CESifo WORKING PAPERS

8947
2021

March 2021

Factor Strengths, Pricing Errors, and Estimation of Risk Premia

M. Hashem Pesaran, Ron P. Smith

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8947*

Factor Strengths, Pricing Errors, and Estimation of Risk Premia

Abstract

This paper examines the implications of pricing errors and factors that are not strong for the Fama-MacBeth two-pass estimator of risk premia and its asymptotic distribution when T is fixed with $n \to \infty$, and when both n and $T \to \infty$, jointly. While the literature just distinguishes strong and weak factors we allow for degrees of strength using a recently developed measure. Our theoretical results have important practical implications for empirical asset pricing. Pricing errors and factor strength matter for consistent estimation of risk premia and subsequent inference, thus an estimate of factor strength is required before attempting to estimate risk. Finally, using a recently developed procedure we provide rolling estimates of factor strengths for the five Fama-French factors, and show that only the market factor can be viewed as strong.

JEL-Codes: C380, G120.

Keywords: factor strength, pricing errors, risk premia, Fama and MacBeth two-pass estimators, Fama-French factors, panel R^2.

M. Hashem Pesaran
University of Southern California
Los Angeles / USA
pesaran@usc.edu

Ron P. Smith
Birkbeck, University of London / UK
r.smith@bbk.ac.uk

March 1, 2021
We are grateful to Natalia Bailey for carrying out the computation of the results presented in the empirical section of the paper. We have also received helpful comments from Natalia Bailey and Zhaoguo Zhan.

1 Introduction

There is a large literature in finance which estimates the risk premia associated with observed risk factors using a two-pass estimation procedure introduced in Fama and MacBeth (1973). The first pass estimates the factor loading, β_{ik}, of factor, f_{kt}, for the security i over the time period $t = 1, 2, ..., T$, by running least squares regressions of returns on each security r_{it}, $i = 1, 2, ..., n$, using K observed factors f_{kt}, $k = 1, 2, ..., K$. The second pass runs a pure cross-section regression of time averages of security returns on the estimated factor loadings to price the factors and obtain estimates of their risk premia, λ_k. The asymptotic properties of this procedure have been investigated by Shanken (1992), Shanken and Zhou (2007), Kan, Robotti and Shanken (2013), and Bai and Zhou (2015), among others. The survey paper by Jagannathan, Skoulakis & Wang (2010) provides further references.

This literature implicitly assumes that (a) all the risk factors used are *strong*, in that they are pervasive, influencing almost all securities, (b) there are *no pricing errors*, (c) the parameters of interest are what Shanken (1992, p6) calls the "ex post" risk premia, which differ from the actual risk premia due to a bias caused by the difference between the means of risk factors and their expected (population) values.

This paper examines the consequences of relaxing the above three restrictions, individually and in combination, for estimation of risk premia and their asymptotic distribution, under different values of n and T and their relative expansion rates. Different issues are involved in the second pass estimation of risk premia from the first pass estimation of factor loadings. We first focus on the second pass and look at the effect of non-strong factors and pricing errors on risk premia estimation with known, rather than estimated, factor loadings, β_{ik}. We then move to the more realistic case where β_{ik} are estimated.

The possibility of weak factors has been discussed in the econometrics literature by Chudik, Pesaran and Tosseti (2011), Onatski (2012), and Kleibergen (2009). Connor and Korajczyk (2019) construct a test statistic for empirically distinguishing strong from semi-strong factors using marginal R squared. Anatolyev and Mikusheva (2020), discussed further below, consider models with strong as well as weak factors in asset pricing, but do not allow for pricing errors or semi-strong factors, and focus on estimation of ex post risk premia.[1]

Unlike the existing literature we do not just distinguish between weak and strong factors, but allow for different degrees of strength. The strength of factor f_{kt}, is measured by the exponent α_k

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

where β_{ik} is the loading of f_{kt} on the i^{th} security, and $\ominus\left(n^{\alpha_k}\right)$ denotes the exact rate at which $\sum_{i=1}^{n} \beta_{ik}^2$ rises with n. Factor k is strong if $\alpha_k = 1$, semi-strong if $1 < \alpha_k < 0.5$, and weak if $\alpha_k \leq 0.5$. [2]

Our theoretical results establish explicit links between the precision with which risk premia can be estimated and the strength of the underlying factors. In turn, the factor strengths can be estimated, as proposed by Bailey, Kapetanios and Pesaran (2021), from the proportion of securities with statistically significant factor loadings, allowing for multiple testing. Whereas a

[1] There is also an emerging literature on unobserved factors that allow for weak as well as strong factors. See, for example, Lettau and Pelger (2020a, 2020b) and Section 4 of Bailey et al. (2021). But for the identification of risk premia, which is the focus of the present paper, the factors must be observed.

[2] The $\ominus(.)$ notation in (1) should not be confused with the standard big O notation, $O(.)$.

strong factor with $\alpha_k = 1$ has non-zero loadings for almost all securities, a factor with $\alpha_k = 1/2$ has non-zero loadings for $\sqrt{n}$ of the n securities, and the share of non-zero loadings tends to zero quite fast as n increases. This means that even with known β_{ik} a large number of loadings used in the second pass regression could be zero, making estimation of the risk premia imprecise. For instance, with $n = 400$ and $\alpha_k = 1/2$, only about 20 securities will have non zero loadings. As $n \to \infty$, the rank of the variance covariance matrix of the loadings is given by the number of strong factors.

For identification of risk premia, λ_k, we also need a sufficient degree of heterogeneity across the loadings as $n \to \infty$, such that

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

where $\bar{\beta}_k = n^{-1}\sum_{i=1}^{n} \beta_{ik}$. This is not necessarily implied by (1).

Pricing errors, denoted by η_i for security i, are likely to be important in practice. In his theoretical derivation of Arbitrage Pricing Theory (APT), Ross (1976) assumed pricing errors to be square summable, namely $\sum_{i=1}^{n} \eta_i^2 < \infty$. We use the weaker condition

$$\sum_{i=1}^{n} |\eta_i| = O\left(n^{\alpha_\eta}\right), \tag{3}$$

with the exponent α_η measuring the pervasiveness of the pricing errors.

If the strength of factor k is α_k and the pervasiveness of the pricing errors is α_η, when factor loadings are known, we need $\alpha_k > \alpha_\eta$ for consistent estimation of risk premia. The associated convergence rate is $n^{-(\alpha_k - \alpha_\eta)/2}$, which can be very slow if $\alpha_k - \alpha_\eta$ is small, requiring n to be very large to get accurate estimates of λ_k.[3] Similarly, even if there are no pricing errors, when α_k is small the number of non zero β_{ik} rises slowly with n.

Our results have important practical implications for Fama-MacBeth type procedures. To accommodate non-zero pricing errors n has to be large even when the first pass factor loadings are known. The need for large n was highlighted by Roll (1977). This argues against the practice of using a small number of constructed portfolios in the second pass regressions and we use individual securities in our derivations and empirical work. Chordia and Subrahmanyam (1998) emphasize the problems for inference created by using portfolios and, more recently, Ang, Liu and Schwarz (2020) show that creating portfolios to reduce estimation error in the factor loadings does not necessarily lead to smaller estimation errors of the factor risk premia.

While n needs to be large, the number of time periods, T, is inevitably small because of the possible instability of factor loadings over time, hence the nearly universal practice in finance is to use regressions estimated over relatively short rolling windows (5 to 10 years). As noted above, in the second pass as $n \to \infty$, the rank of the covariance matrix of the true loadings is given by the number of strong factors. So if there are any factors that are not strong, the population covariance matrix of factor loadings will become rank deficient and the estimator of the risk premia will not be defined. However, when the loadings are estimated with T small relative to n, the two-pass estimator of the risk premia is still defined because it uses the sample covariance matrix of the estimated loadings.

[3] In the case examined in the literature where $\alpha_k = 1$, $\alpha_\eta = 0$, and factor loadings are known, one gets the usual $n^{-1/2}$ rate.

As is well known, for a fixed T, the two pass-estimator of the risk premium is biased because of sampling error in the first stage estimates of β_{ik}. This is usually dealt with using the Shanken (1992) small T bias correction. The bias correction works by shifting the covariance matrix of the estimated loadings back to the covariance of the true loadings. But the correction made to eliminate the bias renders the limiting distribution of the bias-corrected estimator (for a finite T and as $n \to \infty$) ill-conditioned, unless *all* the factors under consideration are strong. In short, the Shanken bias-corrected estimators of expost risk premia are defined for small n, but its limit as $n \to \infty$ exist if and only if all the factors under consideration are strong.

When both n and T are sufficiently large such that as $n, T \to \infty$, $n/T \to \kappa$, with $0 < \kappa < \infty$, and there are no pricing errors, the two-pass estimator of the risk premium associated to factor f_{kt} converges to the "ex post" risk premium at the rate of $n^{-\alpha_k/2}T^{-1/2}$, and the ex post risk premia of all factors, whether strong or not, can be consistently estimated and tested. However, this is not true for the actual risk premia, which is the primary object of interest. When there are no pricing errors, the two-pass estimator of risk premia converges to its true value at the $T^{-1/2}$ rate and does not depend on the factor strengths, α_k. In the more realistic case where there are pricing errors then the two-pass estimator converges to its true value at the rate of $n^{-\frac{1}{2}(\alpha_k - \alpha_\eta)}$, which, as noted above, can be quite slow. Furthermore, although risk premia may be identified with very large n, it is not possible to carry out inference on the weak or even semi-strong factors. In the presence of pricing errors the asymptotic distribution of the two-pass estimator of risk premia is non-degenerate only in the case of strong factors, with the influence of remaining factors waning as n and $T \to \infty$.

A central question in this literature is the ability of factors to explain returns. We show that asymptotically the pooled R-squared of the panel regression of returns on factors is only determined by strong factors. In the limit as $n \to \infty$, factors with $\alpha_k < 1$ do not contribute to explanation of returns. The effects of weak or semi-strong factors on the fit of the return regressions will vanish eventually.

As an empirical application we estimate factor strength for the five Fama and French (2015) factors. We compute 10-year rolling estimates for all the five factors over the period September 1989 to May 2018, a total of 345 rolling estimates for each factor. The factor strengths are very precisely estimated, with the stronger the factor the more precisely its strength is estimated. As might be expected, the market factor is strong, with a strength always very close to one and a time average over the 345 rolling estimates of 0.99. The estimated strengths for the other four factors are much lower, generally below 0.8, with the average of their strengths over the rolling windows never exceeding 0.75.

Our theoretical analysis suggests that in practice it is important to measure the strength of the factors and to focus on estimation of risk premia for strong factors. Semi-strong or weak factors can be included, to reduce cross section dependence, so long as the strength of the weakest factor, α_{min}, exceed $2\alpha_\eta$. However, use of risk factors that are not strong poses important new challenges, since there are likely to be many weak or even semi-strong potential factors in what has been labelled a "factor zoo". Harvey and Liu (2019) list over 400 factors suggested by early 2019 in the literature. Fama and French (2019) discuss the issue of choosing factors while Jensen, Kelly and Pedersen (2021) argue that the large number of factors can be clustered into fewer themes.

The rest of the paper is organized as follows: Section 2 sets up the multi-factor model allowing for differing factor strengths and the APT restrictions. Section 3 considers the estimation of risk premia with known factor loadings. Section 4 introduces the two pass estimator.

Section 5 considers the asymptotic properties of the two-pass estimator with a fixed T and large n. Section 6 examines the asymptotic properties of the two-pass estimator when both n and T are large, under different pricing error scenarios. Section 7 examines the effect of factor strength on the explanatory power of the regressions explaining returns. Section 8 presents the rolling estimates of factor strength for the five Fama-French factors. Section 9 provides some concluding remarks. Detailed mathematical proofs are relegated to the Appendix.

Notations: Generic positive finite constants are denoted by C when large, and c when small. They can take different values at different instances. $\rightarrow^p$ denotes convergence in probability as $n, T \rightarrow \infty$. $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the maximum and minimum eigenvalues of matrix $\boldsymbol{A}$. $\boldsymbol{A} > 0$ denotes that $\boldsymbol{A}$ is a positive definite matrix. $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\boldsymbol{A}\|_F = [Tr(\mathbf{A}'\mathbf{A})]^{1/2}$, $\|\boldsymbol{A}\|_p = (E\|\boldsymbol{A}\|^p)^{1/p}$, for $p \geq 2$ denote spectral, Frobenius, and ℓ_p norm 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 $f_n/g_n \rightarrow 0$ as $n \rightarrow \infty$. Similarly, $f_n = O_p(g_n)$ if f_n/g_n is stochastically bounded, and $f_n = o_p(g_n)$, if $f_n/g_n \rightarrow_p 0$, where $\rightarrow_p$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 $\sup_{n \geq n_0}(f_n/g_n) \leq C$.

2 Return regressions, APT and factor strengths

2.1 A multi-factor model of returns

Following the literature, we assume that returns on security i in period t, r_{it}, are generated by the following linear multi-factor model

$$r_{it} - r^f = \mathrm{a}_i + \sum_{k=1}^{K} \beta_{ik} f_{kt} + u_{it}, \text{ for } i = 1, 2, ..., n, \ t = 1, 2, ..., T, \tag{4}$$

where r^f is the risk free rate, assumed to be fixed; a_i is the return-specific effect; f_{kt}, $k = 1, 2, ..., K$ are the K common factors with associated factor loadings, β_{ik}; and u_{it} are the idiosyncratic components of asset returns. The model can be written more compactly as

$$r_{it} - r^f = \mathrm{a}_i + \boldsymbol{\beta}_i' \mathbf{f}_t + u_{it}, \tag{5}$$

where $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{iK})'$, and $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{Kt})'$.

We make the following standard assumptions about $\mathbf{f}_t$ and u_{it} (the drivers of asset returns):

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

$$\mathbf{f}_t = \boldsymbol{\mu}_f + \sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_\ell \mathbf{v}_{t-\ell}, \tag{6}$$

where $\|\boldsymbol{\mu}_f\| < C$, $\mathbf{v}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$, 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 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, $\hat{\boldsymbol{\Sigma}}_f \rightarrow_p \boldsymbol{\Sigma}_f = E(\mathbf{f}_t - \mu_f)(\mathbf{f}_t - \mu_f)' > \mathbf{0}$, and $\left\|\hat{\boldsymbol{\Sigma}}_f^{-1}\right\| = O_p(1)$.

Assumption 2 *(Idiosyncratic errors) The errors* $\{u_{it},\ i = 1, 2, ..., n;\ t = 1, 2, ..., T\}$ *are serially independent across* t*, with zero means,* $E(u_{it}) = 0$*, and constant covariances,* $E(u_{it}u_{jt}) = \sigma_{ij}$*, such that* $0 < c < \sigma_{ii} < C < \infty$*,*

$$(a): \sup_j \sum_{i=1}^{n} |\sigma_{ij}| < C, \tag{7}$$

and

$$(b):\ n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(u_{it}^2, u_{jt}^2\right) < C. \tag{8}$$

(c) The errors, u_{it} *are distributed independently of the factors* $f_{k,t'}$*, for all* i, t, t' *and* $k = 1, 2, ..., K$*, and their associated loadings* β_{ik}*.*

Assumption 3 *The dependence between loadings, factors and errors is such that*

$$n^{-1/2}\sum_{i=1}^{n}\beta_i'\left[T^{-1/2}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)u_{it}\right] = O_p(1).$$

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

Remark 2 *Let* $\mathbf{V}_u = E\left(\mathbf{u}_{\circ t}\mathbf{u}_{\circ t}'\right) = (\sigma_{ij})$*, where* $\mathbf{u}_{\circ t} = (u_{1t}, u_{2t}, ..., u_{nt})'$*, then condition (7) also ensures that*

$$\lambda_{\max}(\mathbf{V}_u) \leq \|\mathbf{V}_u\|_1 = \sup_j \sum_{i=1}^{n}|E(u_{it}u_{jt})| < C.$$

Condition (7) is in line with the assumptions of the approximate factor models used in the APT literature that require $\lambda_{\max}(\mathbf{V}_u) < C$*. Condition (8) is needed to establish probability limits in Lemma A.1 of the Appendix.*

Remark 3 *Part (a) of Assumption 2 is standard in the literature and allows for errors to be weakly cross correlated. It rules out serial correlation, but can be relaxed to allow for a limited degree of serial correlation when both* n *and* T *are large. But it is required if* T *is fixed and* n *large.*

2.2 Arbitrage pricing theory (APT) restrictions

The main result of the APT can be summarized in the following cross section return regression where (population) return of security i is related to its factor loadings

$$E(r_{it}) = \lambda_0 + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \eta_i, \tag{9}$$

where $\boldsymbol{\lambda}$ is the $K \times 1$ vector of factor risk prices (or risk premia), and η_i is the pricing error of the i^{th} security, assumed to satisfy the APT condition (18) of Ross (1976), namely

$$\sum_{i=1}^{n}\eta_i^2 < C. \tag{10}$$

To impose the APT restrictions on the statistical model, first using (5) and (6) we note that

$$E(r_{it}) = r^f + \mathrm{a}_i + \boldsymbol{\beta}_i' E(\mathbf{f}_t) + E(u_{it}) = r^f + \mathrm{a}_i + \boldsymbol{\beta}_i' \boldsymbol{\mu}_f. \tag{11}$$

