta_{jk} - \bar{\beta}_k\right)\right], \tag{A.24}$$

and by Cauchy–Schwarz inequality and Assumption 2 we have $E\left[\left(\beta_{ik} - \bar{\beta}_k\right)\left(\beta_{jk} - \bar{\beta}_k\right)\right] \leq \left[E\left(\beta_{ik} - \bar{\beta}_k\right)^2\right]^{1/2}\left[E\left(\beta_{jk} - \bar{\beta}_k\right)^2\right]^{1/2} < C$. Hence, $Var\left(c_{nT,k}\right) \leq CT^{-1}n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sigma_{ij}\right|$. Since $\alpha_\gamma < 1/2$, then by (A.7) we have $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sigma_{ij}\right| = O(1)$ and $Var\left(c_{nT,k}\right) = O(T^{-1}n^{-1})$.

[22] Note that $\sup_{ik}E\left|\beta_{ik} - \bar{\beta}_k\right| < C$ follows from $\sup_i E\left|\beta_{ik}\right|^2 < C$, required by Assumption 2.

Thus by Markov inequality it follows that $c_{nT,k} = O_p\left(T^{-1/2}n^{-1/2}\right)$, for $k = 1, 2, ..., K$, and (A.18) follows. To establish (A.19) using $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$ we have

$$n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T = n^{-1}T^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{M}_T\mathbf{F}\right)\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1},$$

and since $\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$ is a positive definite matrix then it is sufficient to consider the probability order of the $K \times K$ matrix $T^{-1}n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{M}_T\mathbf{F} = (q_{kk'})$, where

$$q_{kk'} = T^{-1}n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\beta_{ik} - \bar{\beta}_k\right)\left(f_{k't} - \bar{f}_{k'}\right)u_{it} = n^{-1}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)\psi_{iT},$$

where $\psi_{iT} = T^{-1}\sum_{t=1}^{T}\left(f_{k't} - \bar{f}_{k'}\right)u_{it}$. Thus

$$Var\left(q_{kk'}\left|\boldsymbol{\beta}_{\circ k}\right.\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)\left(\beta_{jk} - \bar{\beta}_k\right)Cov\left(\psi_{iT}, \psi_{jT}\right). \tag{A.25}$$

Also

$$Cov\left(\psi_{iT}, \psi_{jT}\left|\mathbf{F}\right.\right) = E\left(\psi_{iT}\psi_{jT}\left|\mathbf{F}\right.\right) = T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left(f_{k't} - \bar{f}_{k'}\right)\left(f_{k't'} - \bar{f}_{k'}\right)E\left(u_{it}u_{jt'}\right),$$

and $E(u_{it}u_{jt'}) = \sigma_{ij}$ for $t = t'$ and 0 otherwise ($t \neq t'$). Then,

$$Cov\left(\psi_{iT}, \psi_{jT}\left|\mathbf{F}\right.\right) = T^{-2}\sum_{t=1}^{T}\left(f_{k't} - \bar{f}_{k'}\right)^2 E\left(u_{it}u_{jt}\right) = \sigma_{ij}T^{-2}\sum_{t=1}^{T}\left(f_{k't} - \bar{f}_{k'}\right)^2.$$

Using this result in (A.25) now yields

$$Var\left(q_{kk'}\right) = T^{-1}n^{-1}\left[T^{-1}\sum_{t=1}^{T}E\left(f_{k't} - \bar{f}_{k'}\right)^2\right]n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}E\left[\left(\beta_{ik} - \bar{\beta}_k\right)\left(\beta_{jk} - \bar{\beta}_k\right)\right].$$

The first term of the above is bounded since by assumption $\mathbf{f}_t$ is stationary. The second term is bounded as established above (see the derivations below (A.24)). Hence $Var\left(q_{kk'}\right) = O(T^{-1}n^{-1})$ and result (A.19) follows. To establish (A.20) note that $n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n = \left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\left(n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n\right)$, where $\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1} = O_p(1)$. Also the k^{th} element of $n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n$ is given by $p_k = T^{-1}\sum_{t=1}^{T}\left(f_{kt} - \bar{f}_k\right)c_{n,t}$, where $c_{n,t} = n^{-1}\sum_{i=1}^{n}u_{it}\eta_i = n^{-1}\boldsymbol{\eta}_n'\mathbf{u}_{\circ t}$. Under Assumption (4) f_{kt} and $c_{n,t}$ are distributed independently, $E(p_{n,t}) = 0$, and $c_{n,t}$ are also serially uncorrelated we have $Var(p_k) = T^{-2}\sum_{t=1}^{T}E\left(f_{kt} - \bar{f}_k\right)^2 Var(c_{n,t})$. Also noting that $Var(c_{n,t}\left|\eta\right.) = n^{-2}\eta_n'\mathbf{V}_u\eta_n$, then

$$Var(p_k) = \left[T^{-1}\sum_{t=1}^{T}E\left(f_{kt} - \bar{f}_k\right)^2\right]\left(T^{-1}n^{-2}\boldsymbol{\eta}_n'\mathbf{V}_u\boldsymbol{\eta}_n\right).$$

The first term is bounded, and it follows that

$$Var(p_k\left|\boldsymbol{\eta}_n\right.) < CT^{-1}n^{-2}\boldsymbol{\eta}_n'\mathbf{V}_u\boldsymbol{\eta}_n \leq CT^{-1}n^{-2}\left(\boldsymbol{\eta}_n'\boldsymbol{\eta}_n\right)\lambda_{\max}(\mathbf{V}_u) = O\left(T^{-1}n^{-2+\alpha_\eta+\alpha_\gamma}\right),$$

and (A.20) follows. To establish (A.21) note that

$$n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} = \left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\left(n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ}\right),$$

and the k^{th} element of $n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\overline{\mathbf{u}}_{n\circ} = (d_k)$ is given by

$$\begin{aligned} d_k &= \frac{1}{nT}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\sum_{i=1}^{n}\left(u_{it}-\bar{u}_{i\circ}\right)\bar{u}_{i\circ} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(f_{kt}-\bar{f}_k)(u_{it}\bar{u}_{i\circ}-\bar{u}_{i\circ}\bar{u}_{i\circ}) \\ &= \frac{1}{T}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\left[n^{-1}\sum_{i=1}^{n}\left(u_{it}\bar{u}_{i\circ}-T^{-1}\bar{\sigma}_n^2\right)\right] = \frac{1}{T}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)b_{n,t}, \end{aligned}$$

where $b_{n,t} = n^{-1}\sum_{i=1}^{n}\left(u_{it}\bar{u}_{i\circ} - T^{-1}\sigma_i^2\right)$. By assumption u_{it} and $f_{kt'}$ are distributed independently and $b_{n,t}$ are also serially independent. Then $E(d_k) = 0$, and

$$Var(d_k) = \frac{1}{T^2}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2 Var(b_{n,t}).$$

Now using (A.13) in Lemma A.2, $Var(b_{n,t}) = O\left(T^{-1}n^{-1}\right)$, and overall we have $Var(d_k) = O\left(T^{-2}n^{-1}\right)$, and by Markov inequality $d_k = O_p\left(T^{-1}n^{-1/2}\right)$, as required. Finally, consider (A.22) and note that

$$\begin{aligned} &\mathbf{G}'_T\left(n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}-\bar{\sigma}_n^2\mathbf{I}_T\right)\mathbf{G}_T \\ &= T^{-2}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\left[n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}-\bar{\sigma}_n^2\mathbf{I}_T\right]\mathbf{M}_T\mathbf{F}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}. \end{aligned}$$

Since by assumption $\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$ is a positive definite matrix for all T, and K is fixed then it is sufficient to derive the probability order of the (k,k') element of of the $K\times K$ matrix $\mathbf{\Delta}_{nT} = (\delta_{kk'})$

$$\begin{aligned} \mathbf{\Delta}_{nT} &= T^{-2}\mathbf{F}'\mathbf{M}_T\left[n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}-\bar{\sigma}_n^2\mathbf{I}_T\right]\mathbf{M}_T\mathbf{F} \\ &= T^{-1}\left[n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{M}_T\mathbf{F}-\bar{\sigma}_n^2T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right] \\ &= T^{-1}\left(n^{-1}T^{-1}\mathbf{A}\mathbf{M}_n\mathbf{A}'-\bar{\sigma}_n^2T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right) = T^{-1}\left(\mathbf{S}-\mathbf{R}\right), \end{aligned} \tag{A.26}$$

where $\mathbf{A} = \mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT} = (a_{ki})$. Denote the (k,k') elements of $\mathbf{S} = n^{-1}T^{-1}\mathbf{A}\mathbf{M}_n\mathbf{A}'$ and $\mathbf{R} = \bar{\sigma}_n^2T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ by $s_{kk'}$ and $r_{kk'}$, respectively, and note that

$$r_{kk'} = \left(n^{-1}\sum_{i=1}^{n}\sigma_i^2\right)\left[T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)(f_{k't}-\bar{f}_{k'})\right] \tag{A.27}$$

and $s_{kk'} = n^{-1}T^{-1}\sum_{i=1}^{n}\left(a_{ki}-\bar{a}_k\right)a_{k'i}$, where $a_{ki} = \sum_{t=1}^{T}\tilde{f}_{kt}u_{it}$, and $\bar{a}_k = \sum_{t=1}^{T}\tilde{f}_{kt}\bar{u}_{\circ t}$, where $\tilde{f}_{kt} = f_{kt}-\bar{f}_k$. Then

$$\begin{aligned} s_{kk'} &= n^{-1}T^{-1}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\tilde{f}_{kt}\left(u_{it}-\bar{u}_{\circ t}\right)\right]\left[\sum_{t=1}^{T}\tilde{f}_{k't}u_{it},\right] \\ &= n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{i=1}^{n}\left(u_{it}-\bar{u}_{\circ t}\right)u_{it'}\tilde{f}_{kt}\tilde{f}_{k't'} \\ &= n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{i=1}^{n}u_{it}u_{it'}\tilde{f}_{kt}\tilde{f}_{k't'} - T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\bar{u}_{\circ t}\bar{u}_{\circ t'}\tilde{f}_{kt}\tilde{f}_{k't'}. \end{aligned} \tag{A.28}$$

Using this result and $r_{kk'}$ given by (A.27) in (A.26) now yields

$$\begin{aligned}
T\delta_{kk'} &= (s_{kk'} - r_{kk'}) \\
&= n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{i=1}^{n} u_{it}u_{it'}\tilde{f}_{kt}\tilde{f}_{k't'} - T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\bar{u}_{\circ t}\bar{u}_{\circ t'}\tilde{f}_{kt}\tilde{f}_{k't'} \\
&- \bar{\sigma}_n^2\left(T^{-1}\sum_{t=1}^{T}\tilde{f}_{kt}\tilde{f}_{k't}\right) - \bar{\sigma}_n^2\left(T^{-1}\sum_{t=1}^{T}\tilde{f}_{kt}\tilde{f}_{k't}\right).
\end{aligned}$$

Also writing the first two terms as sums of the elements with $t = t'$ and those with $t \neq t'$, we have

$$\begin{aligned}
T\delta_{kk'} &= n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} u_{it}^2\tilde{f}_{kt}\tilde{f}_{k't} - \bar{\sigma}_n^2\left(T^{-1}\sum_{t=1}^{T}\tilde{f}_{kt}\tilde{f}_{k't}\right) - T^{-1}\sum_{t=1}^{T}\bar{u}_{\circ t}^2\tilde{f}_{kt}\tilde{f}_{k't}+ \\
&+ n^{-1}T^{-1}\sum_{t\neq t'}^{T}\sum_{i=1}^{n} u_{it}u_{it'}\tilde{f}_{kt}\tilde{f}_{k't'} - T^{-1}\sum_{t\neq t'}^{T}\bar{u}_{\circ t'}\tilde{f}_{kt}\tilde{f}_{k't'} \\
&= T^{-1}\sum_{t=1}^{T} a_{n,tt}\tilde{f}_{kt}\tilde{f}_{k't} - T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\bar{u}_{\circ t}\bar{u}_{\circ t'}\tilde{f}_{kt}\tilde{f}_{k't} + T^{-1}\sum_{t\neq t'}^{T} a_{n,tt'}\tilde{f}_{kt}\tilde{f}_{k't'}, \\
T\delta_{kk'} &= A_{kk'} - B_{kk'} + C_{kk'}, \qquad \text{(A.29)}
\end{aligned}$$

where $a_{n,tt} = n^{-1}\sum_{i=1}^{n}\left(u_{it}^2 - \sigma_i^2\right)$, and $a_{n,tt'} = n^{-1}\sum_{i=1}^{n} u_{it}u_{it'}$. Since u_{it} and $f_{kt'}$ are distributed independently then $E\left(A_{kk'}\right) = 0$, and

$$Var\left(A_{kk'}\right) = T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(a_{n,tt}a_{n,tt'}\right)E\left(\tilde{f}_{kt}\tilde{f}_{kt'}\tilde{f}_{k't}\tilde{f}_{k't'}\right).$$

Since u_{it} is serially independent, then $E\left(a_{n,tt}a_{n,tt'}\right) = 0$ for $t \neq t'$ and

$$Var\left(A_{kk'}\right) = T^{-2}\sum_{t=1}^{T} E\left(a_{n,tt}^2\right)E\left(\tilde{f}_{kt}^2\tilde{f}_{k't}^2\right) \leq \left[\sup_{k,k',t} E\left(\tilde{f}_{kt}^2\tilde{f}_{k't}^2\right)\right]\left[T^{-2}\sum_{t=1}^{T} E\left(a_{n,tt}^2\right)\right].$$

$\sup_{k,k',t} E\left(\tilde{f}_{kt}^2\tilde{f}_{k't}^2\right) < C$ since by assumption $sup_{t,k}E\left(\tilde{f}_{kt}^4\right) < C$. Also (using (A.8) of Lemma A.2)

$$E\left(a_{n,tt}^2\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(u_{it}^2, u_{jt}^2\right) = O\left(n^{-1}\right),$$

and we have $A_{kk'} = O_p\left(T^{-1}n^{-1/2}\right)$. Now write $B_{kk'}$ as

$$B_{kk'} = T^{-1}\sum_{t=1}^{T}\sum_{t'=1}^{T}\bar{u}_{\circ t}\bar{u}_{\circ t'}\tilde{f}_{kt}\tilde{f}_{k't} = \left(T^{-1/2}\sum_{t=1}^{T}\bar{u}_{\circ t}\tilde{f}_{kt}\right)\left(T^{-1/2}\sum_{t=1}^{T}\bar{u}_{\circ t}\tilde{f}_{k't}\right) = q_{\circ k}q_{\circ k'},$$

where $q_{\circ k} = T^{-1/2}\sum_{t=1}^{T}\bar{u}_{\circ t}\tilde{f}_{kt}$. Also $E(q_{\circ k}) = 0$, and $Var\left(q_{\circ k}\right) = T^{-1}\sum_{t=1}^{T} E\left(\bar{u}_{\circ t}^2\right)E\left(\tilde{f}_{kt}^2\right) = E\left(\bar{u}_{\circ t}^2\right)\left[T^{-1}\sum_{t=1}^{T} E\left(\tilde{f}_{kt}^2\right)\right]$, where $E\left(\bar{u}_{\circ t}^2\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij} = O\left(n^{-1}\right)$. Hence, $T^{-1/2}\sum_{t=1}^{T}\bar{u}_{\circ t}\tilde{f}_{kt} = O_p\left(n^{-1/2}\right)$, and $B_{kk'} = O_p\left(n^{-1}\right)$. Finally, consider $C_{kk'} = T^{-1}\sum_{t\neq t'}^{T} a_{n,tt'}\tilde{f}_{kt}\tilde{f}_{k't'}$, and note that

$E\left(a_{n,tt'}\right) = n^{-1}\sum_{i=1}^{n} E\left(u_{it}u_{it'}\right) = 0$ for all $t \neq t'$,which ensures that $E\left(C_{kk'}\right) = 0$. Further,

$$Var\left(C_{kk'}\right) = T^{-2}\sum_{t\neq t'}^{T}\sum_{s\neq s'}^{T} E\left(a_{n,tt'}a_{n,ss'}\right) E\left(\tilde{f}_{kt}\tilde{f}_{k't'}\tilde{f}_{ks}\tilde{f}_{k's'}\right) = T^{-2}\sum_{t\neq t'}^{T} E\left(a_{n,tt'}^{2}\right) E\left(\tilde{f}_{kt}^{2}\tilde{f}_{k't'}^{2}\right),$$

