The role of factor strength and pricing errors for estimation and inference in asset pricing models*

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

Ron P. Smith
Birkbeck, University of London

October 16, 2019

Abstract

In this paper we are concerned with the role of factor strength and pricing errors in asset pricing models, and their implications for identification and estimation of risk premia. We establish an explicit relationship between the pricing errors and the presence of weak factors that are correlated with stochastic discount factor. We introduce a measure of factor strength, and distinguish between observed factors and unobserved factors. We show that unobserved factors matter for pricing if they are correlated with the discount factor, and relate the strength of the weak factors to the strength (pervasiveness) of non-zero pricing errors. We then show, that even when the factor loadings are known, the risk premia of a factor can be consistently estimated only if it is strong and if the pricing errors are weak. Similar results hold when factor loadings are estimated, irrespective of whether individual returns or portfolio returns are used. We derive distributional results for two pass estimators of risk premia, allowing for non-zero pricing errors. We show that for inference on risk premia the pricing errors must be sufficiently weak. We consider both when n (the number of securities) is large and T (the number of time periods) is short, and the case of large n and T. Large n is required for consistent estimation of risk premia, whereas the choice of short T is intended to reduce the possibility of time variations in the factor loadings. We provide monthly rolling estimates of the factor strengths for the three Fama-French factors over the period 1989-2018.

JEL Classifications: C38, G12

Key Words: Arbitrage Pricing Theory, APT, factor strength, identification of risk premia, two-pass regressions, Fama-French factors.

*We are grateful to Natalia Bailey for carrying out the computation of the results presented in the empirical section of the paper, and for her helpful comments. We would also like to thank Alex Chudik, George Kapetanios, Allan Timmermann and Takashi Yamagata for their constructive comments and suggestions.

1 Introduction

Asset pricing models tend to attribute differences in expected returns to differences in exposure to systematic risk factors. For instance, the arbitrage pricing theory (APT) formalised by Ross (1976), assumes that there are many assets, with returns determined by a small number of factors, and that competitive markets do not permit arbitrage opportunities. Thus returns can be split into two components: non-diversifiable systematic risk from exposure to the common factors, and idiosyncratic risk, which can be eliminated in a well diversified portfolio. Assets with similar risk factors are close substitutes so should have similar returns. In this linear return generating process, expected excess returns are proportional to systematic risk, measured by factor loadings and risk premia are the coefficients of such loadings. Chamberlain and Rothschild (1983) extend the theory to an approximate factor structure and provide a rigorous treatment of the case of infinitely many assets. Wei (1988) links the APT to the capital asset pricing model, CAPM. Within the context of such models, the risk factors and their loadings have to be identified and the risk premia associated with them estimated. The standard procedure is to estimate factor loadings from a first-pass time series regression of excess returns for each asset, and in a second-pass the Fama-MacBeth (1973) type cross section regression is used to price the factors and obtain the risk premia.

In this paper we are concerned with the role of factor strength and pricing errors in asset pricing models, and their implications for identification and estimation of risk premia. We distinguish between strong and weak factors that underlie the APT by relating them to the stochastic discount factor, m_t, used to price securities within the inter-temporal asset pricing models. We introduce a measure of factor strength denoted by δ, and distinguish between observed factors, $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{kt})'$, with strengths $\boldsymbol{\delta}_f = (\delta_{f_1}, \delta_{f_2}, ..., \delta_{f_k})'$, and unobserved factors (included in the idiosyncratic errors) labelled $\mathbf{g}_t = (g_{1t}, g_{2t}, ..., g_{k_g t})'$ with strengths $\boldsymbol{\delta}_g = \left(\delta_{g_1}, \delta_{g_2}, ..., \delta_{g_{k_g}}\right)'$. Non-zero pricing errors arise when the errors in the factor model are correlated with m_t. By decomposing the errors into a part which is correlated with m_t, and a *purely* idiosyncratic part, ε_t, which is not, we show that the APT equilibrium condition only bounds the cross correlation of the first part and does not impose any restrictions on the cross correlation of the purely idiosyncratic part. Such cross correlations could arise at times of financial crises possibly due to herding and other forms of correlated behavior that stem from non-fundamental considerations.

We define the strength of a given factor, say f_t, by the exponent δ_f in the following norm condition

$$\sum_{i=1}^{n} \left(\beta_i - \bar{\beta}_n\right)^2 = \ominus\left(n^{\delta_f}\right), \tag{1}$$

where β_i is the loading of f_t on the i^{th} security, $\bar{\beta}_n = n^{-1}\sum_{i=1}^{n}\beta_i$, and $\ominus\left(n^{\delta_f}\right)$ denotes the expansion rate of the dispersion measure $\sum_{i=1}^{n}\left(\beta_i - \bar{\beta}_n\right)^2$ in terms of n. To motivate the use of δ in the analysis of factor pricing we first generalise Ross's APT equilibrium condition. Ross required pricing errors, denoted by η_i for security i, to be bounded, in the sense that $\sum_{i=1}^{n}\eta_i^2 < \infty$. We show that this condition can be relaxed to

$$\sum_{i=1}^{n} \eta_i^2 = O\left(n^{\alpha}\right), \tag{2}$$

with the exponent α measuring the degree of pervasiveness of pricing errors (which as noted above relates to unobserved common factors that are correlated with the discount factor). The exponent of the pricing errors, α, is to be contrasted with the exponent of factor strength δ. For a factor to be strong it is necessary that its effect is pervasive on all securities. In terms of (1), this requires $\delta_f = 1$. Whilst, the pricing errors must not be pervasive, requiring that $\alpha < 1$. Note that the standard order condition, $O(\cdot)$, used in (2) allows zero pricing errors ($\eta_i = 0$ for all i), and $\ominus(\cdot)$ ensures that the effects of f_t are sufficiently pervasive when $\delta_f = 1$.

Having linked the APT to the inter-temporal asset pricing condition, and given the distinction between $\mathbf{f}_t$ and $\mathbf{g}_t$ factors, we then consider the problem of identification of risk premia, initially assuming known factor loadings, using an approximate linear factor model. We show that for large n the risk premia can be $\sqrt{n}$ consistently estimated if the factors all have maximum strength with $\delta_f = 1$, and the pricing errors are weak such that $\alpha < 1$. In the literature the distinction between risk factors and pricing errors is not always as clear-cut. For instance Stambaugh & Yuan (2017, p. 1273) say "there need not be a clear distinction between mispricing and risk compensation as alternative motivations for factor models of expected returns." They use the example of "noise-trader" sentiment, but if such sentiment is sufficiently pervasive it constitutes a factor that may be priced.

We further establish that these conditions for the identification of the risk premia are unaffected if one uses portfolios as compared to individual securities, as is often done in the empirical literature. There is a belief that the first stage estimation of the betas causes an errors in variables problem in the second stage and that the construction of portfolios mitigates this problem. We regard this as a generated regressor problem, which is rather different, and like Ang et al. (2019) argue that forming portfolios wastes information. Nonetheless we analyse both cases: using individual assets and using portfolios.

We then move to the more interesting case where the factor loadings are unknown, and derive conditions under which the risk premia can be identified. Using the two pass estimator of Fama-MacBeth, we show that risk premia are only identified if both n and $T \to \infty$, such that $n/T \to \kappa$, with $0 < \kappa < \infty$, all factors have maximum strength, and the pricing errors are sufficiently weak. The large T results on estimation of risk premia obtained in the literature, and reviewed for example by Jagannathan, Skoulakis & Wang (2010), only apply if it is assumed that pricing errors are all zero (namely $\eta_i = 0$ for all i). In the presence of non-zero pricing errors, we need $n \to \infty$, and it is not sufficient to consider large T asymptotics. When T is fixed and $n \to \infty$, we consider a bias-corrected version of the two pass estimator and show that in this case risk premia are not identified and one can only consistently estimate what Shanken (1992, p. 6) has termed "ex-post prices of risk". For this result, we still require all factors to have maximum strength and the pricing errors to be sufficiently weak. Finally, we consider the limiting distributions of the bias corrected estimator, centred around the ex post risk premia, for a fixed T and as $n \to \infty$, as well as when $n, T \to \infty$. Under the former we show that the limiting distribution of bias-corrected two pass estimator exists, but need not be Gaussian due to the error cross sectional dependence. In contrast, under joint n and T asymptotics, the estimator is asymptotically normal, and does not depend on the errors of individual securities, but is primarily driven by the time series properties of the factors. In both cases, to ensure that the asymptotic distributions do not depend on the pricing errors, we must have $\sum_{i=1}^{n} |\eta_i| = O(n^{\alpha_*})$, with $\alpha_* < 1/2$, which is much more restrictive than the condition needed on α for consistent estimation of risk premia.

To measure factor strength we use an estimator of δ recently proposed by Bailey, Kapetanios and Pesaran (2019b, BKP), which in turn builds on earlier papers by the same authors (BKP

2016 and BKP 2019a) on measures of cross-sectional dependence in large panels. Whereas BKP (2019b) are concerned with estimation and inference about the strength of a factor, as noted above this paper is concerned with a different issue, the theoretical role of factor strength and pricing errors in asset pricing models and estimation and inference about risk premia in such models.

BKP (2019b) show that the strength of a factor can be estimated from the proportion of statistically significant loadings across a large number of securities. Let n be the number of securities and D_n the number of significant loadings, then $\pi_n = D_n/n$ is the proportion of non-zero loadings and the measure of factor strength is the logarithmic transform: $\delta_f = 1 + ln(\pi_n)/ln(n)$. The critical values of the tests are suitably adjusted to allow for multiple testing and it is shown that this estimator is consistent. Using Monte Carlo experiments, BKP also show that the estimates of the factor strengths are quite accurate when the factors are sufficiently strong, even for moderate sample sizes. It is important to note that for identification of risk premia, the crucial measure is δ_f not π_n. Since $\pi_n = n^{\delta-1}$, then if δ does not equal one, the proportion of the n securities that have non-zero loadings goes to zero with n, albeit rather slowly at log rate. For the proportion of non-zero loadings to be constant as n rises, it is required that $\delta = 1$.

We conclude the paper with an empirical application, using the estimator of δ proposed by BKP (2019b). For each monthly time point ending from September 1989 to May 2018, the stocks in the S&P 500 at that month with a 10 year history are identified. Then a time series regression is estimated regressing the excess return for each stock on a constant and the three Fama & French (1993) factors. These are the market factor, the size factor, small minus big (SMB), and the value factor, high book to market minus low (HML). The estimated strength of the market factor is either one or very close to one. The market factor is always much stronger than the other two factors, whose strength varies substantially over the period considered, and tend to vary between 0.65 and 0.9. The confidence bounds around the estimates of the factor strengths are tight and the outcomes are reasonably robust to the choice of estimation window.

Throughout this paper we assume that the potential factors are known. There is no shortage of suggested factors, Harvey and Liu (2019) document a "factor zoo" of over 400 suggested factors in early 2019. They may be observable macro factors like those suggested by Chen, Roll & Ross (1986); obtained from factor analysis of the returns as discussed in Lehmann & Modest (2005); based on asset pricing anomalies like the Fama-French factors; or obtained in some other way. Fama & French (2018) discuss some issues in choosing factors. We are concerned with establishing the theoretical role of factor strengths and providing estimates of them, not with the issue of how to select factors.

Our analysis has important practical implications for estimating factor risk premia from second-pass cross section regressions of average returns on factor loadings. The factors used in the first-pass time series regressions to estimate the loadings must be sufficiently strong. The strength of the factor ought to be measured using rolling windows estimates of the parameter δ, as illustrated in Section 5. Only factors with an average strength of $\delta \geq 2/3$, should be included in the analysis. Estimates of the risk premia for factors with $\delta = 2/3$ will be consistent at the rate of $n^{1/3}$, and thus will be difficult to price unless n is very large indeed. The situation is clearly worse for $\delta < 2/3$, when estimates of δ also become less reliable due to increased sampling uncertainty. Furthermore, to identify the risk premia of weak factors the pricing errors that arise from unobserved factors correlated with the stochastic discount factor, must be sufficiently weak. (See Remark 9, in Section 4.1).

The rest of the paper is organized as follows. Section 2 sets out the factor model and derives the APT pricing errors by imposing the equilibrium conditions from standard pricing theory. Section 3 discusses the identification of the risk premia for the factors from a cross section when the factor loadings are known. It considers both the cases where the observations are individual securities and where they are portfolios. Section 4 discusses the case where the factor loadings are unknown and have to be estimated in the first-pass time series regressions. Again it considers individual securities and portfolios. Section 5 presents the estimates of time-varying factor strength. Section 6 has some concluding comments. Lemmas, proofs and related results are provided in appendices.

Notation: 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. $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$ and $\|\boldsymbol{A}\|_F = [Tr(\mathbf{A}'\mathbf{A})]^{1/2}$ denote the spectral and Frobenius 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 $n_0 \geq 1$ and positive finite constants C_0 and C_1, such that $\inf_{n \geq n_0}(f_n/g_n) \geq C_0$, and $\sup_{n \geq n_0}(f_n/g_n) \leq C_1$.

2 APT, equilibrium pricing theory and non-zero pricing errors

Non-zero pricing errors have a central role in this paper. This section sets out the factor model and derives the APT pricing errors in terms of the correlation of the discount factor with the idiosyncratic component of excess returns. We achieve this by imposing the equilibrium conditions from standard pricing theory on the linear multi-factor model used by Ross and others in the literature. We show that non-zero pricing errors arise when there are factors in the idiosyncratic (error) part of excess returns that are correlated with the stochastic discount factor.

To establish our main result we follow the literature and assume that returns, $r_{i,t+1}$, $i = 1, 2, ...n$, are generated according to the following linear multi-factor model[1]

$$r_{i,t+1} - r_t^f = \mathrm{a}_{it} + \sum_{j=1}^{k} \beta_{it,j} f_{j,t+1} + u_{i,t+1}, \text{ for } i = 1, 2, ..., n, \tag{3}$$

where r_t^f is the risk free rate; a_{it} are the intercepts in the factor model; $f_{j,t+1}$, $j = 1, 2, ..., k$ are the common factors with associated factor loadings, $\beta_{it,j}$; and $u_{i,t+1}$ is the idiosyncratic component of asset return. The model can be written more compactly as

$$r_{i,t+1} - r_t^f = \mathrm{a}_{it} + \boldsymbol{\beta}_{it}' \mathbf{f}_{t+1} + u_{i,t+1}, \tag{4}$$

[1]Because the number of assets may change through time, we could have a time varying n_t. But to keep the notations simple we assume $n_t = n$.

where $\boldsymbol{\beta}_{it} = (\beta_{it,1}, \beta_{it,2}, ..., \beta_{it,k})'$, and $\mathbf{f}_{t+1} = (f_{1,t+1}, f_{2,t+1}, ..., f_{k,t+1})'$. We assume the errors of the factor model, $u_{i,t+1}$, are martingale differences, $E_t(u_{i,t+1}) = 0$, and have finite conditional variances, and are cross-sectionally weakly correlated. We explore the relationship between $u_{i,t+1}$ and the pricing errors in detail below. But at this stage we view $u_{i,t+1}$ as errors in the statistical factor model used to represent the time series variations of individual excess returns.

The cross-sectional dependence of the individual returns is captured by the common factors, $\mathbf{f}_{t+1}$, and the degree of cross-sectional dependence of the errors, $u_{i,t+1}$. Denoting the $n \times 1$ vector of errors by $\mathbf{u}_{n,t+1} = (u_{1,t+1}, u_{2,t+1}, ..., u_{n,t+1})'$, and its covariance matrix by $\boldsymbol{\Sigma}_{u,nt} = E_t\left(\mathbf{u}_{n,t+1}\mathbf{u}'_{n,t+1}\right)$, an overall measure of cross dependence of the errors is given by the rate at which $\lambda_{\max}(\boldsymbol{\Sigma}_{u,nt})$ rises with n. We denote this rate by $\delta_{t,\sigma}$, and relate it to the pervasiveness of the pricing errors under APT. Ross (1976) assumed $u_{i,t+1}$ were cross sectionally independent, Chamberlain & Rothschild (1983) weakened this to an approximate factor model that requires $\lambda_{\max}(\boldsymbol{\Sigma}_{u,nt}) < C$ in n and t, which corresponds to setting $\delta_{t,\sigma} = 0$ for all t.

In what follows we require that $\delta_{t,\sigma} < 1$, and assume that the factors, $\mathbf{f}_{t+1}$, are strong in the sense that δ_{t,f_j} defined by[2]

$$\sum_{i=1}^{n} \left(\beta_{it,j} - \bar{\beta}_{tj}\right)^2 = \ominus\left(n^{\delta_{t,f_j}}\right), \text{ for } j = 1, 2, ..., k, \tag{5}$$

are equal to unity. The main result of the APT can now be written as the cross section return regression

$$E_t\left(r_{i,t+1} - r_t^f\right) = \boldsymbol{\beta}'_{it}\boldsymbol{\lambda}_t + \eta_{it}, \tag{6}$$

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

$$sup_t \sum_{i=1}^{n} \eta_{it}^2 < C. \tag{7}$$

To relate the pricing errors to the idiosyncratic component of returns we use standard results from intertemporal asset pricing theory that require equilibrium prices, P_{it}, to be set as the expected discounted value of the payoff, $P_{i,t+1} + D_{i,t+1}$. Denoting the holding period return by $r_{i,t+1} = (\Delta P_{i,t+1} + D_{i,t+1})/P_{it}$, the basic equilibrium pricing equation can be written as

$$E_t\left[m_{t+1}(r_{i,t+1} - r_t^f)\right] = 0, \tag{8}$$

where m_{t+1} is the stochastic discount factor used to price all assets in the market, and r_t^f is the risk free rate.[3] We also note that

$$E_t(m_{t+1}) = 1/(1 + r_t^f) > 0. \tag{9}$$

To derive conditions under which the factor model (4) also satisfies the equilibrium conditions, substituting for $r_{i,t+1} - r_t^f$ from (4) in (8), we have

$$\mathrm{a}_{it}E_t(m_{t+1}) + \boldsymbol{\beta}'_{it}E_t(m_{t+1}\mathbf{f}_{t+1}) + E_t(m_{t+1}u_{i,t+1}) = 0.$$

[2] See also (1) and the related discussion in the Introduction.