Therefore, for APT restrictions (9) to hold we must have

$$r^f + \mathrm{a}_i + \boldsymbol{\beta}_i' \boldsymbol{\mu}_f = \lambda_0 + \boldsymbol{\lambda}' \boldsymbol{\beta}_i + \eta_i, \text{ for all } i$$

which in turn requires that

$$\boldsymbol{\lambda} = \boldsymbol{\mu}_f, \text{ and } a_i + r^f = \lambda_0 + \eta_i, \text{ for all } i. \tag{12}$$

Also, since $E(r_{it})$ is unobserved, it is typically replaced by its sample mean, $\bar{r}_{iT} = T^{-1} \sum_{t=1}^{T} r_{it}$, which, using (5), is given by

$$\bar{r}_{iT} = r^f + \mathrm{a}_i + \boldsymbol{\beta}_i' \bar{\mathbf{f}}_T + \bar{u}_{iT}.$$

Imposing the APT restrictions (12) yields:

$$\bar{r}_{iT} = \lambda_0 + \lambda' \boldsymbol{\beta}_i + \eta_i + \zeta_{iT}, \tag{13}$$

where

$$\zeta_{iT} = \left[\bar{\mathbf{f}}_T - E(\mathbf{f}_t) \right]' \boldsymbol{\beta}_i + \bar{u}_{iT}. \tag{14}$$

But under Assumptions 1 and 2 it also follows that

$$\bar{u}_{iT} = \frac{1}{\sqrt{T}} \left(T^{-1/2} \sum_{t=1}^{T} u_{it} \right) = O_p\left(T^{-1/2}\right),$$

$$\bar{\mathbf{f}}_T - E(\mathbf{f}_t) = \frac{1}{\sqrt{T}} \left(T^{-1/2} \sum_{t=1}^{T} [\mathbf{f}_t - E(\mathbf{f}_t)] \right) = O_p\left(T^{-1/2}\right).$$

Hence, under standard assumptions made in the literature, the use of time averages allows the sampling errors in ζ_{iT} to tend to zero at the rate of $T^{-1/2}$, while the pricing errors, η_i, remain. Specifically,

$$\bar{r}_{iT} = \lambda_0 + \boldsymbol{\lambda}' \boldsymbol{\beta}_i + \eta_i + O_p\left(T^{-1/2}\right). \tag{15}$$

Also, $\bar{r}_{iT} \to_p E(r_{it}) = \mu_i = \lambda_0 + \lambda' \boldsymbol{\beta}_i + \eta_i$, hence APT restrictions only hold in the limit with $T \to \infty$. Therefore, one might expect pricing errors, η_i, to play an important role even for large samples or where the factor loadings, $\boldsymbol{\beta}_i$ are treated as known. A formal account is provided below.

2.3 Factor strengths

From the perspective of risk diversification, APT also rules out the possibility of a fully diversified portfolio. This means that there must exist at least one strong factor, otherwise it would be possible to construct a portfolio whose risk vanishes as $n \to \infty$. To see this using (5) a portfolio constructed as a simple average of the returns, $\bar{r}_{\circ t} = n^{-1} \sum_{i=1}^{n} r_{it}$, is given by

$$\bar{r}_{\circ t} - r^f = \bar{\mathrm{a}} + \bar{\boldsymbol{\beta}}_n' \mathbf{f}_t + \bar{u}_{\circ t},$$

and under Assumption 2 (that errors are weakly cross correlated) we have

$$\bar{r}_{ot} = r^f + \bar{a} + \bar{\boldsymbol{\beta}}_n' \mathbf{f}_t + O_p\left(n^{-1/2}\right),$$

where $\bar{\boldsymbol{\beta}}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{\beta}_i = (\bar{\beta}_{n1}, \bar{\beta}_{n2}, ..., \bar{\beta}_{nK})'$. Full diversification occurs if $\bar{\boldsymbol{\beta}}_n \to \mathbf{0}$, as $n \to \infty$. To avoid this outcome, factor loadings should be such that $\bar{\beta}_k = \ominus_p(1)$, for some k.

Also for estimation of $\boldsymbol{\lambda}$, using (15), it is commonly assumed that the covariance matrix of factor loadings defined by

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

is positive definite. For $\boldsymbol{\Sigma}_{\beta\beta}$ to be positive definite it is *necessary* that all the K risk factors under consideration are strong in the sense that

$$plim_{n\to\infty}\left[n^{-1}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)^2\right] > 0, \text{ for } k = 1, 2, ..., K. \quad (16)$$

However, such an assumption is quite restrictive and is unlikely to be satisfied for many risk factors being considered in the literature. BKP (2021) propose a method for estimating factor strength and show that, apart from the market factor, only a handful of 144 factors considered by Feng et al. (2020) in the literature come close to being strong. More importantly, factor strengths vary over time in addition to the familiar variation over time in the factor loadings.

BKP (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 $\boldsymbol{\Sigma}_{\beta\beta}$. (see also Chudik et al. (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*

$$\sum_{i=1}^{n}\beta_{ik}^2 = \ominus(n^{\alpha_k}), \quad (17)$$

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 $0 \leq \alpha_k \leq 1/2$.

In the above definition $\ominus(n^{\alpha_k})$ 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. It is also clear that $\boldsymbol{\Sigma}_{\beta\beta}$ is positive definite *if and only if* $\alpha_k = 1$ for all $k = 1, 2, ..., K$. Condition (17) applies irrespective of whether the loadings, β_{ik}, are viewed as deterministic or stochastic. Under the latter condition (17) can be written as

$$n^{-\alpha_k}\sum_{i=1}^{n}E\left(\beta_{ik}^2\right) > c > 0,$$

which we write more compactly as $\sum_{i=1}^{n}\beta_{ik}^2 = \ominus_p(n^{\alpha_k})$.

3 Estimation of risk premia with known factor loadings

To highlight the importance of pricing errors and factor strengths for estimation of risk premia, we begin with the case where the factor loadings, β_{ik}, and their strength, α_k, $k = 1, 2, ...K$, are known. We also assume that $E(r_{it}) = \mu_i$ is known. We make the following assumption about the loadings. Throughout we assume K is fixed as n and $T \to \infty$.

Assumption 4 *(Factor loadings) The factor loadings β_{ik} for $i = 1, 2, ..., n$ and $k = 1, 2, ..., K$ are (a) either deterministic such that $sup_{ik} |\beta_{ik}| < C$, or stochastically bounded such that $sup_{ik} E\left(\beta_{ik}^2\right) < C$, and for $k = 1, 2, ..., K$*

$$\sum_{i=1}^{n} \beta_{ik}^2 = \ominus_p(n^{\alpha_k}). \tag{18}$$

(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

$$0 < c < \lambda_{min}\left(\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}\right) < \lambda_{max}\left(\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}\right) < C < \infty, \tag{19}$$

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 $\mathbf{D}_n$ is the $n \times n$ diagonal matrix

$$\mathbf{D}_n = Diag(n^{\alpha_1/2}, n^{\alpha_2/2}, ..., n^{\alpha_K/2}). \tag{20}$$

Assumption 5 *(Pricing errors) (a) The pricing errors, η_i, for $i = 1, 2, ..., n$ are deterministic such that $sup_i |\eta_i| < C$, and satisfy*

$$\sum_{i=1}^{n} |\eta_i| = O\left(n^{\alpha_\eta}\right), \tag{21}$$

where $\alpha_\eta \geq 0$ denotes its degree of pervasiveness. (b) The pricing errors, η_i, are stochastic such that

$$\sum_{i=1}^{n} E\left|\eta_i\right| = O\left(n^{\alpha_\eta}\right), \tag{22}$$

and are distributed independently of the factor loadings, β_{jk}.

Remark 4 *Under Assumption 4*

$$\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1} \to_p \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha}) > 0, \tag{23}$$

where $\boldsymbol{\Sigma}_{\beta\beta}(\alpha)$ 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 (19) since for any non-zero $n \times 1$ vector $\mathbf{c}$,

$$\mathbf{c}'\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}\mathbf{c} \geq \left(\mathbf{c}'\mathbf{c}\right)\lambda_{min}\left(\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}\right) > 0.$$

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

Remark 5 *The exponent parameter, α_η, of the pricing condition in (21), 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*

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

and under Assumption (5) it also follows that

$$\|\boldsymbol{\eta}_n\|^2 = \sum_{i=1}^{n} \eta_i^2 = O\left(n^{\alpha_\eta}\right). \quad (24)$$

Remark 6 *Whilst (21) implies (24), 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 *Parts (a) and (b) of Assumption 5 differ in whether the pricing errors and factor loadings are correlated, and as we shall see this can play an important role for the analysis of risk premia.*

Consider the APT equations (9), denote the expected returns on asset i by $\mu_i = E(r_{it})$, and stack the equations for $i = 1, 2, ..., n$, to obtain:

$$\boldsymbol{\mu}_n = \lambda_0 \boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\lambda} + \boldsymbol{\eta}_n, \quad (25)$$

where $\mathbf{B}_n$ is the $n \times k$ matrix of factor loadings, $\boldsymbol{\mu}_n = (\mu_1, \mu_2, ..., \mu_n)'$, $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$ and λ_0 is treated as an unknown constant. Under this setting and assuming $\mathbf{B}_n$ is known, $\boldsymbol{\lambda}$ can be estimated by least squares

$$\hat{\boldsymbol{\lambda}}_n = \left(\mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n\right)^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\mu}_n. \quad (26)$$

To establish the asymptotic properties of $\hat{\boldsymbol{\lambda}}_n$ as $n \to \infty$, when the factors have different strengths, we need to standardize the factor loadings using the diagonal matrix $\mathbf{D}_n$, defined by (20). For any given n, $\hat{\boldsymbol{\lambda}}_n$ can be written equivalently as

$$\mathbf{D}_n \hat{\boldsymbol{\lambda}}_n = \left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right)^{-1} \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\mu}_n,$$

which upon using (25) yields

$$\mathbf{D}_n \left(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) = \left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right)^{-1} \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\eta}_n. \quad (27)$$

First, note that[4]

$$\left\| \left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right)^{-1} \right\| = \lambda_{max} \left[\left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right)^{-1} \right] = \frac{1}{\lambda_{min} \left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right)},$$

and since under Assumption 4 $\lambda_{min} \left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right) > c > 0$, then it follows that

$$\left\| \left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1}\right)^{-1} \right\| < C < \infty. \quad (28)$$

[4]Recall that $\|\mathbf{A}\| = \lambda_{max}^{1/2}(\mathbf{A}'\mathbf{A})$ denotes the spectral norm of $\mathbf{A}$, and when $\mathbf{A}$ is symmetric then $\|\mathbf{A}\| = \lambda_{max}(\mathbf{A})$.

If there are no pricing errors, and $\eta_i = 0$ for all i, it is clear from (27) that $\hat{\boldsymbol{\lambda}}_n = \boldsymbol{\lambda}$; there are no remaining uncertainties and estimates of risk premia do not depend on factor strengths. But it is unrealistic to assume no pricing errors. To take them into account write the second term of (27) as

$$\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n = \left[\frac{\left(\boldsymbol{\beta}_{\circ 1} - \boldsymbol{\tau}_n\bar{\beta}_1\right)'\eta_n}{n^{\alpha_1/2}}, \frac{\left(\boldsymbol{\beta}_{\circ 2} - \tau_n\bar{\beta}_2\right)'\eta_n}{n^{\alpha_2/2}}, ..., \frac{\left(\boldsymbol{\beta}_{\circ K} - \tau_n\bar{\beta}_K\right)'\eta_n}{n^{\alpha_K/2}}\right]'. \tag{29}$$

where $\bar{\beta}_k = n^{-1}\boldsymbol{\tau}_n'\boldsymbol{\beta}_{\circ k}$, and note that that k^{th} element of $\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n$ can be written as

$$\pi_{k,n} = n^{-\alpha_k/2}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)\eta_i. \tag{30}$$

The probability order of $\pi_{k,n}$ critically depends on the nature and the degree of pervasiveness of the pricing errors, η_i. By Cauchy–Schwarz inequality

$$|\pi_{k,n}| \leq \left[n^{-\alpha_k}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)^2\right]^{1/2}\left(\sum_{i=1}^{n}\eta_i^2\right)^{1/2},$$

and under Assumption 4, $n^{-\alpha_k}\sum_{i=1}^{n}\left(\beta_{ik} - \bar{\beta}_k\right)^2 = \ominus(1)$. Also using (24), implied by part (a) of Assumption 5, we have $\sum_{i=1}^{n}\eta_i^2 = O\left(n^{\alpha_\eta}\right)$, and hence $|\pi_{k,n}| = O_p\left(n^{\alpha_\eta/2}\right)$. Given that K is fixed, overall we have

$$\mathbf{D}_n^{-1}\boldsymbol{B}_n'\mathbf{M}_n\eta_n = O_p\left(n^{\alpha_\eta/2}\right). \tag{32}$$

But when part (b) of Assumption 5 holds, the effects of pricing errors on risk premia estimates will be much less pronounced. This is achieved by conditioning on the pricing errors and then exploiting their independence from the factor loadings, namely by noting that[5]

$$E\left(|\pi_{k,n}|\,|\eta_i\right) \leq n^{-\alpha_k/2}\sum_{i=1}^{n}E\left|\beta_{ik} - \bar{\beta}_k\right||\eta_i| \leq \left[sup_{ik}E\left|\beta_{ik} - \bar{\beta}_k\right|\right]n^{-\alpha_k/2}\sum_{i=1}^{n}|\eta_i| = O_p\left(n^{\alpha_\eta - \alpha_k/2}\right).$$

and hence

$$\mathbf{D}_n^{-1}\boldsymbol{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(n^{\alpha_\eta - \alpha_{min}/2}\right), \tag{33}$$

where $\alpha_{min} = min(\alpha_1, \alpha_2, ..., \alpha_K)$. Using this result together with (28) in (27) we now have $\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) = O_p\left(n^{\alpha_\eta/2}\right)$ under part (a) of Assumption 5, and $\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) = O_p\left(n^{\alpha_\eta - \alpha_{min}/2}\right)$, under part (b) of Assumption 5. For the risk premia of the k^{th} factor, these two results can be written equivalently as

$$\hat{\lambda}_{kn} - \lambda_k = O_p\left(n^{-\left(\frac{\alpha_k - \alpha_\eta}{2}\right)}\right), \text{ under part } (a) \text{ of Assumption 5,} \tag{34}$$

$$\hat{\lambda}_{kn} - \lambda_k = O_p\left(n^{-\left(\frac{\alpha_k + \alpha_{min}}{2} - \alpha_\eta\right)}\right), \text{ under part } (b) \text{ of Assumption 5,} \tag{35}$$

[5] Note that under Assumption 4

$$E\left|\beta_{ik} - \bar{\beta}_k\right| < E\left|\beta_{ik}\right| + E\left|\bar{\beta}_k\right| < C.$$

The convergence rate of $\hat{\lambda}_{kn}$ to its true value, λ_k, is governed by both α_k as well as α_η and α_{min}. If only part (a) of the pricing error assumption holds, then for $\hat{\lambda}_{kn} \to_p \lambda_k$ it is required that $\alpha_\eta < \alpha_k$, for all k, which follows if $\alpha_\eta < \alpha_{min} = min(\alpha_1, \alpha_2, ..., \alpha_K)$. This condition weakens to $\alpha_\eta < 1$, if all the factors are strong (i.e. when $\alpha_k = 1$). If in addition the factor loadings and pricing errors are independently distributed (as in part b of Assumption 5), then for $\hat{\lambda}_{kn} \to_p \lambda_k$ it is required that $\alpha_\eta < (\alpha_k + \alpha_{min})/2$. Finally, $\hat{\lambda}_{kn} \to_p \lambda_k$ faster under part b since $\frac{\alpha_k + \alpha_{min}}{2} - \alpha_\eta > \frac{\alpha_k - \alpha_\eta}{2}$, when $\alpha_\eta < \alpha_{min}$. These results show that for identification of risk premia it is essential that the strength of the pricing error is less than the strength of the weakest factor.

The equilibrium pricing condition of Ross (1976), given by (10), is satisfied if $\alpha_\eta = 0$. It is clear that for λ_k to be $\sqrt{n}$-consistent we must have both $\alpha_k = 1$ and $\alpha_\eta = 0$, satisfied. More pervasive pricing errors and/or weaker factors combine to reduce the rate of convergence of $\hat{\lambda}_{kn}$ to λ_k. The above result also establishes that the risk premia associated with a factor whose strength is close to α_η is poorly identified at best. Even in the unlikely case of known factor loadings the rate at which $\hat{\lambda}_{kn}$ converges to λ_k can be painfully slow in the case of factors with strength close to α_η. For example, when only part a of Assumption 5 holds with $\alpha_\eta = 1/4$ (representing a moderate degree of pricing errors), identification of λ_k with strength $\alpha_k = 3/4$ implies a convergence rate of $n^{-1/4}$ which is very slow and requires a very large number of securities for a reasonably accurate estimation and inference.