$E\left(a_{n,tt'}^{2}\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^{2}$. See (A.11) in Lemma A.2. Since by assumption $E\left(\tilde{f}_{kt}^{4}\right) < C$, then we have

$$Var\left(C_{kk'}\right) = \left(n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^{2}\right)\left[T^{-2}\sum_{t\neq t'}^{T} E\left(\tilde{f}_{kt}^{2}\tilde{f}_{k't'}^{2}\right)\right] < C\left(n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^{2}\right),$$

and using (A.7), it follows that $Var\left(C_{kk'}\right) = O\left(n^{-1}\right)$, and hence $C_{kk'} = O_p\left(n^{-1/2}\right)$. Using this result and the ones obtained for $A_{kk'}$ and $B_{kk'}$ in (A.29) now yields $\delta_{kk'} = T^{-1}\left[O_p\left(T^{-1}n^{-1/2}\right) + O_p\left(n^{-1}\right) + O_p\left(n\right.\right.$
$O_p\left(T^{-1}n^{-1/2}\right)$, as required. ■

B Proof of theorems in the paper

B.1 Proof of theorem 1

Consider the two-pass estimator of $\boldsymbol{\lambda}$ defined by (24), which we reproduce here for convenience

$$\hat{\boldsymbol{\lambda}}_{nT} = \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}_{n\circ},$$

where $\hat{\mathbf{B}}_{nT} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, ..., \hat{\boldsymbol{\beta}}_n)'$, $\bar{\mathbf{r}}_{n\circ} = (\bar{r}_{1\circ}, \bar{r}_{2\circ}, ..., \bar{r}_{n\circ})'$, $\bar{r}_{i\circ} = T^{-1}\sum_{t=1}^{T} r_{it}$,

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{B.1}$$

and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under the factor model (21)

$$\mathbf{r}_{i\circ} = \alpha_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{B.2}$$

where $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$, and hence

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}. \tag{B.3}$$

Stacking these results over i yields

$$\hat{\mathbf{B}}_{nT} = \mathbf{B}_n + \mathbf{U}_{nT}\mathbf{G}_T, \tag{B.4}$$

where $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, and

$$\mathbf{G}_T = \mathbf{M}_T\mathbf{F}(\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}. \tag{B.5}$$

Also

$$\mathbf{r}_{\circ t} = \alpha_n + \mathbf{B}_n\mathbf{f}_t + \mathbf{u}_{\circ t}, \tag{B.6}$$

where $\mathbf{u}_{\circ t} = (u_{1t}, u_{2t}, ..., u_{nt})'$, $\alpha_n = (\alpha_1,\alpha_2, ..., \alpha_n)'$. Using (9) in the paper,

$$\alpha_n = c\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\phi}_0 + \boldsymbol{\eta}_n, \tag{B.7}$$

and (B.6) can be written as $\mathbf{r}_{\circ t} = c\boldsymbol{\tau}_n + \mathbf{B}_n(\boldsymbol{\phi}_0 + \mathbf{f}_t) + \mathbf{u}_{\circ t} + \boldsymbol{\eta}_n$. Now averaging over t yields

$$\bar{\mathbf{r}}_{n\circ} = c\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{n\circ} + \boldsymbol{\eta}_n, \tag{B.8}$$

where $\bar{\mathbf{r}}_{n\circ} = T^{-1}\sum_{t=1}^{T}\mathbf{r}_{\circ t} = (\bar{r}_{1\circ}, \bar{r}_{2\circ}, ..., \bar{r}_{n\circ})'$, $\bar{\mathbf{u}}_{n\circ} = T^{-1}\sum_{t=1}^{T}\mathbf{u}_{\circ t} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, and

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0). \tag{B.9}$$

Using (B.8) in (24) we have

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_{nT} &= \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\left(c\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{n\circ} + \boldsymbol{\eta}_n\right) \\ &= \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\left[\hat{\mathbf{B}}_{nT}\boldsymbol{\lambda}_T^* - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{n\circ} + \boldsymbol{\eta}_n\right].\end{aligned} \tag{B.10}$$

Also using (B.9), and recalling that $\boldsymbol{\lambda}_0 = \boldsymbol{\phi}_0 + \boldsymbol{\mu}_0$, we have $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 - (\boldsymbol{\lambda}_T^* - \boldsymbol{\lambda}_0) = \left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0\right) - (\boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T - \boldsymbol{\lambda}_0)$, which yields $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = \hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0)$. Furthermore, $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$,where $\hat{\boldsymbol{\phi}}_{nT}$ is the two-step estimator of $\boldsymbol{\phi}_0$ given by (31). This results follows noting that $\hat{\boldsymbol{\lambda}}_{nT} = \hat{\boldsymbol{\phi}}_{nT} + \hat{\boldsymbol{\mu}}_T$, and $\boldsymbol{\lambda}_T^* = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T$. Therefore,

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = \left(\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0). \tag{B.11}$$

We focus on deriving the asymptotic distribution of $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$ since the panel (cross section) dimension does not apply to the second component, $(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0)$. Now using (B.4) in (B.10) and after some simplifications we have

$$\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)\hat{\boldsymbol{\lambda}}_{nT} = \hat{\mathbf{B}}_{nT}'\mathbf{M}_n\left[\hat{\mathbf{B}}_{nT}\boldsymbol{\lambda}_T^* - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{n\circ} + \boldsymbol{\eta}_n\right],$$

or

$$\left(n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)\left(\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) = \mathbf{p}_{nT}, \tag{B.12}$$

where

$$\begin{aligned}\mathbf{p}_{nT} &= n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1}\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} + n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} \\ &\quad - n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^* - n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*,\end{aligned} \tag{B.13}$$

also

$$\begin{aligned}n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT} &= n^{-1}(\mathbf{B}_n + \mathbf{U}_{nT}\mathbf{G}_T)'\mathbf{M}_n(\mathbf{B}_n + \mathbf{U}_{nT}\mathbf{G}_T) \\ &= n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right) + n^{-1}\left(\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{B}_n\right) + \\ &\quad n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\right) + n^{-1}\left(\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\right),\end{aligned} \tag{B.14}$$

Now using the results in Lemma A.3 for case where all the observed factors are strong, for a fixed T and as $n \to \infty$ we have

$$n^{-1}\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right) = \boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2\mathbf{G}_T'\mathbf{G}_T + o_p(1), \tag{B.15}$$

where, using (B.5),

$$\mathbf{G}_T'\mathbf{G}_T = \frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}. \tag{B.16}$$

Similarly, for the terms on the right hand side of (B.13) we have

$$\mathbf{p}_{nT} = -\frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* + O_p\left(n^{-1+\alpha_\eta}\right) + O_p\left(T^{-1/2}n^{-1/2}\right) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right).$$

It is now easily seen that for a fixed T, and if $\alpha_\eta < 1$ and $\alpha_\gamma < 1/2$, then as $n \to \infty$

$$\mathbf{p}_{nT} \to_p -\frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^*.$$

Also, for a fixed T by Assumption 1, $\frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}$ is a positive definite matrix, and by (B.15) $n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT} \to_p \boldsymbol{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}$ which is also a positive definite matrix, noting that under Assumption 2 $\boldsymbol{\Sigma}_{\beta\beta}$ is a positive definite matrix. Using these results in (B.12) we now have

$$\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 = -\frac{\bar{\sigma}^2}{T}\left[\boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2\frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* + o_p(1), \text{ for a fixed } T \text{ as } n \to \infty.$$

The bias of estimating λ_0 by the two-step estimator also contains the bias of estimating $\boldsymbol{\mu}_0$. Using the above result in (B.11) we now have (for a fixed T and as $n \to \infty$)

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) - \frac{\bar{\sigma}^2}{T}\left[\boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2\frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* + o_p(1),$$

which establishes Theorem 1.

B.2 Proof of theorem 2

Using the expression for $\hat{u}_{it}$ given by (27), we have $\hat{u}_{it} = \alpha_i - \hat{\alpha}_{iT} - \left(\hat{\boldsymbol{\beta}}_{i,T} - \boldsymbol{\beta}_i\right)'\mathbf{f}_t + u_{it}$. Since $\hat{u}_{it}$ are OLS residuals then for each i, we also have $T^{-1}\sum_{t=1}^{T}\hat{u}_{it} = 0$, and the above can be written equivalently as $\hat{u}_{it} = u_{it} - \bar{u}_{i\circ} - \left(\hat{\boldsymbol{\beta}}_{i,T} - \boldsymbol{\beta}_i\right)'(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)$, for $i = 1, 2, ..., n$, and stacking over i now yields

$$\hat{\mathbf{u}}_t = \mathbf{u}_t - \bar{\mathbf{u}} - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) = \mathbf{u}_t - \bar{\mathbf{u}} - \mathbf{U}_{nT}\mathbf{G}_T(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T), \tag{B.17}$$

and stacking over t

$$\hat{\mathbf{U}}_{nT} = \mathbf{U}_{nT}\mathbf{M}_T - \mathbf{U}_{nT}\mathbf{G}_T\mathbf{F}'\mathbf{M}_T = \mathbf{U}_{nT}\left(\mathbf{M}_T - \mathbf{G}_T\mathbf{F}'\mathbf{M}_T\right).$$

But $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$, and we have

$$\hat{\mathbf{U}}_{nT} = \mathbf{U}_{nT}\mathbf{R}_T,\ \mathbf{R}_T = \mathbf{M}_T - \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T,$$

where $\mathbf{R}_T^2 = \mathbf{R}_T = \mathbf{R}_T'$, $Tr\left(\mathbf{R}_T\right) = T - 1 - K$. Then

$$\hat{\bar{\sigma}}_{nT}^2 = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\hat{u}_{it}^2}{n(T-K-1)} = \frac{Tr\left(n^{-1}\hat{\mathbf{U}}_{nT}'\hat{\mathbf{U}}_{nT}\right)}{T-K-1}.$$

Also

$$n^{-1}T^{-1}E\left[Tr\left(\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\right]=n^{-1}T^{-1}E\left(\sum_{t=1}^{T}\sum_{i=1}^{n}u_{it}^{2}\right)=n^{-1}\sum_{i=1}^{n}\sigma_{i}^{2}=\bar{\sigma}_{n}^{2},$$

$$E\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)=n^{-1}\sum_{i=1}^{n}E\left(\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'\right)=\bar{\sigma}_{n}^{2}\mathbf{I}_{T}.$$

Let $v=T-K-1$ and note that

$$\begin{aligned}v\widehat{\bar{\sigma}}_{nT}^{2}&=Tr\left(n^{-1}\hat{\mathbf{U}}_{nT}'\hat{\mathbf{U}}_{nT}\right)=Tr\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\mathbf{R}_{T}\right)=\\&=Tr\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\mathbf{M}_{T}\right)-Tr\left(n^{-1}\mathbf{F}'\mathbf{M}_{T}\mathbf{U}_{nT}'\mathbf{U}_{nT}\mathbf{M}_{T}\mathbf{F}\left(\mathbf{F}'\mathbf{M}_{T}\mathbf{F}\right)^{-1}\right)\\&=Tr\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)-T^{-1}\tau_{T}'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\tau_{T}-Tr\left[\mathbf{Q}'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\mathbf{Q}\right]\end{aligned}$$

where $\mathbf{Q}=\mathbf{M}_{T}\mathbf{F}\left(T^{-1}\mathbf{F}'\mathbf{M}_{T}\mathbf{F}\right)^{-1/2}$. Consider the first term and note that

$$n^{-1}Tr\left(\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)=\sum_{t=1}^{T}\left[n^{-1}\sum_{i=1}^{n}\left(u_{it}^{2}-\sigma_{i}^{2}\right)\right]+T\bar{\sigma}_{n}^{2}. \tag{B.18}$$

Similarly

$$\begin{aligned}T^{-1}\tau_{T}'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\tau_{T}&=T^{-1}n^{-1}\tau_{T}'\left[\mathbf{U}_{nT}'\mathbf{U}_{nT}-E\left(\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\right]\tau_{T}+\bar{\sigma}_{n}^{2},\\Tr\left[\mathbf{Q}'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\mathbf{Q}\right]&=Tr\left[\mathbf{Q}'\left[\mathbf{U}_{nT}'\mathbf{U}_{nT}-E\left(\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\right]\mathbf{Q}\right]+K\bar{\sigma}_{n}^{2}.\end{aligned}$$

Hence

$$\widehat{\bar{\sigma}}_{nT}^{2}-\bar{\sigma}_{n}^{2}=(T/v)\left(a_{nT}+b_{nT}+c_{nT}\right), \tag{B.19}$$

where $a_{nT}=T^{-1}\sum_{t=1}^{T}\left[n^{-1}\sum_{i=1}^{n}\left(u_{it}^{2}-\sigma_{i}^{2}\right)\right]$, $b_{nT}=T^{-2}n^{-1}\tau_{T}'\left[\mathbf{U}_{nT}'\mathbf{U}_{nT}-E\left(\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\right]\tau_{T}$, and $c_{nT}=T^{-1}n^{-1}Tr\left[\mathbf{Q}'\left[\mathbf{U}_{nT}'\mathbf{U}_{nT}-E\left(\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\right]\mathbf{Q}\right]$. Due to the independence of $\mathbf{F}$ and $\mathbf{U}_{nT}$, we have $E(a_{nT})=0$, $E(b_{nT})=0$ and $E(c_{nT})=0$, and for any fixed n and T $E\left(\widehat{\bar{\sigma}}_{nT}^{2}-\bar{\sigma}_{n}^{2}\right)=0$, and for a fixed T, $\lim_{n\to\infty}E\left(\widehat{\bar{\sigma}}_{nT}^{2}\right)=\bar{\sigma}^{2}$, and result (60) follows. To establish the probability order of $\widehat{\bar{\sigma}}_{nT}^{2}-\bar{\sigma}_{n}^{2}$, we consider the probability orders of a_{nT}, b_{nT}, and c_{nT} in turn, noting that that $T/v=T/(T-K-1)=O(1)$. For a_{nT}, using result (A.8) in Lemma A.2, and noting that u_{it} are serially independent we have

$$a_{nT}=O_{p}\left(T^{-1/2}n^{-1/2}\right). \tag{B.20}$$

Consider now b_{nT} and note that $b_{nT}=T^{-2}n^{-1}\sum_{i=1}^{n}\left[\tau_{T}'\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'\tau_{T}-E\left(\tau_{T}'\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'\tau_{T}\right)\right]$. Also $\tau_{T}'\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'\tau_{T}=\sum_{t=1}^{T}\sum_{t'=1}^{T}u_{it}u_{it'}$, and

$$\begin{aligned}b_{nT}&=T^{-2}n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left[u_{it}u_{it'}-E\left(u_{it}u_{it'}\right)\right]\\&=T^{-2}\sum_{t=1}^{T}n^{-1}\sum_{i=1}^{n}\left(u_{it}^{2}-\sigma_{i}^{2}\right)+T^{-2}\sum_{t\neq t'}^{T}\left(n^{-1}\sum_{i=1}^{n}u_{it}u_{it'}\right)\\&=T^{-2}\sum_{t=1}^{T}a_{n,tt}+T^{-2}\sum_{t\neq t'}^{T}a_{n,tt'},\end{aligned}$$

where $a_{n,tt}$ and $a_{n,tt'}$ are both shown in Lemma A.2 to be $O_p(n^{-1/2})$. See equations (A.8) and (A.10). Therefore, given that $a_{n,tt}$ and $a_{n,tt'}$ with $t \neq t'$ are also distributed independently over t we have

$$b_{nT} = O_p(n^{-1/2}T^{-1/2}). \tag{B.21}$$

Denote the k^{th} column of $\mathbf{Q}$ by $\mathbf{q}_k = (q_{k1}, q_{2k}, ..., q_{Tk})'$ (a $T \times 1$ vector) and write c_{nT} as

$$\begin{aligned} c_{nT} &= \sum_{k=1}^{K} T^{-1} \left[n^{-1} \sum_{i=1}^{n} \mathbf{q}_k' \left[\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}' - E\left(\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}' \right) \right] \mathbf{q}_k \right] \\ &= \sum_{k=1}^{K} T^{-1} \left[n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{t'=1}^{T} q_{kt}q_{kt'} \left[u_{it}u_{it'} - E\left(u_{it}u_{it'} \right) \right] \right]. \end{aligned}$$

Consider the k^{th} term of the above sum, and note that

$$c_{nT,k} = T^{-1}n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{t'=1}^{T} q_{kt}q_{kt'} \left[u_{it}u_{it'} - E\left(u_{it}u_{it'} \right) \right] = T^{-1} \sum_{t=1}^{T} \sum_{t'=1}^{T} q_{kt}q_{kt'}a_{n,tt'},$$

where $a_{n,tt'} = n^{-1}\sum_{i=1}^{n} \left[u_{it}u_{it'} - E\left(u_{it}u_{it'} \right) \right]$. is defined by (A.8) and (A.10) in Lemma A.2, with $Var(a_{n,tt'}) = O\left(n^{-1} \right)$ for all t and t'. Also $a_{n,tt'}$ and $a_{n,ss'}$ are distributed independently if t or t' differ from s or s'. Therefore, for all t and t' for all t

$$Var\left(c_{nT,k} \right) = T^{-2} \sum_{t=1}^{T} \sum_{t'=1}^{T} q_{kt}^2 q_{kt'}^2 Var(a_{n,tt'}) \leq O\left(n^{-1} \right) \left(T^{-2} \sum_{t=1}^{T} q_{kt}^2 \right).$$

But it is easily verified that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_K$ which yields $\sum_{t=1}^{T} q_{kt}^2 = 1$, and $Var\left(c_{nT,k} \right) = O\left(T^{-2}n^{-1} \right)$. Hence, $c_{nT,k} = O_p(T^{-1}n^{-1/2})$, for $k = 1, 2, ..., K$, which establishes that $c_{nT} = O_p(T^{-1}n^{-1/2})$. The order result in (61) now follows using this result, (B.20) and (B.21) in (B.19).