[3] Under consumption based asset pricing $m_{t+1} = \rho u'(c_{t+1})/u'(c_t)$, where ρ is the subjective discount factor and $u'(c_t)$ is marginal utility of consumption c_t.

Now noting that $E_t(m_{t+1}) > 0$, (see (9)), then a_{it} can be solved as

$$\mathrm{a}_{it} = -\frac{1}{E_t(m_{t+1})}\left[\boldsymbol{\beta}'_{it}E_t\left(m_{t+1}\mathbf{f}_{t+1}\right) + E_t(m_{t+1}u_{i,t+1})\right]. \tag{10}$$

Imposing this restriction by substituting the above expression for a_{it} back in (3) yields

$$r_{i,t+1} - r_t^f = \boldsymbol{\beta}'_{it}\left(\mathbf{f}_{t+1} - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})}\right) - \frac{E_t(m_{t+1}u_{i,t+1})}{E_t(m_{t+1})} + u_{i,t+1}. \tag{11}$$

Taking conditional expectations of both sides of the above and noting that $E_t\left(u_{i,t+1}\right) = 0$, we now have

$$E_t\left(r_{i,t+1} - r_t^f\right) = \boldsymbol{\beta}'_{it}E_t\left(\mathbf{f}_{t+1} - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})}\right) - \frac{E_t(m_{t+1}u_{i,t+1})}{E_t(m_{t+1})}.$$

Matching this result with the APT equilibrium condition given by (6), we obtain the following expressions for the risk premia and the pricing errors

$$\boldsymbol{\lambda}_t = E_t\left(\mathbf{f}_{t+1} - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})}\right) = -\frac{Cov_t(m_{t+1}, \mathbf{f}_{t+1})}{E_t(m_{t+1})}, \tag{12}$$

and

$$\eta_{it} = -\frac{E_t(m_{t+1}u_{i,t+1})}{E_t(m_{t+1})}, \text{ for } i = 1, 2, ..., n. \tag{13}$$

The pricing errors are zero only if $u_{i,t+1}$ and m_{t+1} are conditionally uncorrelated. Also their identification requires knowledge of the stochastic discount factor as well as the factors. But (13) does not place any restrictions on the relationships between the pricing errors and the factor loadings, which as we shall see is an important consideration for estimation of risk premia.

For a better insight into the factors that could lead to non-zero pricing errors, it is useful to decompose $u_{i,t+1}$ into three mean zero components: (i) a set of k_g *unobserved* factors denoted by $\mathbf{g}_{t+1} = (g_{1,t+1}, g_{2,t+1}, ..., g_{kg,t+1})'$ with $E_t\left(\mathbf{g}_{t+1}\right) = \mathbf{0}$, that are *correlated* with the discount factor, m_{t+1}; (ii) a set of k_h *unobserved* factors $\mathbf{h}_{t+1} = (h_{1,t+1}, h_{2,t+1}, ..., h_{k_h,t+1})'$ with $E_t\left(\mathbf{h}_{t+1}\right) = 0$ that are *uncorrelated* with m_{t+1} ; and (iii) idiosyncratic errors, $\varepsilon_{i,t+1}$, that are also uncorrelated with m_{t+1}, but could be cross-sectionally correlated without having a common factor representation. Such dependencies could arise from local or network spillover effects that are unrelated to m_{t+1}. The error components that do not depend on m_{t+1}, namely $\mathbf{h}_{t+1}$ and $\varepsilon_{i,t+1}$, can still become highly correlated across securities at times of financial crises, for example, possibly due to herding behavior or correlated beliefs that cause asset returns to move together in a way that is unrelated to the economy's fundamentals. The unobserved factors $\mathbf{g}_{t+1}$ and $\mathbf{h}_{t+1}$ are distinguished from the observed risk factors, $\mathbf{f}_t$, in two respects. First, since $\mathbf{g}_{t+1}$ and $\mathbf{h}_{t+1}$ are martingale difference processes thus having zero means, whilst risk factors, $\mathbf{f}_{t+1}$, need not have zero means; and most importantly the two sets of factors could differ in their strengths. As we shall see, the APT theory requires the unobserved factors that are correlated with m_{t+1} must be weak and the risk factors must be strong. It does not impose any restrictions on the degree of cross-sectional dependence of the returns that are due to $\mathbf{h}_{t+1}$ and/or $\varepsilon_{i,t+1}$, the two components of $u_{i,t+1}$ that are unrelated to m_{t+1}.

More specifically let

$$u_{i,t+1} = \sum_{j=1}^{k_g}\phi_{it,j}g_{j,t+1} + \sum_{j=1}^{k_h}\psi_{it,j}h_{j,t+1} + \varepsilon_{i,t+1} = \boldsymbol{\phi}'_{it}\mathbf{g}_{t+1} + \boldsymbol{\psi}'_{it}\mathbf{h}_{t+1} + \varepsilon_{i,t+1}, \tag{14}$$

where $\boldsymbol{\phi}_{it} = (\phi_{it,1}, \phi_{it,2}, ..., \phi_{it,k_g})'$ and $\boldsymbol{\psi}_{it} = (\psi_{it,1}, \psi_{it,2}, ..., \psi_{it,k_h})'$ are $k_g \times 1$ and $\kappa_h \times 1$ vectors of loadings associated to $\mathbf{g}_{t+1}$ and $\mathbf{h}_{t+1}$, respectively. Now using (14) in (13) and noting that $\mathbf{h}_{t+1}$ and $\varepsilon_{i,t+1}$ are uncorrelated with m_{t+1}, we have

$$\eta_{it} = -\sum_{j=1}^{k_g} \theta_{j,t+1}\phi_{it,j} = -\boldsymbol{\theta}_{t+1}^{'}\boldsymbol{\phi}_{it}, \tag{15}$$

where $\boldsymbol{\theta}_{t+1} = Cov_t\left(\mathbf{g}_{t+1}, m_{t+1}\right)/E_t(m_{t+1}) \neq \mathbf{0}$. Therefore, the pricing errors arise solely due to the non-zero correlation of $\mathbf{g}_{t+1}$ and m_{t+1}.[4] Using (15) in (13) we have

$$\sum_{i=1}^{n} \eta_{it}^2 = \boldsymbol{\theta}_{t+1}' \left(\sum_{i=1}^{n} \boldsymbol{\phi}_{it}\boldsymbol{\phi}_{it}' \right) \boldsymbol{\theta}_{t+1}. \tag{16}$$

It is clear that the leading terms of above expression are given by $\sum_{i=1}^{n} \phi_{it,j}^2 = O(n^{\delta_{t,g_j}})$, for $j = 1, 2, ..., k_g$, where δ_{t,g_j} is the strength of factor $g_{j,t+1}$. The strength of the pricing errors, given by $\sum_{i=1}^{n} \eta_{it}^2 = O(n^{\alpha_t})$, is determined by the strongest of the weak factors, $g_{j,t}$ with $\alpha_t = \sup_j(\delta_{t,g_j})$. Below we will see that to identify the risk premia, the factors f_{jt} must be strong with $\delta_{t,f_j} = 1$, for all $j = 1, 2, .., k$, and for estimation and inference on the risk premia the pricing errors must be sufficiently weak, $\alpha_t = \sup_j(\delta_{t,g_j}) < 1/2$.

The above derivations link the pervasiveness of the pricing errors to the presence of weak factors in the idiosyncratic errors, $u_{i,t+1}$, of the return equations. The equilibrium pricing condition of Ross (1976), given by (7), is satisfied if the factors $\mathbf{g}_{t+1}$ that enter the idiosyncratic component of the returns have zero strength, in the sense that $\delta_{t,g_j} = 0$, for all j. Our analysis relaxes this condition by requiring that $\delta_{t,g_j} < 1/2$. Also the components of $u_{i,t+1}$ in (14), $\mathbf{h}_{t+1}$ and $\varepsilon_{i,t+1}$,which do not depend on m_{t+1}, can be more strongly cross-correlated, because of herding or correlated beliefs, than the first component, $\mathbf{g}_{t+1}$ which depends on the fundamentals, and is governed by the APT condition, (7).

3 Identification of risk premia with known factor loadings

Identification of factor risk premia, $\boldsymbol{\lambda}_t$, can be achieved either using individual securities or portfolios of securities. We consider each of these approaches in turn. We make the following assumptions about the idiosyncratic errors, factors and their loadings.

Assumption 1 *(Weak cross sectional error dependence) The errors, $u_{i,t+1}$, are martingale differences with respect to the information set available at time t, $\mathfrak{F}_t$, have finite variances,*

[4] Examples of factors that are martingale difference processes and at the same time are correlated with the stochastic discount rate, include $g_{j,t+1} = m_{t+1}^j - E_t(m_{t+1}^j)$. For $j = 1$ we have $Cov_t\left(g_{1,t+1}, m_{t+1}\right)/E_t(m_{t+1}) = Var_t(m_{t+1})/E_t(m_{t+1})$ which is clearly non-zero. In the case of consumption-based asset pricing $m_{t+1} = \beta u'(c_{t+1})/u'(c_t)$, and $g_{1,t+1}$ takes the following specific form

$$g_{1,t+1} = \frac{\beta\left\{u'(c_{t+1}) - E_t\left[u'(c_{t+1})\right]\right\}}{u'(c_t)},$$

where $u'(c_t)$ is the marginal utility of consumption, and β is the discount factor.

$E_t(u_{i,t+1}^2) = \sigma_{it}^2$, *where* $E_t(.) = E(. \mid \mathfrak{F}_t)$., *and are cross-sectionally weakly correlated such that*

$$\sup_j \sum_{i=1}^{n} |E_t(u_{i,t+1}u_{j,t+1})| = O(n^{\delta_{t,\sigma}}), \tag{17}$$

with $\delta_{t,\sigma} < 1$.

Assumption 2 *(Common factors) The* $k \times 1$ *vector of factors,* $\mathbf{f}_{t+1}$*, has mean* $E_t(\mathbf{f}_{t+1}) = \mu_{ft}$ *and a positive definite matrix* $\mathbf{\Sigma}_{ft} = E_t\left[(\mathbf{f}_{t+1} - \mu_{ft})(\mathbf{f}_{t+1} - \mu_{ft})'\right] > 0$.

Assumption 3 *(Factor loadings) The* $n \times k$ *matrix of factor loadings,* $\mathbf{B}'_{nt} = (\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t}, ..., \boldsymbol{\beta}_{nt})$ *has full column rank and*

$$\lim_{n\to\infty} \left(n^{-1}\mathbf{B}'_{nt}\mathbf{M}_n\mathbf{B}_{nt}\right) = \mathbf{\Sigma}_{t,\beta\beta} > 0, \tag{18}$$

where $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}'_n$, $\boldsymbol{\tau}_n = (1, 1, ..., 1)'$, *and* $\mathbf{\Sigma}_{t,\beta\beta}$ *is a* $k \times k$ *symmetric positive definite matrix.*

Definition 1 *(Factor strengths) The strength of factor* $f_{j,t}$ *is measured by its degree of pervasiveness as defined by the exponent* δ_{t,f_j} *in*

$$\sum_{i=1}^{n} (\beta_{it,j} - \overline{\beta}_{tj})^2 = \ominus(n^{\delta_{t,f_j}}), \text{ for } j = 1, 2, ..., k, \tag{19}$$

where $\overline{\beta}_{tj} = n^{-1}\sum_{i=1}^{n} \beta_{it,j}$*, and* $0 \leq \delta_{t,f_j} \leq 1$*. We refer to* $\left\{\delta_{t,f_j}, j = 1, 2, ..., k\right\}$ *as factor strengths. Factor* $f_{j,t}$ *is said to have maximum strength at time* t *if* $\delta_{t,f_j} = 1$.

Proposition 1 *The asymptotic covariance matrix of factor loadings,* $\mathbf{\Sigma}_{t,\beta\beta}$*, defined by (18) is positive definite only if all the factors have maximum strengths, namely if* $\delta_{t,f_j} = 1$ *for all* $j = 1, 2, ..., k$.

Remark 1 *The above definition of factor strength allows for the possibility of non-zero pricing errors* ($\eta_{it} \neq 0$) *in the theory consistent factor model (11), and in the related APT equilibrium condition (6). In the absence of pricing errors (i.e. when* $\eta_{it} = 0$*), the condition (19) indeed simplifies to* $\sum_{i=1}^{n} \beta_{it,j}^2 = \ominus(n^{\delta_{t,f_j}})$*, for* $j = 1, 2, ..., k$*. In what follows we adopt the more general definition given above and further elaborate on its relevance in relation to our empirical application.*

Remark 2 *Under Assumption 1,* $u_{i,t+1}$ *are serially uncorrelated with zero means.*

Remark 3 *Let* $\mathbf{\Sigma}_{u,nt} = E_t\left(\mathbf{u}_{n,t+1}\mathbf{u}'_{n,t+1}\right)$*, then condition (17) also ensures that*

$$\lambda_{\max}(\mathbf{\Sigma}_{u,nt}) = O(n^{\delta_{t,\sigma}}).$$

This follows since

$$\lambda_{\max}(\mathbf{\Sigma}_{u,nt}) = |\lambda_{\max}(\mathbf{\Sigma}_{u,nt})| \leq \|\mathbf{\Sigma}_{u,nt}\|_1 = \sup_i \sum_{i=1}^{n} |E_t(u_{i,t+1}u_{j,t+1})|.$$

Therefore, condition (17) relaxes the standard assumption of the approximate factor models used in the APT literature that requires $\lambda_{\max}(\mathbf{\Sigma}_{u,nt}) < C$ in n and t, which corresponds to setting $\delta_{t,\sigma} = 0$ for all t.

3.1 Identification using individual securities

Consider the APT equations (6), denote the expected returns on asset i by $\mu_{it} = E_t(r_{i,t+1})$, and stack the equations for $i = 1, 2, ..., n$, to obtain:

$$\boldsymbol{\mu}_{nt} - r_t^f \boldsymbol{\tau}_n = \mathbf{B}_{nt} \boldsymbol{\lambda}_t + \boldsymbol{\eta}_{nt}, \tag{20}$$

where $\mathbf{B}_{nt}$ is the $n \times k$ matrix of factor loadings, $\boldsymbol{\mu}_{nt} = (\mu_{1t}, \mu_{2t}, ..., \mu_{nt})'$ and $\boldsymbol{\eta}_{nt} = (\eta_{1t}, \eta_{2t}, ..., \eta_{nt})'$. In practice r_t^f is not known and is often treated as unknown time effect and (21) is written more generally as

$$\boldsymbol{\mu}_{nt} = \lambda_{0t} + \mathbf{B}_{nt} \boldsymbol{\lambda}_t + \boldsymbol{\eta}_{nt}, \tag{21}$$

where λ_{0t} is treated as an unknown time effect. Under this setting and assuming $\mathbf{B}_{nt}$ is known, $\boldsymbol{\lambda}_t$ is identified if Assumption 3 holds, that is

$$\lim_{n \to \infty} \left(\frac{\mathbf{B}'_{nt} \mathbf{M}_n \mathbf{B}_{nt}}{n} \right) = \boldsymbol{\Sigma}_{t,\beta\beta} > \mathbf{0}, \tag{22}$$

and

$$n^{-1} \mathbf{B}'_{nt} \mathbf{M}_n \boldsymbol{\eta}_{nt} \to_p \mathbf{0}, \tag{23}$$

where $\mathbf{M}_n = \mathbf{I}_n - n^{-1} \boldsymbol{\tau}_n \boldsymbol{\tau}'_n$. Both conditions are likely to be met when all the κ factors are strong, namely the exponent of their factor loadings is unity. The second condition is met under the APT condition of bounded pricing errors, namely if $\sum_{i=1}^n \eta_{it}^2 = O(n^{\alpha_t})$, with $\alpha_t < 1$. This follows since

$$\begin{aligned} n^{-1} \left\| \mathbf{B}'_{nt} \mathbf{M}_n \boldsymbol{\eta}_{nt} \right\|_F &\leq n^{-1} \left\| \mathbf{B}'_{nt} \mathbf{M}_n \right\|_F \left\| \boldsymbol{\eta}_{nt} \right\|_F \\ &= \left[Tr \left(n^{-1} \mathbf{B}'_{nt} \mathbf{M}_n \mathbf{B}_{nt} \right) \right]^{1/2} \left(n^{-1} \sum_{i=1}^n \eta_{it}^2 \right)^{1/2} . \end{aligned}$$

Since k is fixed then $Tr \left(n^{-1} \mathbf{B}'_{nt} \mathbf{M}_n \mathbf{B}_{nt} \right) = \ominus(1)$, and $n^{-1} \sum_{i=1}^n \eta_{it}^2 = O(n^{\alpha_t - 1})$. Hence, $n^{-1} \left\| \mathbf{B}'_{nt} \mathbf{M}_n \boldsymbol{\eta}_{nt} \right\|_F \to 0$ if $\alpha_t < 1$. Under these conditions $\boldsymbol{\lambda}_t$ can be estimated consistently by

$$\boldsymbol{\lambda}_t = \left(\mathbf{B}'_{nt} \mathbf{M}_n \mathbf{B}_{nt} \right)^{-1} \mathbf{B}'_{nt} \mathbf{M}_n \boldsymbol{\mu}_{nt}.$$

In practice, the matrix of factor loadings must also be estimated which entails further restrictions on the stability of the factor loadings to be discussed below.

3.2 Identification using portfolios

Following Fama and MacBeth (1973), it is often argued in the literature that more robust estimates of $\boldsymbol{\lambda}_t$ can be obtained by using portfolios constructed from the individual securities. We consider two types of portfolio weights: (a) a small number of fully diversified portfolios, and (b) a large number of portfolios formed from a small number of securities. In both cases we denote the portfolio weights by the $n \times 1$ vector $\mathbf{w}_{pt} = (w_{1p,t}, w_{2p,t}, ..., w_{np,t})'$, and consider P return portfolios, $\bar{r}_{pt}$, defined by

$$\bar{r}_{pt} = \sum_{i=1}^n w_{ip,t} r_{it} = \mathbf{w}'_{pt} \mathbf{r}_{nt}, \text{ for } p = 1, 2, ..., P. \tag{24}$$

Collecting all the portfolio weights in the $n \times P$ portfolio weights matrix $\mathbf{W}_{Pt} = (\mathbf{w}_{1t}, \mathbf{w}_{2t}, ..., \mathbf{w}_{Pt})$, we also have $\bar{\boldsymbol{\eta}}_{Pt} = \mathbf{W}'_{Pt}\boldsymbol{\eta}_{nt}$, and

$$\bar{\mathbf{r}}_{Pt} = \mathbf{W}'_{Pt}\mathbf{r}_{nt}, \tag{25}$$

where $\bar{\mathbf{r}}_{Pt} = (\bar{r}_{1t}, \bar{r}_{2t}, ..., \bar{r}_{Pt})'$, is the $P \times 1$ vector of portfolio returns.