We now consider how the above results are affected once we allow for estimation uncertainty (often referred as measurement errors) in the factor loadings in the more realistic case where $\mathbf{B}_n$ is estimated using a sample T observations on r_{it}, for $t = 1, 2, ..., T$.

4 Two-pass estimator of risk premia

In practice the factor loadings are unknown when the Fama and MacBeth (1973, FM) two-pass estimation procedure is used. In addition, the second pass regression uses average returns, $\bar{r}_{iT}$, that do not coincide with true mean returns $E(r_{it})$, when T is small. As noted in the introduction, the use of portfolio returns and their associated $\beta' s$ in the second pass does not alleviate the small T bias and in some settings could even accentuate it.

We now derive finite T large n asymptotic properties of the two-pass estimators of risk premia when one or more of the risk factors are not strong, namely when $\alpha_{min} < 1$, where $\alpha_{min} = min(\alpha_1, \alpha_2, ..., \alpha_K)$. The first pass of FM estimator estimates β_{ik} by running ordinary least squares (OLS) regressions of the individual (excess) returns, $\{r_{it} \, ; \, i = 1, 2, ..., n\}$ on an intercept and the same common factors, $\mathbf{f}_t$, over the time periods $t = 1, 2, ..., T$. These individual regressions can be written as

$$\mathbf{r}_{i\circ} = \mathrm{a}_i \boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \text{ for } i = 1, 2, ..., n, \tag{36}$$

where $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, and $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$. True values of the factor risk prices (or risk premia), $\boldsymbol{\lambda}$, are defined by the cross section regressions (CSR)

$$E(r_{it}) = \lambda_0 + \boldsymbol{\beta}_i' \boldsymbol{\lambda} + \eta_i, \text{ for } i = 1, 2, ..., n, \tag{37}$$

where η_i is the pricing error. It is also convenient to combine the individual return regressions as

$$\mathbf{r}_{\circ t} = \mathbf{a}_n + \boldsymbol{B}_n \mathbf{f}_t + \mathbf{u}_{\circ t}, \text{ for } t = 1, 2, ..., T, \tag{38}$$

where $\mathbf{r}_{\circ t} = (r_{1t}, r_{2t}, ..., r_{nt})'$ is an $n \times 1$ vector of excess returns on individual securities during period t, $\mathbf{a}_n = (\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_n)'$, $\mathbf{B}_n = (\boldsymbol{\beta}_{\circ 1}, \boldsymbol{\beta}_{\circ 2}, ..., \boldsymbol{\beta}_{\circ K})$, $\beta_{\circ k} = (\beta_{1k}, \beta_{2k}, ..., \beta_{nk})'$, and $\mathbf{u}_{\circ t} = (u_{1t}, u_{2t}, ..., u_{nt})'$.

The two-pass estimator is given by

$$\hat{\boldsymbol{\lambda}}_{nT} = \left(\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1} \hat{\mathbf{B}}'_{nT}\mathbf{M}_n\bar{\mathbf{r}}_n, \tag{39}$$

where $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}'_n$ as defined above, $\hat{\mathbf{B}}_{nT} = (\hat{\boldsymbol{\beta}}_{1,T}, \hat{\boldsymbol{\beta}}_{2,T}, ..., \hat{\boldsymbol{\beta}}_{n,T})'$, $\bar{\mathbf{r}}_n = (\bar{r}_{1\circ}, \bar{r}_{2\circ}, ..., \bar{r}_{n\circ})'$, $\bar{r}_{iT} = T^{-1}\sum_{t=1}^{T} r_{it}$,

$$\hat{\boldsymbol{\beta}}_{i,T} = (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{40}$$

$\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}'_T$, and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under (36), $\hat{\boldsymbol{\beta}}_{i,T} = \boldsymbol{\beta}_i + (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}$, and hence

$$\hat{\mathbf{B}}_{nT} = \boldsymbol{B}_n + \mathbf{U}_{nT}\mathbf{G}_T, \tag{41}$$

where $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, and $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}(\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}$. Also, averaging the return equations (36) over t for each i, we have

$$\bar{r}_{iT} = \mathrm{a}_i + \boldsymbol{\beta}'_i\bar{\mathbf{f}}_T + \bar{u}_{iT}, \text{ and } E(\bar{r}_i) = \mathrm{a}_i + \boldsymbol{\beta}'_i E\left(\bar{\mathbf{f}}_T\right), \tag{42}$$

where $\bar{\mathbf{f}}_T = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$, and $\bar{u}_{iT} = T^{-1}\sum_{t=1}^{T} u_{it}$. Hence, using the above results together with the APT condition given by (37), we have

$$\begin{aligned}\bar{\mathbf{r}}_n &= \lambda_0\boldsymbol{\tau}_n + \boldsymbol{B}_n(\boldsymbol{\lambda}+\mathbf{d}_{fT}) + \bar{\mathbf{u}}_n + \boldsymbol{\eta}_n \\ &= \boldsymbol{\mu}_n + \boldsymbol{B}_n\mathbf{d}_{fT} + \bar{\mathbf{u}}_n\end{aligned} \tag{43}$$

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

$$\boldsymbol{\mu}_n = \lambda_0\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\lambda} + \boldsymbol{\eta}_n, \tag{44}$$

and

$$\mathbf{d}_{fT} = \bar{\mathbf{f}}_T - E\left(\bar{\mathbf{f}}_T\right) = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right], \tag{45}$$

$\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, and $\boldsymbol{\eta}_n$ is the $n \times 1$ vector of pricing errors. Relations (41) and (43) can now be used in (39) to derive the asymptotic properties of $\hat{\boldsymbol{\lambda}}_n$.

5 Asymptotic properties of two-pass estimator in the case of short T and large n panels

As is well known when T is finite the two-pass estimator is biased due the errors in estimation of factor loadings that do not vanish. In our case the derivations are further complicated since we also allow for some of the factors not to be strong. Even if only one of the factors is not strong, namely if $\alpha_{min} < 1$, then $\lim_{n\to\infty}(n^{-1}\mathbf{B}'_n\mathbf{M}_n\mathbf{B}_n) = \boldsymbol{\Sigma}_{\beta\beta}$ will be rank deficient. It is easily seen that the rank $\boldsymbol{\Sigma}_{\beta\beta}$ will be equal to the number of strong factors. In the extreme case where none of the factors are strong we have $\boldsymbol{\Sigma}_{\beta\beta} = \mathbf{0}$. Interestingly enough, even in this case the two-pass estimator tends to a finite limit so long as T is fixed as $n \to \infty$. The following theorem provides the main asymptotic result when T is short.

Theorem 1 *(Small T bias of two-pass estimator of risk premia) Consider the multi-factor linear return model (38) and the associated risk premia,* $\boldsymbol{\lambda}$*, defined by (37), and suppose that Assumptions 1, 2, 4 and part (a) of Assumption 5, hold such that* $\alpha_\eta < \alpha_{min} = min(\alpha_1, \alpha_2, ..., \alpha_k)$*. Suppose further that* $\boldsymbol{\lambda}$ *is estimated by Fama-MacBeth two-pass estimator based on individual excess returns,* r_{it}*, and the factors,* $\mathbf{f}_t$*, for* $i = 1, 2, ..., n$*, and* $t = 1, 2, ..., T$*. Then for any fixed* $T > K + 1$ *we have (as* $n \to \infty$*)*

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda} \to_p \left[\boldsymbol{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Sigma}_{\beta\beta}\mathbf{d}_{fT} - \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}\right). \quad (46)$$

where $\hat{\boldsymbol{\lambda}}_n$ *is defined by (39) and*

$$\mathbf{d}_{fT} = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right], \ \boldsymbol{\Sigma}_{\beta\beta} = \lim_{n\to\infty}\left(\frac{\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n}{n}\right), \text{ and } \bar{\sigma}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 > 0. \quad (47)$$

The proof is provided in sub-section A.3.1 of the Appendix.

It is clear that the probability limit of $\hat{\boldsymbol{\lambda}}_{nT}$ exists even if $\boldsymbol{\Sigma}_{\beta\beta}$ is rank deficient. As an example, consider a two-factor case where only one of the factors is strong, namely $\alpha_1 = 1$ and $\alpha_2 < 1$. In this case

$$\begin{aligned}\frac{\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n}{n} &= \begin{pmatrix} n^{-1}\sum_{i=1}^{n}(\beta_{i1}-\bar{\beta}_1)^2 & n^{-1}\sum_{i=1}^{n}(\beta_{i1}-\bar{\beta}_1)(\beta_{i2}-\bar{\beta}_2) \\ n^{-1}\sum_{i=1}^{n}(\beta_{i1}-\bar{\beta}_1)(\beta_{i2}-\bar{\beta}_2) & n^{-1}\sum_{i=1}^{n}(\beta_{i2}-\bar{\beta}_2)^2 \end{pmatrix} \\ &= \begin{pmatrix} \ominus(1) & \ominus\left(n^{\frac{\alpha_2-1}{2}}\right) \\ \ominus\left(n^{\frac{\alpha_2-1}{2}}\right) & \ominus\left(n^{\alpha_2-1}\right) \end{pmatrix} \to_p \boldsymbol{\Sigma}_{\beta\beta} = \begin{pmatrix} \sigma_{\beta_1}^2 & 0 \\ 0 & 0 \end{pmatrix},\end{aligned}$$

where $\sigma_{\beta_1}^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(\beta_{i1}-\bar{\beta}_1)^2 > 0$. Let $\mathbf{A}_T = \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} = (a_{ij,T})$, and $\mathbf{d}_{fT} = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right] = (d_{fT,1}, d_{fT,2})'$, then using (46) and after some algebra we obtain

$$\hat{\boldsymbol{\lambda}}_{nT} = \begin{pmatrix} \hat{\lambda}_{1,nT} \\ \hat{\lambda}_{2,nT} \end{pmatrix} \to_p \begin{pmatrix} \frac{a_{22,T}\,\sigma_{\beta_1}^2(\lambda_1 + d_{fT,1})}{|\mathbf{A}_T| + a_{22,T}\sigma_{\beta_1}^2} \\ \frac{-a_{12,T}\sigma_{\beta_1}^2(\lambda_1 + d_{fT,1})}{|\mathbf{A}_T| + a_{22,T}\sigma_{\beta_1}^2} \end{pmatrix}.$$

For a finite T the estimate of risk premia of the weak factor, $\hat{\lambda}_{2,nT}$ tends to $-a_{12,T}\sigma_{\beta_1}^2(\lambda_1 + d_{fT,1})/\left(|\mathbf{A}_T| + a_{22,T}\sigma_{\beta_1}^2\right)$ which is proportional to λ_1, and does not converges to λ_2 even if we consider sufficiently large T. In fact the ratio $\hat{\lambda}_{2,nT}/\hat{\lambda}_{1,nT} \to_p -a_{12,T}/a_{22,T}$ which reduces to $Cov(f_{2t}, f_{1t})/Var(f_{1t})$ for T sufficiently large. In effect the risk premia estimated for the weak factor represent an spill over effect from the strong factor via the correlation of the underlying risk factors. In the special case where the two risk factors are uncorrelated $\hat{\lambda}_{2,nT} \to 0$.

In the extreme case where none of the factors are strong $\boldsymbol{\Sigma}_{\beta\beta} = \mathbf{0}$ and we have

$$\begin{aligned}\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda} &\to_p \left[\frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(-\frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}\right) \\ &= -\boldsymbol{\lambda},\end{aligned}$$

and we have $\hat{\boldsymbol{\lambda}}_{nT} \to_p \mathbf{0}$, irrespective of the true values of the risk premia!

5.1 Shanken bias-corrected two-pass estimator

The small T bias of the two-pass estimator of $\boldsymbol{\lambda}$ has been a source of concern in the empirical literature. As can be seen from (46) and (47) the bias of $\hat{\boldsymbol{\lambda}}_n$ is due to terms that involve $\mathbf{d}_T$ and $\overline{\sigma}^2$. Following Shanken (1992), $\overline{\sigma}^2$ can be consistently estimated (for a fixed $T > k+1$) by[6]

$$\widehat{\overline{\sigma}}^2_{nT} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\hat{u}_{it}^2}{n(T-k-1)}, \tag{48}$$

where $\hat{u}_{it} = r_{it} - \hat{\mathrm{a}}_{iT} - \hat{\boldsymbol{\beta}}'_{i,T}\mathbf{f}_t$, and $\hat{\mathrm{a}}_{iT}$ and $\hat{\boldsymbol{\beta}}_{i,T}$ are the OLS estimators of a_i and $\boldsymbol{\beta}_i$. Using this result the bias-corrected version of the two-pass estimator is given by:

$$\tilde{\boldsymbol{\lambda}}_n = \left[\frac{\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n} - T^{-1}\widehat{\overline{\sigma}}^2_{nT}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\bar{\mathbf{r}}_n}{n}\right). \tag{49}$$

To obtain the probability limit of $\tilde{\boldsymbol{\lambda}}_n$, we first note that since $\widehat{\overline{\sigma}}^2_{nT} \to_p \overline{\sigma}^2$, and $n^{-1}\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\hat{\mathbf{B}}_{nT} \to_p \boldsymbol{\Sigma}_{\beta\beta} + \frac{\overline{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}$, then

$$\frac{\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n} - T^{-1}\widehat{\overline{\sigma}}^2_{nT}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} \to_p \boldsymbol{\Sigma}_{\beta\beta}, \tag{50}$$

and the probability limit of $\tilde{\boldsymbol{\lambda}}_n$ exists if $\boldsymbol{\Sigma}_{\beta\beta}$ is full rank, which requires all the K factors to be strong. Note that the k^{th} diagonal element of $\boldsymbol{\Sigma}_{\beta\beta}$ is given by $plim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(\beta_{ik} - \bar{\beta}_k)^2$, which tends to zero if $\alpha_k < 1$.

Supposing that $\alpha_k = 1$ for all k, then for a fixed $T > K+1$ and as $n \to \infty$, we have (as required)

$$\tilde{\boldsymbol{\lambda}}_n \to_p \boldsymbol{\lambda}^*_T = \boldsymbol{\lambda} + \mathbf{d}_{fT}, \tag{51}$$

where $\mathbf{d}_T$ is defined by (45). Shanken refers to $\boldsymbol{\lambda}^*_T$ as "ex-post" risk premia to be distinguished from $\boldsymbol{\lambda}$, referred to as "ex ante" risk premia. See also section 3.7 of Jagannathan et al. (2010).

