Pesaran, M. Hashem; Smith, Ron P.

Working Paper

Arbitrage Pricing Theory, the Stochastic Discount Factor and Estimation of Risk Premia from Portfolios

CESifo Working Paper, No. 9001

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Pesaran, M. Hashem; Smith, Ron P. (2021) : Arbitrage Pricing Theory, the Stochastic Discount Factor and Estimation of Risk Premia from Portfolios, CESifo Working Paper, No. 9001, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/235371

CESifo WORKING PAPERS

9001
2021

April 2021

Arbitrage Pricing Theory, the Stochastic Discount Factor and Estimation of Risk Premia from Portfolios

M. Hashem Pesaran, Ron P. Smith

CESifo

Impressum:

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

CESifo *Working Paper No. 9001*

Arbitrage Pricing Theory, the Stochastic Discount Factor and Estimation of Risk Premia from Portfolios

Abstract

The arbitrage pricing theory (APT) attributes differences in expected returns to exposure to systematic risk factors, which are typically assumed to be strong. In this paper we consider two aspects of the APT. Firstly we relate the factors in the statistical factor model to a theoretically consistent set of factors defined by their conditional covariation with the stochastic discount factor (m_t) used to price securities within inter-temporal asset pricing models. We show that risk premia arise from non-zero correlation of observed factors with m_t; and the pricing errors arise from the correlation of the errors in the statistical factor model with m_t: Secondly we compare estimates of factor risk premia using portfolios with the ones obtained using individual securities, and show that the identification conditions in terms of the strength of the factor are the same and that, in general, no clear cut ranking of the small sample bias of the two estimators is possible.

JEL-Codes: C380, G120.

Keywords: arbitrage pricing theory, stochastic discount factor, portfolios, factor strength, identification of risk premia, two-pass regressions, Fama-MacBeth.

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

Ron P. Smith
Department of Economics
Birkbeck, University of London
London / United Kingdom
r.smith@bbk.ac.uk

March 30, 2021
This paper draws on material in an earlier working paper entitled "The role of factor strength and pricing errors for estimation and inference in asset pricing models", CESifo working paper 7919, October 2019.

1 Introduction

This paper considers two questions raised by the conventional use of the arbitrage pricing theory (APT) in empirical finance. Firstly, APT is generally regarded as an empirical model and there is a question as to how the risk factors in the APT formulation relate to the inter-temporal model of asset pricing in which a stochastic discount factor represents the fundamentals. Secondly, estimation of risk premia and the construction of the factors tends to be done using portfolios, and there is a question as to whether using portfolios, rather than individual securities, helps identify the prices of risk factors and reduces the small sample bias.

The 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 in equilibrium. Thus returns can be split into two components: a non-diversifiable systematic risk component and an idiosyncratic part which can be eliminated in a well diversified portfolio. Assets with similar risk factors are close substitutes so should have similar expected 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.[1]

The loadings and risk premia are usually estimated using a two-pass procedure suggested by Fama and MacBeth (1973). The first-pass, is a time series regression of excess returns for each asset, r_{it}, $i = 1, 2, ..., n$ on K observed factors, f_{kt}, $k = 1, 2, ..., K$, $t = 1, 2, ..., T$. This is used to estimate the factor loadings, β_{ik}, which are assumed to be stable over the given sample period. The second-pass, cross section regression of average returns on the factor loadings is then used to price the factors and obtain the risk premia, λ_k, for the factor f_{kt}.[2] For estimation, it is assumed that factors are strong in the sense that they impact all security returns, almost surely. In terms of a measure of factor strength, discussed further below, a factor is said to be strong if $\alpha_k = 1$, semi-strong if $1 > \alpha_k > 0.5$, and weak if $\alpha_k \leq 0.5$. 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 potential factors in their paper.

Following the pioneering contribution of Fama and MacBeth, it is conventional in this literature to use mean returns on a relatively small number of portfolios ($P < n$) formed from the underlying securities in the second pass regression rather than the securities themselves. It is argued that the sampling errors in the estimates of the first stage β_{ik} from individual securities can be substantially reduced by using the $\beta's$ of portfolios. We provide a theoretical investigation of this practice and give conditions under which the use of portfolios rather than individual securities could be justified.

More specifically, this paper first relates factors in the statistical factor model to a theoretically consistent set of factors defined by their conditional covariation with the stochastic discount factor, denoted by m_t, used to price securities within inter-temporal asset pricing models. We show that a risk factor is priced only if its conditional covariance with m_t is non-zero. In contrast, pricing errors arise when there is non-zero correlations between the idiosyncratic errors of asset returns and the stochastic discount factor. The APT theory places bounds on the pricing errors, requiring them to be square summable.

[1] Wei (1988) links the APT to the capital asset pricing model, CAPM.

[2] The asymptotic properties of the Fama-MacBeth estimation procedure 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.