In the case of fully diversified portfolios we assume that $\sup_{it,p} \{n |w_{ip,t}|\} < \bar{c} < \infty$ and $\inf_{it,p} \{n |w_{ip,t}|\} > \underline{c} > 0$, which ensures $w_{ip,t} = \ominus(n^{-1})$ and $\|\mathbf{W}_{Pt}\| = \ominus\left(n^{-1/2}\right)$. In the case of non-diversified portfolios, $w_{ip,t}$ is non-zero only for a finite number of securities. The following assumption covers both types of portfolios and is generally applicable.

Assumption 4 *(Portfolio weights) The portfolio weights, w_{ip}, for $i = 1, 2, .., n; p = 1, 2, ..., P$ satisfy the following conditions*

$$\text{(a): } \sum_{i=1}^{n} w_{ip,t} = 1, \text{ (b): } \sup_{p,n} \sum_{i=1}^{n} |w_{ip,t}| < C, \text{ and (c): } \sup_{i,P} \sum_{p=1}^{P} |w_{ip,t}| < C. \tag{26}$$

Remark 4 *The normalization restriction, $\sum_{i=1}^{n} w_{i,pt} = 1$, is made for convenience and is not necessary and other choices such as $\sum_{i=1}^{n} w_{it}^{p} = 0$, can also be entertained. Short sales ($w_{it}^{p} < 0$) are allowed, and it is easily verified that the above assumption applies to a wide variety of portfolios, fully diversified or mutually exclusive portfolios with each security appearing in only one portfolio. Condition (b) of the above assumption follows from the normalization condition if $w_{it} \geq 0$. The important binding condition (c) restricts the frequency with which the same security enters all the P portfolios. Conditions (a) and (b) can also be written as bounds on rows and columns of $\mathbf{W}_{Pt}$, namely $\|\mathbf{W}_{Pt}\|_1 < C$ and $\|\mathbf{W}_{Pt}\|_\infty < C$.*

Remark 5 *For the purpose of the identification analysis that follows, the primary difference between fully diversified and non-diversified portfolios is captured by the rate at which the spectral norm of the portfolio weights matrix, $\|\mathbf{W}_{Pt}\|$, varies with the number of securities included in each portfolio. In the case of fully diversified portfolios we require that $\|\mathbf{W}_{Pt}\| = \ominus\left(n^{-1/2}\right)$, and for non-diversified portfolios we will assume that $\|\mathbf{W}_{Pt}\| = \ominus\left(m^{-1/2}\right)$ where m is the maximum number of securities included in a single portfolio. As an example of the latter note that for mutually exclusive portfolios $\mathbf{w}'_{pt}\mathbf{w}_{p',t} = 0$ for all $p \neq p'$, and $\mathbf{w}'_{pt}\mathbf{w}_{pt} = 1/m$, where m is the integer part of n/P, and $\|\mathbf{W}_{Pt}\| = m^{-1/2}$. In this set up m is fixed and n and $P \to \infty$, such that $n/P \to m \geq 1$. When $m = 1$ portfolios coincide with individual securities.*

Aggregating (3) we have the following expressions for portfolio excess returns (using $\sum_{i=1}^{n} w_{it}^{p} = 1$)

$$\bar{r}_{p,t+1} - r_t^f = \bar{\mathrm{a}}_{pt} + \bar{\boldsymbol{\beta}}'_{pt}\mathbf{f}_{t+1} + \bar{u}_{p,t+1}, \text{ for } p = 1, 2, ..., P, \tag{27}$$

where

$$\bar{\mathrm{a}}_{pt} = \sum_{i=1}^{n} w_{ip,t}\mathrm{a}_{it}, \; \bar{\boldsymbol{\beta}}_{pt} = \sum_{i=1}^{n} w_{pi,t}\boldsymbol{\beta}_{it}, \text{ and } \bar{u}_{p,t+1} = \sum_{i=1}^{n} w_{ip,t}u_{i,t+1}. \tag{28}$$

Then substitute (27) in (8) to give

$$E_t\left[m_{t+1}\left(\bar{\mathrm{a}}_{pt} + \bar{\boldsymbol{\beta}}'_{pt}\mathbf{f}_{t+1} + \bar{u}_{p,t+1}\right)\right] = 0,$$

$$\bar{\mathrm{a}}_{pt}E(m_{t+1}) + \bar{\boldsymbol{\beta}}'_{pt}E_t\left(m_{t+1}\mathbf{f}_{t+1}\right) + E_t\left(m_{t+1}\bar{u}_{p,t+1}\right) = 0,$$

solving for $\bar{\mathrm{a}}_{pt}$ as was done there

$$\bar{\mathrm{a}}_{pt} = -\frac{1}{E(m_{t+1})}\left[\overline{\boldsymbol{\beta}}'_{pt}E_t\left(m_{t+1}\mathbf{f}_{t+1}\right) + E_t\left(m_{t+1}\overline{u}_{p,t+1}\right)\right],$$

and substituting in (27) gives

$$\overline{r}_{p,t+1} - r_t^f = -\frac{1}{E(m_{t+1})}\left[\overline{\boldsymbol{\beta}}'_{pt}E_t\left(m_{t+1}\mathbf{f}_{t+1}\right) + E_t\left(m_{t+1}\overline{u}_{p,t+1}\right)\right] + \overline{\boldsymbol{\beta}}'_{pt}\mathbf{f}_{t+1} + \overline{u}_{p,t+1},$$

and yields the APT condition in portfolio returns corresponding to (11):

$$\overline{r}_{p,t+1} - r_t^f = \overline{\boldsymbol{\beta}}'_{pt}\left(\mathbf{f}_{t+1} - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})}\right) + \bar{\eta}_{pt} + \overline{u}_{p,t+1},$$

where the portfolio pricing errors are given by

$$\bar{\eta}_{pt} = -\frac{E_t(m_{t+1}\overline{u}_{p,t+1})}{E_t(m_{t+1})} = \sum_{i=1}^{n} w_{it}^{p}\eta_{it}.$$

The APT equilibrium condition for portfolios , corresponding to (6), is given by

$$E_t\left(\overline{r}_{p,t+1}\right) = \bar{\mu}_{pt} = r_t^f + \overline{\boldsymbol{\beta}}'_{pt}\boldsymbol{\lambda}_t + \bar{\eta}_{pt},$$

where $\boldsymbol{\lambda}_t$ is defined as before by (12). For identification of $\boldsymbol{\lambda}_t$ (given the portfolio mean returns, $\bar{\mu}_{pt}$, and portfolio factor loadings, $\overline{\boldsymbol{\beta}}_{pt}$, $p = 1, 2, ..., P$), we stack the portfolio return equations to obtain

$$\bar{\boldsymbol{\mu}}_{Pt} = r_t^f\boldsymbol{\tau}_P + \overline{\boldsymbol{B}}_{Pt}\boldsymbol{\lambda}_t + \bar{\boldsymbol{\eta}}_{Pt},$$

where $\bar{\boldsymbol{\mu}}_{Pt} = \left(\bar{\mu}_{1t}, \bar{\mu}_{2t}, ..., \bar{\mu}_{Pt}\right)'$, $\overline{\boldsymbol{B}}'_{pt} = \left(\overline{\boldsymbol{\beta}}_{1t}, \overline{\boldsymbol{\beta}}_{2t}, ..., \overline{\boldsymbol{\beta}}_{Pt}\right)$, $\bar{\boldsymbol{\eta}}_{Pt} = \left(\bar{\eta}_{1t}, \bar{\eta}_{2t}, ..., \bar{\eta}_{Pt}\right)'$. To identify $\boldsymbol{\lambda}_t$ using the portfolio return equations it is now required that

$$P^{-1}\left(\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\overline{\boldsymbol{B}}_{Pt}\right) > \mathbf{0}, \text{ and } P^{-1}\left(\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\bar{\boldsymbol{\eta}}_{Pt}\right) \rightarrow_p \mathbf{0},$$

where $\mathbf{M}_P = \mathbf{I}_P - P^{-1}\boldsymbol{\tau}_P\boldsymbol{\tau}'_P$. To relate the above conditions to those we obtained when using individual securities we first note that

$$\overline{\boldsymbol{\beta}}_{pt} = \sum_{i=1}^{n} w_{ip,t}\boldsymbol{\beta}_{it} = \boldsymbol{B}'_{nt}\mathbf{w}_{pt},$$
$$\bar{\eta}_{pt} = \sum_{i=1}^{n} w_{ip,t}\eta_{it} = \mathbf{w}'_{pt}\boldsymbol{\eta}_{nt},$$

and

$$\begin{aligned}\overline{\boldsymbol{B}}'_{Pt} &= \boldsymbol{B}'_{nt}\left(\mathbf{w}_{1t}, \mathbf{w}_{2t}, ..., \mathbf{w}_{Pt}\right) = \boldsymbol{B}'_{nt}\mathbf{W}_{Pt}, \\ \bar{\boldsymbol{\eta}}_{Pt} &= \mathbf{W}'_{Pt}\boldsymbol{\eta}_{nt}.\end{aligned} \tag{29}$$

Now write the identification conditions when portfolio returns are used as

$$P^{-1}\left(\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\overline{\boldsymbol{B}}_{Pt}\right) = P^{-1}\left(\boldsymbol{B}'_{nt}\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{tP}\mathbf{B}_{nt}\right) > \mathbf{0}, \tag{30}$$

and

$$P^{-1}\left(\boldsymbol{B}'_{nt}\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{Pt}\eta_{nt}\right)\rightarrow_p \mathbf{0}. \tag{31}$$

Comparing these conditions with the corresponding conditions (22) and (23) when using individual securities, it readily follows that using portfolios does not relax the identification condition but requires that the portfolio weights are such that $\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{Pt}$ is a full rank matrix. In fact the factors must have maximum strength irrespective of whether individual securities or portfolios are used for estimation of risk premia. To show that this condition is also necessary when portfolios are used to estimate $\boldsymbol{\lambda}_t$, suppose that $n^{-1}\boldsymbol{B}'_{nt}\mathbf{B}_{nt}\rightarrow\mathbf{0}$, as $n\rightarrow\infty$, and hence $\boldsymbol{\lambda}_t$ cannot be identified using individual securities, and consider the limiting properties of $P^{-1}\left(\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\overline{\boldsymbol{B}}_{Pt}\right)$ given by (30). We have[5]

$$\begin{aligned} P^{-1}\left\|\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\overline{\boldsymbol{B}}_{Pt}\right\| &= P^{-1}\left\|\boldsymbol{B}'_{nt}\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{tP}\mathbf{B}_{nt}\right\| \\ &\leq P^{-1}\left\|\mathbf{B}_{nt}\right\|^2\left\|\mathbf{W}_{Pt}\right\|^2. \end{aligned}$$

Consider the case of non-diversified portfolios and recall that in this case $\left\|\mathbf{W}_{Pt}\right\|^2=\ominus\left(\frac{1}{m}\right)$, and hence

$$P^{-1}\left\|\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\overline{\boldsymbol{B}}_{Pt}\right\|\leq C\ n^{-1}\left\|\mathbf{B}_{nt}\right\|^2,$$

and $P^{-1}\left\|\overline{\boldsymbol{B}}'_{Pt}\mathbf{M}_P\overline{\boldsymbol{B}}_{Pt}\right\|\rightarrow 0$ if $n^{-1}\boldsymbol{B}'_{nt}\mathbf{B}_{nt}\rightarrow\mathbf{0}$. The same result follows in the case of fully diversified portfolios where P is fixed and $\left\|\mathbf{W}_{Pt}\right\|^2=\ominus\left(\frac{1}{n}\right)$. Similarly, condition (31) holds if the associated condition for individual securities given by (23) holds and *vice versa*. To see this, using (31) note that

$$P^{-1}\left\|\boldsymbol{B}'_{nt}\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{Pt}\boldsymbol{\eta}_{nt}\right\|\leq P^{-1}\left\|\mathbf{B}_{nt}\right\|\left\|\mathbf{W}_{Pt}\right\|^2\left\|\boldsymbol{\eta}_{nt}\right\|,$$

and since $\left\|\mathbf{W}_{Pt}\right\|^2=\ominus\left(\frac{1}{m}\right)$, for the non-diversified portfolios, we have (recall that $mP=n$)

$$P^{-1}\left\|\boldsymbol{B}'_{nt}\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{Pt}\boldsymbol{\eta}_{nt}\right\|\leq C\left\|n^{-1/2}\mathbf{B}_{nt}\right\|\left\|n^{-1/2}\boldsymbol{\eta}_{nt}\right\|,$$

and the right hand side of the above tends to zero if $\left\|n^{-1/2}\boldsymbol{\eta}_{nt}\right\|\rightarrow 0$, since $\left\|n^{-1/2}\mathbf{B}_{nt}\right\|<C$. But

$$\left\|n^{-1/2}\boldsymbol{\eta}_{nt}\right\|^2=n^{-1}\boldsymbol{\eta}'_{nt}\boldsymbol{\eta}_{nt}=n^{-1}\sum_{i=1}^{n}\eta_{it}^2=O(n^{\alpha_t-1}),$$

and hence $P^{-1}\left\|\boldsymbol{B}'_{nt}\mathbf{W}_{Pt}\mathbf{M}_P\mathbf{W}'_{Pt}\eta_{nt}\right\|\rightarrow 0$, if $\alpha_t<1$, which is the APT equilibrium condition at the level of individual securities.

4 Identification of risk premia with estimated factor loadings

The above analysis shows that even when the true factor loadings, $\beta_{ij,t}$, are known the factor risk premia could only be identified if the factors have maximum strength, $\delta_{j,t}=1$ such that $\sum_{i=1}^{n}(\beta_{ij,t}-\overline{\beta}_{jt})^2=\ominus(n)$. In practice the factor loadings must be estimated and then additional

[5] Note that since $\mathbf{M}_P$ is an idempotent matrix then $\left\|\mathbf{M}_P\right\|=1$.

restrictions are required. For clarity and to avoid confusion, and as is standard in the literature, it is assumed that $\beta_{ij,t}$ is stable over a given sample period, say $t = 1, 2, ..., T$, and factor loadings are estimated by running least squares regressions of individual security returns on an intercept and the observed factors for a given sample period T. As already noted, in application of this two-pass estimation procedure many researchers have followed Fama and MacBeth (1973) and, rather than using mean returns on individual securities, have used mean returns on a relatively small number of portfolios ($P < n$) formed from the underlying securities. It is argued that the sampling errors in estimation of $\beta's$ of portfolios can be substantially smaller than $\beta's$ estimated using individual securities. To compensate for loss of information from using portfolios as compared to individual securities, it is often recognized that P must be relatively large and the different portfolios not too closely correlated. Fama and MacBeth (1973, p. 615) recommend forming $P = 20$ equal weighted portfolios from ranked values of $\hat{\beta}_{ij}$ estimated over a training sample of four years.

The Fama and MacBeth two-pass estimation procedure is extensively used in the empirical finance literature and its asymptotic properties have been investigated by Shanken (1992), Shanken and Zhou (2007), Kan, Robotti and Shanken (2013), and Bai and Zhou (2015). See also the survey paper by Jagannathan, Skoulakis & Wang (2010) for further references. The two-pass estimates of $\boldsymbol{\lambda}$ is subject to the generated regressor problem also encountered in estimation of certain classes of rational expectations models, See Pagan (1984) and Pesaran (1987). 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. 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. As Ang, Liu and Schwarz (2019) show, creating portfolios to reduce estimation error in the factor loadings does not necessarily lead to smaller estimation errors of the factor risk premia.

In what follows we derive finite T large n bias of two-pass estimators of risk premia, both when individual and portfolio returns are used, and compare their relative performance. We consider a restricted form of the factor model with time-invariant coefficients and make some technical assumption on time series and cross-sectional dependence of the errors and loadings. Specifically, we assume that $\mathrm{a}_{it} = \mathrm{a}_i$, and $\beta_{ij,t} = \beta_{ij}$ for $t = 1, 2, ..., T$, and consider the multi-factor linear model

$$\mathbf{r}_{nt} = \mathbf{a}_n + \boldsymbol{B}_n \mathbf{f}_t + \mathbf{u}_{nt}, \text{ for } t = 1, 2, ..., T, \tag{32}$$

where $\mathbf{r}_{nt} = (r_{1t}, r_{2t}, ..., r_{nt})'$ is an $n \times 1$ vector of excess returns on individual securities during period t, $\mathbf{a}_n = (\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_n)'$, $\boldsymbol{B}_n = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)'$, and $\mathbf{u}_{nt} = (u_{1t}, u_{2t}, ..., u_{nt})'$. Writing the return equations by individual securities we also have

$$\mathbf{r}_{i\circ} = \mathrm{a}_i \boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{33}$$

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)

$$E(r_{it}) = \lambda_0 + \boldsymbol{\beta}_i' \boldsymbol{\lambda} + \eta_i, \text{ for } i = 1, 2, ..., n, \tag{34}$$

where η_i is the pricing error.

We make the following assumptions about the errors and factor loadings:

Assumption 5 *(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)* $\sup_j \sum_{i=1}^{n} |\sigma_{ij}| < C$, *and*

$$\text{(b): } n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(u_{it}^2, u_{jt}^2\right) \to 0, \quad \text{as } n \to \infty.$$

Assumption 6 *(Pricing errors) The pricing errors,* η_i*, defined by (34) have zero means and satisfy the approximate bound*

$$\sum_{i=1}^{n} \eta_i^2 = O\left(n^{\alpha}\right). \tag{35}$$

Assumption 7 *(Common factors) The* $T \times k$ *matrix* $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$ *is full column rank and the* $k \times k$ *matrix* $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ *is positive definite.* k *is a fixed number.*

Assumption 8 *(Factor loadings) (a) The factor loadings* $\boldsymbol{\beta}_i$ *and the errors* u_{jt} *are independently distributed for all* i, j *and* t*. (b)* $\sup_i \|\boldsymbol{\beta}_i\| < C$*, and (c) The* $n \times k$ *matrix of factor loadings,* $\mathbf{B}_n = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)'$*, have full column rank and* $\boldsymbol{\Sigma}_{\beta\beta}$*, defined by*

$$\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right) = \boldsymbol{\Sigma}_{\beta\beta}, \tag{36}$$

is positive definite.