Remark 8 *In short, bias-correction of two-pass estimator is not innocuous, and pre-supposes that all the included factors are strong. The inclusion of a factor whose strength is below unity can lead to ill-defined bias-corrected estimates if a sufficiently large number of securities is considered. Ironically, this deficiency of two-pass bias-corrected estimators starts to show up only when n is sufficiently large and one or more of the factors is not strong.*

6 Asymptotic properties of the two-pass estimator when both n and T are large

It is clear that the large n asymptotic distribution of the two-pass estimators, whether bias-corrected or not, is not correctly centred when T is small. Furthermore, as argued above the validity of the bias-corrected two-pass estimator requires all the risk factors under consideration to be strong. Here we consider the asymptotic distribution of the two-pass estimator (without

[6] A simple proof of n consistency of of $\widehat{\overline{\sigma}}^2_{nT}$ for $\overline{\sigma}^2$ is provided in sub-section A.3.2 of the Appendix.

bias correction) when both n and T are large, whilst at the same time allowing for the possibility that one or more of the risk factors are not strong. Accordingly, we consider the following scaled version of the two-pass estimator given (39):

$$\mathbf{D}_n \hat{\boldsymbol{\lambda}}_{nT} = \mathbf{S}_{nT}^{-1} \mathbf{q}_{nt}, \tag{52}$$

where

$$\mathbf{S}_{nT} = \mathbf{D}_n^{-1} \hat{\mathbf{B}}_{nT}' \mathbf{M}_n \hat{\mathbf{B}}_{nT} \mathbf{D}_n^{-1},\ \mathbf{q}_{nT} = \left(\mathbf{D}_n^{-1} \hat{\mathbf{B}}_{nT}' \mathbf{M}_n \bar{\mathbf{r}}_n \right),$$

and $\mathbf{D}_n$ is the $K \times K$ diagonal matrix defined by (20). Write $\mathbf{S}_{nT}$ as

$$\mathbf{S}_{nT} = \mathbf{D}_n^{-1} \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_{nT} + \mathbf{B}_{nT} \right)' \mathbf{M}_n \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_{nT} + \mathbf{B}_{nT} \right) \mathbf{D}_n^{-1},$$

and recall from (41) that $\hat{\mathbf{B}}_{nT} - \boldsymbol{B}_n = \mathbf{U}_{nT} \mathbf{G}_T$, where $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, and $\mathbf{G}_T = \mathbf{M}_T \mathbf{F} \left(\mathbf{F}' \mathbf{M}_T \mathbf{F} \right)^{-1}$. Then we have

$$\begin{aligned} \mathbf{S}_{nT} = &\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} + \mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} \\ &+ \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \mathbf{D}_n^{-1} + \mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \mathbf{D}_n^{-1}. \end{aligned}$$

Similarly, using (43) we also have

$$\begin{aligned} \mathbf{q}_{nT} = &\left[\left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} \right) + \left(\mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} \right) \right] \mathbf{D}_n \boldsymbol{\lambda}_T^* \\ &+ \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\eta}_n + \mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \boldsymbol{\eta}_n \\ &+ \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \bar{\mathbf{u}} + \mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \bar{\mathbf{u}}_n. \end{aligned}$$

To derive the probability order of the two-pass estimator we need the following results established in Lemma A.3:

$$\mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} = O_p \left(T^{-1/2} n^{-a_{min}/2} \right), \tag{53}$$

$$\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \bar{\mathbf{u}}_n = O_p \left(T^{-1/2} \right), \tag{54}$$

$$\mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \bar{\mathbf{u}}_n = O_p \left(\frac{n^{(1-\alpha_{min})/2}}{T^{3/2}} \right) + O_p \left(n^{-1-\alpha_{min}/2} \right), \tag{55}$$

$$\mathbf{D}_n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \boldsymbol{\eta}_n = \left(\sqrt{\frac{n}{T}} \right) O_p \left[n^{-(\alpha_{min} - \alpha_\eta)/2} \right], \tag{56}$$

$$\mathbf{D}_n^{-1} \mathbf{G}_T' \left(\mathbf{U}_{nT}' \mathbf{M}_n \mathbf{U}_{nT} \right) \mathbf{G}_T \mathbf{D}_n^{-1} = O_p \left(\frac{n^{1-\alpha_{min}}}{T} \right). \tag{57}$$

Using the above results we have

$$\mathbf{S}_{nT} = \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} + O_p \left(T^{-1/2} n^{-a_{min}/2} \right) + O_p \left(\frac{n^{1-\alpha_{min}}}{T} \right), \tag{58}$$

and

$$\begin{aligned} \mathbf{q}_{nT} = &\left[\left(\mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n \mathbf{D}_n^{-1} \right) + O_p \left(T^{-1/2} n^{-a_{min}/2} \right) \right] \mathbf{D}_n \boldsymbol{\lambda}_T^* \\ &+ \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\eta}_n + \mathbf{D}_n^{-1} \mathbf{B}_n' \mathbf{M}_n \bar{\mathbf{u}}_n + O_p \left(\frac{n^{(1-\alpha_{min})/2}}{T^{3/2}} \right) \\ &+ O_p \left(n^{-1-\alpha_{min}/2} \right) + \left(\sqrt{\frac{n}{T}} \right) O_p \left[n^{-(\alpha_{min} - \alpha_\eta)/2} \right]. \end{aligned} \tag{59}$$

Also by Assumption 4, $\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1} \to_p \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha}) > \mathbf{0}$, as $n, T \to \infty$, where $\boldsymbol{\alpha}= (\alpha_1, \alpha_2, ..., \alpha_K)'$. Hence it follows that as n and $T \to \infty$, and $\mathbf{S}_{nT} \to \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha}) > \mathbf{0}$, and overall we have

$$\begin{aligned}\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right) &= \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n + \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_n \\ &+ O_p\left(n^{-1-\alpha_{min}/2}\right) + O_p\left(\frac{n^{(1-\alpha_{min})/2}}{T^{3/2}}\right) \\ &+ \left(\sqrt{\frac{n}{T}}\right) O_p\left(n^{-(\alpha_{min}-\alpha_\eta)/2}\right),\end{aligned} \tag{60}$$

where

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda} + \mathbf{d}_{fT} = \boldsymbol{\lambda} + \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]. \tag{61}$$

6.1 Asymptotic distribution of two-pass estimator under zero pricing errors

The asymptotic distribution of two-pass estimator critically depends on assumptions made about the pricing errors, η_n. In the case where pricing errors are ignored the expression for $\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right)$ reduces to

$$\begin{aligned}\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right) &= \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\bar{\mathbf{u}}_n + O_p\left(\frac{n^{(1-\alpha_{min})/2}}{T^{3/2}}\right) + O_p\left(n^{-1-\alpha_{min}/2}\right) \\ &= o_p(1).\end{aligned}$$

where $o_p(1)$ here refers to an amalgam of terms that tend to 0 in probability. To obtain a non-degenerate distribution we need scale up both sides by $\sqrt{T}$ which yields

$$\sqrt{T}\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right) = \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\alpha)\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\sqrt{T}\bar{\mathbf{u}}_n\right) + O_p\left(\frac{\sqrt{T}}{n^{1+\alpha_{min}/2}}\right) + O_p\left(\frac{n^{(1-\alpha_{min})/2}}{T}\right). \tag{62}$$

The last two $O_p(.)$ terms tend to zero as $n, T \to \infty$, so long as $n/T \to \kappa$ where $c < \kappa < C$. Also note that

$$\left\|\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\right\| = \lambda_{max}^{1/2}\left[\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\right] \to_p \lambda_{max}^{1/2}\left[\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\right] < C,$$

and hence

$$T^{1/2}n^{\alpha_k/2}\left(\hat{\lambda}_{k,nT} - \lambda_{kT}^*\right) = O_p(1), \text{ for } k = 1, 2, ..., K. \tag{63}$$

This is a generalization of the second part of Theorem 1 in Anatolyev and Mikusheva (2020). They use a drifting parameter model similar to that used for weak instruments in which the loadings, β_{ik}, of the weak factor drift to zero at rate $\sqrt{T}$ to model the near degenerate rank. Since they also assume n grows with T, this is equivalent to $\sqrt{n}$ and corresponds to the special case of our models with $\alpha_k = 0.5$. They only consider strong ($\alpha_k = 1$) and weak ($\alpha_k = 1/2$) factors, though they also have other terms that come from missing factors.

Under our assumptions and in absence of any pricing errors we obtain

$$\sqrt{T}\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT}-\boldsymbol{\lambda}_T^*\right)\to_d N\left(\mathbf{0},\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{H}_{\beta\beta}(\boldsymbol{\alpha})\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\right), \tag{64}$$

where

$$\mathbf{H}_{\beta\beta}(\boldsymbol{\alpha})=plim_{n,T\to\infty}\left(\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}\right),\text{ and }\mathbf{V}_u=(\sigma_{ij}). \tag{65}$$

The above result holds for different degrees of factor strength and reduces to familiar results when $\alpha_k=1$, namely

$$\sqrt{nT}\left(\hat{\boldsymbol{\lambda}}_{nT}-\boldsymbol{\lambda}_T^*\right)\to_d N\left(\mathbf{0},\boldsymbol{\Sigma}_{\beta\beta}^{-1}\mathbf{H}_{\beta\beta}\boldsymbol{\Sigma}_{\beta\beta}^{-1}\right),$$

where $\mathbf{H}_{\beta\beta}=plim_{n,T\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\right)$.

A consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\lambda}}_{nT}-\boldsymbol{\lambda}_T^*$ is also given by

$$\widehat{Var}\left(\hat{\boldsymbol{\lambda}}_{nT}-\boldsymbol{\lambda}_T^*\right)=T^{-1}\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\left(\hat{\mathbf{B}}_n'\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_n\right)\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}, \tag{66}$$

where $\tilde{\mathbf{V}}_u=(\tilde{\sigma}_{ij})$ represents a consistent estimator of $\mathbf{V}_u$. Since $\mathbf{V}_u$ is assumed to be row (column) bounded, it can be consistently estimated using the various thresholding procedures advanced in the statistical literature. See, for example, Bickel and Levina (2008a, 2008b) and Cai and Liu (2011). Here we suggest the threshold estimator proposed by Bailey, Pesaran and Smith (2019, BPS), which does not require cross-validation and is shown to have desirable small sample properties. Accordingly, we propose the following threshold estimator

$$\begin{aligned}&\tilde{\sigma}_{ii}=\hat{\sigma}_{ii}\\&\tilde{\sigma}_{ij}=\hat{\sigma}_{ij}\mathbf{1}\left[|\hat{\rho}_{ij}|>T^{-1/2}c_\alpha(n)\right],\ i=1,2,\ldots,n-1,\ j=i+1,\ldots,n,\end{aligned} \tag{67}$$

where

$$\hat{\sigma}_{ij}=\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt},\ \hat{\rho}_{ij}=\frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}},\quad \hat{u}_{it}=r_{it}-\hat{a}_{i,T}-\hat{\boldsymbol{\beta}}_{i,T}'\mathbf{f}_t, \tag{68}$$

and $c_p(n)=\Phi^{-1}\left(1-\frac{p}{2n^\delta}\right)$, is a normal critical value function, p is the the nominal size of testing of $\sigma_{ij}=0$, $(i\neq j)$ and δ 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 $\delta=2$. The variance estimator given by (66) does not require a knowledge of the factor strength and applies to risk factors of differing degrees. It can be shown that $\left\|\tilde{\mathbf{V}}_u-\mathbf{V}_u\right\|=O_p\left(\frac{ln(n)}{\sqrt{T}}\right)$.

As it stands the asymptotic distribution in (64) is very encouraging, but it is important to note that the fast $\sqrt{n^{\alpha_k}T}$ convergence rate obtained for $\left(\hat{\lambda}_{k,nT}-\lambda_{kT}^*\right)$ does not carry over to the risk premia, λ_k, which is the primary object of interest. To this end, using (61) in (62) we observe that

$$\begin{aligned}\sqrt{T}\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT}-\boldsymbol{\lambda}\right)=&\ \mathbf{D}_n\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathbf{f}_t-E(\mathbf{f}_t)\right]\right)+\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\sqrt{T}\bar{\mathbf{u}}_n\right)\\&+O_p\left(\frac{\sqrt{T}}{n^{1+\alpha_{min}/2}}\right)+O_p\left(\frac{n^{(1-\alpha_{min})/2}}{T}\right),\end{aligned} \tag{69}$$

where under Assumption 1 we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right] = O_p(1)$. Now that the distribution of the two-pass estimator is centered correctly around $\boldsymbol{\lambda}$, the first term on the right hand side of (69) is unbounded in n, and will eventually dominate the remaining terms. To avoid this outcome we need to multiply both sides of (69) by $\mathbf{D}_n^{-1}$ to obtain

$$\sqrt{T}\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right] + \mathbf{D}_n^{-1}\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\alpha)\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\sqrt{T}\overline{\mathbf{u}}_n\right) + o_p(1), \tag{70}$$

where

$$\begin{aligned}\left\|\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\alpha)\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\sqrt{T}\overline{\mathbf{u}}_n\right)\right\| &\leq \left\|\mathbf{D}_n^{-1}\right\|\left\|\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\alpha)\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\left(\sqrt{T}\overline{\mathbf{u}}_n\right)\right\| \\ &= O_p\left(n^{-\alpha_k/2}\right),\end{aligned}$$

and hence we end up with

$$\sqrt{T}\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right] + o_p(1).$$

Interestingly enough, when pricing errors are ignored the asymptotic distribution of the two-pass estimator around $\boldsymbol{\lambda}$ is not affected by factor strengths and is primarily governed by the distribution of the risk factors around their means. The asymptotic distribution of $T^{-1/2}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]$ can be obtained using standard results from covariance stationary multivariate time series literature. Also see Theorem 2 below.

6.2 Asymptotic distribution of two-pass estimator allowing for pricing errors

In the more realistic case where there are pricing errors, using (60) and noting that $\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\overline{\mathbf{u}}_n = O_p\left(T^{-1/2}\right)$, we have

$$\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right) = \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n + o_p(1), \tag{71}$$

assuming $\alpha_{min} > \alpha_\eta$, and if $n/T \to \kappa < C$ as $n, T \to \infty$. The probability order of $\mathbf{D}_n^{-1}\mathbf{B}_n'\mathbf{M}_n\boldsymbol{\eta}_n$, is given by (32) and (33), depending on whether part (a) or part (b) of the pricing error Assumption 5 is met. The weaker part (a) allows consistent estimation of the risk premia, λ_k, but at the rather slow rate of $n^{-\left(\frac{\alpha_k-\alpha_\eta}{2}\right)}$, assuming $\alpha_{min} > \alpha_\eta$ and if n and $T \to \infty$ such that n/T converges to a finite constant (inclusive of zero). The convergence of $\hat{\boldsymbol{\lambda}}_{nT}$ to $\boldsymbol{\lambda}$ is ensured since under Assumption 1 we have $\mathbf{d}_{ft} = O_p\left(T^{-1/2}\right)$. However, the asymptotic distribution of $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*$ will be dominated by the pricing error, unless stronger conditions in part (b) of Assumption 5 is made. Under part (b) of Assumption 5 we have (using (33))

$$\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right) = O_p\left(n^{\alpha_\eta - \alpha_{min}/2}\right) + o_p(1), \tag{72}$$

and the effects of the pricing errors on the asymptotic distribution of the two-pass estimator of risk premia vanish if $\alpha_\eta < \alpha_{min}/2$. By definition (see (61) we have

$$\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^*\right) = \mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}\right) - T^{-1}\mathbf{D}_n\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right],$$

which in conjunction with (72) yields

$$\mathbf{D}_n\left(\hat{\boldsymbol{\lambda}}_{nT}-\boldsymbol{\lambda}\right)=T^{-1}\mathbf{D}_n\sum_{t=1}^{T}\left[\mathbf{f}_t-E(\mathbf{f}_t)\right]+o_p(1).$$

Hence, for the two-step estimator of the k^{th} risk premia we have

$$n^{\alpha_k/2}\left(\hat{\lambda}_{k,nT}-\lambda_k\right)=\left(\sqrt{\frac{n^{\alpha_k}}{T}}\right)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[f_{kt}-E(f_{kt})\right]\right]+o_p(1). \tag{73}$$

Therefore, under joint n and T asymptotic $\hat{\lambda}_{k,nT}$ is correctly centred but its limiting distribution depends on the strength of its underlying risk factor. When $\alpha_k<1$, then $\frac{n^{a_k}}{T}\to 0$ as n and $T\to\infty$ and we end up with a degenerate distribution for $\hat{\lambda}_{k,nT}$. Only for strong factors with $\alpha_k=1$ the asymptotic distribution is non-degenerate, and can be written as

$$n^{1/2}\left(\hat{\lambda}_{k,nT}-\lambda_k\right)=\left(\sqrt{\frac{n}{T}}\right)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[f_{kt}-E(f_{kt})\right]\right]+o_p(1),$$

or equivalently (after post multiplication by $\sqrt{T/n}$)

$$\sqrt{T}\left(\hat{\lambda}_{k,nT}-\lambda_k\right)=\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[f_{kt}-E(f_{kt})\right]\right]+o_p(1).$$

The following theorem gives the asymptotic distribution for the case where a subset of K_s factors, denoted by $\mathbf{f}_{st}$, are strong.

Theorem 2 *Consider the multi-factor linear return model (38) and the associated risk premia, $\boldsymbol{\lambda}$, defined by (37), and suppose that K_s of K factors are strong, and the remaining $K-K_s$ factors are weak, such that $\boldsymbol{\alpha}=(\boldsymbol{\tau}'_{K_s},\alpha_{K_s+1},\alpha_{K_s+2},....,\alpha_K)'$, where $\boldsymbol{\tau}_{K_s}$ is a $K_s\times 1$ of ones, and $\alpha_k<1$ for $k=K_s+1,K_s+2,....,K$. Further suppose that Assumptions 1, 2, 4, and part (b) of Assumption 5 hold, and $\alpha_\eta<a_{min}/2$, where $\alpha_{min}=min(\alpha_{K_s+1},\alpha_{K_s+2},....,\alpha_K)$. Denote the strong factors by $\mathbf{f}_{st}$ and let $\mathbf{V}_s(p)=E\left[(\mathbf{f}_{st}-\mu_{sf})(\mathbf{f}_{s,t-p}-\mu_{sf})'\right]$, such that the long run covariance matrix defined by*

$$\mathbf{V}_s=\mathbf{V}_s(0)+\sum_{p=1}^{\infty}\left(\mathbf{V}_s(p)+\mathbf{V}'_s(p)\right), \tag{74}$$

is positive definite. Denote the two-pass estimator of the risk premia associated to the K_s strong factors by $\hat{\boldsymbol{\lambda}}_{s,nT}$, and further suppose that n and $T\to\infty$ such that $n/T\to\kappa$, with $0<\kappa<C$. Then

$$\sqrt{T}\left(\hat{\boldsymbol{\lambda}}_{s,nT}-\boldsymbol{\lambda}_s\right)\to_d N(\mathbf{0},\mathbf{V}_s). \tag{75}$$

The proof of this theorem follows directly from (73) and the application of standard results from stationary time series processes applied to $T^{-1/2}\sum_{t=1}^{T}\left[\mathbf{f}_{st}-E(\mathbf{f}_{st})\right]$. Also $\mathbf{V}_s$ can be consistently estimated by

$$\hat{\mathbf{V}}_s=\hat{\mathbf{V}}_s(0)+\sum_{p=1}^{m}b(p,m)\left(\hat{\mathbf{V}}_s(p)+\hat{\mathbf{V}}'_s(p)\right),$$

where $\hat{\mathbf{V}}_s(p) = T^{-1}\sum_{t=p+1}^{T}\left(\mathbf{f}_{st} - \bar{\mathbf{f}}_{sT}\right)\left(\mathbf{f}_{s,t-p} - \bar{\mathbf{f}}_{sT}\right)'$, $\bar{\mathbf{f}}_{sT} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{st}$, $b(p,m)$ is the kernel or lag window, and m is the bandwidth. This is a standard HAC estimator where the kernel and bandwidth must be chosen carefully to ensure that m/T tends to zero at a sufficiently fast rate, and $\hat{\mathbf{V}}_s$ is invertible.