B.3 Proof of theorem 3

The bias-corrected estimator of $\boldsymbol{\phi}_0$ is given by (32) which we reproduce here and re-write as

$$\mathbf{H}_{nT} \left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 \right) = \mathbf{s}_{nT}, \tag{B.22}$$

where

$$\mathbf{s}_{nT} = \frac{\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{a}}_{nT}}{n} + T^{-1}\hat{\bar{\sigma}}_{nT}^2 \left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T} \right)^{-1} \hat{\mu}_T - \mathbf{H}_{nT}\ \boldsymbol{\phi}_0, \tag{B.23}$$

$$\mathbf{H}_{nT} = \frac{\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n} - T^{-1}\hat{\bar{\sigma}}_{nT}^2 \left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T} \right)^{-1}. \tag{B.24}$$

Also $\hat{\mathbf{a}}_{nT} = \bar{\mathbf{r}}_{n\circ} - \hat{\mathbf{B}}_{nT}\hat{\mu}_T$, and $\bar{\mathbf{r}}_{n\circ} = c\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{n\circ} + \boldsymbol{\eta}_n$ (see (B.8)). Using these results and noting that $\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T$, we have

$$\hat{\alpha}_{nT} = c\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\phi}_0 + \bar{\mathbf{u}}_{n\circ} + \boldsymbol{\eta}_n - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) \hat{\boldsymbol{\mu}}_T$$

and $\bar{\mathbf{u}}_{n\circ} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$. Then,

$$\begin{aligned} \hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{a}}_{nT} &= \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\mathbf{B}_n \right) \boldsymbol{\phi}_0 + \hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} \\ &\quad + \hat{\mathbf{B}}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n - \hat{\mathbf{B}}_{nT}'\mathbf{M}_n \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) \hat{\boldsymbol{\mu}}_T. \end{aligned} \tag{B.25}$$

Also

$$\hat{\mathbf{B}}_{nT} = \mathbf{B}_n + \mathbf{U}_{nT}\mathbf{G}_T, \tag{B.26}$$

where $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$ and $\mathbf{G}_T$ is defined by (B.16). Using these results together with (B.24), the right hand side of (B.22) can be written as

$$\mathbf{s}_{nT} = \frac{\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n}{n} + \frac{\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n}{n} + \frac{\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ}}{n} - \frac{\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*}{n}$$
$$- \mathbf{G}_T'\left(\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} - \widehat{\bar{\sigma}}_{nT}^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^*,$$

where the last term can be decomposed as

$$\mathbf{G}_T'\left(\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} - \widehat{\bar{\sigma}}_{nT}^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^* = \mathbf{G}_T'\left(\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^* - \left(\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2\right)\mathbf{G}_T'\mathbf{G}_T\boldsymbol{\lambda}_T^*.$$

Similarly, using (B.14), we have

$$\mathbf{H}_{nT} = n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right) + n^{-1}\left(\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{B}_n\right) +$$
$$n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\right) + \mathbf{G}_T'\left(\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2\right)\mathbf{G}_T - T^{-1}\left(\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}.$$

Using Theorem 2 and since by assumption $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ is positive definite, then

$$\left(\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} = O_p\left(T^{-1/2}n^{-1/2}\right).$$

Also using results in Lemma A.3, we have

$$\mathbf{H}_{nT} = n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right) + O_p\left(T^{-1/2}n^{-1/2}\right). \tag{B.27}$$

Hence, $\mathbf{H}_{nT} \to_p \boldsymbol{\Sigma}_{\beta\beta}$ for a fixed T as $n \to \infty$, so long as $\alpha_\gamma < 1/2$. Note also that by Assumption $\boldsymbol{\Sigma}_{\beta\beta}$ is positive definite. Further

$$\mathbf{s}_{nT} = n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right) + n^{-1}\left(\mathbf{B}_n' + \mathbf{G}_T'\mathbf{U}_{nT}'\right)\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} \tag{B.28}$$
$$- \mathbf{G}_T'\left(\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^* + T^{-1}\left(\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^*.$$

Using (A.17) and (A.20) (in Lemma A.3) we have

$$n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(n^{-1+\alpha_\eta}\right) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right),$$

and (A.21) and (A.22)

$$n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{G}_T'\left(\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^* = O_p\left(T^{-1}n^{-1/2}\right).$$

Further by (61)

$$T^{-1}\left(\widehat{\bar{\sigma}}_{nT}^2 - \bar{\sigma}_n^2\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* = O_p\left(T^{-3/2}n^{-1/2}\right). \tag{B.29}$$

Hence

$$\mathbf{s}_{nT} = n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right) + O_p\left(n^{-1+\alpha_\eta}\right) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{-1}n^{-1/2}\right) + O_p\left(T^{-3/2}n^{-1/2}\right). \tag{B.30}$$

Using this result and (B.27) in (B.22) now yields (62), as required. To derive the asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$ since by assumption $\alpha_\eta + \alpha_\gamma < 1$, then the dominant term of $\mathbf{s}_{nT}$ is given by

$$n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right) = O_p\left(T^{-1/2}n^{-1/2}\right), \tag{B.31}$$

and to ensure that we end up with a non-degenerate, stable limiting distribution, $\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right)$ needs to be scaled by $\sqrt{nT}$ with n and $T \to \infty$, jointly. To this end we first note that when T is fixed $\mathbf{H}_{nT} \to_p \boldsymbol{\Sigma}_{\beta\beta}$, as $n \to \infty$ and we have

$$\sqrt{nT}\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) \overset{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}\left(\sqrt{nT}\mathbf{s}_{nT}\right). \tag{B.32}$$

Using (B.30)

$$\sqrt{nT}\mathbf{s}_{nT} = \boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2}+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{1/2}n^{-1/2+\alpha_\eta}\right) + O_p\left(T^{-1/2}\right),$$

where

$$\boldsymbol{\xi}_{nT} = T^{1/2}n^{-1/2}\left(\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right). \tag{B.33}$$

Using the above results in (B.32) we now have

$$\sqrt{nT}\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) \overset{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}\left[\boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2}+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{1/2}n^{-1/2+\alpha_\eta}\right) + O_p\left(T^{-1/2}\right)\right]. \tag{B.34}$$

Therefore, when condition $(T/n)^{1/2}n^{\alpha_\eta} \to 0$, as n and $T \to \infty$, is met we have

$$\sqrt{nT}\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) \overset{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}\boldsymbol{\xi}_{nT} + o_p(1).$$

To derive the asymptotic distribution of $\boldsymbol{\xi}_{nT}$, we note that $\bar{\mathbf{u}}_{n\circ} = T^{-1}\mathbf{U}_{nT}\boldsymbol{\tau}_T$, and $\mathbf{G}_T\boldsymbol{\lambda}_T^* = T^{-1}\mathbf{M}_T\mathbf{F}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\boldsymbol{\lambda}_T^*$. Then, using these results in (B.33)

$$\boldsymbol{\xi}_{nT} = (\xi_{k,nT}) = n^{-1/2}T^{-1/2}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{a}_T, \tag{B.35}$$

where $\mathbf{a}_T = \boldsymbol{\tau}_T - \mathbf{M}_T\mathbf{F}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\boldsymbol{\lambda}_T^* = (a_t)$. Also

$$s_{a,T}^2 = T^{-1}\sum_{t=1}^{T}a_t^2 = T^{-1}\mathbf{a}_T'\mathbf{a}_T = 1 + \boldsymbol{\lambda}_T^{*\prime}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\boldsymbol{\lambda}_T^*, \tag{B.36}$$

where $\boldsymbol{\lambda}_T^* = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0)$, and $(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) = O_p\left(T^{-1/2}\right)$. Further

$$s_{a,T}^2 \geq 1 \text{ and } s_{a,T}^2 \leq 1 + \left(\boldsymbol{\lambda}_T^{*\prime}\boldsymbol{\lambda}_T^*\right)\lambda_{\max}\left[(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\right] < C, \tag{B.37}$$

and $s_a^2 = \lim_{T\to\infty} s_{a,T}^2 = 1+\boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0'$. The k^{th} element of $\boldsymbol{\xi}_{nT}$ is given by $\xi_{k,nT} = n^{-1/2}T^{-1/2}\sum_{i=1}^n \sum_{t=1}^T a_t(\beta_{ik} - \bar{\beta}_k)u_{it}$, and using (42) we have

$$\xi_{k,nT} = \left(T^{-1/2}\sum_{t=1}^T a_t g_t\right)\left[n^{-1/2}\sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k)\gamma_i\right] + n^{-1/2}T^{-1/2}\sum_{i=1}^n\sum_{t=1}^T a_t(\beta_{ik} - \bar{\beta}_k)v_{it}.$$

Under Assumption 3 g_t is distributed independently of $\boldsymbol{f}_t$ (and hence of a_t), as well as being serially independent. Also $Var(T^{-1/2}\sum_{t=1}^T a_t g_t) = s_{a,T}^2$ (recall that $E(g_t) = 0$ and $E(g_t^2) = 1$), and we have $T^{-1/2}\sum_{t=1}^T a_t g_t = O_p(1)$. Further $E\left|n^{-1/2}\sum_{i=1}^n(\beta_{ik} - \bar{\beta}_k)\gamma_i\right| \leq \sup_{i,k} E\left|\beta_{ik} - \bar{\beta}_k\right|\left(n^{-1/2}\sum_{i=1}^n |\gamma_i|\right) = O(n^{-1/2+\alpha_\gamma})$. Hence

$$\begin{aligned}\xi_{k,nT} &= n^{-1/2}T^{-1/2}\sum_{t=1}^T\sum_{i=1}^n a_t(\beta_{ik} - \bar{\beta}_k)v_{it} + O(n^{-1/2+\alpha_\gamma}) \\ &= T^{-1/2}\sum_{t=1}^T a_t h_{nt} + O(n^{-1/2+\alpha_\gamma}), \qquad \text{(B.38)}\end{aligned}$$

where $h_{nt} = n^{-1/2}\sum_{i=1}^n(\beta_{ik} - \bar{\beta}_k)v_{it}$. Under Assumption 2, $h_{nt} = n^{-1/2}\sum_{i=1}^n(\beta_{ik} - \bar{\beta}_k)v_{it} \to_d N(0, \omega_k^2)$, for $k = 1, 2, ..., K$, where

$$\omega_k^2 = p\lim_{n\to\infty} n^{-1}\sum_{i=1}^n\sum_{j=1}^n(\beta_{ik} - \bar{\beta}_k)(\beta_{jk} - \bar{\beta}_k)\sigma_{v,ij} > 0,$$

and $\omega_k^2 \leq \sup_{i,k} E(\beta_{ik} - \bar{\beta}_k)^2 \lim_{n\to\infty} n^{-1}\sum_{i=1}^n\sum_{j=1}^n |\sigma_{v,ij}| < C$. Also, since v_{it} are serially independent then there exists T_0 such that for any fixed $T > T_0$ and as $n \to \infty$

$$T^{-1/2}\sum_{t=1}^T a_t h_{nt} \to_d N\left(0, \omega_k^2\left(1 + s_{aT}^2\right)\right),$$

where s_{aT}^2 is defined by (B.36). Using this result in (B.38) and noting that $\alpha_\gamma < 1/2$, we also have for any fixed T and as $n \to \infty$,

$$\xi_{k,nT} \to_d N\left(0, \omega_k^2\left(1 + s_{aT}^2\right)\right), \text{ for a fixed } T > T_0 \text{ and as } n \to \infty.$$

This result extends readily to the case where n and $T \to \infty$, jointly. In this case

$$\xi_{k,nT} \to_d N\left(0, \omega_k^2\left(1 + s_a^2\right)\right), \text{ where } s_a^2 = 1 + \boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0.$$

Similarly, We have (using $u_{it} = \gamma_i g_t + v_{it}$)

$$Cov\left(\xi_{k,nT}, \xi_{k',nT}\right) = n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}a_t^2(\beta_{ik} - \bar{\beta}_k)(\beta_{jk'} - \bar{\beta}_{k'})E\left(u_{it}u_{jt}\right) =$$

$$\left(1 + s_{aT}^2\right)\left[n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i\gamma_j(\beta_{ik} - \bar{\beta}_k)(\beta_{jk'} - \bar{\beta}_{k'})\right]$$

$$+ \left(1 + s_{aT}^2\right)\left[n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{v,ij}(\beta_{ik} - \bar{\beta}_k)(\beta_{jk'} - \bar{\beta}_{k'})\right].$$

But

$$E\left|n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i\gamma_j(\beta_{ik} - \bar{\beta}_k)(\beta_{jk'} - \bar{\beta}_{k'})\right| \leq \sup_{i,k,k'}\left|(\beta_{ik} - \bar{\beta}_k)(\beta_{jk'} - \bar{\beta}_{k'})\right| E(\beta_{ik} - \bar{\beta}_k)^2\left(n^{-1/2}\sum_{j=1}^{n}|\gamma_i|\right)$$

$$\leq \sup_{i,k} E(\beta_{ik} - \bar{\beta}_k)^2\left(n^{-1/2}\sum_{j=1}^{n}|\gamma_i|\right)^2 = O\left(n^{-1+2\alpha_\gamma}\right).$$

and similarly

$$\left|n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{v,ij}(\beta_{ik} - \bar{\beta}_k)(\beta_{jk'} - \bar{\beta}_{k'})\right| \leq \sup_{i,k} E(\beta_{ik} - \bar{\beta}_k)^2 n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{v,ij}| < C.$$

Hence $Cov\left(\xi_{k,nT}, \xi_{k',nT}\right) < C$, for all k and k'. Using the above results in (B.35, and noting that K is fixed, we have $\boldsymbol{\xi}_{nT} \rightarrow_d N\left(\mathbf{0}, \mathbf{V}_\xi\right)$, as n and $T \rightarrow \infty$,where $\mathbf{V}_\xi = \left(1 + \boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0\right) p\lim_{n\rightarrow\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\right.$ noting that $s_{a,T}^2 \rightarrow_p 1 + \boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0$, where $s_{a,T}^2$ is given by (B.36). Also recall from (B.34) that $\sqrt{nT}\left(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0\right) = \boldsymbol{\Sigma}_{\beta\beta}^{-1}\boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2}+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{1/2}n^{-1/2+\alpha_\eta}\right) + O_p\left(T^{-1/2}\right)$. Hence, result (66) follows since by assumption $\alpha_\eta < 1/2$, $\alpha_\gamma < 1/2$, and $(T/n)^{1/2}n^{\alpha_\gamma} \rightarrow 0$.