Remark 6 *Assumption 8 can be relaxed in two respects. When the risk free rate is fixed and known, then* λ_0 *in the CRS (34) is set to the risk free rate, and for* $\sqrt{n}$ *consistent estimation of the risk premia,* $\boldsymbol{\lambda}$*, instead of (36) it is required that* $\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{B}_n\right)$ *is positive definite. If we were willing to settle for a slower rate of convergence, and the factor strengths,* δ_j *for factors* f_{tj}*,* $j = 1, 2, ..., k$ *are known, then condition (36) can be further relaxed by requiring that* $\lim_{n\to\infty}\left(\mathbf{D}_n\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n\right)$ *is positive definite where* $\mathbf{D}_n$ *is a* $k \times k$ *diagonal matrix with elements* $n^{-\delta_j/2}$*, for* $j = 1, 2, ..., k$*.*

Part (a) of Assumption 5 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.

Assumption 6 is more general than is assumed in the literature which either ignores the pricing errors, setting $\eta_i = 0$, or assumes a very limited degree of pricing errors by setting $\alpha = 0$. Note also that the above assumptions do allow for correlations between pricing errors and the factor loadings.

Assumptions 7 and 8 are also standard in the literature.

4.1 Estimation of risk premia using individual returns

The two-pass estimator of risk premia, $\boldsymbol{\lambda}$, based on individual returns is given by[6]

$$\hat{\boldsymbol{\lambda}}_n = \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}_n, \tag{37}$$

[6] The two-pass estimator depends on T as well as on n. We omit the subscript T for convenience, but keep n to highlight the direct use of individual returns in the computation of the estimator.

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

$$\hat{\boldsymbol{\beta}}_{i,T} = (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{38}$$

$\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}_T'$, and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under (33), $\hat{\boldsymbol{\beta}}_{i,T} = \boldsymbol{\beta}_i + (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}$, and hence

$$\hat{\mathbf{B}}_{nT} = \boldsymbol{B}_n + \mathbf{U}_n\mathbf{G}_T, \tag{39}$$

where $\mathbf{U}_n = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, and $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$. Also, averaging the return equations (33) over t for each i, we have

$$\bar{r}_{i\circ} = \mathrm{a}_i + \boldsymbol{\beta}_i'\bar{\mathbf{f}}_T + \bar{u}_{i\circ}, \text{ and } E\left(\bar{r}_i\right) = \mathrm{a}_i + \boldsymbol{\beta}_i'E\left(\bar{\mathbf{f}}_T\right), \tag{40}$$

where $\bar{\mathbf{f}}_T = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$, and $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$. Hence, using the above results together with the APT condition given by (34), we have

$$\bar{\mathbf{r}}_n = \lambda_0\boldsymbol{\tau}_n + \boldsymbol{B}_n\left(\boldsymbol{\lambda}+\mathbf{d}_T\right) + \bar{\mathbf{u}}+\boldsymbol{\eta}, \tag{41}$$

where

$$\mathbf{d}_T = \bar{\mathbf{f}}_T - E\left(\bar{\mathbf{f}}_T\right) = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right], \tag{42}$$

$\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, and $\boldsymbol{\eta}$ is the $n \times 1$ vector of pricing errors. Relations (39) and (41) can now be used in (37) to derive the asymptotic properties of $\hat{\boldsymbol{\lambda}}_n$. The following theorem gives the small T bias of $\hat{\boldsymbol{\lambda}}_n$ as an estimator of $\boldsymbol{\lambda}$, as $n \to \infty$.

Theorem 1 *(Small T bias of the risk premia using individual returns) Consider the multi-factor linear return model (32) and the associated risk premia, $\boldsymbol{\lambda}$, defined by (34), and suppose that Assumptions (5), (6), (7) and (8) hold, and the pricing errors are bounded such that $\alpha < 1$. 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 $T > k$ we have (as $n \to \infty$)*

$$\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda} \to_p \left[\boldsymbol{\Sigma}_{\beta\beta} + \frac{\overline{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Sigma}_{\beta\beta}\mathbf{d}_T - \frac{\overline{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}\right). \tag{43}$$

where $\hat{\boldsymbol{\lambda}}_n$ is defined by (37) and

$$\mathbf{d}_T = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right], \; \boldsymbol{\Sigma}_{\beta\beta} = \lim_{n\to\infty}\left(\frac{\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n}{n}\right), \text{ and } \overline{\sigma}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 > 0. \tag{44}$$

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

It is interesting to note that the above theorem holds under fairly general conditions. It allows for weak error cross-sectional dependence, does not impose any restrictions on the distribution of factors when T is finite, and accommodates a wide range of APT models with moderately large pricing errors, η_i, which are allowed to have any degree of dependence with the factor loadings. However, to obtain the standard $\sqrt{n}$ convergence rate, the more restrictive condition of $\alpha < 1/2$ on the degree of pervasiveness of the pricing errors is needed. Nevertheless, this condition is still weaker than $\alpha = 0$, assumed by Ross (1976). It is also interesting that the probability limit of $\hat{\boldsymbol{\lambda}}_n$ exists even if $\boldsymbol{\Sigma}_{\beta\beta}$ is singular, so long as T is finite. But for a consistent estimation of risk factors, where n and $T \to \infty$, $\boldsymbol{\Sigma}_{\beta\beta}$ must be non-singular, which in turn requires that all the k factors must be strong, as discussed earlier.

Example 1 *As a simple example, consider a two-factor case with one strong and one weak factor ($\delta_{\beta_1} = 1$ and $\delta_{\beta_2} < 1$). In this case*

$$\boldsymbol{\Sigma}_{\beta\beta} = \begin{pmatrix} \sigma^2_{\beta_1} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\sigma^2_{\beta_1} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(\beta_{i1} - \bar{\beta}_1)^2 > 0$. Let $\mathbf{A}_T = \frac{\overline{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} = (a_{ij,T})$, and $\mathbf{b}_T = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right] = (b_{i,T})$, then it is easily seen that

$$\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda} \to_p \begin{pmatrix} \frac{a_{22,T}b_{1,T}\sigma^2_{\beta_1} - \lambda_1|\mathbf{A}_T|}{|\mathbf{A}_T| + a_{22,T}\sigma^2_{\beta_1}} \\ \frac{-a_{12,T}b_{2,T}\sigma^2_{\beta_1}}{|\mathbf{A}_T| + a_{22,T}\sigma^2_{\beta_1}} - \lambda_2 \end{pmatrix}.$$

For a finite T the estimate of risk premia of the weak factor tends to $\frac{-a_{12,T}b_{2,T}\sigma^2_{\beta_1}}{|\mathbf{A}_T| + a_{22,T}\sigma^2_{\beta_1}}$ whose sign depends on the sign of $a_{12,T}b_{2,T}$. In the special case where $\mathbf{b}_T$ is relatively small and negligible we have $\hat{\lambda}_2 \to 0$, and

$$\hat{\lambda}_1 - \lambda_1 \to_p \frac{-\lambda_1}{\left(1 + \frac{a_{22,T}\sigma^2_{\beta_1}}{|\mathbf{A}_T|}\right)}.$$

Since $|\mathbf{A}_T| = O(T^{-2})$ whilst $a_{22,T} = O(T^{-1})$, then $\hat{\lambda}_1$ will still be unbiased for T large. This is important since it means that erroneously adding weak factors to the cross section regression does not affect the consistency of the risk premia of the strong factor so long as T is sufficiently large.

4.1.1 A bias-corrected two-pass estimator of risk premia

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 (43) and (44) 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[7]

$$\hat{\overline{\sigma}}^2_{nT} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\hat{u}^2_{it}}{n(T-k-1)}, \tag{45}$$

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 we now have the following bias-corrected version of the two-pass estimator:

$$\tilde{\boldsymbol{\lambda}}_n = \left[\frac{\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n} - T^{-1}\hat{\overline{\sigma}}^2_{nT}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\hat{\mathbf{B}}'_{nT}\mathbf{M}_n\bar{\mathbf{r}}_n}{n}\right). \tag{46}$$

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

It is now easily seen that under the assumptions of Theorem 1, for a fixed $T > k+1$ and as $n \to \infty$, then

$$\tilde{\boldsymbol{\lambda}}_n \to_p \boldsymbol{\lambda}_T^* = \boldsymbol{\lambda} + \mathbf{d}_T, \tag{47}$$

where $\mathbf{d}_T$ is defined by (42). The bias of estimating $\boldsymbol{\lambda}$ is now reduced to $\mathbf{d}_T$, which could be small for moderate values of T (say 60 months often used in the literature), if $\mathbf{f}_t$ is stationary and not too persistent. 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).

4.1.2 Asymptotic distribution of the bias corrected estimator

It is clear that the large n asymptotic distribution of the two-pass estimators, whether bias-corrected or not, are not correctly centred when T is small. There is also the additional difficulty that due to the error cross-sectional dependence, the asymptotic distribution need not be normally distributed and in general depends on the error covariances. To see this consider the asymptotic distribution of the bias-corrected estimator around $\boldsymbol{\lambda}_T^*$, we first note that since $\widehat{\overline{\sigma}}_{nT}^2 \to_p \overline{\sigma}^2$, 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

$$\frac{\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}}{n} - T^{-1}\widehat{\overline{\sigma}}_{nT}^2\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} \to_p \boldsymbol{\Sigma}_{\beta\beta}, \tag{48}$$

where $\boldsymbol{\Sigma}_{\beta\beta}$ defined by (44). Hence, using (39) and (41) in (46), and by Slutsky's theorem we have

$$\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_T^*\right) \overset{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}\left[\frac{\mathbf{G}_T'\mathbf{U}_n'\mathbf{M}_n\boldsymbol{B}_n\boldsymbol{\lambda}_T^*}{\sqrt{n}} + \frac{\left(\boldsymbol{B}_n' + \mathbf{G}_T'\mathbf{U}_n'\right)\mathbf{M}_n\left(\bar{\mathbf{u}} + \eta\right)}{\sqrt{n}}\right]. \tag{49}$$

Consider first the dependence of the asymptotic distribution on the pricing errors, η, and note that

$$n^{-1/2}\boldsymbol{B}_n'\mathbf{M}_n\eta = n^{-1/2}\boldsymbol{B}_n'\eta - \left(\frac{\boldsymbol{B}_n'\tau_n}{n}\right)\left(\frac{\tau_n'\eta}{\sqrt{n}}\right) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\beta}_i\eta_i - \left(\frac{\sum_{i=1}^{n}\boldsymbol{\beta}_i}{n}\right)\left(\frac{\sum_{i=1}^{n}\eta_i}{\sqrt{n}}\right),$$

and hence

$$n^{-1/2}\boldsymbol{B}_n'\mathbf{M}_n\eta \leq 2sup_i\left\|\boldsymbol{\beta}_i\right\|\left(n^{-1/2}\sum_{i=1}^{n}|\eta_i|\right). \tag{50}$$

But by Assumption 8, $sup_i\left\|\boldsymbol{\beta}_i\right\| < C$, and measuring the pervasiveness of the pricing errors by the exponent α_* defined by[8]

$$\sum_{i=1}^{n}|\eta_i| = O\left(n^{\alpha_*}\right), \tag{51}$$

we have $n^{-1/2}\boldsymbol{B}_n'\mathbf{M}_n\eta = O\left(n^{\alpha_*-1/2}\right)$, and $n^{-1/2}\boldsymbol{B}_n'\mathbf{M}_n\eta \to \mathbf{0}$, if $\alpha_* < 1/2$.

Similarly, $n^{-1/2}\mathbf{U}_n'\mathbf{M}_n\eta \to \mathbf{0}$ if $\alpha_* < 1/2$, namely the asymptotic distribution of $\tilde{\boldsymbol{\lambda}}_n$ will not depend on the pricing errors if such errors are sufficiently bounded. This condition is, for example, met if the pricing errors are absolute summable, assumed by Ross. It is also interesting to note that this condition is much more stringent that the condition $\alpha < 1$ needed for the consistent estimation of $\boldsymbol{\lambda}_T^*$ by $\tilde{\boldsymbol{\lambda}}_n$.

[8] The exponent α_* defined here is to be distinguished from α defined by (35), although they coincide when $|\eta_i|$ take dichotomous zero and non-zero values. They differ, for example, if $|\eta_i| = c/i$, for $i = 1, 2, ..., n$.

Consider now the remaining terms in (49) and note that

$$\boldsymbol{\Sigma}_{\beta\beta}\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n-\boldsymbol{\lambda}_T^*\right)\overset{a}{\sim}\mathbf{a}_{nT}+\mathbf{b}_{nT}+\mathbf{c}_{nT}+O_p\left(n^{\alpha_*-1/2}\right), \tag{52}$$

where

$$\begin{aligned}\mathbf{a}_{nT}&=n^{-1/2}\mathbf{U}_n'\mathbf{M}_n\boldsymbol{B}_n\boldsymbol{\lambda}_T^*=n^{-1/2}\sum_{i=1}^{n}\theta_{iT}\left(\mathbf{G}_T'\mathbf{u}_{i\circ}\right),\\ \mathbf{b}_{nT}&=n^{-1/2}\boldsymbol{B}_n'\mathbf{M}_n\bar{\mathbf{u}}=n^{-1/2}\sum_{i=1}^{n}\bar{u}_i\left(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}}_n\right),\\ \mathbf{c}_{nT}&=n^{-1/2}\mathbf{G}_T'\mathbf{U}_n'\mathbf{M}_n\bar{\mathbf{u}}=n^{-1/2}\sum_{i=1}^{n}\left(\bar{u}_i-\bar{u}\right)\mathbf{G}_T'\mathbf{u}_{i\circ},\end{aligned}$$

where $\theta_{iT}=\left(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}}_n\right)'\boldsymbol{\lambda}_T^*$, $\mathbf{G}_T'=\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T$, $\bar{u}_i=T^{-1}\sum_{t=1}^{T}u_{it}$, $\bar{u}=n^{-1}\sum_{i=1}^{n}\bar{u}_i$, and $\mathbf{u}_{i\circ}=(u_{i1},u_{i2},...,u_{iT})'$. Under Assumptions (5), (7) and (8) and for a fixed T, $\mathbf{a}_{nT}$, $\mathbf{b}_{nT}$ and $\mathbf{c}_{nT}$ have limiting distributions as $n\to\infty$, but they need not be Gaussian when T is fixed. The limiting distributions exist due to the assumption of weakly correlated errors that ensure $sup_j\left[n^{-1}\sum_{i=1}^{n}|\sigma_{ij}|\right]<C$, and the boundedness of the factor loadings and existence of $\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$. To see this note that for given factor loadings

$$Var\left(\mathbf{a}_{nT}\right)=\frac{1}{T}\left(n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\sigma_{ij}\theta_{iT}\theta_{jT}\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} \tag{53}$$

$$Var\left(\mathbf{b}_{nT}\right)=\frac{1}{T}\left[n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\sigma_{ij}\left(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}}_n\right)\left(\boldsymbol{\beta}_i-\bar{\boldsymbol{\beta}}_n\right)'\right]. \tag{54}$$

It is now easily seen that under our assumptions, $Var\left(\sqrt{T}\mathbf{a}_{nT}\right)=O(1)$ and $Var\left(\sqrt{T}\mathbf{b}_{nT}\right)=O(1)$. To derive the variance of $\mathbf{c}_{nT}$ we first note that $\bar{u}_i=T^{-1}\mathbf{u}_{i\circ}'\tau_T$, $\mathbf{G}_T'\tau_T=\mathbf{0}$, and $\bar{u}=O_p\left(n^{-1/2}T^{-1/2}\right)$. Hence

$$\begin{aligned}\mathbf{c}_{nT}&=n^{-1/2}T^{-1}\sum_{i=1}^{n}\mathbf{G}_T'\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'\tau_T++O_p\left(n^{-1/2}T^{-1/2}\right)\\ &=T^{-1}\mathbf{G}_T'\left[n^{-1/2}\sum_{i=1}^{n}\left(\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'-\sigma_i^2\mathbf{I}_T\right)\right]\tau_T+O_p\left(n^{-1/2}T^{-1/2}\right).\end{aligned} \tag{55}$$

For a fixed T the limiting distribution of $\mathbf{c}_{nT}$ is determined by the limiting distributions of $n^{-1/2}\sum_{i=1}^{n}\left(u_{it}^2-\sigma_i^2\right)$ and $n^{-1/2}\sum_{i=1}^{n}u_{it}u_{it'}$, which exist under Assumption (5). However, the limiting distributions of $\mathbf{a}_{nT}$, $\mathbf{b}_{nT}$ and $\mathbf{c}_{nT}$ need not be Gaussian due to the fact that u_{it} and u_{it}^2 are cross sectionally dependent and for a fixed T the application of standard Central Limit Theorems to these terms would not be valid. Even under Gaussian errors, the limiting distribution of $\mathbf{c}_{nT}$ need not be Gaussian when the errors are cross correlated. Limiting Gaussian distribution follows if the errors are cross sectionally independent, which requires exact factor pricing and could be restrictive in practice. In general where the errors are cross correlated the limiting distribution of $\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n-\boldsymbol{\lambda}_T^*\right)$ is unknown and depends on σ_{ij} as well as on $Cov(u_{it}^2,u_{jt}^2)$ which are problematic to estimate when T is fixed.

It is instructive to place the above large n asymptotic analysis in the context of the literature. Originally the analysis of Fama-MacBeth, and its further developments by Shanken (1992) and Shanken and Zhou (2007), focussed on the case of n fixed and $T \to \infty$, under the assumption of zero pricing errors (i.e. $\eta_i = 0$, for all i). In this case the asymptotic normality of the two-pass estimator is ensured due to the assumption of serial error independence and zero pricing errors, and cross sectional error dependence does not present any difficulties. In contrast when T is fixed and n tends to infinity the reverse is true; namely it is error cross-sectional dependence that pose difficulties.