7 Asymptotic properties of pooled R squared

A common procedure to judge the effectiveness of a factor model is to look at the fit, measured by R^2, of regressions explaining returns. There are a number of ways this can be done. One possibility is to use the average the R^2 of the first pass regression. Fama and French (2015p12) report average R^2 for the 5 factor model for various ways of constructing portfolios. But averages are sensitive to outliers. Another possibility is the R^2 of the second pass Fama-MacBeth regression. Kleibergen and Zhan (2015) consider the large sample distributions of this R^2 when the observed proxy factors are weakly correlated with the true unobserved factors. This implies an unexplained factor structure in the first pass residuals; a large estimation error in the estimated beta's; and possibly large spurious values of the second pass R^2. A third possibility is the pooled R^2 which we consider.

Consider first stage regressions of returns r_{it} on a vector of K factors $\mathbf{f}_t$, for securities $i = 1, 2, ..., n$ and time periods $t = 1, 2, ..., T$

$$r_{it} = \mathrm{a}_i + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it}$$

average these over time, where $\bar{r}_{iT} = \sum_{t=1}^{T} r_{it}$ and $\bar{u}_{iT} = \sum_{t=1}^{T} u_{it}$:

$$\bar{r}_{iT} = \mathrm{a}_i + \boldsymbol{\beta}_i'\bar{\mathbf{f}}_T + \bar{u}_{iT}.$$

The pooled R squared is then given by

$$PR^2 = 1 - \frac{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} u_{it}^2}{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\left(r_{it} - \bar{r}_{iT}\right)^2}.$$

Under our assumptions, as shown in Appendix A.3.3, it follows that (for both n and T large)

$$PR_{nT}^2 = \frac{n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_{fT}\boldsymbol{\beta}_i/\bar{\sigma}_n^2 + o_p(1)}{1 + n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_{fT}\boldsymbol{\beta}_i/\bar{\sigma}_n^2 + o_p(1)}.$$

Hence, the order of PR_{nT}^2 is governed by the pooled signal-to-noise ratio defined by

$$s_{nT}^2 = \frac{n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_{fT}\boldsymbol{\beta}_i}{\bar{\sigma}_n^2},$$

and has the same expansion rate as $\sum_{k=1}^{K}\left(n^{-1}\sum_{i=1}^{n}\beta_{ik}^2\right)$. Therefore, the limit of PR_{nT}^2 is dominated by the strongest factor. For example, in the case where only one of the factors is strong, say f_{1t}, then $\alpha_1 = 1$ and $\alpha_k < 1$ for $k = 2, 3, ...K$, and we have

$$PR_{nT}^2 \to \frac{\left(\frac{Var(f_{1t})}{\bar{\sigma}_n^2}\right)\left[n^{-1}lim_{n\to\infty}\sum_{i=1}^{n}\beta_{i1}^2\right]}{1 + \left(\frac{Var(f_{1t})}{\bar{\sigma}_n^2}\right)\left[n^{-1}lim_{n\to\infty}\sum_{i=1}^{n}\beta_{i1}^2\right]} > 0.$$

where by assumption $n^{-1}lim_{n\to\infty}\sum_{i=1}^{n}\beta_{i1}^2 > 0$. This finding is in line with the result of Theorem 2 that only risk premia of strong factors can be identified. The effects of factors with strength below unity will vanish eventually. Therefore, the use of semi-strong or weak factors in estimation of risk premia can be justified only by appeal to market inefficiency that might prevail when the number of available securities is not sufficiently large and the effects of semi-strong factors do not vanish and could be contributing to the pooled R^2.

8 Empirical evidence on the strength of the Fama-French 5 risk factors

The earlier sections demonstrated the importance of knowing the strength of the factors both for assessing the relative importance of pricing errors and for the convergence properties of the estimators of the risk premia in terms of n and T. This section uses the procedure suggested by Bailey, Kapetanios and Pesaran (2021, BKP) to estimate the strength of the five factors suggested by Fama and French (2015), which have been used extensively in the finance literature. We first provide a brief overview of the BKP estimation method.

8.1 Estimation Procedure

BKP propose estimating the strength of a factor from the proportion of the n securities in which the factor loading is significant in the first pass time-series regression. Denote by t_{ik} the t-statistic of the estimated loading of factor k for security i and consider the proportion of regressions where the estimated factor loadings, β_{ik}, is statistically significant:

$$\hat{\pi}_k = \frac{\sum_{i=1}^{n}\mathbf{1}\left[|t_{ik}| > c_p(n)\right]}{n}, \tag{76}$$

where $\mathbf{1}(A) = 1$ if $A > 0$, and zero otherwise. To control for the multiple testing problem, the critical value function, $c_p(n) = \Phi^{-1}\left(1 - \frac{p}{2n^c}\right) = \ominus\left(\sqrt{ln(n)}\right)$, is used, where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the standard normal distribution, p is the nominal size of the multiple tests, and c is a small positive constant that controls the overall size of the multiple tests and ensures the consistency of the estimator of α_k. BKP use $c = 0.25$ and $p = 0.1$. Monte Carlo results indicate that the estimates of factor strength do not seem to be very sensitive to the choice of c, and even less so to the choice of p. Note that $lim_{n\to\infty}c_p^2(n)/ln(n) = 2c$, and does not depend on p.

The estimator of the strength of factor k, is given by

$$\hat{\alpha}_k = 1 + ln(\hat{\pi}_k)/ln(n). \tag{77}$$

BKP derive its asymptotic distribution and give analytical expressions for its asymptotic standard errors for values of α_k in the range $1/2 < \alpha_k < 1$. The confidence intervals become quite narrow as α_k gets closer to unity. When $\alpha_k = 1$, the distribution of $\hat{\alpha}_k$ tends to its true value of 1, at an exponential rate, a convergence rate they label as ultra consistency.

The relationship between α_k and π_k in (77) can be used to examine the association between α_k and the amount of information in the sample about the factor. With $n = 400$ securities the condition $\alpha_k < 1/2$ implies that the factor loading is non-zero in around 20 out of 400 securities.

The fact that around 380 out of the 400 beta estimates are zero in the second pass regression makes it clear that factors with $\alpha_k < 1/2$ cannot be distinguished from the idiosyncratic errors. In comparison, in the case of a factor with strength $\alpha_k = 0.9$, it will have non-zero loadings in around 220 of the 400 securities under consideration. As α_k approaches unity, almost all securities will be affected by the factor.

Factor strength is determined by the number as well as the size of non-zero loadings, β_i. While BKP focus on the number of non zero factors, other papers assume that

$$\beta_i = \beta_{in} = \delta_i / n^{(1-\alpha)/2} \tag{78}$$

with $\delta_i > c > 0$ for all i. See, for example, Kleibergen (2009) and Onatski (2012) who consider the above specification with $\alpha = 1/2$. This specification requires *all* factor loadings to decline at the *same* rate with n, when $\alpha < 1$. But uniformly declining values for β_{in}, as n increases, make little empirical sense and there does not seem to be any suggestions in the literature on how to estimate α under (78). The BKP procedure, of using the number of non-zero factor loadings to determine factor strength, is empirically more sensible and easier to implement than (78).

8.2 The strength of the Fama-French factors

Our empirical application uses the BKP procedure to obtain rolling estimates of the strength of the five Fama and French (2015) factors, FF5. The loadings are obtained from regressions in which all five factors are included. We use all stocks in the S&P 500 portfolio that have at least 10 years of return history, for each month from September 1989 to May 2018. The list is updated monthly and includes at least 400 stocks, with an average number of 442 stocks. This procedure avoids the possible survivorship bias caused by the changing composition of S&P 500 portfolio.

The data for the FF5 factors are taken from Kenneth French's web pages.[7] The factors are: market, size, value, profitability, and investment. The market factor, the excess market return, differs from the average of the roughly 400 stocks we consider. In particular, it is value weighted and has a much wider coverage. It includes all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have data for that month. The risk free rate is the one-month Treasury bill rate.

In addition to the market factor, the original three Fama and French (1993) factors included a size factor, SMB, small minus big measured as the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, and a value factor, HML, high minus low measured as the return on a portfolio of high book to market stocks minus the return on a low portfolio. To these three, Fama and French (2015) added a profitability factor, RMW measured as the difference between the returns on portfolios of stocks with robust and weak profitability, and an investment factor, CMA measured as the difference between portfolios of low and high investment firms, which they call conservative and aggressive.

The time series regressions are of the form:

$$r_{it} - r_t^f = \mathrm{a}_i + \sum_{k=1}^{5} \beta_{ik} f_{kt} + u_{it}, \tag{79}$$

[7] The FF factors are obtained from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Excess returns on individual securities were originally compiled by Takashi Yamagata and extended to May 2018 by Natalia Bailey. For further details see Appendix C of Pesaran and Yamagata (2018).

where the excess return for each stock, $i = 1, 2, ...n$, is regressed on a constant and the five factors over rolling ten year samples, $\tau = 1, 2, ..., 345$, where τ denotes the last month of the ten year rolling sample from September 1989 to May 2018. The t statistics for the hypotheses $\beta_{ik} = 0$ are then used to calculate (76) and (77).

Figure 1 plots the 10 year rolling estimates, $\widehat{\alpha}_{\tau k,}$ for k = Mkt, SMB, HML, RMW, and CMA. The market factor always has a strength that is either one or very close to it, with little variation over time. The other four factors are much weaker with a great deal of variation over time. The time average of the rolling estimates of the strength of five factors in order of strength are 0.994 for the market factor, 0.725 for SMB, 0.739 for HML, 0.722 for RMW, and 0.622 for CMA. The lowest value for the strength of the market factor is 0.982 in August 2001, the highest value for the strength of any other factor is 0.821 for RMW in October 2001.

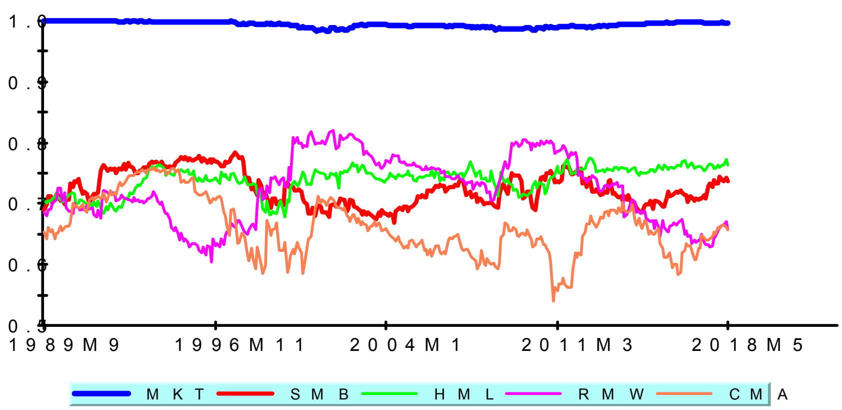

Figure 1: Rolling estimates of factor strength for the five Fama-French factors

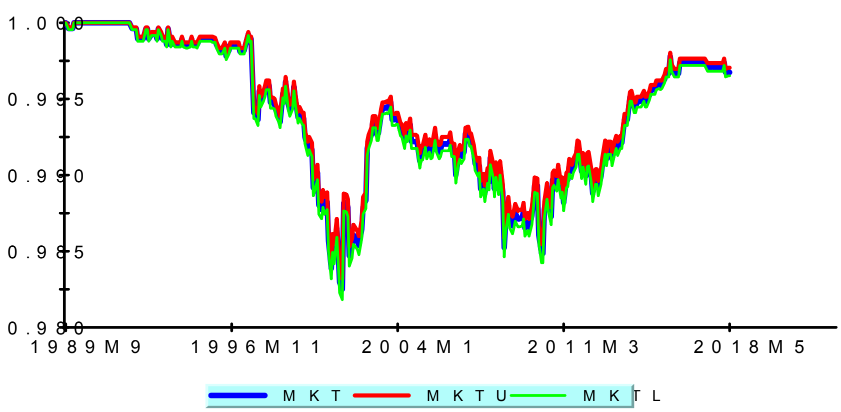

Figure 2: Rolling estimates of factor strength for the market factor with its 95 percent error band

The time profile of the rolling estimates of factor strengths together with their 95 per cent standard error bands are given in Figures 2-6. As can be seen, the error band for the market factor is very tight indeed, such that in many instances the upper and lower bounds almost coincide, reflecting the ultra-consistency property of the estimator when its true value is close to unity. The error bands for other factors are wider given that they are weaker in strength, nevertheless they are still reasonably tight. This suggest that the factor strengths are estimated fairly precisely, and their variations cannot be attributed to sampling variation alone. The results also suggest that the importance of the additional four factors is subject to substantial structural change, which adds another dimension to risk analysis. For instance the lowest strength for any factor is for CMA in January 2011 at 0.54, with a 95% confidence interval between 0.51 and 0.57; a small interval compared to its time variation. In general, higher values of strength are associated with smaller confidence intervals.

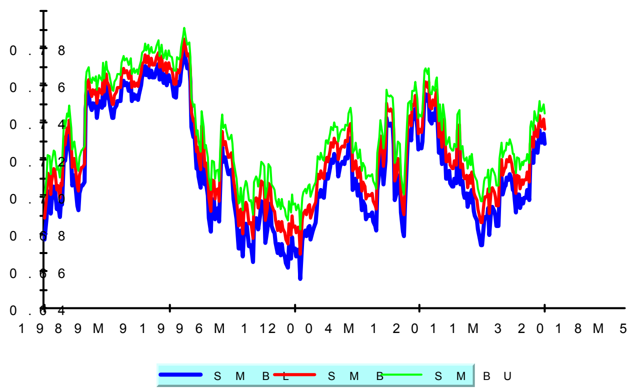

Figure 3: Rolling estimates of factor strength for the SMB factor with its 95 percent error band

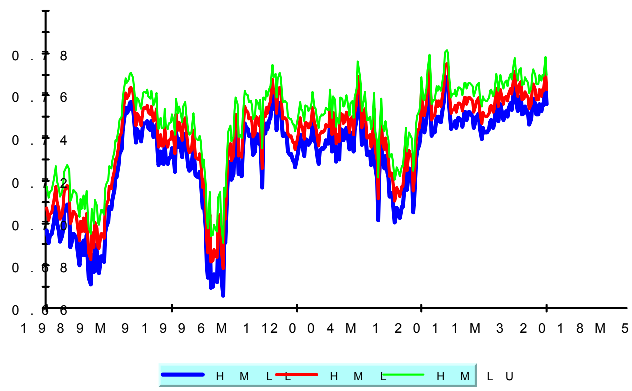

Figure 4 Rolling estimates of factor strength for the HML factor with its 95 percent error band

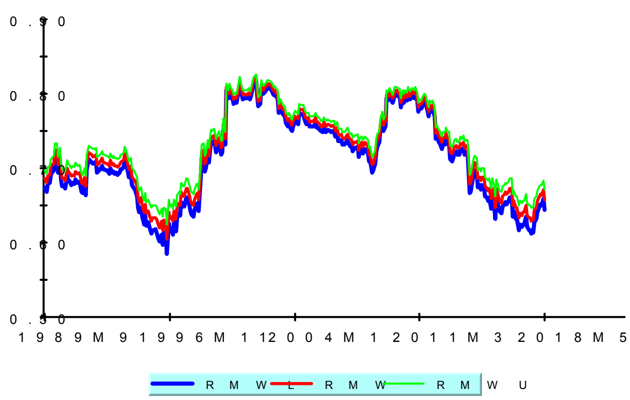

Figure 5: Rolling estimates of factor strength for the RMW factor with its 95 percent error band

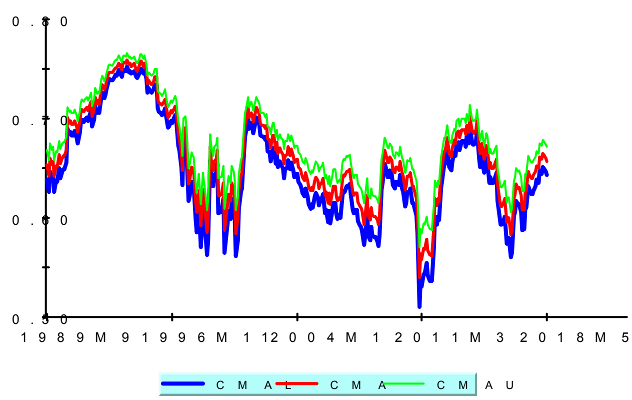

Figure 6: Rolling estimates of factor strength for the CMA factor with its 95 percent error band

BKP follow a similar procedure but use a different first pass regression. Different first pass regressions will produce different proportions of significant coefficients and thus different estimates of factor strength.[8] Whereas we use all five factors in the first pass regression, they used just two factors in each first pass regression. These were the market factor and one other factor taken in turn from the 145 factors considered by Feng et al. (2020). The estimates they obtain are very similar. BKP also find that the market factor is the only strong factor, never falling below 0.95 in strength, while none of the other 144 factors considered has strength exceeding 0.95. After the market factor, with a time averaged strength of 0.99, the next strongest factor is leverage, which has a time averaged strength of 0.827.