B.4 Proof of theorem 4

Using (70) and (69) and replacing $T^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$ by $\mathbf{G}_T'\mathbf{G}_T$ we have (see (B.5))

$$\mathbf{H}_{nT}\left(\boldsymbol{\alpha}\right)\mathbf{D}_\alpha\left(\tilde{\boldsymbol{\phi}}_{nT}\left(\boldsymbol{\alpha}\right) - \boldsymbol{\phi}_0\right) = \mathbf{q}_{nT}\left(\boldsymbol{\alpha}\right), \tag{B.39}$$

where $\mathbf{q}_{nT}\left(\boldsymbol{\alpha}\right) = \mathbf{D}_\alpha^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{a}}_{nT} + n\widehat{\bar{\sigma}}_{nT}^2\mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{G}_T\hat{\mu}_T - \mathbf{H}_{nT}\left(\boldsymbol{\alpha}\right)\mathbf{D}_\alpha\boldsymbol{\phi}_0$, and $\mathbf{H}_{nT}\left(\boldsymbol{\alpha}\right) = \mathbf{D}_\alpha^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\mathbf{D}_\alpha^{-1} - n\widehat{\bar{\sigma}}_{nT}^2\mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{G}_T\mathbf{D}_\alpha^{-1}$. But

$$\mathbf{H}_{nT}\left(\boldsymbol{\alpha}\right) = n\mathbf{D}_\alpha^{-1}\mathbf{H}_{nT}\mathbf{D}_\alpha^{-1}, \text{ and } \mathbf{q}_{nT}\left(\boldsymbol{\alpha}\right) = n\mathbf{D}_\alpha^{-1}\mathbf{s}_{nT}, \tag{B.40}$$

where $\mathbf{s}_{nT}$ and $\mathbf{H}_{nT}$ are already defined by (B.23) and (B.24). Consider first the limiting property of $\mathbf{H}_{nT}\left(\boldsymbol{\alpha}\right)$, and using (B.26) note that

$$\mathbf{H}_{nT}\left(\boldsymbol{\alpha}\right) = \mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1} +$$
$$\mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\mathbf{D}_\alpha^{-1} - n\widehat{\bar{\sigma}}_{nT}^2\mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{G}_T\mathbf{D}_\alpha^{-1},$$

or

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\mathbf{D}_\alpha^{-1}$$
$$+ n\mathbf{D}_\alpha^{-1}\left[\mathbf{G}_T'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\right)\mathbf{G}_T - \hat{\tilde{\sigma}}_{nT}^2\mathbf{G}_T'\mathbf{G}_T\right]\mathbf{D}_\alpha^{-1}.$$

Further

$$n\mathbf{D}_\alpha^{-1}\left[\mathbf{G}_T'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\right)\mathbf{G}_T - \hat{\tilde{\sigma}}_{nT}^2\mathbf{G}_T'\mathbf{G}_T\right]\mathbf{D}_\alpha^{-1}$$
$$= n\mathbf{D}_\alpha^{-1}\left[\mathbf{G}_T'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\right)\mathbf{G}_T - \bar{\sigma}_n^2\mathbf{G}_T'\mathbf{G}_T\right]\mathbf{D}_\alpha^{-1} - \left(\hat{\tilde{\sigma}}_{nT}^2 - \bar{\sigma}_n^2\right)n\mathbf{D}_\alpha^{-1}\mathbf{G}_T'\mathbf{G}_T\mathbf{D}_\alpha^{-1}.$$

But by (61) $\hat{\tilde{\sigma}}_{nT}^2 - \bar{\sigma}_n^2 = O_p(T^{-1/2}n^{-1/2})$, and $\left\|\mathbf{D}_\alpha^{-1}\right\| = \lambda_{max}^{1/2}\left(\mathbf{D}_\alpha^{-2}\right) = n^{-\alpha_{\min}/2}$. Then, using results in Lemma A.3 we have

$$\left\|\mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\mathbf{D}_\alpha^{-1}\right\| \leq n\left\|\mathbf{D}_\alpha^{-1}\right\|^2\left\|n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\right\| = O_p\left(T^{-1/2}n^{-\alpha_{\min}+1/2}\right),$$

and

$$\left\|n\mathbf{D}_\alpha^{-1}\mathbf{G}_T'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT} - \hat{\tilde{\sigma}}_{nT}^2\mathbf{I}_T\right)\mathbf{G}_T\mathbf{D}_\alpha^{-1}\right\|$$
$$\leq n\left\|\mathbf{D}_\alpha^{-1}\right\|\left\|\mathbf{G}_T'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT} - \hat{\tilde{\sigma}}_{nT}^2\mathbf{I}_T\right)\mathbf{G}_T\right\| = O_p\left(T^{-1}n^{-\alpha_{\min}+1/2}\right).$$

Hence

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1} + O_p\left(T^{-1}n^{-\alpha_{\min}+1/2}\right),$$

and $\mathbf{H}_{nT}(\boldsymbol{\alpha}) \to_p \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$, as $n \to \infty$, for a fixed T, so long as $\alpha_{\min} > 1/2 > \alpha_\gamma$. By Assumption 2, $\lim_{n\to\infty}\left(\mathbf{D}_\alpha^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1}\right) = \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$ is a positive definite matrix. Using this result in (B.39) we have

$$\mathbf{D}_\alpha\left(\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0\right) \overset{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{q}_{nT}(\boldsymbol{\alpha}), \tag{B.41}$$

and (since $\boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$ is positive definite, $\left\|\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\right\| < C$)

$$\left\|\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0\right\| \leq \left\|\mathbf{D}_\alpha^{-1}\right\|\left\|\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\right\|\left\|\mathbf{q}_{nT}(\boldsymbol{\alpha})\right\| \leq Cn^{-\alpha_{\min}/2}\left\|\mathbf{q}_{nT}(\boldsymbol{\alpha})\right\|.$$

Using (B.40)

$$\left\|\mathbf{q}_{nT}(\boldsymbol{\alpha})\right\| = \left\|n\mathbf{D}_\alpha^{-1}\mathbf{s}_{nT}\right\| \leq n^{1-\alpha_{\min}/2}\left\|\mathbf{s}_{nT}\right\|. \tag{B.42}$$

Also from (B.30) we have

$$\mathbf{s}_{nT} = n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right) + O_p\left(n^{-1+\alpha_\eta}\right) \tag{B.43}$$
$$+ O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{-1}n^{-1/2}\right).$$

Also using (B.31))

$$n^{-1}\left(\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}_n'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*\right) = O_p\left(T^{-1/2}n^{-1/2}\right). \tag{B.44}$$

Substituting (B.44) in (B.43) and using the result in (B.42) we have

$$\left\|\mathbf{q}_{nT}(\boldsymbol{\alpha})\right\| = O_p\left(n^{-\alpha_{\min}/2+1/2}T^{-1/2}\right) + O_p\left(T^{-1/2}n^{\frac{-\alpha_{\min}+\left(\alpha_\eta+\alpha_\gamma\right)}{2}}\right) \tag{B.45}$$
$$+ O_p\left(n^{-\alpha_{\min}/2+\alpha_\eta}\right) + O_p\left(n^{-\alpha_{\min}/2+1/2}T^{-1}\right).$$

Denote the k^{th} element of $\mathbf{q}_{nT}(\boldsymbol{\alpha})$ by $q_{k,nT}(\boldsymbol{\alpha})$, we also have

$$q_{k,nT}(\boldsymbol{\alpha}) = O_p\left(n^{-\alpha_{\min}/2+1/2}T^{-1/2}\right) + O_p\left(T^{-1/2}n^{\frac{-\alpha_{\min}+\left(\alpha_{\eta}+\alpha_{\gamma}\right)}{2}}\right) + O_p\left(n^{-\alpha_{\min}/2+\alpha_{\eta}}\right) + O_p\left(n^{-\alpha_{\min}/2+1/2}T^{-1}\right).$$

Also note that the k^{th} element of $\mathbf{D}_{\alpha}\left(\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0\right)$ is given by $n^{\alpha_k/2}\left(\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k}\right)$. Hence, in view of (B.41) and since $\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})$ is a positive definite matrix then the probability order of $n^{\alpha_k/2}\left(\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k}\right)$ must be the same as that of $q_{k,nT}$, and hence (as required)

$$\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} = O_p\left(n^{-(\alpha_k+\alpha_{\min})/2+1/2}T^{-1/2}\right) + O_p\left(n^{\frac{-(\alpha_k+\alpha_{\min})+\left(\alpha_{\eta}+\alpha_{\gamma}\right)}{2}}T^{-1/2}\right) + O_p\left(n^{-(\alpha_k+\alpha_{\min})/2+\alpha_{\eta}}\right) + O_p\left(n^{-(\alpha_k+\alpha_{\min})/2+1/2}T^{-1}\right).$$

B.5 Proof of theorem 5

Using (81) and noting that $\hat{s}_{nT} \to_p \boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0$, then

$$\begin{aligned}\hat{\mathbf{V}}_{\xi,nT} - \mathbf{V}_{\xi} &= \left(1+\boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0\right)\left[n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT} - p\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\right)\right] + o_p(1) \\ &= \left(1+\boldsymbol{\lambda}_0'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}_0\right)\left[n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT} - n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\right] + o_p(1).\end{aligned} \quad \text{(B.46)}$$

Also using (B.4) we have

$$n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT} = n^{-1}\left(\mathbf{B}_n+\mathbf{U}_{nT}\mathbf{G}_n\right)'\mathbf{M}_n\left(\tilde{\mathbf{V}}_u - \mathbf{V}_u + \mathbf{V}_u\right)\mathbf{M}_n\left(\mathbf{B}_n+\mathbf{U}_{nT}\mathbf{G}_n\right),$$

which, after some algebra, yields

$$n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_{nT} - n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n = \sum_{j=1}^{7}\mathbf{A}_{j,nT},$$

where

$$\begin{aligned}&\mathbf{A}_{1,nT} = n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\tilde{\mathbf{V}}_u - \mathbf{V}_u\right)\mathbf{M}_n\mathbf{B}_n, \quad \mathbf{A}_{2,nT} = n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\left(\tilde{\mathbf{V}}_u - \mathbf{V}_u\right)\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n, \\ &\mathbf{A}_{3,nT} = n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n, \quad \mathbf{A}_{4,nT} = n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\left(\tilde{\mathbf{V}}_u - \mathbf{V}_u\right)\mathbf{M}_n\mathbf{B}_n, \\ &\mathbf{A}_{5,nT} = n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n, \quad \mathbf{A}_{6,nT} = n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\tilde{\mathbf{V}}_u - \mathbf{V}_u\right)\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n, \\ &\mathbf{A}_{7,nT} = n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n.\end{aligned}$$

Considering the above terms in turn we note that

$$\left\|\mathbf{A}_{1,nT}\right\| \le n^{-1}\left\|\mathbf{B}_n'\mathbf{M}_n\right\|^2\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\| = \lambda_{\max}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right)\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\|.$$

Also, under Assumption 2 $\lambda_{\max}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right) < C$ and using (78) we have $\|\mathbf{A}_{1,nT}\| = O_p\left(n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right)$. Similarly

$$\|\mathbf{A}_{2,nT}\| \leq n^{-1}\left\|\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\right\|^2\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\| = \lambda_{\max}\left(n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n\right)\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\|.$$

Then using (A.22) $n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n \to_p \bar{\sigma}_n^2\mathbf{G}_T'\mathbf{G}_T = T^{-1}\bar{\sigma}_n^2(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} = O(T^{-1})$, and it follows that $\|\mathbf{A}_{2,nT}\| = O_p\left(T^{-1}n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right)$. Turning to the third term

$$\|\mathbf{A}_{3,nT}\| \leq n^{-1}\left\|\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n\right\|\|\mathbf{V}_u\| = \lambda_{\max}\left(n^{-1}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n\right)\|\mathbf{V}_u\|.$$

Also, by Lemma A.2 $\|\mathbf{V}_u\| = O(n^{\alpha_\gamma})$, and hence $\|\mathbf{A}_{3,nT}\| = O_p\left(T^{-1}n^{\alpha_\gamma}\right)$, and $\|\mathbf{A}_{4,nT}\| \leq \left\|n^{-1/2}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\right\|\left\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\right\|\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\|$. Also, as shown above $\left\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\right\| = O_p(1)$, $\left\|n^{-1/2}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\right\| = O_p(T^{-1/2})$, then

$$\|\mathbf{A}_{4,nT}\| = O_p\left(T^{-1/2}n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right),$$

$$\|\mathbf{A}_{5,nT}\| \leq \left\|n^{-1/2}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\right\|\left\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\right\|\|\mathbf{V}_u\| = O_p\left(T^{-1/2}n^{\alpha_\gamma}\right),$$

$$\|\mathbf{A}_{6,nT}\| \leq \left\|n^{-1/2}\mathbf{G}_n'\mathbf{U}_{nT}'\mathbf{M}_n\right\|\left\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\right\|\left\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\right\| = O_p\left(T^{-1/2}n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right),$$

and $\|\mathbf{A}_{7,nT}\| \leq \left\|n^{-1/2}\mathbf{B}_n'\mathbf{M}_n\right\|\|\mathbf{V}_u\|\left\|n^{-1/2}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n\right\| = O_p\left(T^{-1/2}n^{\alpha_\gamma}\right)$. Overall,

$$\left\|n^{-1}\hat{\mathbf{B}}_n'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_n - n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\right\| = O_p\left(n^{\alpha_\gamma}\sqrt{\frac{\ln(n)}{T}}\right),$$

which if used in (B.46) establishes (80), as required.

Online Supplement A: Data Sources and Calibration of Monte Carlo Designs

To

Identifying and exploiting alpha in linear asset pricing models with strong, semi-strong, and latent factors

by

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

Ron P. Smith
Birkbeck, University of London

April 2024

A Introduction

This online supplement provides the details of data sources for the risk factors and the excess returns on securities used to calibrate the Monte Carlo designs and to carry out the empirical applications reported in Sections 4 and 6 of the main paper. Section B describes the data used to calibrate the Monte Carlo (MC) designs, and Section C provides the estimates that formed the basis of the calibration of the parameters of the MC designs. Section D provides evidence for the choice of 14 blocks used in calibration of error covariances in the Monte Carlo experiments. Section E describes the data for factors and excess returns used in the empirical applications, and Section F derives the relationship between pooled R^2 of return regressions and the factor strengths used in the discussion of the empirical results.

B Data used to calibrate the Monte Carlo designs

B.1 Factors

To calibrate the parameters of the three factor model used in the Monte Carlo experiments we used monthly Fama-French three factor data series over the long sample 1963m8-2021m12, downloaded from Kenneth French's webpage.[1] The factors are the market return minus the risk free rate, denoted by MKT, the value factor (high book to market minus low portfolios, HML) and the size factor (small minus big portfolios, SMB). The risk free rate is also downloaded from French's webpage. First order autoregressions, AR(1), were estimated for all the three factors using the full data set, 1963m8-2021m12. Then GARCH(1,1) models were then estimated on the residuals from the fitted AR(1) regressions.