When both n and T are large, the outcome depends on the relative expansion rates of n and T. The interesting case is when n and T rise at the same rate such that $n/T \to \kappa$, with $0 < c < \kappa < C$. The cases $\kappa = 0$ and $\kappa = \infty$, can be viewed as the cases of n fixed with $T \to \infty$, and T fixed with $n \to \infty$, already considered. In the case that κ is a finite non-zero constant, using (52) we have

$$\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) - \sqrt{n}\left(\boldsymbol{\lambda}_T^* - \boldsymbol{\lambda}\right) \overset{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}\left(\mathbf{a}_{nT} + \mathbf{b}_{nT} + \mathbf{c}_{nT}\right) + O_p\left(n^{\alpha_* - 1/2}\right),$$

and using (53), (54) and (55), and recalling that by assumption $\boldsymbol{\Sigma}_{\beta\beta}$ is a positive definite matrix, then

$$\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) = \sqrt{n}\left(\boldsymbol{\lambda}_T^* - \boldsymbol{\lambda}\right) + O_p\left(T^{-1/2}\right) + O_p\left(n^{\alpha_* - 1/2}\right) + +O_p\left(n^{-1/2}T^{-1/2}\right).$$

Also using (47), $(\boldsymbol{\lambda}_T^* - \boldsymbol{\lambda}) = \mathbf{d}_T = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]$, and we obtain

$$\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) = \sqrt{\frac{n}{T}}\left\{T^{-1/2}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]\right\} + O_p\left(T^{-1/2}\right) + O_p\left(n^{\alpha_* - 1/2}\right) + O_p\left(n^{-1/2}T^{-1/2}\right). \tag{56}$$

Therefore, under joint n and T asymptotics $\tilde{\boldsymbol{\lambda}}_n$ is correctly centred and its asymptotic distribution is governed by that of the risk factors. This result is summarized in the following theorem:

Theorem 2 *Consider the multi-factor linear return model (32) and the associated risk premia, $\boldsymbol{\lambda}$, defined by (34), and suppose that Assumptions (5), (6), (7) and (8) hold, the factor loading parameters are stable, the pricing errors satisfy the boundedness condition*

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

the $k \times 1$ vector of factors, $\mathbf{f}_t$, follows a stationary process with mean $E(\mathbf{f}_t) = \mu_f$, and the autocovariance matrices, $\mathbf{V}_s = E\left[(\mathbf{f}_t - \mu_f)(\mathbf{f}_{t-s} - \mu_f)'\right]$, such that the long run covariance matrix defined by

$$\mathbf{V} = \mathbf{V}_0 + \sum_{s=1}^{\infty}\left(\mathbf{V}_s + \mathbf{V}_s'\right), \tag{57}$$

is positive definite. Consider the bias-corrected two-pass estimator, $\tilde{\boldsymbol{\lambda}}_n$, given by (46), and further suppose that n and $T \to \infty$ such that $n/T \to \kappa$, with $0 < c < \kappa < C$, $\alpha_* < 1/2$. Then

$$\sqrt{n}\left(\tilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) \to_d N(\mathbf{0}, \kappa\mathbf{V}). \tag{58}$$

The proof of this theorem follows directly from (56) and the application of standard results from stationary time series processes applied to $T^{-1/2}\sum_{t=1}^{T}[\mathbf{f}_t - E(\mathbf{f}_t)]$. Also $\kappa\mathbf{V}$ can be consistently estimated by

$$\mathbf{S}_T = (n/T)\left[\hat{\mathbf{V}}_0 + \sum_{s=1}^{m} b(s,m)\left(\hat{\mathbf{V}}_s + \hat{\mathbf{V}}_s'\right)\right],$$

where $\hat{\mathbf{V}}_s = T^{-1}\sum_{t=s+1}^{T}\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)\left(\mathbf{f}_{t-s} - \bar{\mathbf{f}}_T\right)'$, $\bar{\mathbf{f}}_T = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$, $b(s,m)$ 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 $\mathbf{S}_T$ is invertible.

Remark 7 *We have included Theorem 2 to highlight the importance of allowing for pricing errors when deriving the asymptotic distribution of the two-pass estimator of $\boldsymbol{\lambda}$, and to show that the large T theory developed in the literature is applicable only if pricing errors are zero. It is clear from (56) that even if $\alpha_* = 0$, namely the more restrictive APT condition derived by Ross holds, we still need $n \to \infty$. On its own T large does not eliminate the pricing errors.*

Remark 8 *It is also interesting to note that the precision with which risk premia are obtained depends on the cross section dimension, n, and for a given ratio of n/T does not improve further when T is increased, which contrasts with the $\sqrt{nT}$ convergence rate obtained in the literature for slope parameters in some homogenous panel data models. Multiplying both sides of (56) by $\sqrt{T}$ yields*

$$\sqrt{nT}\left(\tilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) = \sqrt{n}\left\{T^{-1/2}\sum_{t=1}^{T}[\mathbf{f}_t - E(\mathbf{f}_t)]\right\} + O_p(1) + O_p\left(T^{1/2}n^{\alpha_\eta - 1/2}\right) + O_p\left(n^{-1/2}\right),$$

whose first term explodes with n, and $\sqrt{nT}$ convergence is not a possibility. The large T asymptotic for $\tilde{\boldsymbol{\lambda}}_n$ (or $\hat{\boldsymbol{\lambda}}_n$), derived in the literature assumes n is fixed and as noted assumes zero pricing errors, with $\eta_i = 0$, for all $i = 1, 2, ..., n$.

Remark 9 *As noted earlier our focus in this paper has been on $\sqrt{n}$ consistency, but our analysis can be readily extended to semi-strong factors with strengths above $1/2$ but below unity. In general, the risk premia associated to a factor with strength δ_j, is $n^{\delta_j/2}$ consistent so long as $\delta_j > \alpha$. For inference we require the stronger condition $\delta_j > 2\alpha$. For example, if $\delta_j = 2/3$, we can only achieve $n^{1/3}$ consistency, with $\alpha < 1/3$ for inference. Tests on δ_j are further complicated by the fact that we also need to allow for the sampling uncertainty of δ_j since in practice δ_j is unknown and must be estimated.*

4.2 Estimation of risk premia using portfolio returns

In line with the discussion of Section 3.2, we consider estimates of $\boldsymbol{\lambda}$ based on portfolio returns defined by (25). See also the associated cross section model (27). The risk premia can be estimated either forming portfolio betas, as in (28), or basing the two-pass regressions on portfolio returns, $\bar{r}_{pt} = \sum_{i=1}^{n} w_{ip} r_{it} = \mathbf{w}_p'\mathbf{r}_{nt},$, for $t = 1, 2, ..., T$ and $p = 1, 2, ..., P$, where here we are assuming the portfolio weights, w_{ip}, are fixed and do not depend on the factor loadings or the errors. The resultant estimates will be identical. Denoting the portfolio estimate of $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\lambda}}_P$ we have

$$\hat{\lambda}_P = \left(\overline{\hat{\mathbf{B}}}'_{PT}\mathbf{M}_P\overline{\hat{\mathbf{B}}}_{PT}\right)^{-1}\left(\overline{\hat{\mathbf{B}}}'_{PT}\mathbf{M}_P\bar{\mathbf{r}}_P\right), \tag{59}$$

where $\bar{\mathbf{r}}_P = (\ \bar{r}_1, \bar{r}_2, ..., \bar{r}_P)'$, $\bar{r}_p = T^{-1}\sum_{t=1}^{T}\bar{r}_{pt}$, $\overline{\hat{\mathbf{B}}}_{PT} = (\overline{\hat{\boldsymbol{\beta}}}_{1,T}, \overline{\hat{\boldsymbol{\beta}}}_{2,T}, ..., \overline{\hat{\boldsymbol{\beta}}}_{P,T})'$,

$$\overline{\hat{\boldsymbol{\beta}}}_{p,T} = \sum_{i=1}^{n} w_{ip}\hat{\boldsymbol{\beta}}_{i,T} = \left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\sum_{i=1}^{n} w_{ip}\mathbf{r}_{i,T} = \left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\bar{\mathbf{r}}_P.$$

To relate $\hat{\boldsymbol{\lambda}}_P$ to the estimator, $\hat{\boldsymbol{\lambda}}_n$, based on the individual securities, we note that $\overline{\hat{\mathbf{B}}}_{PT} = \mathbf{W}'_P\hat{\boldsymbol{B}}_{nT}$, and $\bar{\mathbf{r}}_P = \mathbf{W}'_P\bar{\mathbf{r}}_n$, where $\mathbf{W}_P = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P)$, with $\hat{\mathbf{B}}_{nT}$ and $\bar{\mathbf{r}}_P$ defined as before, (see (4.1). Using these results $\boldsymbol{\lambda}_P$ can now be written equivalently as

$$\hat{\boldsymbol{\lambda}}_P = \left(\hat{\boldsymbol{B}}'_{nT}\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\hat{\boldsymbol{B}}_{nT}\right)^{-1}\left(\hat{\boldsymbol{B}}'_{nT}\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{r}}_n\right). \tag{60}$$

It is clear that the limiting properties of $\hat{\boldsymbol{\lambda}}_P$ depend on the choice of $\mathbf{W}_P$, and reduces to $\hat{\boldsymbol{\lambda}}_n$ only if $P = n$ and $\mathbf{W}_P = \mathbf{I}_n$. In what follows we shall consider the asymptotic properties of $\hat{\boldsymbol{\lambda}}_P$ when $\mathbf{W}_p$ (or w_{ip}) satisfy the normalization and the summability conditions of Assumption 4. The asymptotic properties of $\hat{\boldsymbol{\lambda}}_P$ can now be derived using (39) and (41) in (60) under the following identification assumption:

Assumption 9 *(Portfolio factor loadings) (a) The $k \times 1$ vector of portfolio loadings, $\bar{\boldsymbol{\beta}}_p = \sum_{i=1}^{n} w_{ip}\boldsymbol{\beta}_i$ and the portfolio errors, $u_{p't} = \sum_{i=1}^{n} w_{ip'}u_{it}$ are independently distributed for all $p, p' = 1, 2, ..., P$ and $t = 1, 2, ..., T$. (b) $\sup_p \left\|\bar{\boldsymbol{\beta}}_p\right\| < C$, and (c) The $n \times k$ matrix of factor loadings, $\boldsymbol{B}_n = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)'$, have full column rank and $\boldsymbol{\Sigma}_{\beta\beta,w}$ defined by*

$$\lim_{P\to\infty}\left(P^{-1}\boldsymbol{B}'_n\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{B}_n\right) = \boldsymbol{\Sigma}_{\beta\beta,w} > 0, \tag{61}$$

is positive definite.

Remark 10 *When portfolio weights, w_{ip}, satisfy the bounds in (26), then it is readily seen that part (b) of the above assumption follows from part (b) of Assumption 8, and it is therefore somewhat weaker. Similarly, part (a) of the above assumption follows from part (a) of Assumption 8. The weaker conditions in parts (a) and (b) of the above assumption is party due to the implicit assumption that the portfolio weights, w_{ip}, are given and known. Part (c) of the above assumption is more demanding as compared to part (c) of Assumption 8, and also imposes further restrictions on the portfolio weights.*

Remark 11 *As an example, suppose $k = 1$, with $\mathbf{B}_n = (\beta_1, \beta_2, ..., \beta_n)'$, and note that $\boldsymbol{B}'_n\mathbf{W}_P = \left(\bar{\beta}_1, \bar{\beta}_2, ..., \bar{\beta}_P\right)'$, where $\bar{\beta}_p = \sum_{i=1}^{n} w_{ip}\beta_i$. Suppose further that $\sum_{i=1}^{n} w_{ip}^2 = O\left(m^{-1}\right)$, and β_i follows the random coefficient specification $\beta_i = \beta + \xi_i$, where ξ_i have zero means and a finite variance, σ_ξ^2, and are cross sectionally independent as well as being distributed independently of the weights w_{jp} for all i and j. Under the normalization $\sum_{i=1}^{n} w_{ip} = 1$, $\bar{\beta}_p = \beta + \bar{\xi}_p$, where $\bar{\xi}_p = \sum_{i=1}^{n} w_{ip}\xi_i$, and $\boldsymbol{B}'_n\mathbf{W}_P = \beta\boldsymbol{\tau}'_P + \bar{\boldsymbol{\xi}}'_P$ with $\bar{\boldsymbol{\xi}}_P = \left(\bar{\xi}_1, \xi_2, ..., \bar{\xi}_P\right)'$, and we have*

$$P^{-1}\boldsymbol{B}'_n\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{B}_n = P^{-1}\sum_{p=1}^{P}\bar{\boldsymbol{\xi}}'_p\mathbf{M}_P\bar{\boldsymbol{\xi}}_p \leq P^{-1}\sum_{p=1}^{P}\bar{\boldsymbol{\xi}}'_p\bar{\boldsymbol{\xi}}_p.$$

Also since $\xi_i \sim IID(0, \sigma_\xi^2)$, *and* $Var\left(\bar{\xi}_p\right) = \sigma_\xi^2\left(\mathbf{w}_p'\mathbf{w}_p\right) = O\left(m^{-1}\right)$, *then* $\bar{\boldsymbol{\xi}}_P = O_p\left(m^{-1/2}\right)$ *and we have*

$$P^{-1}\mathbf{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}_n = O_p\left(m^{-1}\right).$$

Therefore, for identification m *must be finite, which rules out using diversified portfolio weights with* $w_{ip} = O\left(n^{-1}\right)$. *In this example, the use of portfolios in estimation of risk premia can be justified only if* m *is fixed with* $P \to \infty$.

The small T bias of $\hat{\boldsymbol{\lambda}}_P$ for a fixed m and $P \to \infty$, is given in the following theorem:

Theorem 3 *(Small* T *bias of portfolio estimator of risk premia) Consider the multi-factor linear return model (32) and the associated risk premia,* $\boldsymbol{\lambda}$*, defined by (34), and suppose that Assumptions (5), (6), (7), and (9) hold, and* $\alpha < 1$. *Suppose further that* $\boldsymbol{\lambda}$ *is estimated by Fama-MacBeth two-pass estimator based on portfolio excess returns,* $\bar{r}_{pt} = \mathbf{w}_P'\mathbf{r}_{tn}$*, for* $p = 1, 2, ..., P$*, and the factors,* $\mathbf{f}_t$*, for* $i = 1, 2, ..., n$*, and* $t = 1, 2, ..., T$. *Then under Assumption 4 and assuming that portfolio weights are sufficiently bounded, namely* $\|\mathbf{W}_P\| = \ominus\left(m^{-1/2}\right)$ *where* $\mathbf{W}_P = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P)$*, and* m *is the maximum number of securities included in a single portfolio, then for any fixed* $T > k$ *we have (as* $P \to \infty$*)*

$$\hat{\boldsymbol{\lambda}}_P - \boldsymbol{\lambda} \to_p \left[\boldsymbol{\Sigma}_{\beta\beta,w} + \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left[\boldsymbol{\Sigma}_{\beta\beta,w}\mathbf{d}_T - \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\lambda\right]. \tag{62}$$

where $\hat{\boldsymbol{\lambda}}_n$ *is defined by (37),* $\mathbf{d}_T = T^{-1}\sum_{t=1}^T\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]$,

$$\boldsymbol{\Sigma}_{\beta\beta,w} = \lim_{P\to\infty}\left(\frac{\mathbf{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}_n}{P}\right), \text{ and } \bar{\omega}^2 = \lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^P\left(\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p\right) > 0, \tag{63}$$

where $\boldsymbol{\Sigma}_u = (\sigma_{ij})$.

A proof is provided in sub-section A.3.3 of the Appendix.

It is clear that the small T bias continues to be present when using portfolio returns to estimate $\boldsymbol{\lambda}$. Whether the bias can be reduced using portfolio returns instead of individual security returns is unclear and in a complicated way depends on the within portfolio correlations, as characterised by $\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p$, and the relative norms of $\boldsymbol{\Sigma}_{\beta\beta}$ and $\boldsymbol{\Sigma}_{\beta\beta,w}$.