9 Concluding remarks

The Fama-MacBeth two-pass estimator has been routinely applied to multi-factor models to estimate risk premia. This literature typically assumes that all the risk factors under consideration are strong, and that there are no pricing errors. In this paper we extend the analysis by considering risk factors that are not strong, whilst at the same time allowing for a less restrictive degree of pricing errors than the bounded condition assumed by Ross in his theoretical derivations of APT.

The main message of our analysis for empirical asset pricing is that pricing errors and factor strength matter for consistent estimation of risk premia and subsequent inference, thus an estimate of factor strength is required before attempting to estimate risk premia by the two pass method. The method advanced by BKP (2021) can be used to estimate factor strengths, which then allows investigators to focus their analysis on the strong factors, whilst including additional semi-strong factors to reduce residual cross-correlations. It is possible to allow for a moderate degree of pricing error so long as the risk factors under consideration are sufficiently strong.

The theoretical analysis highlights the importance for the estimation of risk premia of having panels with sufficiently large n and T. This is likely to pose important challenges in practice, since factor loadings are likely to be subject to a significant degree of time variation. We have assumed K the number of possible factors is given and have not discussed how one selects factors, but given our results it is clear that factor selection and measurement of factor strength need to be done jointly.

[8] However, the Monte Carlo experiment 4 in BKP suggests that the estimates of α_k for strong factors will remain close to unity irrespective of the model specification.

A Mathematical Appendix

A.1 Introduction

We first state and establish a number of lemmas and provide a proof of Theorem 1 in the paper.

A.2 Statement of lemmas and their proofs

Lemma A.1 *Consider the errors* $\{u_{it}, i = 1, 2, ..., n; t = 1, 2, ..., T\}$ *in the factor model defined by (38), and suppose that Assumption 2 holds. Then for any* t *and* t' *(as* $n \to \infty$*)*

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

$$b_{n,t} = \frac{1}{n}\sum_{i=1}^{n}\left(u_{it}^2 - \sigma_i^2\right) \to_p 0, \text{ for } t = t', \tag{A.2}$$

$$c_{n,t} = \frac{1}{n}\sum_{i=1}^{n}(u_{it}\bar{u}_{i\circ} - \frac{1}{T}\sigma_i^2) \to_p 0, \tag{A.3}$$

and

$$Var(c_{n,t}) = \frac{1}{T^2}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(u_{it}^2, u_{jt}^2)\right] = O\left(n^{-1}T^{-2}\right), \tag{A.4}$$

where $\sigma_i^2 = E(u_{it}^2)$*, and* $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$.

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

$$\begin{aligned} Var(a_{n,tt'}) &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} E\left(u_{it}u_{it'}u_{jt}u_{jt'}\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\sigma_{ij}^2 \leq \frac{1}{n^2}\sum_{j=1}^{n}\left(\sum_{i=1}^{n}|\sigma_{ij}|\right)^2. \end{aligned}$$

Since by Assumption 2, $\sup_j \sum_{i=1}^{n}|\sigma_{ij}| < C$, then

$$Var(a_{n,tt'}) \leq \frac{1}{n}\sup_j \sum_{i=1}^{n}|\sigma_{ij}| = O\left(n^{-1}\right),$$

which establishes (A.1). Similarly, since $E\left(u_{it}^2 - \sigma_i^2\right) = 0$, then $E(b_{n,t}) = 0$ and

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

which tends to zero by part (b) of Assumption 2, and result (A.2) is established. To prove (A.5) set $z_{it} = u_{it}\overline{u}_{i\circ} - \frac{1}{T}\sigma_i^2$, and note that $u_{it}\overline{u}_{i\circ} = \frac{1}{T}\sum_{s=1}^{T} u_{it}u_{is}$, and given that $\{u_{it}\}$ is serially uncorrelated then $E(z_{it}) = 0$ and we have $E(c_{n,t}) = 0$. Also

$$Var(c_{n,t}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(z_{it}, z_{jt}) \tag{A.5}$$

and

$$Cov(z_{it}, z_{jt}) = E(u_{it}u_{jt}\overline{u}_{i\circ}\overline{u}_{j\circ}) - \frac{1}{T^2}\sigma_i^2\sigma_j^2.$$

Further

$$E\left(u_{it}u_{jt}\overline{u}_{i\circ}\overline{u}_{j\circ}\right) = \frac{1}{T^2}E\left(u_{it}u_{jt}\sum_{s=1}^{T}\sum_{s'=1}^{T} u_{is}u_{js'}\right),$$

and since $\{u_{it}\}$ is serially uncorrelated, then $E\left(u_{it}u_{jt}\overline{u}_{i\circ}\overline{u}_{j\circ}\right) = \frac{1}{T^2}E(u_{it}^2u_{jt}^2)$, which yields

$$Cov(z_{it}, z_{jt}) = \frac{1}{T^2}\left[E(u_{it}^2u_{jt}^2) - \frac{1}{T^2}\sigma_i^2\sigma_j^2\right] = \frac{1}{T^2}Cov(u_{it}^2u_{jt}^2).$$

Using this result in (A.5) we have

$$Var(c_{n,t}) = \frac{1}{T^2}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(u_{it}^2, u_{jt}^2)\right].$$

Therefore, under part (b) of Assumption (2), it follows that $Var(c_{n,t}) = O\left(n^{-1}T^{-2}\right)$, as required. Result (A.3) also follows by Markov inequality. ■

Lemma A.2 *Consider the* $n\times T$ *error matrix* $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, *where* $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$ *,and the* $n \times 1$ *vector of pricing errors* $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$*, and suppose that Assumptions 2 and 5 hold, with* $\alpha_\eta < 1/2$*. Then for a fixed* T *and as* $n \to \infty$ *we have*

$$\frac{\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}}{n} \to_p \overline{\sigma}^2\mathbf{I}_T, \tag{A.6}$$

$$\frac{\mathbf{U}_{nT}'\mathbf{M}_n\overline{\mathbf{u}}}{n} \to_p \frac{\overline{\sigma}^2}{T}\boldsymbol{\tau}_T, \tag{A.7}$$

$$\frac{\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n}{n} \to_p \mathbf{0} \tag{A.8}$$

where $\mathbf{M}_n = \mathbf{I}_n - \frac{1}{n}\boldsymbol{\tau}_n\boldsymbol{\tau}_n'$, $\overline{\mathbf{u}} = (\overline{u}_{1\circ}, \overline{u}_{2\circ}, ..., \overline{u}_{n\circ})'$, $\overline{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$, $\overline{\sigma}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$, *and* $\boldsymbol{\tau}_n$ *and* $\boldsymbol{\tau}_T$ *are, respectively,* $n \times 1$ *and* $T \times 1$ *vectors of ones.*

Proof. To establish (A.6) note that

$$n^{-1}\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT} = \boldsymbol{n}^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT} - \left(\frac{\mathbf{U}_{nT}'\boldsymbol{\tau}_n}{n}\right)\left(\frac{\boldsymbol{\tau}_n'\mathbf{U}_{nT}}{n}\right)$$

and $n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT} = n^{-1}\sum_{i=1}^{n}\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'$, where $\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}' = (u_{it}u_{it'})$, for $t, t' = 1, 2, ...T$. We first note that the t^{th} element of $n^{-1}\mathbf{U}'\boldsymbol{\tau}_n$ is given by $\overline{\mathbf{u}}_{\circ t} = n^{-1}\sum_{i=1}^{n} u_{it}$ and under Assumption (2), $\overline{\mathbf{u}}_{\circ t} \to_p 0$, and we have

$$n^{-1}\mathbf{U}'\boldsymbol{\tau}_n \to_p 0. \tag{A.9}$$

Also, by results (A.1) and (A.2) of lemma A.1, it follows that $n^{-1}\mathbf{U}'_{nT}\mathbf{U}_{nT}\rightarrow_p \overline{\sigma}^2\mathbf{I}_T$, and in conjunction with (A.9) yields (A.6) as required. To establish (A.7) note that

$$n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}} = n^{-1}\mathbf{U}'_{nT}\bar{\mathbf{u}} - \left(\frac{\mathbf{U}'_{nT}\boldsymbol{\tau}_n}{n}\right)\left(\frac{\boldsymbol{\tau}'_n\mathbf{U}_{nT}}{n}\right), \tag{A.10}$$

where $n^{-1}\mathbf{U}'_{nT}\bar{\mathbf{u}} = (\phi_{1,n}, \phi_{2,n}, ..., \phi_{T,n})'$, with $\phi_{t,n} = \frac{1}{n}\sum_{i=1}^n u_{it}\bar{u}_{iT}$, which can be written equivalently as

$$\phi_{t,n} = \frac{1}{n}\sum_{i=1}^n (u_{it}\bar{u}_{iT} - \frac{1}{T}\sigma_i^2) + \frac{1}{T}\bar{\sigma}_n^2,$$

where $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$. Hence, by result (A.3) of Lemma A.1, $\phi_{t,n} \rightarrow_p \frac{1}{T}\bar{\sigma}^2$, which in turn establishes that $n^{-1}\mathbf{U}'_{nT}\bar{\mathbf{u}} \rightarrow_p \frac{1}{T}\bar{\sigma}^2\boldsymbol{\tau}_T$. Also by (A.9) the second term of (A.10) tends to zero in probability and (A.7) follows. Finally to establish (A.8), note that

$$\left\|\frac{\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n}{n}\right\| \leq \left\|\frac{\mathbf{U}'_{nT}\mathbf{M}_n}{\sqrt{n}}\right\|\left\|\frac{\boldsymbol{\eta}_n}{\sqrt{n}}\right\| = \lambda_{\max}^{1/2}\left(\frac{\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}}{n}\right)\left(\frac{\boldsymbol{\eta}'_n\boldsymbol{\eta}_n}{n}\right)^{\frac{1}{2}}.$$

Also using (A.6) it follows that $\lambda_{\max}^{1/2}\left(\frac{\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}}{n}\right) \rightarrow_p \overline{\sigma}^2 < C$, and by part (a) of Assumption 5 $n^{-1}\boldsymbol{\eta}'_n\boldsymbol{\eta}_n = O\left(n^{\alpha_\eta - 1}\right)$, and as required $\left\|\frac{\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n}{n}\right\| \rightarrow_p 0$, if $\alpha_\eta < 1/2$. ■

Lemma A.3 *Consider the* $n\times T$ *error matrix* $\boldsymbol{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, *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* $\beta_{\circ k} = (\beta_{1k}, \beta_{2k}, ..., \beta_{nk})'$, *the* $n \times 1$ *vector of pricing errors* $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$, *with a pervasiveness coefficient,* α_η, $\mathbf{D}_n = diag(n^{\alpha_1/2}, n^{\alpha_2/2}, ..., n^{\alpha_K/2})$, $\alpha_{min} = min(\alpha_1, \alpha_2, ..., \alpha_K)$, $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$, $\mathbf{M}_n = \mathbf{I}_n - \frac{1}{n}\boldsymbol{\tau}_n\boldsymbol{\tau}'_n$, $\bar{\mathbf{u}} = (\overline{u}_{1\circ}, \overline{u}_{2\circ}, ..., \overline{u}_{n\circ})'$, $\overline{u}_{i\circ} = T^{-1}\sum_{t=1}^T u_{it}$, *and* $\boldsymbol{\tau}_n$ *and* $\boldsymbol{\tau}_T$ *are, respectively,* $n \times 1$ *and* $T \times 1$ *vectors of one. Suppose that Assumptions 1, 2 and 4 hold. Then*

$$\mathbf{D}_n^{-1}\mathbf{B}'_n\mathbf{M}_n\bar{\mathbf{u}} = O_p\left(T^{-1/2}\right), \tag{A.11}$$

$$\mathbf{D}_n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_n = O_p\left(\frac{n^{(1-\alpha_{min})/2}}{T^{3/2}}\right) + O_p\left(n^{-1-\alpha_{min}/2}\right). \tag{A.12}$$

$$\mathbf{D}_n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n = \left(\sqrt{\frac{n}{T}}\right)O_p\left[n^{-(\alpha_{min}-\alpha_\eta)/2}\right], \tag{A.13}$$

$$\mathbf{D}_n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1} = O_p\left(T^{-1/2}n^{-a_{min}/2}\right) \tag{A.14}$$

$$\mathbf{D}_n^{-1}\mathbf{G}'_T\left(\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\right)\mathbf{G}_T\mathbf{D}_n^{-1} = O_p\left(\frac{n^{1-\alpha_{min}}}{T}\right). \tag{A.15}$$

Proof. To establish (A.11), noting that

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

then it follows that the k^{th} element of $\mathbf{D}_n^{-1}\mathbf{B}'_n\mathbf{M}_n\bar{\mathbf{u}}$ is given by $c_{k,nT} = n^{-\alpha_k/2}\sum_{i=1}^n\left(\beta_{ik} - \bar{\beta}_k\right)\bar{u}_{iT}$, and we have $E\left(c_{k,nT}\right) = 0$, and

$$Var\left(c_{k,nT}\right) = n^{-\alpha_k}\sum_{i=1}^n\sum_{i=1}^n\left(\beta_{ik} - \bar{\beta}_k\right)\left(\beta_{jk} - \bar{\beta}_k\right)E\left(\overline{u}_{i\circ}\overline{u}_{j\circ}\right),$$

with $E(\bar{u}_{i\circ}\bar{u}_{j\circ}) = T^{-1}\sigma_{ij}$. Hence (recalling that $\mathbf{V}_u = (\sigma_{ij})$)

$$\begin{aligned}
Var(c_{nT,k}) &= T^{-1}n^{-\alpha_k}\sum_{i=1}^{n}\sum_{i=1}^{n}\sigma_{ij}\left(\beta_{ik}-\bar{\beta}_k\right)\left(\beta_{jk}-\bar{\beta}_k\right) \\
&= T^{-1}n^{-\alpha_k}\left(\boldsymbol{\beta}_{\circ k}-\tau_n\bar{\beta}_k\right)'\mathbf{V}_u\left(\boldsymbol{\beta}_{\circ k}-\tau_n\bar{\beta}_k\right) \\
&\leq T^{-1}n^{-\alpha_k}\left(\boldsymbol{\beta}_{\circ k}-\tau_n\bar{\beta}_k\right)'\left(\boldsymbol{\beta}_{\circ k}-\tau_n\bar{\beta}_k\right)\lambda_{max}\left(\mathbf{V}_u\right).
\end{aligned}$$

But under Assumptions (1) and (2) $\lambda_{max}(\mathbf{V}_u) < C$, and $\left(\boldsymbol{\beta}_{\circ k}-\tau_n\bar{\beta}_k\right)'\left(\boldsymbol{\beta}_{\circ k}-\tau_n\bar{\beta}_k\right) = \ominus(n^{\alpha_k})$, and overall $Var(c_{nT,k}) = O(T^{-1})$. Thus by Markov inequality it follows that $c_{nT,k} = O_p\left(T^{-1/2}\right)$, for $k = 1, 2, ..., K$, and result (A.11) follows. Similarly, to establish (A.12) using $\mathbf{G}_T' = \left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T$ we note that

$$\mathbf{D}_n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}} = \mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}\right). \tag{A.16}$$

By assumption $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ is a positive definite matrix and $\|\mathbf{D}_n^{-1}\| = \lambda_{max}(\mathbf{D}_n^{-1}) = n^{-\alpha_{min}/2}$, where $\alpha_{min} = min(\alpha_1, \alpha_2, ..., \alpha_K)$. Also

$$\begin{aligned}
T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}} &= \frac{1}{T}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\bar{\mathbf{u}} - \mathbf{F}'\mathbf{M}_T\left(\frac{\mathbf{U}_{nT}'\boldsymbol{\tau}_n}{T}\right)\left(\frac{\boldsymbol{\tau}_n'\mathbf{U}_{nT}}{n}\right), \\
&= \mathbf{q}_{a,nT} - \mathbf{q}_{b,nT} \qquad\qquad (A.17)
\end{aligned}$$

Further $n^{-1}\mathbf{U}_{nT}'\bar{\mathbf{u}} = (\phi_{1,n}, \phi_{2,n}, ..., \phi_{T,n})'$, where

$$\phi_{t,n} = \frac{1}{n}\sum_{i=1}^{n}u_{it}\bar{u}_{iT} = c_{n,t} + \frac{1}{T}\bar{\sigma}_n^2,$$