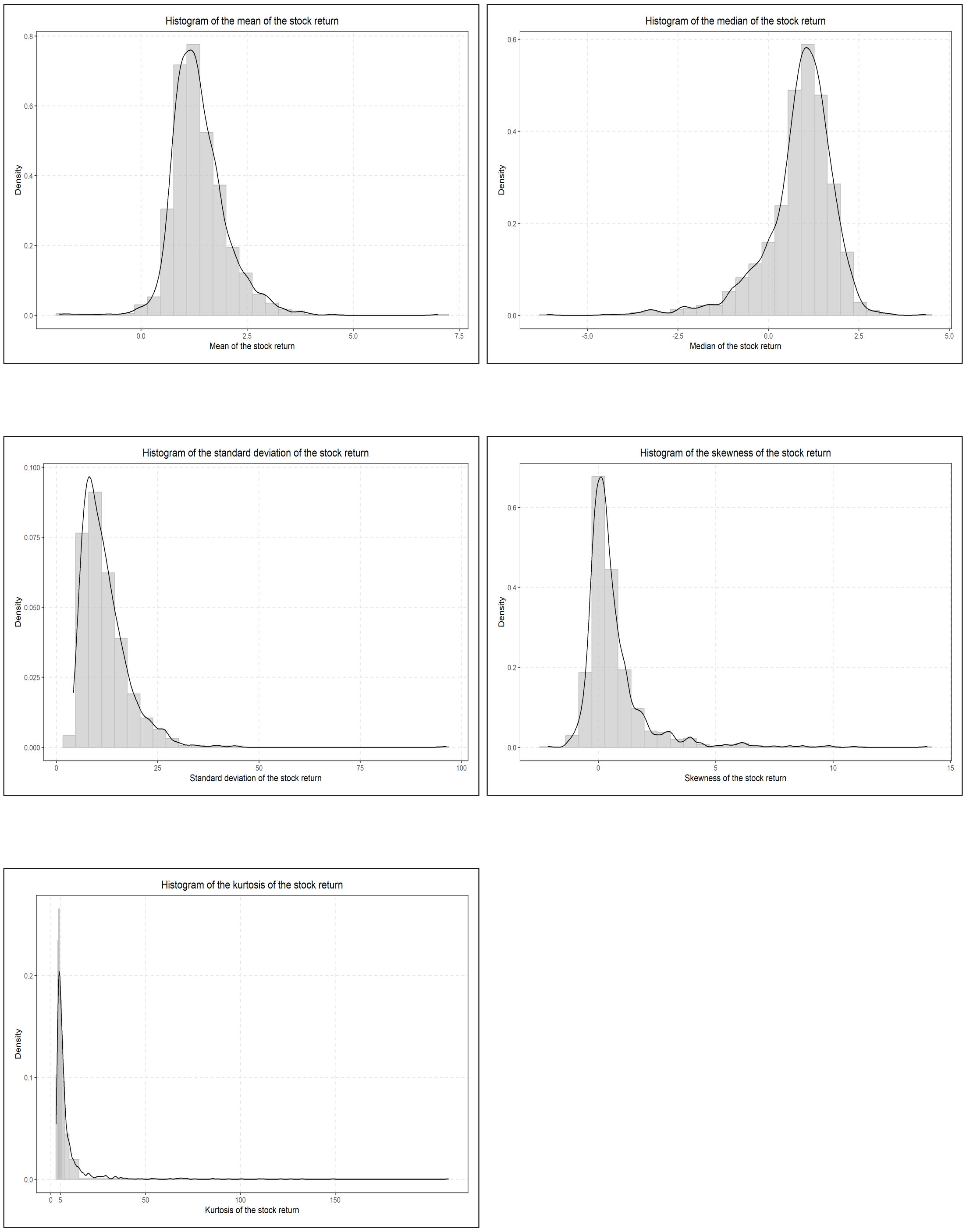
B.2 Excess returns

To calibrate the factor loadings and other parameters of the excess return regressions we used the shorter sample over the 20 years 2002m1 - 2021m12 ($T = 240$). Monthly returns (inclusive of dividend payment) over 2002m1 - 2021m12 ($T = 240$) for NYSE and NASDAQ stocks with share codes of 10 and 11 from CRSP were downloaded from Wharton Research Data Services and transformed to firm-specific excess returns using the risk free rate from French's webpage, and measured in percent, per month. Only stocks with data over the period 2002m1-2021m12 were used to arrive at a balanced panel with $T = 240$ monthly observations and a total number of $n = 1,289$, securities.

To avoid extreme outliers influencing the estimates we first computed mean, median, standard deviation, skewness and kurtosis for each security over 2002m1 - 2021m12. Table 1 reports the mean, standard deviations and interquartile range of these statistics over all the $1,289$ securities in our sample. The histograms of these summary statistics (mean, median, standard deviation, skewness and kurtosis) of the individual stock returns for the full sample over 2002m1 - 2021m12 are shown in Figure 1. As can be seen from these summary statistics, there are outlier security returns with very large standard deviations. This is clear from the cross firm standard deviation of 13.97 which is much larger than the kurtosis of 9.24 (See Table 1), resulting from a number of extreme outliers also seen from the long right tail of the histogram for the distribution of kurtosis across firms.

[1] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Figure 1: Histogram and density function of the individual stock returns ($n = 1,289$) over 2002m1-2021m12 ($T = 240$)



B.3 Stocks with the kurtosis less than or equal to 14 and less than 16

To reduce the influence of outlier returns on our results we considered dropping stock excess returns having kurtosis in excess of 14 and 16. Excluding stocks with kurtosis less than or equal to 14, resulted in a sample with $n = 1,148$ securities, whilst if we use the cut off point of 16 we ended up with $n = 1,175$ securities. The mean, median, standard deviation, skewness and kurtosis for each security over 2002m1 - 2021m12 of the whole and two sub-samples, $(1,289$, $1,148$ and $1,175)$. Then the average, standard deviation and interquartile range (IQR) for each summary statistic of the stocks from the two sub-sample are summarized respectively in Table 1.

Table 1: The average, standard deviation and interquartile range of the summary statistics of the individual stocks over 2002m1 - 2021m12 ($T = 240$)

	Average	Standard deviation	Interquartile range
Panel A: All stocks ($n = 1289$)			
stock.mean	1.3666	0.6593	0.7511
stock.median	0.8360	0.9970	0.9310
stock.standard deviation	11.9030	6.0082	6.5248
stock.skewness	0.6411	1.3766	0.9290
stock.kurtosis	9.2407	13.9786	4.0183
Panel B: Kurtosis ≤ 14 ($n = 1148$)			
stock.mean	1.3303	0.5898	0.7032
stock.median	0.9474	0.8816	0.8654
stock.standard deviation	10.8837	4.3944	5.6429
stock.skewness	0.2976	0.5996	0.7270
stock.kurtosis	5.9885	2.3932	2.7762
Panel B: Kurtosis ≤ 16 ($n = 1175$)			
stock.mean	1.3336	0.5922	0.7081
stock.median	0.9363	0.8878	0.8764
stock.standard deviation	10.9943	4.4695	5.6836
stock.skewness	0.3219	0.6285	0.7607
stock.kurtosis	6.1947	2.7222	2.9819

Histograms for mean, median, standard deviation, skewness and kurtosis of the individual stock returns for two sub-samples are shown in Figure 2 for the sub-sample with the kurtosis less than or equal to 14 over 2002m1 - 2021m12 ($n = 1,148$), and Figure 3 are for the sub-sample with the kurtosis less than 16 over 2002m1 - 2021m12 ($n = 1,175$).

Figure 2: Histogram and density function of the individual stock returns with kurtosis less than or equal to 14 over 2002m1-2021m12 ($T = 240$) and $n = 1,148$

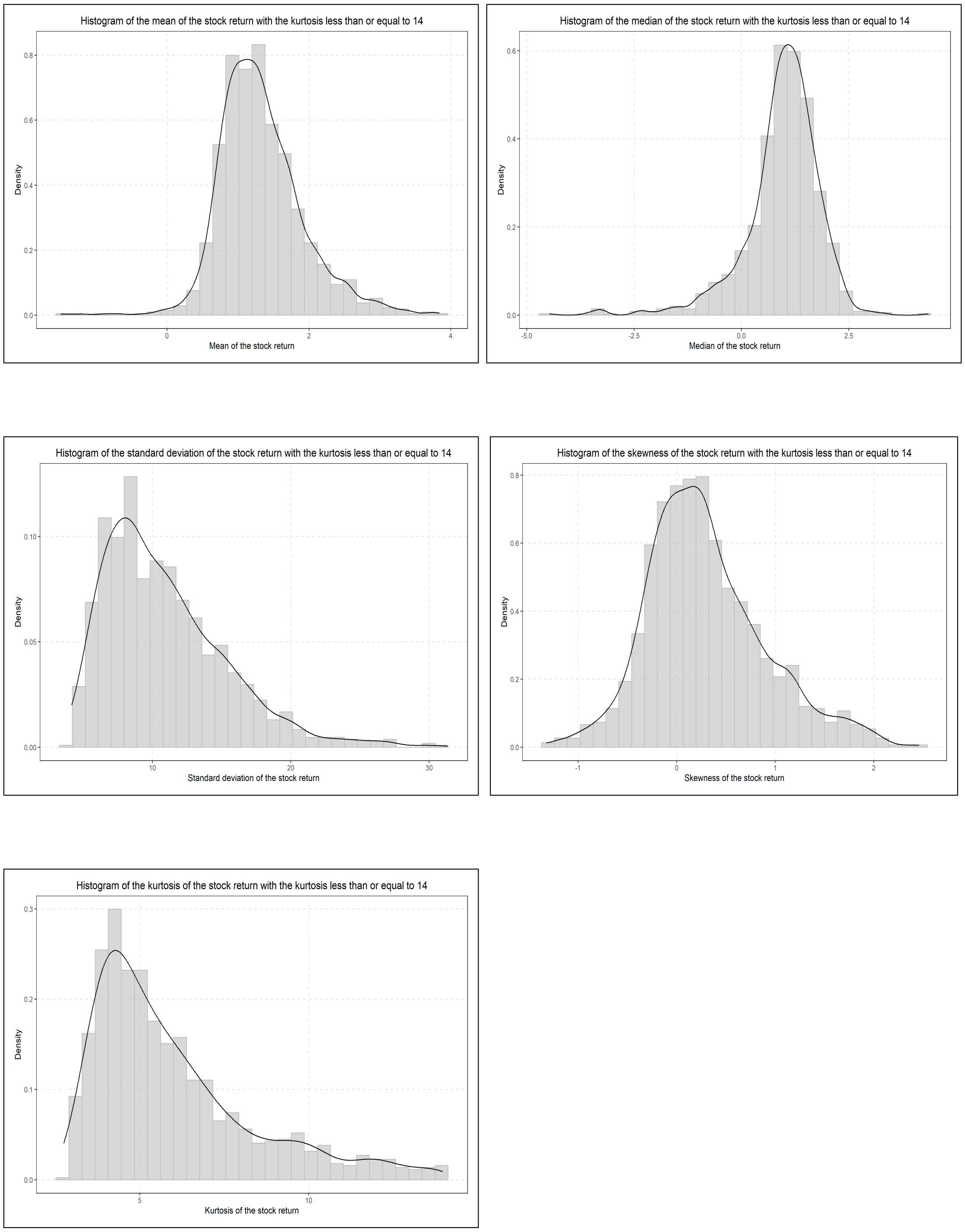
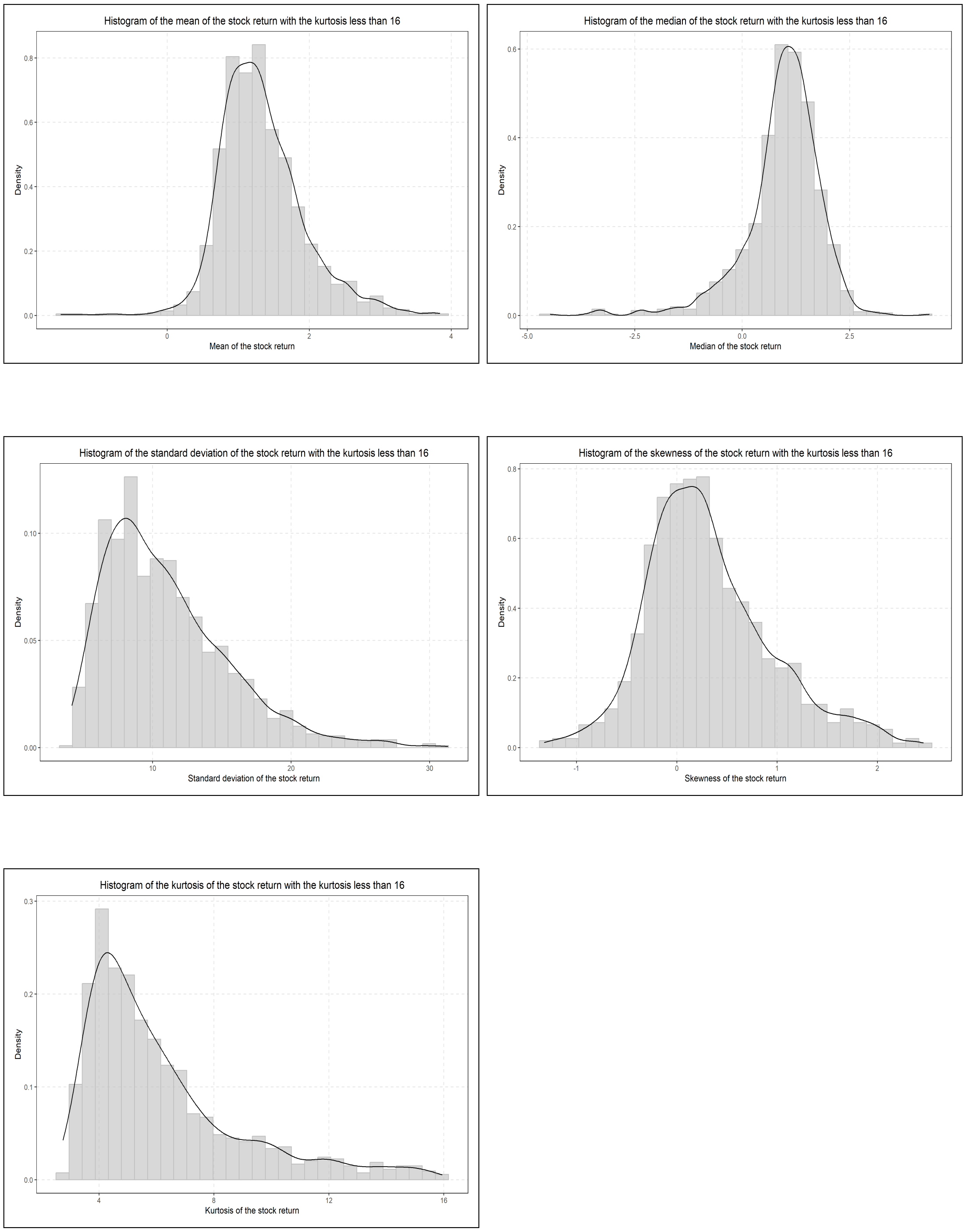


Figure 3: Histogram and density function of the individual stock returns with kurtosis less than or equal to 16 over 2002m1-2021m12 ($T = 240$) and $n = 1,148$



C The MC design and its calibration

C.1 Generation of returns

Excess returns, r_{it}, are generated as

$$r_{it} = \alpha_i + \sum_{k=M,H,S} \beta_{ik} f_{kt} + u_{it}, \tag{C.1}$$

$$= \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + u_{it},$$

for $i = 1, 2, .., n$; $t = 1, 2, ..., T$, with

$$\alpha_i = c + \boldsymbol{\beta}_i' \boldsymbol{\phi} + \eta_i, \tag{C.2}$$

where $\mathbf{f}_t = (f_{Mt}, f_{Ht}, f_{St})'$, and $\boldsymbol{\beta}_i = (\beta_{Mi}, \beta_{Hi}, \beta_{\text{Si}})'$, and $\boldsymbol{\phi} = (\phi_M, \phi_H, \phi_S)'$.