Example 2 *Suppose that* T *is sufficiently large such that* $\mathbf{d}_T$ *is negligible, and* $k = 1$*, so that the risk premia,* λ*, is a scalar. Also assume that* $\lambda > 0$*, then the bias of the estimator of* λ*, whether based on individual securities or portfolios is negative and the magnitude of the bias of the estimator based on portfolios relative to the estimator based on individual securities is given by the ratio(using (43) and (62))*

$$\frac{\bar{\omega}^2\left[\sigma_{\beta\beta}^2 + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{f}'\mathbf{M}_T\mathbf{f}}{T}\right)^{-1}\right]}{\bar{\sigma}^2\left[\sigma_{\beta\beta,w}^2 + \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{f}'\mathbf{M}_T\mathbf{f}}{T}\right)^{-1}\right]},$$

and for $\hat{\lambda}_P$ *to be less biased as compared to the estimator based on individual securities,* $\hat{\lambda}_n$*, we must have*

$$\sigma_{\beta\beta,w}^2 > \left(\frac{\bar{\omega}^2}{\bar{\sigma}^2}\right)\sigma_{\beta\beta}^2,$$

which can be written equivalently as the limit $(n, P \to \infty)$ *of the following inequality*

$$\frac{\beta_n' \mathbf{W}_P \mathbf{M}_P \mathbf{W}_P' \beta_n}{P} > \left(\frac{\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}_p' \boldsymbol{\Sigma}_u \mathbf{w}_p\right)}{\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2} \right) \frac{\beta_n' \mathbf{M}_n \beta_n}{n}. \tag{64}$$

It is clear that the answer will depend on the choice of the portfolio weights. Consider P *equally weighted, mutually exclusive portfolios, each with* m *securities, such that* $n = mP$. *In this case* $\mathbf{w}_p = m^{-1}(\mathbf{0}_m', \mathbf{0}_m', ..., \mathbf{0}_m', \boldsymbol{\tau}_m', \mathbf{0}_m', ..., \mathbf{0}_m')'$, *where* τ_m *is an* $m \times 1$ *vector of ones. Suppose that the allocation of securities to portfolios are done randomly, and without loss of generality assume that the first* m *securities form the first portfolio,* $p = 1$, *the second* m *securities the second portfolio,* $p = 2$, *and so on. Then*

$$\bar{r}_{1t} = m^{-1}\sum_{i=1}^{m} r_{it},\ \bar{r}_{2t} = m^{-1}\sum_{i=m+1}^{2m} r_{it}, ..., \bar{r}_{Pt} = m^{-1}\sum_{i=(P-1)m+1}^{n} r_{it},$$

Also to simplify the exposition suppose that $k = 1$ *(single factor) and note similarly that*

$$\bar{\beta}_1 = \mathbf{w}_1'\boldsymbol{\beta} = m^{-1}\sum_{i=1}^{m}\beta_i,\ \bar{\beta}_2 = \mathbf{w}_2'\boldsymbol{\beta} = m^{-1}\sum_{i=m+1}^{2m}\beta_i, ..., \bar{\beta}_P = \mathbf{w}_P'\boldsymbol{\beta} = m^{-1}\sum_{i=(P-1)m+1}^{n}\beta_i, \tag{65}$$

and the sample average of $\bar{\beta}_p$ *across* p *gives*

$$\ddot{\beta}_P = P^{-1}\sum_{p=1}^{P}\bar{\beta}_p = P^{-1}\sum_{p=1}^{P}\mathbf{w}_p'\boldsymbol{\beta} = n^{-1}\sum_{i=1}^{n}\beta_i = \bar{\beta},$$

and hence

$$P^{-1}\mathbf{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}_n = P^{-1}\sum_{p=1}^{P}(\bar{\beta}_p - \bar{\beta})^2.$$

Similarly we have (noting that $n = mP$*)*

$$\begin{aligned}
n^{-1}\boldsymbol{\beta}_n'\mathbf{M}_n\boldsymbol{\beta}_n &= n^{-1}\sum_{i=1}^{n}(\beta_i - \bar{\beta})^2 = n^{-1}\sum_{p=1}^{P}\sum_{i=(p-1)m+1}^{mp}(\beta_i - \bar{\beta})^2 \\
&= n^{-1}\sum_{p=1}^{P}\sum_{i=(p-1)m+1}^{mp}(\beta_i - \bar{\beta}_p + \bar{\beta}_p - \bar{\beta})^2 \\
&= n^{-1}\sum_{p=1}^{P}\sum_{i=(p-1)m+1}^{mp}\left[(\beta_i - \bar{\beta}_p)^2 + (\bar{\beta}_p - \bar{\beta})^2 + 2(\beta_i - \bar{\beta}_p)(\bar{\beta}_p - \bar{\beta})\right] \\
&= \frac{1}{P}\sum_{p=1}^{P}\left[m^{-1}\sum_{i=(p-1)m+1}^{mp}(\beta_i - \bar{\beta}_p)^2\right] + \frac{1}{P}\sum_{p=1}^{P}(\bar{\beta}_p - \bar{\beta})^2,
\end{aligned}$$

which decomposes the total cross variations of individual β*'s into within and between portfolio variations. To rank order the bias of the two estimators we also need to consider within and* *between error covariances. We note that* $\mathbf{w}_p'\mathbf{\Sigma}_u\mathbf{w}_p = m^{-2}\boldsymbol{\tau}_m'\mathbf{\Sigma}_{p,u}\tau_m$, *where* $\mathbf{\Sigma}_{p,u}$ *is the* $m \times m$ *covariance of the errors of the returns included in the* p^{th} *portfolio, and*

$$\bar{\omega}_n^2 = \frac{1}{Pm^2}\sum_{p=1}^{P}\boldsymbol{\tau}_m'\mathbf{\Sigma}_{p,u}\boldsymbol{\tau}_m.$$

It is now easily seen that $\bar{\omega}_n^2 = m^{-1}\bar{\sigma}_n^2$, *when* $\mathbf{\Sigma}_{p,u}$ *is diagonal, namely when within portfolio errors are uncorrelated, although between portfolio errors are still allowed to be correlated. Under this additional restriction and using the above results in (64), then for* $\hat{\lambda}_P$ *to be less biased than* $\hat{\lambda}_n$, *we require*

$$P^{-1}\sum_{p=1}^{P}(\bar{\beta}_p - \bar{\beta})^2 > \frac{1}{m}\left\{P^{-1}\sum_{p=1}^{P}\left[m^{-1}\sum_{i=(p-1)m+1}^{mp}(\beta_i - \bar{\beta}_p)^2\right] + \frac{1}{P}\sum_{p=1}^{P}(\bar{\beta}_p - \bar{\beta})^2\right\},$$

or equivalently if

$$\psi_P(\beta) = (m-1)\left[P^{-1}\sum_{p=1}^{P}(\bar{\beta}_p - \bar{\beta})^2\right] - P^{-1}\sum_{p=1}^{P}\left[m^{-1}\sum_{i=(p-1)m+1}^{mp}(\beta_i - \bar{\beta}_p)^2\right] > 0.$$

This condition is met if dispersion of β_i *within a given portfolio is small relative to the dispersion of* $\bar{\beta}_p$ *across the portfolios. Introducing non-zero within portfolio error covariances leads to further reduction in relative bias of* $\hat{\lambda}_P$ *when on average these covariances are negative and vice versa, when they are positive. Therefore, to achieve bias reduction the portfolio approach should be capable of identifying securities with similar* β*'s whose errors are negatively correlated. It is also important that these differences do not vanish as* $n \to \infty$. *For instance, when* β_i *follow the random coefficient model,* $\beta_i = \beta + \xi_i$, *with* $\xi_i \sim IID(0, \sigma_\xi^2)$, *then (also see Remark 11)*

$$\psi_P(\beta) = P^{-1}\sum_{p=1}^{P}\left[(m-1)(\bar{\xi}_p - \bar{\xi})^2 - m^{-1}\sum_{i=(p-1)m+1}^{mp}(\xi_i - \bar{\xi}_p)^2\right],$$

and

$$\begin{aligned}\frac{E[\psi_P(\beta)]}{\sigma_\xi^2} &= (m-1)P^{-1}\sum_{p=1}^{P}\left(\frac{1}{p} + \frac{1}{n} - \frac{2}{pn}\right) - P^{-1}\sum_{p=1}^{P}\left(1 - \frac{1}{p}\right) \\ &= -1 + m\left(P^{-1}\sum_{p=1}^{P}p^{-1}\right) - \frac{2(m-1)}{mP}\left(P^{-1}\sum_{p=1}^{P}p^{-1}\right) + \frac{(m-1)}{mP}.\end{aligned}$$

Since $\sum_{p=1}^{P}p^{-1} \approx ln(P)$, *then* $ln(P)/P \to 0$, *as* $P \to \infty$, *and therefore* $E[\psi_n(\beta)] \to -\sigma_\xi^2$. *Hence, in this random setting* $\hat{\lambda}_n$, *which uses individual securities is likely to be less biased as compared to* $\hat{\lambda}_P$, *for* n *sufficiently large. This example highlights that using portfolio returns to estimate the risk premia can be justified if there are a priori known stock characteristics that could be used to sort the returns into groups with systematically different* $\bar{\beta}_p$ *across* p.

Remark 12 *In the literature, following Fama-MacBeth, it is standard to rank securities by estimated* $\widehat{\beta}_i$ *and then form portfolios based on similar estimated betas. The focus here is on the bias since issues of efficiency are likely to be of second order of importance. Whereas a bias correction can be constructed for individual securities one does not seem available for portfolios because* $\bar{\omega}^2 = \lim_{P\to\infty} \frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}_p'\mathbf{\Sigma}_u\mathbf{w}_p\right)$*, defined by (63) depends on the covariances of the different securities included in the portfolio.*

5 Estimates of factor strength

The earlier sections demonstrated the theoretical importance of factor strength. This section uses the estimator introduced in Bailey, Kapetanios and Pesaran, BKP, (2019b) to provide 10-year rolling estimates of δ for the three Fama-French (1993), FF, factors used extensively in the finance literature. For each month from September 1989 to May 2018, all stocks in the S&P 500 portfolio that have at least 10 years of return history are selected. The list is updated monthly and always includes at least 400 stocks, with an average of 440 stocks. This procedure avoids the possible survivorship bias caused by the changing composition of the index. Time-series regressions of the excess return for each stock are then run on a constant and the three FF factors. The data for the FF factors are taken from Kenneth French's web pages.[9] The FF market factor, the excess market return, differs from the average of the roughly 400 stocks we consider. In particular it is value weighted and much wider. 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. The size factor is small minus big, SMB $f_{SMB,t}$. The value factor is high minus low book to market, HML, $f_{HML,t}$.

It is important that the market factor is always included in the return regressions. To see why this should be so, consider the regressions:

$$r_{it} - r_t^f = \mathrm{a}_i + \sum_{j=1}^{k} \beta_{ij} f_{jt} + u_{it}, \tag{66}$$

for $i = 1, 2, ..., n$; over $t = 1, 2, ..., T$. Section 3, showed that consistent estimation of risk premia, λ_j, associated with factor f_{jt}, required that its strength, δ_j, must be unity. The strength of factor j is defined by $\sum_{i=1}^{n}\left(\beta_{ij} - \bar{\beta}_j\right)^2 = \ominus\left(n^{\delta_j}\right)$. See also the discussion that surrounds equation (1) in the introduction. Therefore, one must measure the loadings relative to their means. To achieve this, average both sides of (66) over i, and subtract the result from (66) to obtain:[10]

$$r_{it} - \overline{r}_t = \tilde{\mathrm{a}}_i + \sum_{j=1}^{k} (\beta_{ij} - \overline{\beta}_j) f_{jt} + \widetilde{u}_{it}, \tag{67}$$

where $\overline{r}_t = n^{-1}\sum_{i=1}^{n} r_{it}$, $\tilde{\mathrm{a}}_i = \mathrm{a}_i - n^{-1}\sum_{i=1}^{n}\mathrm{a}_i$, and $\widetilde{u}_{it} = u_{it} - n^{-1}\sum_{i=1}^{n} u_{it}$. Suppose that one of the factors, say f_{1t}, is the market factor measured by r_t^m, which approximates $\overline{r}_t$, when n is

[9] 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).

[10] We could also consider using weighted averages, but when n is large simple and weighted averages tend to yield similar results assuming the weights used are of order $1/n$.

sufficiently large. Then (67) can be equivalently written as:

$$r_{it} = \tilde{\mathrm{a}}_i + \widetilde{\beta}_{i1} r_t^m + \sum_{j=2}^{k} \widetilde{\beta}_{ij} f_{jt} + \widetilde{u}_{it}. \tag{68}$$

where $\widetilde{\beta}_{i1} = \beta_{i1} - \overline{\beta}_1 + 1$ and $\widetilde{\beta}_{ij} = (\beta_{ij} - \overline{\beta}_j)$. Thus, once the market factor is included the coefficients give us the deviations of the factor loadings from their means as required. Note also that since u_{it} are weakly cross-correlated then $\widetilde{u}_{it} = u_{it} + O_p(n^{-1/2})$. The estimation of factor strengths can now be based on the return regression (68). It is also worth noting that OLS estimates of $\widetilde{\beta}_{i1}$ add up over i to one.

To estimate the strength of the FF factors (relative to the market) we run the following excess return regressions

$$r_{it} - r_t^f = \mathrm{a}_i + \beta_i^M r_t^M + \tilde{\beta}_i^{SMB} f_{SMB,t} + \tilde{\beta}_i^{HML} f_{HML,t} + u_{it}. \tag{69}$$

Denote by t_{ij} the t-statistic corresponding to the estimated loading of factor j for security i and consider the proportion of regressions where the coefficients for factor j, $\tilde{\beta}_{ij}$, is statistically significant:

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

where $\mathbf{1}(A) = 1$ if $A > 0$, and zero otherwise. There is clearly a multiple testing problem and to control for this, the critical value function, $c_p(n)$, defined by $c_p(n) = \Phi^{-1}\left(1 - \frac{p}{2n^c}\right)$, 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 δ_j. No allowance is made for the multiple testing problem if we set $c = 0$. In the application below, we set $c = 0.25$ and $p = 0.1$. Using a larger value of c or a smaller value of p would tend to increase the critical values and reduce the estimate of π_j and hence δ_j. However, the results do not seem to be that sensitive to these values and estimates using $c = 0.5$ and $p = 0.05$, for example, are qualitatively similar to those based on $c = 0.25$ and $p = 0.1$.

The estimator of the strength of factor j, is defined by $\hat{\delta}_j = 1 + ln(\hat{\pi}_j)/ln(n)$. The properties of this estimator are examined in detail in BKP (2019b), who also derive its asymptotic distribution and give analytical expressions for its asymptotic standard errors for values of δ_j in the range $1/2 < \delta_j < 1$. The confidence intervals tend to become quite narrow as δ_j gets closer to unity, and the asymptotic distribution of $\hat{\delta}_j$ (as as $n, T \to \infty$) in fact becomes degenerate when $\delta_j = 1$, a kind of ultra-consistency result. This property is illustrated in the very narrow confidence bands that we obtain for the estimates of the strength of FF factors reported below.

Figure 1, plots the 10 yearly rolling estimates, $\delta_{j\tau}$ for the three FF factors ($j = MKT, SMB, HML$), where τ denotes the last month of the estimation sample, from May 1989 to May 2018. The strength of the market factor is always one or very close to one for all rolling windows. It only falls below 0.99 for the ten year samples ending between January 2001 and February 2002 and then again for the samples ending between May 2007 and May 2010, both periods of market turmoil. The lowest estimate of the strength of the market factor is 0.9848, for the sample ending on August 2008, with a 90% confidence interval from 0.9842 to 0.9854. The other two factors have similar strengths for the period 1989 to 1999, in the range about 0.7 to 0.8, with the strength of the size factor, SMB, being greater than the strength of the value factor, HML, except for a short period at the beginning. Then in January 1999 HML overtakes SMB, increasing sharply and exceeding 0.9 in September 2001. It then stays high until September 2009 when it falls below 0.9. At the end of the period the strength of the SMB factor was 0.7 and the HML 0.77. The 90% confidence intervals are not large even for the lower estimates of δ observed. The lowest estimate for SMB is 0.652 in February 2015, with a 90% confidence interval from 0.638 to 0.667. For HML, the minimum is 0.685 in January 1991 with a 90% confidence interval from 0.673 to 0.697.

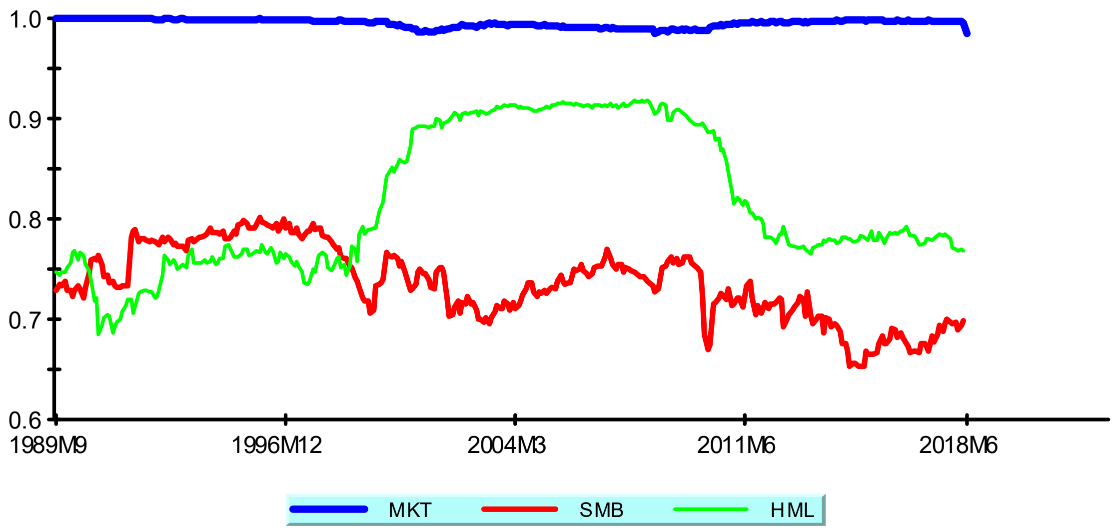

Figure 1: Ten year rolling estimates of factor strength for market factor (top), and HML and SMB factors

Figure 2 plots the estimate of δ and the 90% confidence interval for the strength of SMB and Figure 3 for HML, again using 10 year rolling windows, $c = 0.25$ and $p = 0.1$.

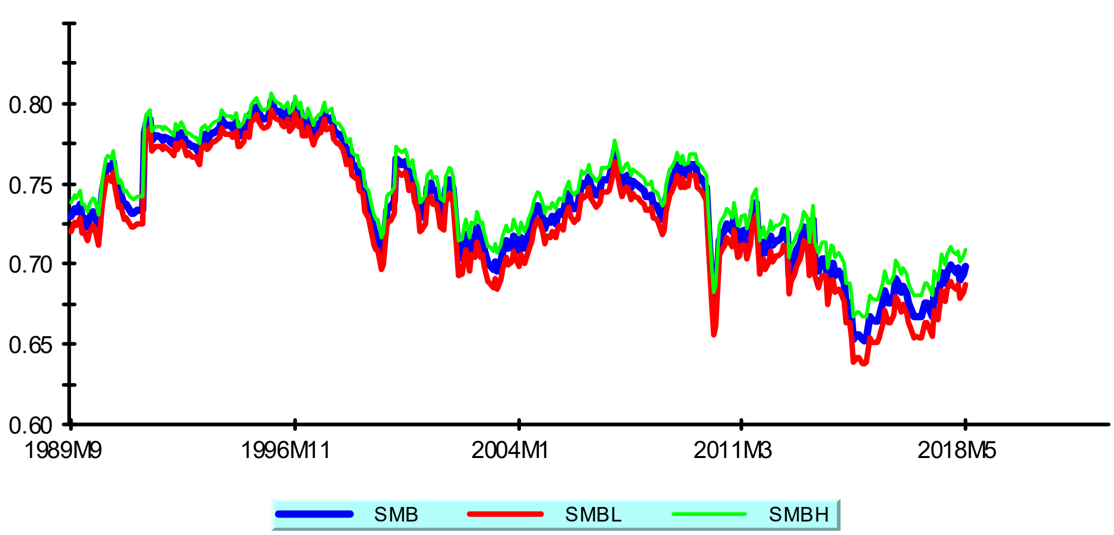

Figure 2: Ten year rolling estimates of factor strength for the SMB factor with its 90% confidence band

The results from the 5 year rolling window are qualitatively similar but show more fluctuations and lower minima, because fewer of the time series coefficients are significant. For the market the lowest estimate is 0.935 in May 2008 with a confidence interval from 0.934 to 0.937. For SMB the lowest estimate is 0.566 in May 2005 with a confidence interval of 0.541 to 0.591. For HML the lowest estimate is 0.594 in December 2007 with a confidence interval of 0.573 to 0.615. Whereas on the 10 year rolling samples HML stays quite strong, above 0.8, from samples ending in November 1999 to December 2011, on the 5 year measure it weakens between February 2006 and September 2008 dropping to values below 0.7 before recovering. It is instructive to note that with 400 securities condition $\alpha < 1/2$ implies that at most 20 securities out of the 400 could have non-zero pricing errors, as compared to 220 for $\alpha = 0.90$.