$c_{n,t} = n^{-1}\sum_{i=1}^{n}z_{it}$, $z_{it} = u_{it}\bar{u}_{i\circ} - \frac{1}{T}\sigma_i^2$, and $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$. Then the k^{th} element of $\mathbf{q}_{nT}^a$ can be written as

$$\begin{aligned}
q_{k,nT}^a &= \left(\frac{n}{T}\right)\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\phi_{tn} = \left(\frac{n}{T}\right)\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\left(c_{n,t}+\frac{1}{T}\bar{\sigma}_n^2\right) \\
&= \left(\frac{n}{T}\right)\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)c_{n,t}.
\end{aligned}$$

Under Assumption 2 $E(z_{it}) = 0$, and $c_{n,t}$ and $(f_{kt} - \bar{f}_k)$ are independently distributed (since u_{it} and f_{kt} are independently distributed). Then $E\left(q_{k,nT}^a\right) = 0$ and since u_{it} are serially uncorrelated we also have

$$Var\left(q_{k,nT}^a\,|\mathbf{f}_{k\circ}\right) = \left(\frac{n}{T}\right)^2\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2 Var\left(c_{n,t}\right),$$

and using result (A.4) in Lemma A.1 we have

$$Var\left(q_{k,nT}^a\,|\mathbf{f}_{k\circ}\right) = \left(\frac{n^2}{T}\right)\left\{T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2\frac{1}{T^2}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(u_{it}^2, u_{jt}^2)\right]\right\}.$$

However, under part b of Assumption 2 $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(u_{it}^2,u_{jt}^2)<C$, and it follows that (recall that $T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2<C$)

$$Var\left(q_{k,nT}^{a}\left|\mathbf{f}_{k\circ}\right.\right)\leq C\left(\frac{n^2}{T}\right)\left(\frac{1}{nT^2}\right)\left\{T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2\right\}=O\left(\frac{n}{T^3}\right). \tag{A.18}$$

Hence $q_{k,nT}^{a}=O_p\left(\frac{\sqrt{n}}{T^{3/2}}\right)$, and since K is fixed establishes that

$$\mathbf{q}_{nT}^{a}=O_p\left(\frac{\sqrt{n}}{T^{3/2}}\right). \tag{A.19}$$

Consider now the second term of (A.17) and note that

$$\left\|\mathbf{q}_{nT}^{b}\right\|\leq\left\|\frac{\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\boldsymbol{\tau}_n}{T}\right\|\left\|\frac{\boldsymbol{\tau}_n'\mathbf{U}_{nT}}{n}\right\|, \tag{A.20}$$

where $\frac{1}{n}\boldsymbol{\tau}_n'\mathbf{U}_{nT}=(\bar{u}_{\circ 1},\bar{u}_{\circ 2},...,\bar{u}_{\circ T})$, and

$$\left\|\frac{\boldsymbol{\tau}_n'\mathbf{U}_{nT}}{n}\right\|=\left(\sum_{t=1}^{T}\bar{u}_{\circ t}^2\right)^{1/2}=O_p\left(\sqrt{\frac{T}{n}}\right). \tag{A.21}$$

Also the k^{th} element of $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\boldsymbol{\tau}_n$ is given by

$$s_{k,nT}=T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\bar{u}_{\circ t},$$

Since $\bar{u}_{\circ t}$ is serially uncorrelated and distributed independently of $f_{kt'}$ for all t and t', it then follows that $E\left(s_{k,nT}\right)=0$, and

$$Var\left(s_{k,nT}\left|\mathbf{f}_{k\circ}\right.\right)=T^{-2}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2E\left(\bar{u}_{\circ t}^2\right)=T^{-1}\left(T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2\right)\left(n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\right).$$

But under Assumption 2, $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{ij}|<C$, and under Assumption 1, $T^{-1}\sum_{t=1}^{T}E\left[(f_{kt}-\bar{f}_k)^2\right]<C$, overall $Var\left(s_{k,nT}\right)=O\left(\frac{1}{nT}\right)$, which establishes that $s_{k,nT}=O_p\left(n^{-1/2}T^{-1/2}\right)$, and hence $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\boldsymbol{\tau}_n=O_p\left(n^{-1/2}T^{-1/2}\right)$ considering that K is fixed. Now using this result and (A.21) in (A.20), now yields $\mathbf{q}_{nT}^{b}=O_p\left(n^{-1}\right)$. Using this result together with (A.19) in (A.17) we have

$$T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\bar{\mathbf{u}}=O_p\left(\frac{\sqrt{n}}{T^{3/2}}\right)+O_p\left(n^{-1}\right),$$

which if used in (A.16) establishes (A.12). Consider now (A.13) and note that

$$\mathbf{D}_n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n=T^{-1}\mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\left(\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n\right), \tag{A.22}$$

and as established already $\left\|\mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\right\|=O_p\left(n^{-\alpha_{min}/2}\right)$. Also

$$\left\|\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\boldsymbol{\eta}_n\right\|_2\leq\left\|\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\right\|_2\left\|\mathbf{M}_n\boldsymbol{\eta}_n\right\|_2,$$

where $\|\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\|_2 = \left[E\|\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\|^2\right]^{1/2}$ and $\|\mathbf{M}_n\boldsymbol{\eta}_n\|_2 = \left[E\|\mathbf{M}_n\boldsymbol{\eta}_n\|^2\right]^{1/2}$. Consider the first term and note that its (k,i) element of $\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}$ is given by $\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)u_{it}$. Then

$$\begin{aligned}\|\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\|^2 \leq \|\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\|_F^2 &\leq \sum_{k=1}^{K}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)u_{it}\right)^2 \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}(f_{kt}-\bar{f}_k)(f_{kt'}-\bar{f}_k)u_{it}u_{it'},\end{aligned}$$

and since u_{it} are serially correlated we have (note that K is fixed)

$$E\|\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\|^2 \leq \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2\sigma_i^2 = nTK\left(T^{-1}K^{-1}\sum_{k=1}^{K}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2\right)\left(n^{-1}\sum_{i=1}^{n}\sigma_i^2\right) = O\left(nT\right).$$

Also

$$E\|\mathbf{M}_n\boldsymbol{\eta}_n\|^2 = \sum_{i=1}^{n}E\left(\eta_i-\bar{\eta}\right)^2 \leq \sum_{i=1}^{n}E\left(\eta_i^2\right) = O\left(n^{\alpha_\eta}\right).$$

Hence

$$\|\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n\| \leq O\left(n^{\alpha_\eta/2}\right)O_p\left(\sqrt{nT}\right)$$

Using this result in (A.22) now yields

$$\begin{aligned}\mathbf{D}_n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n &= T^{-1}O_p\left(n^{-\alpha_{min}/2}\right)O\left(n^{\alpha_\eta/2}\right)O_p\left(\sqrt{nT}\right) \\ &= \left(\sqrt{\frac{n}{T}}\right)O_p\left[n^{-(\alpha_{min}-\alpha_\eta)/2}\right],\end{aligned}$$

as required by (A.13). To establish (A.14) using $\mathbf{G}'_T = \left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T$ we note that

$$\mathbf{D}_n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1} = T^{-1}\mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}. \tag{A.23}$$

Since $\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$ is bounded (by assumption), we focus on the $K\times K$ matrix $\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}$ and note that its (k,k') element is given by

$$p_{nT}\left(k,k'\right) = \sum_{i=1}^{n}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\left(\frac{\beta_{ik'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right)u_{it}.$$

Since K is fixed to obtain the probability order of $\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}$ it is sufficient to consider the probability order of $p_{nT}\left(k,k'\right)$ for given $k,k'=1,2,...,K$. To this end we note that since u_{it} is distribute independently of $f_{kt'}$ and β_{jk} for all $i,j,t$and t', then $E\left[p_{nT}\left(k,k'\right)\right]=0$ and (conditional on β_{ik} f_{kt}) we have

$$\begin{aligned}Var\left[p_{nT}\left(k,k'\right)\right] &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(u_{it}u_{jt'}\right)(f_{kt}-\bar{f}_k)(f_{kt'}-\bar{f}_k)\left(\frac{\beta_{ik'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right)\left(\frac{\beta_{jk'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sigma_{ij}(f_{kt}-\bar{f}_k)^2\left(\frac{\beta_{ik'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right)\left(\frac{\beta_{jk'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right) \\ &= T\left[T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2\right]\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\left(\frac{\beta_{ik'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right)\left(\frac{\beta_{jk'}-\bar{\beta}_{k'}}{n^{\alpha_{k'}/2}}\right).\end{aligned} \tag{A.24}$$

But $T^{-1}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)^2<C$, and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\left(\frac{\beta_{ik}-\bar{\beta}_k}{n^{\alpha_k/2}}\right)\left(\frac{\beta_{jk}-\bar{\beta}_k}{n^{\alpha_k/2}}\right)=\delta_{k,n}'\mathbf{V}_u\delta_{k,n},$$

where $\delta_k=\left[\left(\frac{\beta_{1k}-\bar{\beta}_k}{n^{\alpha_k/2}}\right),\left(\frac{\beta_{2k}-\bar{\beta}_k}{n^{\alpha_k/2}}\right),...,\left(\frac{\beta_{nk}-\bar{\beta}_k}{n^{\alpha_k/2}}\right)\right]'$, and $\mathbf{V}_u=(\sigma_{ij})$. Also $\delta_k'\mathbf{V}_u\delta_k\leq(\delta_k'\delta_k)\lambda_{max}(\mathbf{V}_u)$, and under Assumptions 2 and 1, $\lambda_{max}(\mathbf{V}_u)<C$ and

$$\delta_k'\delta_k=n^{-\alpha_k}\sum_{i=1}^{n}\left(\beta_{ik}-\bar{\beta}_k\right)^2<C.$$

Using this result in (A.24) it now follows that that $Var\left[p_{nT}(k,k')\right]=O(T)$, and since $E\left[p_{nT}(k,k')\right]=0$, then $p_{nT}(k,k')=O_p\left(\sqrt{T}\right)$. From this it also follows that $\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}=O_p\left(\sqrt{T}\right)$, which if used in (A.23) yields $\mathbf{D}_n^{-1}\mathbf{G}_T'\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n^{-1}=O_p\left(T^{-1/2}n^{-a_{min}/2}\right)$ as required by (A.14). Consider now (A.15) and replacing $\mathbf{G}_T$ in terms of $\mathbf{F}$ we have

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

Let $\mathbf{A}=\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n=\left[a_{nT}(k,i)\right]$, and note that

$$a_{nT}(k,i)=\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)\left(u_{it}-\bar{u}_{iT}\right).$$

Then $\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{M}_T\mathbf{F}=\mathbf{A}\mathbf{A}'=\left[s_{nT}(k,k')\right]$, where

$$\begin{aligned}s_{nT}(k,k')&=\sum_{i=1}^{n}a_{nT}(k,i)\,a_{nT}(k',i)\\&=\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}(f_{kt}-\bar{f}_k)(f_{k't'}-\bar{f}_{k'})\left(u_{it}-\bar{u}_{iT}\right)\left(u_{it'}-\bar{u}_{iT}\right)\end{aligned}$$

$$\begin{aligned}E\left[s_{nT}(k,k')\,|\mathbf{F}\right]&=\sum_{i=1}^{n}\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)(f_{k't}-\bar{f}_{k'})E\left(u_{it}-\bar{u}_{iT}\right)^2\\&=n\left(1-\frac{1}{T}\right)\left(n^{-1}\sum_{i=1}^{n}\sigma_i^2\right)\sum_{t=1}^{T}(f_{kt}-\bar{f}_k)(f_{k't}-\bar{f}_{k'}).\end{aligned}$$

Hence

$$E\left(\mathbf{F}'\mathbf{M}_T\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\mathbf{M}_T\mathbf{F}\,|\mathbf{F}\right)=n(T-1)\,\bar{\sigma}_n^2\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right),$$

and

$$\begin{aligned}&T^{-2}\mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\left(\mathbf{U}_{nT}'\mathbf{M}_n\mathbf{U}_{nT}\right)\mathbf{M}_T\mathbf{F}'\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{D}_n^{-1}\\&=n(T-1)\,\bar{\sigma}_n^2T^{-2}\mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{D}_n^{-1}\\&=n(T-1)\,\bar{\sigma}_n^2T^{-2}\mathbf{D}_n^{-1}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{D}_n^{-1}\\&=O_p\left(\frac{n}{T}n^{-\alpha_{min}}\right),\end{aligned}$$

as required. ■

A.3 Proof of theorems and related results

A.3.1 Proof of Theorem 1

Consider the two-pass estimator of $\boldsymbol{\lambda}$ defined by (39), and to simplify notations, write it as

$$\hat{\boldsymbol{\lambda}}_{nT} = \left(\frac{\hat{\mathbf{B}}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n}\right)^{-1}\left(\frac{\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\bar{\mathbf{r}}}{n}\right), \tag{A.25}$$

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

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{A.26}$$

and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under the factor model (38)

$$\mathbf{r}_{i\circ} = \alpha_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{A.27}$$

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

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}. \tag{A.28}$$

Stacking these results over i yields:

$$\hat{\mathbf{B}}_{nT} = \mathbf{B}_n + \mathbf{U}_{nT}\mathbf{G}_T \tag{A.29}$$

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

$$\mathbf{G}_T = \mathbf{M}_T\mathbf{F}(\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} \tag{A.30}$$

Also using result (43) in the paper we have (in terms of the simplified notations used here)

$$\bar{\mathbf{r}} = \lambda_0\boldsymbol{\tau}_n + \mathbf{B}_n\boldsymbol{\lambda}^*_T + \bar{\mathbf{u}} + \boldsymbol{\eta} \tag{A.31}$$

where

$$\boldsymbol{\lambda}^*_T = \boldsymbol{\lambda} + \mathbf{d}_T, \text{ and } \mathbf{d}_{fT} = \bar{\mathbf{f}}_T - E\left(\bar{\mathbf{f}}_T\right). \tag{A.32}$$

and $\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$. To derive the asymptotic limit of $\hat{\boldsymbol{\lambda}}_{nT}$ as $n \to \infty$, when T is fixed, we first consider the probability limits of $n^{-1}\left(\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)$ and $n^{-1}\left(\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\bar{\mathbf{r}}\right)$. Using (41) and (A.31) we have

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

$$\begin{aligned} n^{-1}\left(\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\bar{\mathbf{r}}\right) &= n^{-1}\left(\mathbf{B}'_n\mathbf{M}_n\mathbf{B}_n\right)\boldsymbol{\lambda}^*_T + n^{-1}\left(\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{B}_n\right)\boldsymbol{\lambda}^*_T \\ &+ n^{-1}\left(\mathbf{B}'_n\mathbf{M}_n\bar{\mathbf{u}}\right) + n^{-1}\left(\mathbf{B}'_n\mathbf{M}_n\boldsymbol{\eta}_n\right) \\ &+ n^{-1}\left(\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}\right) + n^{-1}\left(\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n\right). \end{aligned}$$

Now using the results in Lemma A.2, under Assumptions 2 and 1 we have

$$n^{-1}\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right) \rightarrow_p \mathbf{\Sigma}_{\beta\beta} + \bar{\sigma}^2\mathbf{G}_T'\mathbf{G}_T,$$
$$n^{-1}\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}\right) \rightarrow_p \mathbf{\Sigma}_{\beta\beta}\boldsymbol{\lambda}_T^* + \frac{\bar{\sigma}^2}{T}\mathbf{G}_T'\boldsymbol{\tau}_T.$$

But using (A.30) we also have

$$\mathbf{G}_T'\mathbf{G}_T = \frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}, \tag{A.33}$$

and $\mathbf{G}_T'\boldsymbol{\tau}_T = \mathbf{0}$. Hence

$$n^{-1}\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right) \rightarrow_p \mathbf{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}\mathbf{M}_T\mathbf{F}}{T}\right)^{-1},$$
$$n^{-1}\left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}\right) \rightarrow_p \mathbf{\Sigma}_{\beta\beta}\boldsymbol{\lambda}_T^*.$$

When T is fixed, 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 Assumption 4 $\mathbf{\Sigma}_{\beta\beta}$ is a semi-positive definite (could be zero), then for a fixed T the probability limit of $n^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}$ is non-singular and using (A.25) by the Slutsky Theorem we have

$$\hat{\boldsymbol{\lambda}}_n \rightarrow_p \left[\mathbf{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\mathbf{\Sigma}_{\beta\beta}\boldsymbol{\lambda}_T^*,$$

which in view of (A.32) can be written equivalently in the form stated in Theorem 1.