C.2 Generation of factors

Factors are generated as first-order autoregressive, AR(1), processes with GARCH(1,1) effects:

$$f_{kt} = \mu_k(1 - \rho_k) + \rho_k f_{k,t-1} + \left(1 - \rho_k^2\right)^{1/2} \sigma_{kt} \zeta_{kt}, \tag{C.3}$$

$$\sigma_{kt}^2 = (1 - b_k - c_k)\sigma_k^2 + b_k \sigma_{k,t-1}^2 + c_k \sigma_{k,t-1}^2 \zeta_{k,t-1}^2, \tag{C.4}$$

for $k = M, H, S$, starting from $t = -49, ...0, 1, 2, ..., T$, with $f_{k,-50} = 0$ (and $\sigma_{k,-50} = 0$ in the case where $c_k \neq 0$) to minimize the effects of the initial values on the sample $f_{kt}, t = 1, 2, ..., T$ used in the simulations.

The data generating process for the factors is calibrated using the full set of Fama-French three factor data set covering the period 1963m8-2021m12. The calibrated parameter values are $\boldsymbol{\mu} = (\mu_M, \mu_H, \mu_S)' = (0.59, 0.27, 0.23)'$, $\boldsymbol{\sigma} = (\sigma_M, \sigma_H, \sigma_S)' = (4.45, 2.86, 3.03)'$, and $\boldsymbol{\rho} = (\rho_M, \rho_H, \rho_S)' = (0.06, 0.17, 0.07)'$. Note that $Var(f_{kt}) = \sigma_k^2$. The parameters of (C.2) are also estimated using the bias-corrected procedure and are set as $c = 0.83$ and $\boldsymbol{\phi} = (-0.49, -0.35, 0.16)'$. To ensure that correlation across the three factors match the Fama-French data we generated $\boldsymbol{\zeta}_t = (\zeta_{Mt}, \zeta_{Ht}, \zeta_{St})'$ as $\boldsymbol{\zeta}_t = \mathbf{Q}_\zeta \boldsymbol{\omega}_t$, where $\mathbf{Q}_\zeta$ is the Cholesky factor of $\mathbf{R}_\zeta$, the correlation matrix of $\mathbf{f}_t$ given by

$$\mathbf{R}_\zeta = \begin{pmatrix} 1 & -0.21 & 0.28 \\ -0.21 & 1 & -0.02 \\ 0.27 & -0.02 & 1 \end{pmatrix}.$$

We consider both Gaussian and non-Gaussian errors and generate $\boldsymbol{\omega}_t$ as $IID(\mathbf{0}, \mathbf{I}_3)$, as well as a multivariate t with 5 degrees of freedom, namely $t(\mathbf{0}, \mathbf{I}_3, 5)$. The remaining parameters are set as $b_k = c_k = 0$ to generate homoskedastic errors, and $b_k = 0.8$ and $c_k = 0.1$ for $k = M, H, S$ to generate GARCH effects.

C.3 Estimation of factor models

Consider the AR(1) processes with a GARCH(1,1) errors

$$f_{kt} = \mu_k(1 - \rho_k) + \rho_k f_{k,t-1} + \left(1 - \rho_k^2\right)^{1/2} \sigma_{kt} \zeta_{kt}, \tag{C.5}$$

$$\sigma_{kt}^2 = (1 - b_k - c_k)\sigma_k^2 + b_k \sigma_{k,t-1}^2 + c_k \sigma_{k,t-1}^2 \zeta_{k,t-1}^2, \tag{C.6}$$

where f_{kt}, $k = M, H, S$ and $t = 1, ..., T$, denote the the values of three factors MKT, HML, SMB in month t, respectively. The estimates of GARCH parameters obtained using the sample 2002m1 - 2021m12 ($T = 240$) are summarized in Table 2.

Table 2: GARCH parameters for the models of three Fama-French factors for the sample over 2002m1 - 2021m12 ($T = 240$)

	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}$	$\hat{b}$	$\hat{c}$
MKT	0.8030	0.0711	4.5703	0.6781	0.2395
	(0.3035)	(0.0648)	$(\cdot)$	(0.0854)	(0.0627)
HML	–0.0805	0.1816	3.1513	0.7582	0.1987
	(0.2194)	(0.0639)	$(\cdot)$	(0.0986)	(0.0686)
SMB	0.1718	–0.0267	2.5815	0.8353	0.0680
	(0.1651)	(0.0649)	$(\cdot)$	(0.1845)	(0.0606)

The correlation matrix of three factors MKT, HML, SMB over 2001m1-2021m9 ($T = 240$) is

$$\begin{pmatrix} 1.00 & 0.20 & 0.35 \\ 0.20 & 1.00 & 0.35 \\ 0.35 & 0.35 & 1.00 \end{pmatrix}.$$

C.4 Factor loadings estimates

For each of the securities $i = 1, 2, ..., 1175$ (with kurtosis below 16), and $t = 1, 2, ..., 240$, OLS regressions excess returns y_{it} for security i was run on an intercept and the three FF factors

$$y_{it} = r_{it} - r_t^f = a_i + \sum_{k \in \{M,H,S\}} \beta_{ik} f_{kt} + u_{it},$$

where r_{it} is the return of i^{th} security at time t, inclusive of dividend (if any), and r_t^f is the risk free rate. The sample mean and standard deviation of the excess return for each individual stock, denoted as $\bar{y}_{iT}$ and s_{yiT} are computed as

$$\bar{y}_{iT} = T^{-1} \sum_{t=1}^{T} y_{it}, \tag{C.7}$$

$$sd_{iT}(y) = \sqrt{(T-1)^{-1} \sum_{t=1}^{T} (y_{it} - \bar{y}_{iT})^2}. \tag{C.8}$$

The estimates $\hat{a}_{i,T}$, $\hat{\beta}_{ik,T}$, $k = M, H, S$ are given by

$$\left(\hat{a}_{i,T}, \hat{\beta}_{iM,T}, \hat{\beta}_{iH,T}, \hat{\beta}_{iS,T}\right)' = (\boldsymbol{F}_0' \boldsymbol{F}_0)^{-1} \boldsymbol{F}_0' \boldsymbol{y}_{i\circ}, \tag{C.9}$$

where $\boldsymbol{F}_0 = (\boldsymbol{\tau}_T, \mathbf{F})$, $\mathbf{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_T)'$, $\boldsymbol{f}_t = (f_{Mt}, f_{Ht}, f_{St})'$ and $\mathbf{y}_{i\circ} = (y_{i1}, y_{i2}, ..., y_{iT})'$. The standard error of the i^{th} regression, denoted as s_{iT}, is given by

$$\hat{\sigma}_{iT}^2 = (T - K - 1)^{-1} \sum_{t=1}^{T} \hat{u}_{it}^2, \tag{C.10}$$

where $\hat{u}_{it} = y_{it} - \hat{a}_{iT} - \sum_{k\in\{M,H,S\}} \hat{\beta}_{ik,T} f_{kt}$. The coefficient of determination of the i^{th} regression, denoted by R^2_{iT}, is given by

$$R^2_{iT} = 1 - \frac{\sum_{t=1}^{T} \hat{u}_{it}^2}{\sum_{t=1}^{T} (y_{it} - \bar{y}_{iT})^2}. \tag{C.11}$$

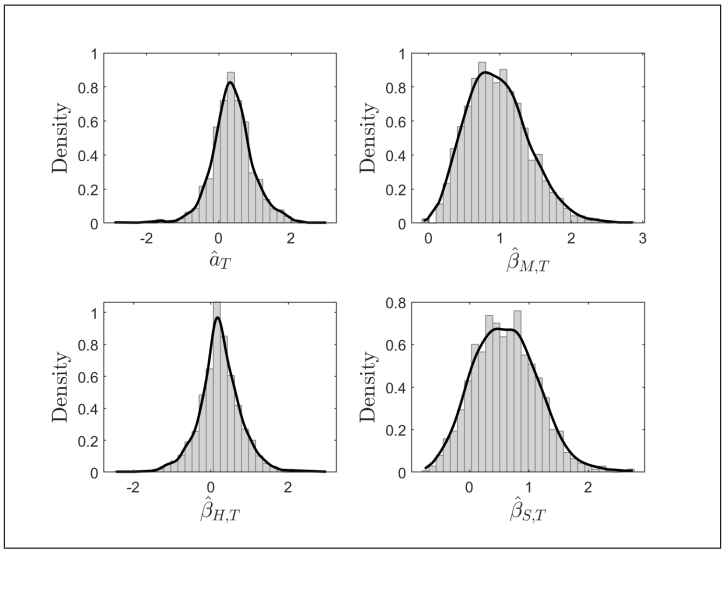
We compute the summary statistics: mean, median, standard deviation (S.D.), skewness, kurtosis, interquartile range, minimum, maximum for the sample mean and standard deviation of the excess returns, the estimates and the corresponding standard error, R-squared of the regressions over 2002m1 - 2021m12 ($T = 240$), for the $n = 1,175$ securities: $\bar{y}_{iT}$, $sd_{i,T}(y)$, $\hat{a}_{i,T}$, $\hat{\beta}_{iM,T}$, $\hat{\beta}_{iH,T}$, $\hat{\beta}_{iS,T}$, $\hat{\sigma}^2_{iT}$ and R^2_{iT} for $i = 1, 2, ..., 1175$, computed using (C.7)-(C.11). The results are summarized in Table 3.

Table 3: The summary statistics of the estimates, standard error and R-squared of the panel regression over 2002m1 - 2021m12 ($T = 240$) and $n = 1,175$

	mean	median	S.D.	skewness	kurtosis	IQR	min	max
$\bar{y}$	1.2366	1.1572	0.5922	0.6219	4.9697	0.7085	–1.5971	3.7434
$sd(y)$	11.0006	10.1115	4.4695	1.1241	4.4498	5.6826	4.1706	31.3322
$\hat{a}$	0.3749	0.3562	0.5810	–0.1980	5.5804	0.6493	–2.8882	2.9736
$\hat{\beta}_M$	0.9714	0.9362	0.4279	0.5352	3.3349	0.5810	–0.0689	2.8590
$\hat{\beta}_H$	0.2235	0.2093	0.5276	–0.0916	4.7150	0.5915	–2.4498	2.9788
$\hat{\beta}_S$	0.6061	0.5790	0.5381	0.3552	3.2197	0.7495	–0.7504	2.7474
$\hat{\sigma}$	9.4123	8.2297	4.2794	1.3292	5.0599	5.3811	3.4523	30.6953
R^2	0.2840	0.2782	0.1383	0.2172	2.2941	0.2131	0.0050	0.6814

The histogram for the $\hat{a}_T$ and the $\hat{\beta}_{k,T}$ for $k = M, H, S$ over 2002m1 - 2021m12 ($T = 240$), each using 1175 data points $\hat{a}_{i,T}$ and the $\hat{\beta}_{ik,T}$ for $k = M, H, S$, $i = 1, 2, ..., 1175$, is shown in the Figure 4.

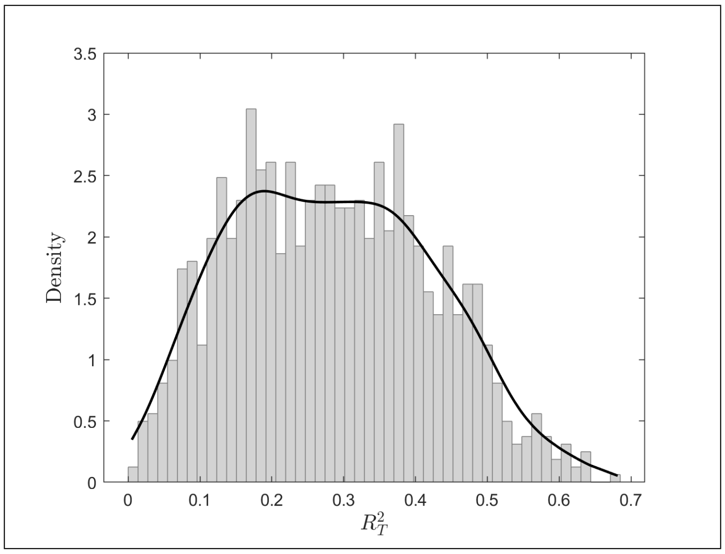
Figure 4: Histogram and density function of the coefficients of the panel regression over 2002m1-2021m12 ($T = 240$) and $n = 1,775$



The histogram for the R^2_T over 2002m1 - 2021m12 ($T = 240$), using 1,175 data points R^2_{iT}

for $i = 1, 2, ..., 1175$, is shown in the Figure 5.

Figure 5: Histogram and density function of the R-squared of the panel regression over 2002m1-2021m12 ($T = 240$) and $n = 1,775$



C.5 Calibrating the fit of return regressions

To see how κ controls the regression fit, note that the n return processes (C.1) can be written more compactly in vector form as

$$\mathbf{r}_t = \mathbf{a} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t,$$

where $\mathbf{u}_t = \boldsymbol{\gamma} g_t + \kappa\hat{\mathbf{S}}\boldsymbol{\varepsilon}_t$, $\mathbf{a} = (\alpha_1, \alpha_2, ..., \alpha_n)'$, with a_i given by $a_i = c + \boldsymbol{\beta}_i'\boldsymbol{\phi} + \eta_i$, and

$$\hat{\mathbf{S}} = Diag(\hat{\mathbf{S}}_b,\ b = 1, 2, ..., B).$$

Overall, the DGP for the return regressions can be written compactly as

$$\mathbf{r}_t = c\boldsymbol{\tau}_n + \mathbf{B}\left(\mathbf{f}_t + \boldsymbol{\phi}\right) + \boldsymbol{\gamma} g_t + \kappa\hat{\mathbf{S}}\boldsymbol{\varepsilon}_t + \boldsymbol{\eta}_n,$$

where $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$. We abstract from pricing errors and weak latent factor and $\eta_i = 0$, $g_t = 0$, and set κ such that the pooled R^2 (PR^2) of return regressions can be controlled to be around $R_0^2 = 0.30$. We have

$$PR_{nT}^2 = 1 - \frac{n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} E\left(u_{it}^2\right)}{n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} Var(r_{it})}.$$

$$Var(r_{it}) = Var\left(\boldsymbol{\beta}_i'\mathbf{f}_t\right) + Var(u_{it}) = E\left(\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_f\boldsymbol{\beta}_i\right) + E(u_{it}^2)$$

$$= \text{Tr}\left[\boldsymbol{\Sigma}_f E\left(\boldsymbol{\beta}_i\boldsymbol{\beta}_i'\right)\right] + E(u_{it}^2).$$

Denote the k^{th} element of $\boldsymbol{\beta}_i$ by β_{ik}, then if $\beta_{ik} \sim IIDN(\mu_{\beta_k}, \sigma_{\beta_k}^2)$, and β_{ik} are distributed independently over $k = 1, 2, ..., K$, we have

$$E\left(\boldsymbol{\beta}_i\boldsymbol{\beta}_i'\right) = Diag\left(\mu_{\beta_k}^2 + \sigma_{\beta_k}^2, \text{ for } k = 1, 2, ..., K\right).$$

Then,

$$Var(r_{it}) = Tr\left[\mathbf{\Sigma}_f E\left(\boldsymbol{\beta}_i\boldsymbol{\beta}_i'\right)\right] + E(u_{it}^2) = \sum_{k=1}^{K}\sigma_{fk}^2\left(\mu_{\beta_k}^2 + \sigma_{\beta_k}^2\right) + E(u_{it}^2).$$

Also,

$$\sum_{i=1}^{n} E(u_{it}^2) = Tr\left[E\left(\mathbf{u}_t\mathbf{u}_t'\right)\right],$$

and (when $g_t = 0$) we have

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{V}_u = \kappa^2\hat{\mathbf{S}}E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\hat{\mathbf{S}}' = \kappa^2\hat{\mathbf{S}}\mathbf{V}_\varepsilon\hat{\mathbf{S}}',$$

where $\mathbf{V}_\varepsilon = Diag(\mathbf{V}_{b\varepsilon}^{(r)}, b = 1, 2, ..., B)'$. Hence,

$$n^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} E\left(u_{it}^2\right) = n^{-1}Tr(\mathbf{V}_u) = \kappa^2 n^{-1}Tr\left(\hat{\mathbf{S}}\mathbf{V}_\varepsilon\hat{\mathbf{S}}'\right)$$

and

$$PR_{nT}^2 = 1 - \frac{\kappa^2 n^{-1}Tr\left(\hat{\mathbf{S}}\mathbf{V}_\varepsilon\hat{\mathbf{S}}'\right)}{\sum_{k=1}^{K}\sigma_{fk}^2\left(\mu_{\beta_k}^2 + \sigma_{\beta_k}^2\right) + \kappa^2 n^{-1}Tr\left(\hat{\mathbf{S}}\mathbf{V}_\varepsilon\hat{\mathbf{S}}'\right)}.$$

To achieve $lim_{n\to\infty}PR_{nT}^2 = R_0^2$, we need to set (assuming all K factors are strong)

$$\kappa^2 = \frac{\sum_{k=1}^{K}\sigma_{fk}^2\left(\mu_{\beta_k}^2 + \sigma_{\beta_k}^2\right)}{n^{-1}Tr\left(\hat{\mathbf{S}}\mathbf{V}_\varepsilon\hat{\mathbf{S}}'\right)}\left(\frac{1 - R_0^2}{R_0^2}\right). \quad \text{(C.12)}$$

When there are no idiosyncratic error dependence, namely when $\hat{\mathbf{S}} = \mathbf{I}_n$, the above expression simplifies to

$$\kappa^2 = \frac{\sum_{k=1}^{K}\sigma_{fk}^2\left(\mu_{\beta_k}^2 + \sigma_{\beta_k}^2\right)}{n^{-1}Tr\left(\mathbf{V}_\varepsilon\right)}\left(\frac{1 - R_0^2}{R_0^2}\right). \quad \text{(C.13)}$$

If we only use the market factor, we have

$$\kappa^2 = \frac{\sigma_M^2\left(\mu_{\beta_M}^2 + \sigma_{\beta_M}^2\right)}{n^{-1}Tr\left(\mathbf{V}_\varepsilon\right)}\left(\frac{1 - R_0^2}{R_0^2}\right). \quad \text{(C.14)}$$

We expect that $n^{-1}Tr\left(\mathbf{V}_\varepsilon\right) \to 1$, if $E(\sigma_{ii}) = 1$, as under our DGP.