The APT theory requires the risk factors, $\mathbf{f}_t$, to be strong and the factors that drive the pricing errors, denoted by $\mathbf{g}_t$ in the theoretical section, to be weak. The strength of $\mathbf{g}_t$ factors could be quite high, so long as they lie below unity. SMB and HML are examples of such factors, whose strengths are estimated to lie between 0.65 and 0.90, but never reach the strength required to consider them as risk factors. Nevertheless, their inclusion along with the market factor could be justified to control for semi-strong factors and help ensure that the remaining errors are sufficiently weak so that the condition $\alpha < 1/2$ is met. See also Example 1 where it is shown that the consistency of the two pass estimator of the risk premia associated to the strong factor is not affected by erroneous inclusion of additional factors that are weak.

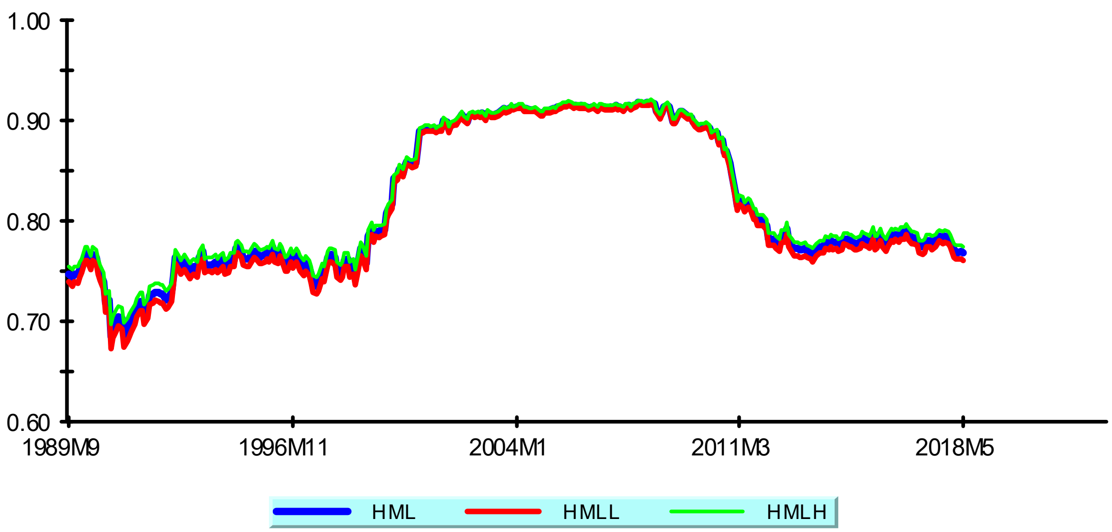

Figure 3: Ten year rolling estimates of factor strength for the HML factor with its 90% confidence band

We have only considered three factors for illustration. Currently there are a large number of suggested factors. Harvey and Liu (2019) document a "factor zoo" of over 400 suggested factors in early 2019. BKP (2019b) provide rolling estimates of factor strength for over 145 factors recently considered by Feng et al. (2019), but find that none of these factors come close to being strong. Our theoretical analysis suggests that before a potential factor is allowed into the zoo, its strength should be estimated in the way we propose. This can be done one factor at a time with the market factor always included in the time series regression.[11]

6 Concluding remarks

Factor strength and pricing errors play central roles in the empirical analysis of APT. If a factor is not strong its influence can be eliminated with a suitably diversified portfolio and it will not command a risk premium. This paper highlights the importance of estimating factor strengths, and proposes the use of an estimator of factor strength, δ, constructed from the proportion of significant factor loadings in the times series regressions of excess returns on the factor. We consider the properties of the two pass estimators of the risk premia obtained from the cross section regression of average returns, for individual securities as well as for portfolios. We show that conditions for identification of the risk premia are unaffected if one uses portfolios or individual securities. For consistency of risk premia it is required that factors are strong with $\delta = 1$ and pricing errors are weak. We also derive asymptotic distribution of the estimated risk premia allowing for non-zero pricing errors. We consider both the most relevant case of large n and fixed T, and the case of large n and T which is required for $\sqrt{n}$ consistent estimation of risk premia when factor loadings are unknown.

In an empirical application, we present rolling estimates of the factor strength parameter δ_j using S&P 500 monthly returns over September 1989 to May 2018. We estimate δ_j from time series regressions for the market factor and the two Fama-French size and value factors (SMB and HML). The market factor is always strong with its δ estimated to be close to unity, whilst the strength of SMB and HML factors vary substantially over time between, lie between 0.65 and 0.90, and at no time come close to being unity. A related empirical analysis carried out by Bailey, Kapetanios and Pesaran (2019b) consider a large number of other factors proposed in the literature and arrive at a similar conclusion that *only* market factor can be viewed as strong.

[11] When a large number of potential factors are included in the time series regressions, multiple testing is a major problem. It is only relatively recently that model selection procedures that adjust for this have been used; for instance Feng et al. (2019). The OCMT procedure suggested by Chudik, Kapetanios and Pesaran (2018) could also be used.

A Mathematical Appendix

A.1 Introduction

We first state and establish a number of lemmas that we shall then use to prove Theorems 1 and 3 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 (32), and suppose that Assumption 5 holds. Then for any t and t' (as $n \to \infty$)*

$$a_{n,tt'} = \frac{1}{n}\sum_{i=1}^{n} u_{it}u_{it'} \to_p 0, \;\; \textit{if } t \neq t', \tag{A.1}$$

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

and

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

where

$$\sigma_i^2 = E(u_{it}^2), \;\; \overline{u}_{i\circ} = \frac{1}{T}\sum_{t=1}^{T} u_{it}.$$

Proof. Since $\{u_{it}\}$ is serially uncorrelated then $E(u_{it}u_{it'}) = 0$ for $t \neq t'$ and

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

Since by Assumption 5, $\sup_j \sum_{i=1}^{n}|\sigma_{ij}| < C$, then

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

Hence $a_{n,tt'}$ (for $t \neq t'$) converges in mean square error to to its mean which is zero, and (A.1) is established. 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 5, and result (A.2) is established. To prove (A.3), set $z_{it} = u_{it}\overline{u}_{i.} - \frac{1}{T}\sigma_i^2$, 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$E(z_{it}) = 0$. Also

$$Var(c_{n,t}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(z_{it}, z_{jt}) \tag{A.4}$$

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.4) 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, by assumption (5), for any fixed T, $Var(c_{n,t}) \to 0$, as $n \to \infty$, and (A.3) is established since $E(c_{n,t}) = 0$. ■

Lemma A.2 *Consider the* $n \times T$ *error matrix* $\mathbf{U} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, *where* $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$, *the* $n \times k$ *matrix of factor loadings,* $\mathbf{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)$, *the* $n \times 1$ *vector of pricing errors* $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_n)'$, *and suppose that assumptions 5, 6 and part (b) of 8 hold, and* $\alpha < 1$.[12] *Then*

$$\frac{\mathbf{B}'\mathbf{M}_n\mathbf{U}}{n} \to_p \mathbf{0}, \tag{A.5}$$

$$\frac{\mathbf{B}'\mathbf{M}_n\overline{\mathbf{u}}}{n} \to_p \mathbf{0}, \tag{A.6}$$

$$\frac{\mathbf{B}'\mathbf{M}_n\boldsymbol{\eta}}{n} \to_p \mathbf{0}, \tag{A.7}$$

$$\frac{\mathbf{U}'\mathbf{M}_n\mathbf{U}}{n} \to_p \overline{\sigma}^2\mathbf{I}_T, \tag{A.8}$$

$$\frac{\mathbf{U}'\mathbf{M}_n\overline{\mathbf{u}}}{n} \to_p \frac{\overline{\sigma}^2}{T}\boldsymbol{\tau}_T, \tag{A.9}$$

$$\frac{\mathbf{U}'\mathbf{M}_n\boldsymbol{\eta}}{n} \to_p \mathbf{0} \tag{A.10}$$

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

[12] As compared to the notation in the body of the paper, we have dropped the subscript n from $\mathbf{B}_n$ as defined by (32).

Proof. Consider (A.5) and note that

$$n^{-1}\mathbf{B}'\mathbf{M}_n\mathbf{U} = \frac{1}{n}\mathbf{B}'\mathbf{U} - \frac{1}{n^2}\mathbf{B}'\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\mathbf{U}, \tag{A.11}$$

where $n^{-1}\mathbf{B}'\ \boldsymbol{U} = n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i\boldsymbol{u}_{i\circ}'$, and $\boldsymbol{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$. Since T is finite, it is sufficient to consider the t^{th} column of $n^{-1}\mathbf{B}'\mathbf{U}$, which is $n^{-1}\sum_{i=1}^{n}u_{it}\boldsymbol{\beta}_i$. Since by assumption $\boldsymbol{\beta}_i$ and u_{it} are distributed independently and $E(\boldsymbol{\beta}_i u_{it}) = 0$, for all i and t. Then conditional on $\boldsymbol{\beta}_i$

$$Var\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i u_{it}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\beta}_i\boldsymbol{\beta}_j'\sigma_{ij},$$

and since by assumption $\sup\|\boldsymbol{\beta}_i\| < C$, then under Assumptions 5 and 8

$$\begin{aligned} Var\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i u_{it}\right) &\leq \left(\sup_i\|\boldsymbol{\beta}_i\|\right)^2\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{ij}|\right) \\ &< \frac{(\sup|\boldsymbol{\beta}_i|)^2\sup_j\sum_{i=1}^{n}|\sigma_{ij}|}{n} = O(\frac{1}{n}), \end{aligned}$$

and $n^{-1}\mathbf{B}'\mathbf{U} \to_p 0$. Now consider the second term of (A.11) and note that

$$\frac{1}{n^2}\|\mathbf{B}'\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\mathbf{U}\| \leq \left\|n^{-1}\mathbf{B}'\boldsymbol{\tau}_n\right\|\left\|n^{-1}\boldsymbol{\tau}_n'\mathbf{U}\right\|.$$

But $n^{-1}\mathbf{B}'\boldsymbol{\tau}_n = \overline{\beta}_n = n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_i$, and since $\sup_{ij}|\beta_{ij}| < C$, then $\|n^{-1}\mathbf{B}'\boldsymbol{\tau}_n\| < C$. Also the t^{th} element of $n^{-1}\mathbf{U}'\boldsymbol{\tau}_n$ is given by $\bar{\mathbf{u}}_{\circ t} = n^{-1}\sum_{i=1}^{n}u_{it}$ and under Assumption (5), $\bar{\mathbf{u}}_{\circ t} \to_p 0$, and we have

$$n^{-1}\mathbf{U}'\boldsymbol{\tau}_n \to_p 0. \tag{A.12}$$

Hence both components of (A.11) converge to zero in probability so (A.5) is established. Result (A.6) follows similarly. To establish (A.7), note that

$$\left\|\frac{\mathbf{B}'\mathbf{M}_n\boldsymbol{\eta}}{n}\right\| \leq \lambda_{\max}^{1/2}\left(\frac{\mathbf{B}'\mathbf{M}_n\mathbf{B}}{n}\right)\left(\frac{\boldsymbol{\eta}'\boldsymbol{\eta}}{n}\right)^{\frac{1}{2}},$$

and by part (c) of Assumption (8), $\lambda_{\max}^{1/2}\left(\frac{\mathbf{B}'\mathbf{M}_n\mathbf{B}}{n}\right) \to_p \lambda_{\max}^{1/2}(\boldsymbol{\Sigma}_{\beta\beta}) < C$, and by Assumption 6 $\left(\frac{\boldsymbol{\eta}'\boldsymbol{\eta}}{n}\right)^{\frac{1}{2}} = O\left(n^{\frac{\alpha-1}{2}}\right)$. Hence $\left\|\frac{\mathbf{B}'\mathbf{M}_n\boldsymbol{\eta}}{n}\right\| \to_p 0$, since $\alpha < 1$, and (A.7) follows. Consider now (A.8) and note that

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

and $n^{-1}\mathbf{U}'\mathbf{U} = n^{-1}\sum_{i=1}^{n}\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}'$, where $\mathbf{u}_{i\circ}\mathbf{u}_{i\circ}' = (u_{it}u_{it'})$, for $t, t' = 1, 2, ...T$. Hence, by results (A.1) and (A.2) of lemma A.1, it follows that $\boldsymbol{n}^{-1}\mathbf{U}'\mathbf{U} \to_p \overline{\sigma}^2\mathbf{I}_T$, and in conjunction with (A.12) yields (A.8) as required. To establish (A.9) note that

$$n^{-1}\mathbf{U}'\mathbf{M}_n\bar{\mathbf{u}} = n^{-1}\mathbf{U}'\bar{\mathbf{u}} - \left(\frac{\mathbf{U}'\boldsymbol{\tau}_n}{n}\right)\left(\frac{\boldsymbol{\tau}_n'\mathbf{U}}{n}\right), \tag{A.13}$$

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

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

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 $n^{-1}\mathbf{U}'\bar{\mathbf{u}} \rightarrow_p \frac{1}{T}\bar{\sigma}^2\boldsymbol{\tau}_T$. Also by (A.12) the second term of (A.13) tends to zero in probability and (A.9) follows. Finally to establish (A.10), note that

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

Also using (A.8) it follows that $\lambda_{\max}^{1/2}\left(\frac{\mathbf{U}'\mathbf{M}_n\mathbf{U}}{n}\right) \rightarrow_p \bar{\sigma}^2 < C$, and by Assumption 6 $n^{-1}\boldsymbol{\eta}'\,\boldsymbol{\eta} = O\left(n^{\alpha-1}\right)$, and as required $\left\|\frac{\mathbf{U}'\mathbf{M}_n\boldsymbol{\eta}}{n}\right\| \rightarrow_p 0$, since $\alpha < 1$. ■

Lemma A.3 *Consider the $n\times T$ error matrix $\boldsymbol{U} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, where $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$, the $n \times k$ matrix of factor loadings, $\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)$, the $n \times 1$ vector of pricing errors $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_n)'$, and the $n \times P$ matrix of portfolio weights, $\mathbf{W}_P = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P)'$, $\mathbf{w}_p = (w_{1p}, w_{2p}, ..., w_{np})'$. Suppose that Assumptions 4, 5, 6 and 8 hold, $\alpha < 1$, and $\|\mathbf{W}_P\| = \ominus\left(m^{-1/2}\right)$. Then for a fixed m, k and T, and as $P \rightarrow \infty$, such that $P/n \rightarrow \pi$, $(0 < \pi < 1)$, then we have*

$$\frac{\mathbf{U}'\mathbf{W}_P\tau_P}{P} \rightarrow_p \mathbf{0}, \tag{A.14}$$

$$\frac{\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{U}}{P} \rightarrow_p \mathbf{0}, \tag{A.15}$$

$$\frac{\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{u}}}{P} \rightarrow_p \mathbf{0}, \tag{A.16}$$

$$\frac{\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\eta}}{P} \rightarrow_p \mathbf{0}, \tag{A.17}$$

$$\frac{\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\eta}}{P} \rightarrow_p \mathbf{0}, \tag{A.18}$$

$$\frac{\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{U}}{P} \rightarrow_p \bar{\omega}^2\mathbf{I}_T, \tag{A.19}$$

$$\frac{\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{u}}}{P} \rightarrow_p \frac{\bar{\omega}^2}{T}\boldsymbol{\tau}_T, \tag{A.20}$$

where $\mathbf{M}_P = \mathbf{I}_P - \frac{1}{P}\boldsymbol{\tau}_P\boldsymbol{\tau}'_P$, $\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$, $\bar{\omega}^2 = \lim_{P\rightarrow\infty}\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}'_p\boldsymbol{\Sigma}_u\mathbf{w}_p\right)$, and $\boldsymbol{\Sigma}_u = (\sigma_{ij})$. Note that $\boldsymbol{\tau}_P$ and $\boldsymbol{\tau}_T$ are, respectively, $P \times 1$ and $T \times 1$ vectors of ones.