A.3.2 Proof of n consistency of $\hat{\bar{\sigma}}_{nT}^2$ for $\bar{\sigma}^2$

Consider the expression for $\hat{\bar{\sigma}}_{nT}^2$ given by (48) and note that under (36) we have

$$\hat{u}_{it} = \alpha_i - \hat{\alpha}_{iT} - \left(\hat{\boldsymbol{\beta}}_{i,T} - \boldsymbol{\beta}_i\right)'\mathbf{f}_t + u_{it},$$

and since $\hat{u}_{it}$ are OLS residuals then for each i, we also have $T^{-1}\sum_{t=1}^T \hat{u}_{it} = 0$. Using this result

$$\hat{u}_{it} = u_{it} - \bar{u}_i - \left(\hat{\boldsymbol{\beta}}_{i,T} - \boldsymbol{\beta}_i\right)'\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right), \text{ for } i = 1, 2, ..., n,$$

and stacking over i now yields $\hat{\mathbf{u}}_t = \mathbf{u}_t - \bar{\mathbf{u}} - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)$. Hence

$$\begin{aligned}
T^{-1}n^{-1}\sum_{t=1}^T\sum_{i=1}^n \hat{u}_{it}^2 &= T^{-1}\sum_{t=1}^T n^{-1}\hat{\mathbf{u}}_t'\hat{\mathbf{u}}_t \\
&= T^{-1}\sum_{t=1}^T n^{-1}\left(\mathbf{u}_t - \bar{\mathbf{u}}\right)'\left(\mathbf{u}_t - \bar{\mathbf{u}}\right) \\
&+ T^{-1}\sum_{t=1}^T \left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)' n^{-1}\left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)'\left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right) \\
&- 2T^{-1}\sum_{t=1}^T n^{-1}\left(\mathbf{u}_t - \bar{\mathbf{u}}\right)'\left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right) \\
&= a_{nT} + b_{nT} + c_{nT}
\end{aligned} \tag{A.34}$$

Consider each of the three terms in the above expression in turn. For the first term we have

$$a_{nT} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} u_{it}^2}{nT} - \frac{\sum_{i=1}^{n} \bar{u}_i^2}{n}.$$

Under Assumption 2 u_{it} and $\bar{u}_i$ are weakly cross correlated and for each t, $n^{-1}\sum_{i=1}^{n} u_{it}^2 \to_p lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(u_{it}^2\right) = lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_i^2 = \bar{\sigma}^2$. Similarly, $n^{-1}\sum_{i=1}^{n}\bar{u}_i^2 \to_p T^{-1}\bar{\sigma}^2$, and (for a fixed T and as $n \to \infty$)

$$a_{nT} \to_p \left(1 - \frac{1}{T}\right)\bar{\sigma}^2. \tag{A.35}$$

Now using (A.29)

$$b_{nT} = T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)' \mathbf{G}_T' \left(\frac{\mathbf{U}_{nT}'\mathbf{U}_{nT}}{n}\right)\mathbf{G}_T\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right),$$

and in view of (A.6) we have (as $n \to \infty$)

$$b_{nT} \to_p \bar{\sigma}^2 T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)'\mathbf{G}_T'\mathbf{G}_T\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right) = \bar{\sigma}^2 Tr\left[\mathbf{G}_T'\mathbf{G}_T T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)'\right]$$
$$= \bar{\sigma}^2 Tr\left[\mathbf{G}_T'\mathbf{G}_T\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)\right].$$

But by (A.33) $\mathbf{G}_T'\mathbf{G}_T = \frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}$, and it follows that

$$b_{nT} \to_p \frac{k}{T}\bar{\sigma}^2. \tag{A.36}$$

Finally, again using (A.29)

$$c_{nT} = -2T^{-1}\sum_{t=1}^{T} n^{-1}\left(\mathbf{u}_t - \bar{\mathbf{u}}\right)'\mathbf{U}_{nT}\mathbf{G}_T\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right) = -2Tr\left[n^{-1}\mathbf{U}_{nT}\mathbf{G}_T T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)\mathbf{u}_t'\right],$$

and noting that

$$T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)\mathbf{u}_t' = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t\mathbf{u}_t' - \bar{\mathbf{f}}_T\bar{\mathbf{u}}',$$
$$T^{-1}\sum_{t=1}^{T}\mathbf{f}_t\mathbf{u}_t' = T^{-1}\mathbf{F}'\mathbf{U}_{nT}'$$

we have$\frac{\bar{\sigma}^2}{T}\boldsymbol{\tau}_T$

$$c_{nT} = -2T^{-1}Tr\left[\mathbf{G}_T\mathbf{F}'\left(n^{-1}\mathbf{U}_{nT}'\mathbf{U}_{nT}\right)\right] + 2Tr\left[\mathbf{G}_T\bar{\mathbf{f}}_T\left(n^{-1}\bar{\mathbf{u}}'\mathbf{U}_{nT}\right)\right].$$

Now using (A.6) and (A.7) it follows that

$$c_{nT} \to_p -2\bar{\sigma}^2 T^{-1}Tr\left(\mathbf{G}_T\mathbf{F}'\right) + 2Tr\left[\mathbf{G}_T\bar{\mathbf{f}}_T\frac{\bar{\sigma}^2}{T}\boldsymbol{\tau}_T'\right]$$
$$= -2\bar{\sigma}^2 T^{-1}Tr\left(\mathbf{F}'\mathbf{G}_T\right) + 2\bar{\sigma}^2 T^{-1}Tr\left[\bar{\mathbf{f}}_T\boldsymbol{\tau}_T'\mathbf{G}_T\right].$$

But using (A.30) it is readily seen that $\mathbf{F}'\mathbf{G}_T = \mathbf{I}_k$ and $\boldsymbol{\tau}_T'\mathbf{G}_T = \mathbf{0}$, and therefore

$$c_{nT} \to_p -\frac{2k}{T}\bar{\sigma}^2. \tag{A.37}$$

Now using (A.35), (A.36) and (A.37) in (A.34)

$$T^{-1}n^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\hat{u}_{it}^2 \to_p \left(1-\frac{1}{T}\right)\bar{\sigma}^2 + \frac{k}{T}\bar{\sigma}^2 - \frac{2k}{T}\bar{\sigma}^2 = \left(\frac{T-k-1}{T}\right)\bar{\sigma}^2,$$

which establishes

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

as required.

Lemma A.4 *Under Assumptions 2, 1 and 4*

$$c_{nT} = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{\beta}_i'\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)u_{it} = O_p\left(1\right).$$

Proof. Since by assumption $\boldsymbol{\beta}_i$, u_{it} and $\mathbf{f}_t$ are distributed independently and u_{it} is serially uncorrelated with zero means then $E\left(c_{nT}\right) = 0$, and it is sufficient to show that $Var\left(c_{nT}\right) < C$. Let $\mathbf{B}_n = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)'$, and note that

$$\begin{aligned} Var\left(c_{nT}\left|\boldsymbol{B}_n\right.\right) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}\boldsymbol{\beta}_i'E\left[\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)u_{it}u_{jt'}\left(\mathbf{f}_{t'} - \bar{\mathbf{f}}_T\right)'\right]\boldsymbol{\beta}_j \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\boldsymbol{\beta}_i'E\left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)'\right]\boldsymbol{\beta}_j \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_f\boldsymbol{\beta}_j \end{aligned}$$

Hence

$$\begin{aligned} \left|Var\left(c_{nT}\left|\boldsymbol{B}_n\right.\right)\right| &\leq \left\|\boldsymbol{\Sigma}_f\right\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sigma_{ij}\right|\left\|\boldsymbol{\beta}_i\right\|\left\|\boldsymbol{\beta}_j\right\| \\ &\leq \left\|\boldsymbol{\Sigma}_f\right\|\left(sup_i\left\|\boldsymbol{\beta}_i\right\|\right)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sigma_{ij}\right| \\ &\leq \left\|\boldsymbol{\Sigma}_f\right\|\left(sup_i\left\|\boldsymbol{\beta}_i\right\|\right)sup_i\sum_{j=1}^{n}\left|\sigma_{ij}\right|. \end{aligned}$$

Since $K = dim(\beta_i)$ is fixed, then $\left\|\boldsymbol{\Sigma}_f\right\|\left(sup_i\left\|\boldsymbol{\beta}_i\right\|\right) < C$, given that by assumption $\sup_{ij}\left|\beta_{ij}\right| < C$ and $\left\|\boldsymbol{\Sigma}_f\right\| < C$. Also under Assumption 2 u_{it} is weakly cross-correlated and we have

$sup_i \sum_{j=1}^{n} |\sigma_{ij}| < C$. Hence, overall $\|Var(c_{nT} |\boldsymbol{B}_n)\| < C$ which establishes the desired result for any given values of factor loadings. When β_i are treated as random variables then $Var(c_{nT}) = E[Var(c_{nT} |\boldsymbol{B}_n)]$, since $E(c_{nT} |\boldsymbol{B}_n) = 0$. In this case

$$Var(c_{nT}) = E[Var(c_{nT} |\boldsymbol{B}_n)] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}Tr\left[\boldsymbol{\Sigma}_f E\left(\boldsymbol{\beta}_j\boldsymbol{\beta}_i'\right)\right],$$

and

$$|Var(c_{nT})| \leq sup_{ij}\left\|Tr\left[\boldsymbol{\Sigma}_f E\left(\boldsymbol{\beta}_j\boldsymbol{\beta}_i'\right)\right]\right\| sup_i \sum_{j=1}^{n}|\sigma_{ij}|.$$

and the desired result follows since $sup_{ik}E(\beta_{ik}^2) < C$, noting that $\boldsymbol{\Sigma}_f$ is bounded and by assumption if $sup_i \sum_{j=1}^{n} |\sigma_{ij}| < C$. ■

A.3.3 Properties of pooled R squared

Lemma A.5 *Consider the factor model*

$$r_{it} = a_i + \sum_{k=1}^{K}\beta_{ik}f_{kt} + u_{it} = a_i + \boldsymbol{\beta}_i'\mathbf{f}_t + u_{it}, \text{ for } i = 1, 2, ..., n;\ t = 1, 2, ..., T, \tag{A.38}$$

and consider the following pooled measure of fit for known values of factor loadings, $\boldsymbol{\beta}_i$:

$$PR^2 = 1 - \frac{(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(u_{it}-\bar{u}_{iT})^2}{(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(r_{it}-\bar{r}_{iT})^2}, \tag{A.39}$$

where $\bar{r}_{iT} = T^{-1}\sum_{t=1}^{T} r_{it}$, and $\bar{u}_{iT} = T^{-1}\sum_{t=1}^{T} u_{it}$. Then under Under Assumptions 2, 1 and 4 we have

$$PR_{nT}^2 = \sum_{k=1}^{K}\ominus\left(n^{\alpha_k - 1}\right) + O_p(n^{-1/2}T^{-1/2}) + O_p(T^{-1/2}). \tag{A.40}$$

where α_k is the strength of factor f_{tk}.

Proof. Averaging (A.38) over t and forming deviations of r_{it} from its time average, $\bar{r}_{iT}$, we have

$$r_{it} - \bar{r}_{iT} = u_{it} - \bar{u}_{iT} + \beta_i'\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right).$$

Using this result we have

$$\begin{aligned}(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(r_{it}-\bar{r}_{iT})^2 = {} & (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(u_{it}-\bar{u}_{iT})^2 \\ & + n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_{fT}\boldsymbol{\beta}_i \\ & - 2(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(u_{it}-\bar{u}_{iT})\boldsymbol{\beta}_i'\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right),\end{aligned}$$

where

$$\boldsymbol{\Sigma}_{fT} = T^{-1} \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right)'$$

But

$$\begin{aligned}
(nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(u_{it} - \bar{u}_{iT} \right) \boldsymbol{\beta}_i' \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) &= (nT)^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) \left(u_{it} - \bar{u}_{iT} \right) \\
&= (nT)^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) u_{it} - (nT)^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) \bar{u}_{iT} \\
&= (nT)^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) u_{it} - (nT)^{-1} \sum_{i=1}^{n} \bar{u}_{iT} \boldsymbol{\beta}_i' \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) \\
&= (nT)^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \sum_{t=1}^{T} \left(\mathbf{f}_t - \bar{\mathbf{f}}_T \right) u_{it} = c_{nT}.
\end{aligned}$$

Now using Lemma A.4 we have

$$\begin{aligned}
(nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(r_{it} - \bar{r}_{iT} \right)^2 &= (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(u_{it} - \bar{u}_{iT} \right)^2 \\
&\quad + n^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \boldsymbol{\Sigma}_{fT} \boldsymbol{\beta}_i + O_p\left(n^{-1/2} T^{-/12} \right).
\end{aligned}$$

Also since under Assumption 2, u_{it} are serially uncorrelated we have

$$T^{-1} \sum_{t=1}^{T} \left(u_{it} - \bar{u}_{iT} \right)^2 = \sigma_i^2 + O_p(T^{-1/2}),$$

and

$$(nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(u_{it} - \bar{u}_{iT} \right)^2 = \bar{\sigma}_n^2 + O_p(T^{-1/2}),$$

with $\bar{\sigma}_n^2 > 0$. Using the above results in (A.39) we now have

$$PR_{nT}^2 = \frac{n^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \boldsymbol{\Sigma}_{fT} \boldsymbol{\beta}_i / \bar{\sigma}_n^2 + O_p(n^{-1/2} T^{-1/2}) + O_p(T^{-1/2})}{1 + n^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \boldsymbol{\Sigma}_{fT} \boldsymbol{\beta}_i / \bar{\sigma}_n^2 + O_p(n^{-1/2} T^{-1/2}) + O_p(T^{-1/2})}. \tag{A.41}$$

Hence, the order of PR_{nT}^2 is governed by the pooled signal-to-noise ratio defined by

$$s_{nT}^2 = \frac{n^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \boldsymbol{\Sigma}_{fT} \boldsymbol{\beta}_i}{\bar{\sigma}_n^2}.$$

However, under Assumption 1

$$\lambda_{min}(\boldsymbol{\Sigma}_{fT}) \frac{n^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \boldsymbol{\beta}_i}{\bar{\sigma}_n^2} \leq s_{nT}^2 \leq \lambda_{max}(\boldsymbol{\Sigma}_{fT}) \frac{n^{-1} \sum_{i=1}^{n} \boldsymbol{\beta}_i' \boldsymbol{\beta}_i}{\bar{\sigma}_n^2}, \tag{A.42}$$

where $c < \lambda_{min}(\mathbf{\Sigma}_{fT}) < \lambda_{max}(\mathbf{\Sigma}_{fT}) < C$. Hence

$$c\left(\frac{n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\beta}_i}{\bar{\sigma}_n^2}\right) \leq s_{nT}^2 \leq C\left(\frac{n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\beta}_i}{\bar{\sigma}_n^2}\right),$$

and it must be that

$$s_{nT}^2 = \ominus\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i'\boldsymbol{\beta}_i\right) = \ominus\left[\sum_{k=1}^{K}\left(n^{-1}\sum_{i=1}^{n}\beta_{ik}^2\right)\right].$$

Also, under Assumption 4 $n^{-1}\sum_{i=1}^{n}\beta_{ik}^2 = \ominus\left(n^{\alpha_k-1}\right)$. Hence

$$s_{nT}^2 = \sum_{k=1}^{K}\ominus\left(n^{\alpha_k-1}\right),$$

which in view of (A.41) now yields (A.40), as desired. ■

References

Anatolyev, S., and A. Mikusheva (2020) Factor models with many assets: Strong factors, weak factors, and the two-pass procedure, *Journal of Econometrics* forthcoming.

Ang, A., J. Liu and K. Schwarz (2020) Using stocks or portfolios in tests of factor models, *Journal of Financial and Quantitative Analysis, 55, 709-750.*

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, 929-1196.

____ (2019) Exponent of cross-sectional dependence for residuals, *Sankya B. The Indian Journal of Statistics*, 81, 46-102.

____ (2021) Measurement of Factor Strength: Theory and Practice, forthcoming in *Journal of Applied Econometrics*.

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.

Bickel P.J., and E. Levina (2008a) Regularised estimation of large covariance matrices. *The Annals of Statistics*, 36, 199–227.

____ (2008b) Covariance regularisation by thresholding, *The Annals of Statistics*, 36, 2577–2604.

Brennan, M.J., T. Chordia and A. Subrahmanyam (1998) Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics*, 49, 345-373.

Cai T., and W. Liu (2011) Adaptive thresholding for sparse covariance matrix estimation, *Journal of American Statistical Association*, 106, 672–684.

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 (2019) Semi-strong factors in asset returns. Working paper.

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.

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

Harvey, C. R., and Y. Liu (2019) A census of the factor zoo. https://ssrn.com/abstract=3341728.

Jagannathan, R., G. Skoulakis and Z. Wang (2010) The analysis of the cross section of security returns, ch14 of Y Ait-Sahalia & L.P. Hansen (eds) *Handbook of Financial Economics*, vol 2, Elsevier.

Jensen, T.I., B.T. Kelly and L.H. Pedersen (2021) Is there a replication crisis in finance? NBER working paper 28432.

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.

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

Kleibergen, F., and Z. Zhan (2015) Unexplained factors and their effects on second pass R-squared's, *Journal of Econometrics*, 189, 101-116.

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.

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

Pesaran, M. H., and T. Yamagata, (2018), Testing for alpha in linear factor pricing models with a large number of securities, unpublished manuscript, revised version of the paper presented at AFA 2013 San Diego Meetings.

Roll R., (1977) A critique of the asset pricing theorys tests: on past and potential testability of theory. *Journal of Financial Economics* 4, 129-176.

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.