C.6 Estimation of FF factor strengths

Denote by $t_{ik,T} = \hat{\beta}_{ik,T}\,/\,\text{s.e.}\left(\hat{\beta}_{ik,T}\right)$, the t-statistic corresponding to β_{ik}. The total number of factor loadings of factor k, that are statistically significant over $i = 1, 2, \ldots, n$, $n = 1,175$, is:

$$\hat{D}_{nT,k} = \sum_{i=1}^{n}\hat{d}_{ik,nT} = \sum_{i=1}^{n}\mathbf{1}\left[|t_{ik,T}| > c_p(n)\right],$$

where $\mathbf{1}(A) = 1$ if $A > 0$, and zero otherwise, and the critical value function that allows for the multiple testing nature of the problem, $c_p(n, \delta)$, is given by

$$c_p(n, \delta) = \Phi^{-1}\left(1 - \frac{p}{2n^\delta}\right), \quad \text{(C.15)}$$

where p is the nominal size, set, following Bailey, Kapetanios, and Pesaran (2021, BKP), as $p = 0.1$, $\delta > 0$ is the critical value exponent, set $\delta = 0.25$, and $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal distribution. Let $\hat{\pi}_{nT,k}$ be the fraction of significant loadings of factor k, and note that $\hat{\pi}_{nT,k} = \hat{D}_{nT,k}/n$. The strength of factor k, denoted by α_{k0}, for $k = 1, 2, \ldots, K$, $K = 3$, is estimated by

$$\hat{\alpha}_k = \begin{cases} 1 + \frac{\ln \hat{\pi}_{nT,k}}{\ln n}, \text{ if } \hat{\pi}_{nT,k} > 0, \\ 0, \text{ if } \hat{\pi}_{nT,k} = 0. \end{cases} \tag{C.16}$$

The variance of the estimated strength of factor k is given by

$$Var(\hat{\alpha}_k) = (\ln n)^{-2} \psi_n(\alpha_{k0}),$$

where

$$\psi_n(\alpha_{k0}) = p(n - n^{\alpha_{k0}}) n^{-\delta - 2\alpha_{k0}} \left(1 - \frac{p}{n^\delta}\right).$$

So the standard error of the estimated strength of factor k can be computed by:

$$s.e.(\hat{\alpha}_k) = \frac{\sqrt{\psi_n(\hat{\alpha}_k)}}{\ln n}. \tag{C.17}$$

Mean and variance of the loadings associated with factor k are given by

$$\hat{\mu}_{\beta_{kT}}(\hat{\alpha}_k) = \frac{\sum_{i=1}^n \mathbf{1}\left[|t_{ikT}| > c_p(n)\right] \hat{\beta}_{ikT}}{\sum_{i=1}^n \mathbf{1}\left[|t_{ikT}| > c_p(n)\right]}, \tag{C.18}$$

$$\hat{\sigma}^2_{\beta_k}(\hat{\alpha}_k) = \frac{\sum_{i=1}^n \mathbf{1}\left[|t_{ikT}| > c_p(n)\right] \left(\hat{\beta}_{ikT} - \hat{\mu}_{\beta_{kT}}(\hat{\alpha}_k)\right)^2}{\sum_{i=1}^n \mathbf{1}\left[|t_{ikT}| > c_p(n)\right]}. \tag{C.19}$$

In the case where a factor is strong, namely $\alpha_k = 1$, then it must be that $\mathbf{1}\left[|t_{ikT}| > c_p(n)\right] = 1$ for all i. The estimated factor strengths $\hat{\alpha}_k$ and corresponding standard errors for $k = M, H, S$, using the sample over 2002m1 - 2021m12 ($T = 240$) and $n = 1,175$ are reported in 4.

Table 4: Strength of three FF factors estimated over 2001m1-2021m9 ($T = 240$ and $n = 1,175$)

	M	H	S
$\hat{\alpha}$	0.9941	0.8373	0.9023
	(0.0001)	(0.0014)	(0.0008)

Note: This table reports the estimates of the factor strength using (C.16) and the standard errors that are reported in () using (C.17), for three factors MKT, HML and SMB, using the sample over 2001m1-2021m9 ($T = 240$) and $n = 1,175$, $K = 3$.

C.7 Estimates of ϕ for the FF3 factors

Table 5: The bias-corrected estimates of c_0 (intercept) and ϕ_M, ϕ_H, and ϕ_S for the sample 2002m1 - 2021m12 ($T = 240$) with $n = 1175$ securities

$\hat{c}_0$	$\hat{\phi}_M$	$\hat{\phi}_H$	$\hat{\phi}_S$
0.8423	–0.4808	–0.3259	0.1195
	(0.0971)	(0.0780)	(0.0846)

C.8 Rolling estimates used in portfolio construction

The rolling estimates for month t are computed using analogous expressions to those provided in sub-sections 2.3 and 3.2 of the main paper with a rolling window size of $T = 240$. For ease of replication, the algorithms used to estimate the rolling estimates for $t = 2015m12, ..., 2022m11$ are set out below:

$$\hat{\beta}_{it|T} = \left[\sum_{\tau=t-T+1}^{t} \left(\mathbf{f}_\tau - \bar{\mathbf{f}}_{t|T}\right)\left(\mathbf{f}_\tau - \bar{\mathbf{f}}_{t|T}\right)' \right]^{-1} \sum_{\tau=t-T+1}^{t} \left(\mathbf{f}_\tau - \bar{\mathbf{f}}_{t|T}\right) r_{i\tau},$$

$$\bar{\mathbf{f}}_{t|T} = T^{-1} \sum_{\tau=t-T+1}^{t} \mathbf{f}_\tau,$$

$$\tilde{\boldsymbol{\phi}}_{t|T} = \hat{\mathbf{H}}_{t|T}^{-1} \left[\frac{\hat{\mathbf{B}}_{t|T}' \mathbf{M}_n \hat{\alpha}_{t|T}}{n} + T^{-1} \widehat{\bar{\sigma}}_{t|T}^2 \left(\frac{\mathbf{F}_{t|T}' \mathbf{M}_T \mathbf{F}_{t|T}}{T} \right)^{-1} \bar{\mathbf{f}}_{t|T} \right],$$

$$\hat{\mathbf{B}}_{t|T} = \left(\hat{\beta}_{1t|T}, \hat{\beta}_{2t|T} ...., \hat{\beta}_{nt|T}\right)',$$

$$\hat{\alpha}_{t|T} = \bar{\mathbf{r}}_{t|T} - \hat{\mathbf{B}}_{t|T} \bar{\mathbf{f}}_{t|T},$$

$$\hat{\mathbf{H}}_{t|T} = \frac{\hat{\mathbf{B}}_{t|T}' \mathbf{M}_n \hat{\mathbf{B}}_{t|T}}{n} - T^{-1} \widehat{\bar{\sigma}}_{t|T}^2 \left(\frac{\mathbf{F}_{t|T}' \mathbf{M}_T \mathbf{F}_{t|T}}{T} \right)^{-1},$$

$$\widehat{\bar{\sigma}}_{t|T}^2 = \frac{\sum_{t=\tau-T+1}^{t} \sum_{i=1}^{n} \hat{u}_{i,\tau|T}^2}{n(T-K-1)},$$

$$\hat{u}_{i,\tau|T} = r_{i\tau} - \hat{\alpha}_{i,\tau|T} - \hat{\beta}_{i\tau|T}' \mathbf{f}_\tau \text{ , for } \tau = t, t-1, ..., t-T+1,$$

$$\bar{\mathbf{r}}_{t|T} = \left(\bar{r}_{1,t|T}, \bar{r}_{2,t|T}, ..., \bar{r}_{nt|T}\right)', \ \bar{r}_{i,t|T} = T^{-1} \sum_{\tau=t-T+1}^{t} r_{i\tau},$$

$$\widehat{Var\left(\tilde{\boldsymbol{\phi}}_{t|T}\right)} = T^{-1} n^{-1} \mathbf{H}_{t|T}^{-1} \ \hat{\mathbf{V}}_{\xi,t|T} \mathbf{H}_{t|T}^{-1} \ ,$$

,

$$\hat{\mathbf{V}}_{\xi,t|T} = \left(1 + \hat{s}_{t|T}\right) \left(n^{-1} \hat{\mathbf{B}}_{t|T}' \mathbf{M}_n \tilde{\mathbf{V}}_{u,t|T} \mathbf{M}_n \hat{\mathbf{B}}_{t|T} \right),$$

$$\hat{s}_{t|T} = \tilde{\boldsymbol{\lambda}}_{t|T}' \left(\frac{\mathbf{F}_{t|T}' \mathbf{M}_T \mathbf{F}_{t|T}}{T} \right)^{-1} \tilde{\boldsymbol{\lambda}}_{t|T},$$

$$\tilde{\boldsymbol{\lambda}}_{t|T} = \tilde{\boldsymbol{\phi}}_{t|T} + \bar{\mathbf{f}}_{t|T}\,,$$

$$c_{t|T} = \left(\boldsymbol{\tau}'_n\boldsymbol{\tau}_n\right)^{-1}\boldsymbol{\tau}'_n\hat{\alpha}_{t|T} - \left(\boldsymbol{\tau}'_n\boldsymbol{\tau}_n\right)^{-1}\boldsymbol{\tau}'_n\hat{\mathbf{B}}_{t|T}\,\tilde{\boldsymbol{\phi}}_{t|T}\,,$$

where $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}'_T$, $\boldsymbol{\tau}_T$ is a T-dimensional vector of ones, $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}'_n$, and $\boldsymbol{\tau}_n$ is an n-dimensional vector of ones. Finally,

$$\tilde{\mathbf{V}}_{u,t|T} = \left(\tilde{\sigma}_{ij,t|T}\right),$$

$$\tilde{\sigma}_{ij,t|T} = \hat{\sigma}_{ii,t|T}$$

$$\tilde{\sigma}_{ij,t|T} = \hat{\sigma}_{ij,t|T}\,\mathbf{1}\left[\left|\hat{\rho}_{ij,t|T}\right| > T^{-1/2}c_{\alpha}(n,\delta)\right],\ i = 1,2,\ldots,n-1,\ j = i+1,\ldots,n,$$

where

$$\hat{\sigma}_{ij,t|T} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{i,\tau|T}\,\hat{u}_{j,\tau|T}\,,\ \hat{\rho}_{ij,t|T} = \frac{\hat{\sigma}_{ij,t|T}}{\sqrt{\hat{\sigma}_{ii,t|T}\,\hat{\sigma}_{ii,t|T}}},$$

$\hat{u}_{i,\tau|T} = r_{i\tau} - \hat{\alpha}_{i,\tau|T} - \hat{\beta}'_{i\tau|T}\mathbf{f}_{\tau}$, and $c_p(n,d) = \Phi^{-1}\left(1 - \frac{p}{2n^d}\right)$, is a normal critical value function, p is the the nominal size of testing of $\sigma_{ij} = 0$, $(i \neq j)$ and $d = 2$ is chosen to take account of the $n(n-1)/2$ multiple tests being carried out.

D Grouping of securities by their pair-wise correlations

The $T = 240$ sample ending in 2021 was used to estimate the pair-wise correlations of the residuals from the $n = 1,168$ returns regressions using the Fama-French three factors. Then all the statistically insignificant correlations were set to zero. Significance was determined allowing for the multiple testing nature of the tests, using the critical value $c_p(n,\delta) = \Phi^{-1}\left(1 - \frac{p}{2n^{\delta}}\right)$, with $p = 0.05$ and $\delta = 2$, since the number of pair-wise correlations is of order $O(n^2)$. See also Bailey, Pesaran, and Smith (2019). For the majority of securities (668 out of the 1,168), the pair-wise return correlations were not statistically significant. The securities with a relatively large number of non-zero correlations were either in the banking or energy related industries.

Initially, the securities were grouped using the two digit codes from the 1987 standard industrial classification (SIC 1987). But this gave too many groups, 62. Many of the groups had very few members: only one security for 3 out of the 62 groups and less than 10 for 36 groups. However, code 60 (banking) had 145 securities. Therefore, it was decided to work with the industrial classification based on a one digit level, and to aggregate the codes with a small number of securities, taking out two digit codes where there were large numbers in that code. We ended up with 14 contiguous groups ranging in size from 33 to 145. Average correlations were low overall, but the average absolute correlation of securities within the groups was around 10 times that with firms outside the group. See Table 6, which gives averages of pair-wise correlations without thresholding. These estimates suggested that a block diagonal structure with 14 blocks was a reasonable characterization which is used in the Monte Carlo analysis. See Section 4 of the main paper.

Table 6: Sector groupings by SIC codes and within and outside sector average pair-wise correlations.

	SIC codes	Number of stocks	Average correlations within	Average correlations outside
Agriculture & mining	0-17	51	0.0848	–0.0043
Food processing etc	20-27	82	0.0382	0.0074
Chemicals & refining	28-29	78	0.0234	0.0027
Metals	30-34	65	0.0439	0.0070
Machinery & equipment	35	72	0.0380	0.0037
Electrical Equipment	36	77	0.0601	0.0001
Transport equipment	37	33	0.0800	0.0105
Misc. manufacturing	38-39	78	0.0155	0.0027
Transport etc.	40-49	108	0.0810	0.0035
Wholesale & retail trade	50-59	122	0.0357	0.0034
Banking	60	145	0.1240	–0.0037
Other finance	61-67	98	0.0273	0.0061
Commercial Services	70-79	114	0.0149	0.0031
Professional Services	80-89	45	0.0172	0.0023
Total		1168		

Note: This table gives the average correlations within and between 14 groups selected based on one and two digits SIC codes. It shows the number of stocks in each sector, the average pair-wise correlation of stock returns within the sector as well as the average pair-wise correlations of returns of stocks in a given sector with those outside the sector.

E Data used in the empirical application

E.1 Security excess returns

Monthly returns (inclusive of dividends) for NYSE and NASDAQ stocks from CRSP with codes 10 and 11 were downloaded on July 2 2022 from Wharton Research Data Services. They were converted to excess returns by subtracting the risk free rate, which was taken from Kenneth French's data base. To obtain balanced panels of stock returns and factors only variables where there was data for the full sample under consideration were used. Excess returns are measured in percent per month. To avoid outliers influencing the results, stocks with a kurtosis greater than 16 were excluded.