Proof. To establish result (A.14) first note that the t^{th} element of $P^{-1}\mathbf{U}'\mathbf{W}_P\tau_P$ is given by $P^{-1}\sum_{i=1}^n \bar{w}_{iP}u_{it}$, where $\bar{w}_{iP}=\sum_{p=1}^P w_{ip}$. Also $E\left(P^{-1}\sum_{i=1}^n \bar{w}_{iP}u_{it}\right)=0$, and

$$
\begin{aligned}
Var\left(P^{-1}\sum_{i=1}^n \bar{w}_{iP}u_{it}\right) &= P^{-2}\sum_{i=1}^n\sum_{j=1}^n \bar{w}_{iP}\bar{w}_{jP}\sigma_{ij} \\
&\leq \left(\sup_{i,P}|\bar{w}_{iP}|\right)^2 P^{-2}\sum_{i=1}^n\sum_{j=1}^n|\sigma_{ij}| \\
&\leq \left(\frac{1}{P/n}\right)\left(\frac{1}{P}\right)\left(\sup_{i,P}|\bar{w}_{iP}|\right)^2 \sup_i\sum_{j=1}^n|\sigma_{ij}|,
\end{aligned}
$$

which tends to zero as $P\to\infty$, since under Assumptions 4 and 5, $\sup_{i,P}|\bar{w}_{iP}|<C$, and $\sup_i\sum_{j=1}^n|\sigma_{ij}|<C$, and $1>P/n>0$. Hence, the elements of $P^{-1}\mathbf{U}'\mathbf{W}_P\tau_P$ all tend to zero in mean square and hence in probability. Consider now A.15 and note that

$$
P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\boldsymbol{U}=P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{W}_P'\ \boldsymbol{U}-\left(P^{-1}\mathbf{B}'\mathbf{W}_P\tau_P\right)\left(P^{-1}\tau_P'\mathbf{W}_P\boldsymbol{U}\right), \tag{A.21}
$$

Also $\mathbf{B}'\mathbf{W}_P=\left(\bar{\boldsymbol{\beta}}_1,\bar{\boldsymbol{\beta}}_2,...,\bar{\boldsymbol{\beta}}_P\right)$, $\mathbf{B}'\mathbf{W}_P\tau_P=\sum_{i=1}^P\bar{\boldsymbol{\beta}}_p$, where $\bar{\boldsymbol{\beta}}_p=\sum_{i=1}^n w_{ip}\boldsymbol{\beta}_i$, and by Assumption 9 $\sup_p\left\|\bar{\boldsymbol{\beta}}_p\right\|<C$. Hence,

$$
\begin{aligned}
\left\|\left(P^{-1}\mathbf{B}'\mathbf{W}_P\tau_P\right)\left(P^{-1}\tau_P'\mathbf{W}_P\boldsymbol{U}\right)\right\| &\leq \left\|P^{-1}\mathbf{B}'\mathbf{W}_P\tau_P\right\|\left\|P^{-1}\tau_P'\mathbf{W}_P\boldsymbol{U}\right\| \\
&\leq \left(P^{-1}\sum_{i=1}^P\left\|\bar{\boldsymbol{\beta}}_p\right\|\right)\left\|P^{-1}\tau_P'\mathbf{W}_P\boldsymbol{U}\right\|\leq C\left\|P^{-1}\tau_P'\mathbf{W}_P\boldsymbol{U}\right\|,
\end{aligned}
$$

and in view of (A.14), it follows that

$$
P^{-2}\mathbf{B}'\mathbf{W}_P\tau_P\tau_P'\mathbf{W}_P\ \boldsymbol{U}\to_p\mathbf{0}. \tag{A.22}
$$

The first term of (A.21) can be written as

$$
\begin{aligned}
P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{W}_P'\boldsymbol{U} &= P^{-1}\left(\sum_{p=1}^P\mathbf{B}'\mathbf{w}_p\mathbf{w}_p'\boldsymbol{U}\right)=P^{-1}\left(\sum_{p=1}^P\sum_{i=1}^n w_{ip}\bar{\boldsymbol{\beta}}_p\mathbf{u}_{i\circ}'\right) \\
&= P^{-1}\left(\sum_{i=1}^n\phi_{iP}\mathbf{u}_{i\circ}'\right),
\end{aligned}
$$

where $\phi_{iP}=\sum_{p=1}^P w_{ip}\bar{\boldsymbol{\beta}}_p=\left(\phi_{i1,P},\phi_{i2,P},...\phi_{ik,P}\right)'$, and $\phi_{is,P}=\sum_{p=1}^P w_{ip}\bar{\beta}_{sp}$. Since T and k are fixed, then it is sufficient to consider the limiting property of a typical element of $P^{-1}\left(\sum_{i=1}^n\phi_{iP}\mathbf{u}_{i\circ}'\right)$, namely $c_{st,P}=P^{-1}\left(\sum_{i=1}^n\phi_{is,P}u_{it}\right)$. We note that $E(c_{sP})=0$, and

$$
Var\left(c_{st,P}\right)=P^{-2}\sum_{i=1}^n\sum_{j=1}^n\phi_{isP}\phi_{js,P}\sigma_{ij}\leq\left(\sup_{i,s,P}|\phi_{isP}|\right)^2\left(\frac{n}{P^2}\right)\sup_i\sum_{j=1}^n|\sigma_{ij}|.
$$

Also $|\phi_{is,P}|\leq\sup_{s,p}\left|\bar{\beta}_{sp}\right|\sum_{p=1}^P|w_{ip}|<C$ and $\sup_i\sum_{j=1}^n|\sigma_{ij}|<C$, by Assumptions 4, 5,and 9. Hence, it follows that $Var\left(c_{st,P}\right)\to 0$, for all $s=1,2,..,k$ and $t=1,2,...,T$, and hence $P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{W}_P'\ \boldsymbol{U}\to_p\mathbf{0}$. Using this result together with (A.22) in (A.21) now establishes

(A.15). To prove (A.16) we first note that since $\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})' = T^{-1}\boldsymbol{U}\tau_T$, where $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^{T} u_{it}$, and hence $\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{u}} = T^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{U}\tau_T$, and

$$\begin{aligned}\left\|P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{u}}\right\| &\le \left\|P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{U}\right\| \left\|T^{-1}\tau_T\right\| \\ &= T^{-1/2}\left\|P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{U}\right\|,\end{aligned}$$

and tends to zero in probability by virtue of result (A.15). To prove (A.17) we note that

$$P^{-1}\left\|\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\eta}\right\| \le \left\|P^{-1/2}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\right\| \left\|P^{-1/2}\mathbf{W}'_P\boldsymbol{\eta}\right\|.$$

But $\lim_{P\to\infty}\left\|P^{-1/2}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\right\|^2 = \lim_{P\to\infty}\lambda_{\max}\left(P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{B}\right) < C$, by Assumption 9, and

$$P^{-1}\left\|\mathbf{W}'_P\boldsymbol{\eta}\right\|^2 \le P^{-1}\left\|\mathbf{W}_P\right\|^2\left\|\boldsymbol{\eta}\right\|^2 = P^{-1}\left\|\mathbf{W}_P\right\|^2\left(\sum_{i=1}^{n}\eta_i^2\right).$$

Also, since by Assumption $\left\|\mathbf{W}_P\right\|^2 = \ominus\left(m^{-1}\right)$, $P/n \to \pi$, then $P^{-1}\left\|\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\eta}\right\| = \ominus\left(n^{-1}\sum_{i=1}^{n}\eta_i^2\right) = \ominus\left(n^{\alpha-1}\right)$, which tends to zero since $\alpha < 1$. Result (A.18) follows similarly. To establish (A.19), in view of (A.14) it is sufficient to establish the probability limit of $P^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}'_P\boldsymbol{U}$. To this end we note that

$$\begin{aligned}P^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}'_P\boldsymbol{U} = P^{-1}\sum_{p=1}^{P}\mathbf{U}'\mathbf{w}_p\mathbf{w}'_p\ \boldsymbol{U} &= P^{-1}\sum_{p=1}^{P}\left(\sum_{i=1}^{n}w_{ip}\mathbf{u}_{i\circ}\right)\left(\sum_{i=1}^{n}w_{jp}\mathbf{u}'_{j\circ}\right) \\ &= P^{-1}\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ip}w_{jp}\mathbf{u}_{i\circ}\mathbf{u}'_{j\circ}.\end{aligned}$$

Therefore, a typical (t,t') element of the $T\times T$ matrix $\mathbf{B}_P = P^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}'_P\boldsymbol{U}$ is given by $b_{tt',P} = P^{-1}\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ip}w_{jp}u_{it}u_{jt'}$ and we have

$$\begin{aligned}&E\left(b_{tt',P}\right) = P^{-1}\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ip}w_{jp}\sigma_{ij} = P^{-1}\sum_{p=1}^{P}\mathbf{w}'_p\boldsymbol{\Sigma}_u\mathbf{w}_p, \text{ if } t = t', \\ &E\left(b_{tt',P}\right) = 0, \text{ if } t \neq t',\end{aligned}$$

and hence $E\left(\mathbf{B}_P\right) = \bar{\omega}_P^2\mathbf{I}_T$, where $\bar{\omega}_P^2 = P^{-1}\sum_{p=1}^{P}\mathbf{w}'_p\boldsymbol{\Sigma}_u\mathbf{w}_p$. The convergence in probability follows by considering $E\left(b_{tt',P}^2\right)$ when $t \neq t'$ and $E\left(b_{tt',P} - \bar{\omega}_P^2\right)^2$ when $t = t'$, and following the approach used to establish results (A.1) and (A.2) in Lemma A.1. The details are tedious and will be omitted to save space. Finally, result (A.20) follows from (A.19), noting that $\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{u}} = T^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}'_P\boldsymbol{U}\tau_T$. ■

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 (37), and to simplify notations, write it as

$$\hat{\boldsymbol{\lambda}}_n = \left(\frac{\hat{\mathbf{B}}'\mathbf{M}_n\hat{\mathbf{B}}}{n}\right)^{-1}\left(\frac{\hat{\mathbf{B}}'\mathbf{M}_n\bar{\mathbf{r}}}{n}\right), \tag{A.23}$$

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

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{A.24}$$

and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under the factor model (32)

$$\mathbf{r}_{i\circ} = \alpha_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{A.25}$$

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

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

Stacking these results over i yields:

$$\hat{\mathbf{B}} = \mathbf{B} + \mathbf{U}\mathbf{G}_T \tag{A.27}$$

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

$$\mathbf{G}_T = \mathbf{M}_T\mathbf{F}(\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} \tag{A.28}$$

Also using result (41) in the paper we have (in terms of the simplified notations used here)

$$\bar{\mathbf{r}} = \lambda_0\boldsymbol{\tau}_n + \mathbf{B}\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}} + \boldsymbol{\eta} \tag{A.29}$$

where

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda} + \mathbf{d}_T, \text{ and } \mathbf{d}_T = \bar{\mathbf{f}}_T - E\left(\bar{\mathbf{f}}_T\right). \tag{A.30}$$

and $\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$. To derive the asymptotic limit of $\hat{\boldsymbol{\lambda}}_n$ as $n \to \infty$, when T is fixed, we first consider the probability limits of $n^{-1}\left(\hat{\mathbf{B}}'\mathbf{M}_n\hat{\mathbf{B}}\right)$ and $n^{-1}\left(\hat{\mathbf{B}}'\mathbf{M}_n\bar{\mathbf{r}}\right)$. Using (39) and (A.29) we have

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

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

Now using the results in Lemma A.2, under Assumptions 5, 6 and 8 we have

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

But using (A.28) we also have

$$\mathbf{G}_T'\mathbf{G}_T = \frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}, \tag{A.31}$$

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

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

Since by assumption 8, $\boldsymbol{\Sigma}_{\beta\beta}$ is a positive definite matrix, then for any T (including $T \to \infty$) the probability limit of $n^{-1}\hat{\mathbf{B}}'\mathbf{M}_n\hat{\mathbf{B}}$ is non-singular and using (A.23) by the Slutsky Theorem we have

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

which in view of (A.30) can be written equivalently in the form stated in Theorem 1.

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

Consider the expression for $\widehat{\bar{\sigma}}_{nT}^2$ given by (45) and note that under (33) 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 $\hat{\mathbf{u}}_t = \mathbf{u}_t - \bar{\mathbf{u}} - \left(\hat{\mathbf{B}} - \mathbf{B}\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)$. Hence

$$\begin{aligned}
T^{-1}n^{-1}\sum_{t=1}^T\sum_{i=1}^n \hat{u}_{it}^2 &= T^{-1}\sum_{t=1}^T n^{-1}\hat{\mathbf{u}}_t'\hat{\mathbf{u}}_t \\
&= T^{-1}\sum_{t=1}^T n^{-1}\left(\mathbf{u}_t - \bar{\mathbf{u}}\right)'\left(\mathbf{u}_t - \bar{\mathbf{u}}\right) \\
&+ T^{-1}\sum_{t=1}^T \left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)' n^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}\right)'\left(\hat{\mathbf{B}} - \mathbf{B}\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right) \\
&- 2T^{-1}\sum_{t=1}^T n^{-1}\left(\mathbf{u}_t - \bar{\mathbf{u}}\right)'\left(\hat{\mathbf{B}} - \mathbf{B}\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right) \\
&= a_{nT} + b_{nT} + c_{nT} \qquad \text{(A.32)}
\end{aligned}$$

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 5 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 $n \to \infty$)

$$a_{nT} \to_p \left(1 - \frac{1}{T}\right)\bar{\sigma}^2. \qquad \text{(A.33)}$$

Now using (A.27)

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

and in view of (A.8) we have (as $n \to \infty$)

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

But by (A.31) $\mathbf{G}_T'\mathbf{G}_T = \frac{1}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}$, and it follows that

$$b_{nT} \to_p \frac{k}{T}\bar{\sigma}^2. \tag{A.34}$$

. Finally, again using (A.27)

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

and noting that

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

$$T^{-1} \sum_{t=1}^{T} \mathbf{f}_t \mathbf{u}_t' = T^{-1} \mathbf{F}' \mathbf{U}'$$

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

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

Now using (A.8) and (A.9) it follows that

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

But using (A.28) it is readily seen that $\mathbf{F}'\mathbf{G}_T = \mathbf{I}_k$ and $\boldsymbol{\tau}_T'\mathbf{G}_T = \mathbf{0}$, and therefore

$$c_{nT} \to_p -\frac{2k}{T}\bar{\sigma}^2. \tag{A.35}$$

Now using (A.33), (A.34) and (A.35) in (A.32)

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

which establishes

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

as required.

A.3.3 Proof of Theorem 3

The proof is similar to that of Theorem 1. Consider the portfolio estimator $\boldsymbol{\lambda}$ given by (60) and write it simply as

$$\hat{\boldsymbol{\lambda}}_P = \left(P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\hat{\boldsymbol{B}}\right)^{-1}\left(P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{r}}\right), \tag{A.36}$$

Substituting $\hat{\boldsymbol{B}}$ and $\bar{\mathbf{r}}$ using (A.27) and (A.29) respectively, we have

$$\begin{aligned} P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\hat{\boldsymbol{B}} &= P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{B}+P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{U}\mathbf{G}_T \\ &+ P^{-1}\mathbf{G}'_T\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{B}+P^{-1}\mathbf{G}'_{\mathbf{T}}\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{U}\mathbf{G}_T, \end{aligned}$$

and

$$\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{r}} = (\mathbf{B}+\mathbf{U}\mathbf{G}_T)'\,\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\left(\alpha_i\boldsymbol{\tau}_n+\mathbf{B}\boldsymbol{\lambda}^*_T+\bar{\mathbf{u}}+\boldsymbol{\eta}\right),$$

and recall that $\boldsymbol{\lambda}^*_T$ is defined by (A.30). Also, note that since $\Sigma_{i=1}^n w_{ip} = 1$, for all p, then $\mathbf{W}'_P\boldsymbol{\tau}_n = \boldsymbol{\tau}_P$ and $\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\tau}_n = \mathbf{M}_P\boldsymbol{\tau}_P = \mathbf{0}$. Hence,

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

Under Assumptions 5, 6 and 9, and using the results of Lemma A.3, we have (as $P \to \infty$, for a fixed m, T and k):

$$\begin{aligned} P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\hat{\boldsymbol{B}} &\to_p \boldsymbol{\Sigma}_{\beta\beta,\omega} + \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}, \\ P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{r}} &\to_p \boldsymbol{\Sigma}_{\beta\beta,\omega}\boldsymbol{\lambda}^*_T, \end{aligned}$$

where $\bar{\omega}^2$ and $\boldsymbol{\Sigma}_{\beta\beta,\omega}$ are defined by (63). Result (62) then follows by using the above in (A.36), and writing the outcome in terms of $\hat{\boldsymbol{\lambda}}_P - \boldsymbol{\lambda}$.

References

Ang, A., J. Liu & K. Schwarz (2019) Using stocks or portfolios in tests of factor models, forthcoming *Journal of Financial and Quantitative Analysis.*

Bai, Jushan, & Guofu Zhou (2015) Fama-MacBeth two-pass regressions: improving risk premia estimates, *Finance Research Letters*, 15, 31-40.

Bailey, Natalia, George Kapetanios & M. Hashem Pesaran (2016) Exponent of Cross-sectional dependence: estimation and inference, *Journal of Applied Econometrics*, 31(1) 929-1196.

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

____ (2019b) Measures of factor strength in macroeconomics and finance: estimation and inference, unpublished manuscript.

Chamberlain, Gary, & Michael Rothschild (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica*, 51(5) 1281-1304.

Chen, Nai-Fu, Richard Roll & Stephen A. Ross (1986) Economic Forces and the Stock Market, *Journal of Business*, 59(3), 383-403.

Chudik, Alexander, George Kapetanios & M. Hashem Pesaran (2018) A one-covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models, *Econometrica*, 86(4), 1479-1512.

Fama, Eugene F. & Kenneth R. French (1993) Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, 33, 3-56.

Fama, Eugene F. & Kenneth R. French (2018) Choosing Factors, *Journal of Financial Economics*, 128(2) 234-252.

Fama, Eugene F. & James D. MacBeth (1973) Risk, return and equilibrium: empirical tests, *Journal of Political Economy* 81(3) 607-636.

Feng, Guanhao, Stefano Giglio & Dacheng Xiu (2019) Taming the factor Zoo, NBER working paper 25481.

Harvey, Campbell R. and Liu, Yan (2019) A Census of the Factor Zoo. Available at SSRN: https://ssrn.com/abstract=3341728.

Jagannathan, Ravi, Georgios Skoulakis & Zhenyu 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.

Kan, Raymond, Cesare. Robotti & Jay Shanken (2013) Pricing model performance and the two-pass cross-sectional regression methodology, *The Journal of Finance*, 68(6) 2617-2649.

Lehmann, Bruce N., & David M. Modest (2005) Diversification and the optimal construction of basis portfolios, *Management Science*, 51(4) 581-598.

Pagan, Adrian (1984) Econometric issues in the analysis of generated regressors *International Economic Review* 25(1) 221-247.

Pesaran, M. Hashem (1987) *The Limits to Rational Expectations*, Basil Blackwell, Oxford.

Pesaran, M. Hashem, & Takashi 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.

Ross, Stephen A. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13, 341-360.

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

Shanken, Jay & Guofu Zhou (2007) Estimating and testing beta-pricing models: alternative methods and their performance in simulations, *Journal of Financial Economics*, 84, 40-86.

Stambaugh, Robert F., & Yu Yuan (2018) Mispricing factors, *The Review of Financial Studies*, 30(4) 1270-1315.

Wei, K.C. John (1988) An asset pricing theory unifying the CAPM and APT, *The Journal of Finance*, 43(4), 881-892.