Four main samples were considered, each had 20 years of data, $T = 240$, ending in 2015m12, 2017m12, 2019m12, 2021m12. Thus the earliest observation used is for 1996m1. Sub-samples of the main samples of size $T = 120$ and $T = 60$ ending at the same dates were also examined. Table 7 gives averages of the summary statistics across the individual stocks for the various samples. These are very similar over these four periods. Mean returns were high and substantially greater than the median reflecting the skewness of returns, which was slightly less in the last period. Filtering out the stocks with very high kurtosis removed about 100 of the roughly 1,200 stocks in each period and reduced mean return, standard deviation, skewness as well as kurtosis.

The 5 ($T = 60$) and 10 ($T = 120$) year sub-samples of the main samples, ending at the same dates, showed very similar patterns. Because of a requirement for a balanced panel, the shorter the sample the more stocks will be eligible for inclusion. Compared with around 1,200 in the 20 year sample there were around 2,000 in the 10 year sub-sample and around 2,500 in the 5 year sub-sample. Again filtering by kurtosis reduced the number of stocks by about 100. There is more variation in means and medians in the shorter sub-samples and the shorter the sample the lower the average kurtosis is.

Table 7: Summary statistics for monthly returns in percent for NYSE and NASDAQ stocks code 10 and 11 for 20 year ($T = 240$), 10 year ($T = 120$) and 5 year ($T = 60$) samples ending at end of specified year

	All stocks				Stocks with kurtosis < 16			
End date	2015	2017	2019	2021	2015	2017	2019	2021
Panel A: 20 year period, $T = 240$								
Mean	1.38	1.35	1.30	1.37	1.33	1.28	1.25	1.33
Median	0.74	0.70	0.76	0.84	0.80	0.78	0.85	0.94
S.D.	12.44	12.61	12.22	11.90	11.60	11.64	11.31	10.99
Skewness	0.69	0.74	0.72	0.64	0.42	0.44	0.40	0.32
Kurtosis	8.68	9.08	9.22	9.24	6.10	6.29	6.29	6.19
n	1181	1243	1276	1289	1090	1132	1143	1175
Panel B: 10 year period, $T = 120$								
Mean	1.00	1.22	1.24	1.55	0.98	1.17	1.23	1.49
Median	0.52	0.70	0.78	0.89	0.59	0.77	0.84	1.01
S.D.	12.36	12.68	10.70	11.83	11.67	11.80	10.27	10.70
Skewness	0.48	0.50	0.44	0.50	0.29	0.30	0.35	0.28
Kurtosis	6.93	7.08	5.32	7.02	5.57	5.62	4.64	5.33
n	2045	2024	1925	1871	1929	1907	1873	1766
Panel C: 5 year period, $T = 60$								
Mean	0.98	1.39	0.82	1.60	0.96	1.35	0.80	1.47
Median	0.42	0.77	0.20	0.43	0.49	0.82	0.29	0.61
S.D.	10.89	10.62	11.88	14.91	10.41	10.15	11.28	13.28
Skewness	0.44	0.46	0.38	0.48	0.35	0.38	0.29	0.28
Kurtosis	4.84	4.74	4.68	6.21	4.34	4.29	4.20	5.04
n	2600	2425	2497	2512	2541	2373	2439	2388

Note: This table shows the average values of the summary statistics of individual stocks described in Section 3, and the number of stocks (n), for each sample.

E.2 Risk factors

For the empirical applications we combined the 5 Fama-French factors with the 207 factors of Chen and Zimmermann (2022), both downloaded on 6 July 2022. The available risk factors at the end of each of the four 240 months samples ending in 2015, 2017, 2019 and 2021 were then screened and factors whose correlations (in absolute value) with the market factor were larger than 0.70 were dropped. The basic idea was to remove factors that were closely correlated with market factor. However, the application of this filter only reduced the number of factors in the active set by around 9-11. See Table 4 of the main paper. The summary statistics for the factors in the active set are summarized in Table E.2. It reports mean, median, pair-wise correlation, standard deviation (S.D.), skewness and kurtosis of the statistics indicated in the sub-headings of the tables for the K factors included in the active set for samples of size $T = 240$

months ending in December of 2015, 2017, 2019 and 2021. The summary statistics reported for the "Mean" on the left panel of Table E.2 are based on the time series means of the individual factors, those under "Median" are based on the time series medians of the individual factors, and so on.

Table 8: Summary statistics for mean, median, standard deviation, pairwise correlations, skewness and kurtosis of factors for $T = 240$ samples at the end of specified years (2015, 2017, 2019 and 2021)

End date	2015	2017	2019	2021	2015	2017	2019	2021
m(# Factors)	190	191	190	178	190	191	190	178
	Mean				Median			
Mean	0.51	0.47	0.41	0.32	0.42	0.37	0.31	0.28
Median	0.45	0.39	0.33	0.26	0.31	0.28	0.22	0.20
S.D.	0.40	0.40	0.38	0.30	0.53	0.50	0.50	0.44
Skewness	1.88	2.50	2.28	0.76	1.59	1.89	1.58	1.38
Kurtosis	10.54	16.35	14.42	3.82	8.99	10.37	9.16	6.68
	Standard deviation				Pair-wise correlation			
Mean	3.97	3.90	3.78	3.33	0.22	0.22	0.22	0.21
Median	3.51	3.47	3.37	2.99	0.18	0.17	0.17	0.16
S.D.	2.21	2.16	2.12	1.68	0.17	0.17	0.17	0.17
Skewness	1.15	1.23	1.29	1.05	0.86	0.88	0.88	0.92
Kurtosis	4.25	4.69	4.96	4.17	2.88	2.93	2.93	2.98
	Skewness				Kurtosis			
Mean	0.14	0.21	0.23	0.04	8.51	8.79	9.42	7.03
Median	0.13	0.18	0.14	0.06	6.44	6.66	7.09	5.86
S.D.	1.15	1.21	1.27	1.01	5.76	6.55	7.12	5.28
Skewness	0.15	0.55	0.61	-1.03	2.34	3.18	2.96	5.01
Kurtosis	3.98	5.28	5.24	7.82	9.65	17.71	15.70	37.21

Note: The T=240 sample was used to select factors and only factors where the absolute correlation coefficient with the market factor is less than 0.70 are included.

F Pooled R squared and factor strengths

Lemma F.1 *Consider the factor model*

$$r_{it} = a_i + \sum_{k=1}^{K} \beta_{ik} f_{kt} + u_{it} = a_i + \boldsymbol{\beta}_i' \mathbf{f}_t + u_{it}, \text{ for } i = 1, 2, ..., n; \ t = 1, 2, ..., T, \tag{F.1}$$

and consider the following adjusted pooled measure of fit

$$\overline{PR}^2 = 1 - \frac{\widehat{\bar{\sigma}}^2_{nT}}{(nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (r_{it} - \bar{r}_{i\circ})^2}, \tag{F.2}$$

where $\widehat{\bar{\sigma}}^2_{nT}$ *is the bias-corrected estimator of* $(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T E\left(u_{it}^2\right) = n\sum_{i=1}^n \sigma_i^2 = \bar{\sigma}_n^2$*, given by (59)* $\bar{r}_{i\circ} = T^{-1}\sum_{t=1}^T r_{it}$*, and* $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^T u_{it}$*. Then under Under Assumptions 4, 1 and 2 we have*

$$\overline{PR}^2_{nT} = \sum_{k=1}^K \ominus\left(n^{\alpha_k - 1}\right) + O_p\left(T^{-1/2} n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right), \tag{F.3}$$

where α_k *is the strength of factor* f_{tk}*,* $\alpha_{max} = max_k(\alpha_k)$*, and* α_γ *is the strength of the missing factor.*

Proof. Using (61) we first recall that

$$\widehat{\bar{\sigma}}^2_{nT} - \bar{\sigma}^2_n = O_p\left(T^{-1/2} n^{-1/2}\right). \tag{F.4}$$

Now averaging (F.1) over t and forming deviations of r_{it} from its time average, $\bar{r}_{i\circ}$, we have (note that $\hat{\boldsymbol{\mu}}_T = T^{-1}\sum_{t=1}^T \mathbf{f}_t$)

$$r_{it} - \bar{r}_{i\circ} = u_{it} - \bar{u}_{i\circ} + \boldsymbol{\beta}_i'\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right).$$

Using this result we have

$$\begin{aligned}(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T \left(r_{it} - \bar{r}_{i\circ}\right)^2 &= (nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T \left(u_{it} - \bar{u}_{i\circ}\right)^2 + n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i \\ &\quad - 2(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T \left(u_{it} - \bar{u}_{i\circ}\right)\boldsymbol{\beta}_i'\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right),\end{aligned} \tag{F.5}$$

where $\hat{\boldsymbol{\Sigma}}_f = T^{-1}\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)'$. For the first term we have

$$(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T \left(u_{it} - \bar{u}_{i\circ}\right)^2 - \bar{\sigma}^2_n = O_p(n^{-1/2}T^{-1/2}), \tag{F.6}$$

which follows from the proof of Theorem 2 by setting $\boldsymbol{\beta}_i = \mathbf{0}$ and $\mathbf{f}_t = \mathbf{0}$ in Section B.2. For the cross product term we have

$$\begin{aligned}&(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T \left(u_{it} - \bar{u}_{i\circ}\right)\boldsymbol{\beta}_i'\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right) \\ &= (nT)^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)\left(u_{it} - \bar{u}_{i\circ}\right) = (nT)^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)u_{it} \\ &- (nT)^{-1}\sum_{i=1}^n \bar{u}_{i\circ}\boldsymbol{\beta}_i'\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right) = (nT)^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)u_{it} = p_{nT}.\end{aligned}$$

Also, using (42),

$$p_{nT} = (nT)^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)\left(\gamma_i g_t + v_{it}\right) = p_{1,nT} + p_{2,nT},$$

where

$$p_{1,nT} = \left(n^{-1}\sum_{i=1}^n \gamma_i\boldsymbol{\beta}_i'\right)\left[T^{-1}\sum_{t=1}^T\left(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T\right)g_t\right],$$

and

$$p_{2,nT} = (nT)^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \sum_{t=1}^{T} (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) v_{it}.$$

Under Assumption 3, $(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)$ and g_t are distributed independently and g_t are serially independent with $E(g_t) = 0$ and $E(g_t^2) = 1$, and it follows that

$$Var\left[T^{-1}\sum_{t=1}^{T} (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) g_t\right] = E\left[T^{-2}\sum_{t=1}^{T} (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)^2 E\left(g_t^2\right)\right] = T^{-1}E\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right) = O\left(T^{-1}\right).$$

Hence, $T^{-1}\sum_{t=1}^{T} (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) g_t = O_p\left(T^{-1/2}\right)$. Also (see (45) and (56))

$$\begin{aligned}
\left\| n^{-1}\sum_{i=1}^{n} \gamma_i \boldsymbol{\beta}_i' \right\| &\leq n^{-1}\sum_{i=1}^{n} |\gamma_i| \, \|\boldsymbol{\beta}_i\| = n^{-1}\sum_{i=1}^{n} |\gamma_i| \left(\boldsymbol{\beta}_i'\boldsymbol{\beta}_i\right)^{1/2} \\
&\leq \left(n^{-1}\sum_{i=1}^{n} |\gamma_i|^2\right)^{1/2} \left(n^{-1}\sum_{i=1}^{n} \boldsymbol{\beta}_i'\boldsymbol{\beta}_i\right)^{1/2} = O_p\left(n^{\frac{-1+\alpha_\gamma}{2}}\right) O_p\left(n^{\frac{-1+\alpha_{max}}{2}}\right) \\
&= O_p\left(n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right).
\end{aligned}$$

Hence, $p_{1,nT} = O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)$. Consider $p_{2,nT}$ and recall that under Assumption 4 v_{it} are serially independent, have zero means and are distributed independently of $(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)$ and $\boldsymbol{\beta}_i$. Then $E\left(p_{2,nT}\right) = 0$ and

$$Var\left(p_{2,nT} \mid \mathbf{F}\right) = \frac{1}{nT}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij,v}\boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i\right),$$

where $E(v_{it}v_{jt}) = \sigma_{ij,v}$. Also,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij,v}\boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i\right\| \leq \left(sup_i \|\boldsymbol{\beta}_i\|\right)^2 \left\|\hat{\boldsymbol{\Sigma}}_f\right\| \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} |\sigma_{ij,v}|\right),$$

and by assumption $sup_i \|\boldsymbol{\beta}_i\| < C$, $E\left\|\hat{\boldsymbol{\Sigma}}_f\right\| < C$, and $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} |\sigma_{ij,v}| = O(1)$. Hence, $Var\left(p_{2,nT}\right) = O_p(n^{-1}T^{-1})$, and it follows that $p_{2,nT} = O_p(T^{-1/2}n^{-1/2})$, and overall (since $\alpha_\gamma < 1/2$ and $\alpha_{max} \leq 1$)

$$p_{nT} = O_p(T^{-1/2}n^{-1/2}) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right) = O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right). \tag{F.7}$$

Using (F.6) and (F.7) in (F.5), we now have

$$(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} (r_{it} - \bar{r}_{i\circ})^2 = \bar{\sigma}_n^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right).$$

Using this result and (F.4) in (F.2) yields

$$\overline{PR}_{nT}^2 = 1 - \frac{\bar{\sigma}_n^2 + O_p(T^{-1/2}n^{-1/2})}{\bar{\sigma}_n^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i + O_p(T^{-1/2}n^{-1/2}) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)}.$$

Since $O_p(T^{-1/2}n^{-1/2})$ is dominated by $\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)$, we end up with

$$\overline{PR}^2_{nT} = \frac{n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i/\bar{\sigma}_n^2 + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)}{1+n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i/\bar{\sigma}_n^2 + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)}. \tag{F.8}$$

Hence, the order of $\overline{PR}^2_{nT}$ is governed by the pooled signal-to-noise ratio defined by

$$s^2_{nT} = \frac{n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\hat{\boldsymbol{\Sigma}}_f\boldsymbol{\beta}_i}{\bar{\sigma}_n^2}.$$

However, under Assumption 1

$$\lambda_{min}(\hat{\boldsymbol{\Sigma}}_f)\frac{n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\boldsymbol{\beta}_i}{\bar{\sigma}_n^2} \leq s^2_{nT} \leq \lambda_{max}(\hat{\boldsymbol{\Sigma}}_f)\frac{n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\boldsymbol{\beta}_i}{\bar{\sigma}_n^2}, \tag{F.9}$$

where $c < \lambda_{min}(\hat{\boldsymbol{\Sigma}}_f) < \lambda_{max}(\hat{\boldsymbol{\Sigma}}_f) < C$. Hence,

$$c\left(\frac{n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\boldsymbol{\beta}_i}{\bar{\sigma}_n^2}\right) \leq s^2_{nT} \leq C\left(\frac{n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\boldsymbol{\beta}_i}{\bar{\sigma}_n^2}\right),$$

and it must be that

$$s^2_{nT} = \ominus\left(n^{-1}\sum_{i=1}^n \boldsymbol{\beta}_i'\boldsymbol{\beta}_i\right) = \ominus\left[\sum_{k=1}^K\left(n^{-1}\sum_{i=1}^n \beta_{ik}^2\right)\right].$$

Also, under Assumption 2, $n^{-1}\sum_{i=1}^n \beta_{ik}^2 = \ominus\left(n^{\alpha_k-1}\right)$. Hence,

$$s^2_{nT} = \sum_{k=1}^K \ominus\left(n^{\alpha_k-1}\right),$$

which in view of (F.8) now yields (F.3), as desired. ■

References

Bailey, N., G. Kapetanios and M. H. Pesaran (2021). "Measurement of factor strength: theory and practice", *Journal of Applied Econometrics,* 36, 587-613.

Bailey, N., M. H. Pesaran, and L.V. Smith (2019). "A multiple testing approach to the regularisation of sample correlation matrices", *Journal of Econometrics,* 208, 507-534.

Chen, A. Y. and T. Zimmermann (2022). "Open source cross-sectional asset pricing", *Critical Finance Review*, 11, 207-264.