Secondly, having linked the APT to the inter-temporal asset pricing condition, we compare the use of portfolios and individual securities in the identification of risk premia using an approximate linear factor model. The assumptions made about portfolio weights cover a wide variety of portfolio types, both fully diversified and non-diversified portfolio weights are considered. We show that the use of portfolio returns and their associated $\beta's$ in the second pass does not alleviate the small T bias and in some settings could even accentuate it. We begin by assuming known factor loadings. This allows us to focus on the identification of risk premia in the second pass regression, without the complications arising from the first pass estimation of factor loadings, β_{ik}. If one allows for pricing errors, which are often ignored in the literature on estimation, n has to be large. We show, that for known loadings and large n the risk premia can be $\sqrt{n}$ consistently estimated if all the factors are strong (namely if $\alpha_k = 1$ for all k), and the pricing errors are sufficiently weak (namely if a measure of their strength $\alpha_\eta = 0$). This applies whether individual securities or portfolios are used.[3]

In the more realistic case where the first-pass loadings are estimated, there is a small T bias on the second-pass risk premia estimates, whether individual securities or portfolios are used. We obtain an expression for the small T bias of risk premia using portfolios which corresponds to a similar result obtained by Shanken (1992) using individual securities. But when using portfolio returns the small T bias depends on the portfolio weights and error covariances, and can be estimated consistently only for large n and T, even if it is assumed that the errors are weakly cross-correlated. This contrasts with the result obtained for individual securities that does not depend on error covariances. We also compare the bias of estimating the risk premia based on portfolio returns with the one based on individual securities and show that in general no clear cut ranking of the two estimators is possible. This is illustrated in the case of a simple example, where it is shown that the use of portfolio returns can be justified only when returns can be sorted *a priori* into groups with systematically different loadings.

The rest of the paper is organized as follows. Section 2 relates the statistical factor model to the theory consistent factor model in terms of the stochastic discount factor to derive the APT risk premia and pricing errors. Section 3 analyses the effect of using portfolios for the identification of the risk premia for the factors from a cross section when the factor loadings are known. Section 4 analyses the effect of using portfolios when the factor loadings are unknown and provides a Shanken type bias correction formula. Section 5 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. $\to^p$ denotes convergence in probability as $n, T \to \infty$. $\lambda_{\max}(\boldsymbol{A})$ 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. $\|\boldsymbol{A}\| = \lambda^{1/2}_{\max}(\mathbf{A}'\mathbf{A})$ 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 $f_n/g_n \to 0$ as $n \to \infty$. Similarly, $f_n = O_p(g_n)$ if f_n/g_n is stochastically bounded, and $f_n = o_p(g_n)$, if $f_n/g_n \to_p 0$, where $\to_p$denotes convergence in probability. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \ominus(g_n)$ if there exists $n_0 \geq 1$ and positive finite constants C_0 and C_1, such that $\inf_{n \geq n_0}(f_n/g_n) \geq c > 0$, and $\sup_{n \geq n_0}(f_n/g_n) \leq C < \infty$.

[3]The significance of factor strengths for the estimation of risk premia when using individual securities is investigated in a companion paper (Pesaran and Smith, 2021).

2 Statistical factor models, the stochastic discount factor and the APT

This section sets out the statistical factor model, imposes the equilibrium conditions from standard pricing theory in terms of the stochastic discount factor to derive the associated theoretically consistent factor model. This is then interpreted in terms of the risk premia and pricing errors of the APT model used by Ross and others in the literature. The implications of the theory consistent model for portfolios is then examined.

2.1 Statistical and theory consistent factor models

Following the literature we suppose that returns, $r_{i,t+1}$, on securities $i = 1, 2, ...n$ are generated by a linear multi-factor model of the form:

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

where r_t^f is the risk free rate; a_{it} are the intercepts in the factor model; $f_{k,t+1}$, $k = 1, 2, ..., K$ are the observed common factors with associated factor loadings, β_{ik}.[4] The error $u_{i,t+1}$ is a mean zero serially uncorrelated idiosyncratic component of returns.[5] The model can be written more compactly as

$$r_{i,t+1} - r_t^f = \mathrm{a}_{it} + \boldsymbol{\beta}_i' \mathbf{f}_{t+1} + u_{i,t+1}, \tag{2}$$

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

We now examine the restrictions that the standard inter-temporal asset pricing theory imposes on the above 'statistical' factor model in order to interpret the risk premia and pricing errors in terms of a theory consistent model. As it is well known, the inter-temporal equilibrium pricing condition is given by

$$E_t \left[m_{t+1} (r_{i,t+1} - r_t^f) \right] = 0, \tag{3}$$

for all i, 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, and $E_t(\circ)$ stands for conditional expectations with respect to the information set, $\mathcal{I}_t$. More specifically $E_t(m_{t+1}) = E\left(m_{t+1} | \mathcal{I}_t\right)$. In addition,

$$E_t(m_{t+1}) = 1/(1 + r_t^f) > 0. \tag{4}$$

To derive conditions under which the statistical factor model (2) also satisfies the equilibrium pricing condition, substitute for $r_{i,t+1} - r_t^f$ from (2) in (3), to give

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

Since $E_t(m_{t+1}) > 0$, then a_{it} can be solved as

$$\mathrm{a}_{it} = -\frac{1}{E_t(m_{t+1})} \left[\boldsymbol{\beta}_i' E_t\left(m_{t+1} \mathbf{f}_{t+1}\right) + E_t(m_{t+1} u_{i,t+1}) \right]. \tag{5}$$

[4] We assume that the loadings, β_{ik}, are constant, but this could easily be relaxed at the cost of a more complex notation. In practice, the time-varying loadings are estimated by rolling regressions.

[5] Ross (1976) assumed $u_{i,t+1}$ were cross sectionally independent, Chamberlain & Rothschild (1983) weakened this to an approximate factor model that requires the maximal eigenvalue of the covariance matrix of $u_{i,t+1}$ is bounded.

Imposing this restriction by substituting (5) back into (1) yields the following theory consistent factor model

$$r_{i,t+1} - r_t^f = \boldsymbol{\beta}_i' \mathbf{g}_{t+1} + \eta_{it} + u_{i,t+1}, \tag{6}$$

where

$$\mathbf{g}_{t+1} = \mathbf{f}_{t+1} - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})} \tag{7}$$

and

$$\eta_{it} = -\frac{E_t(m_{t+1}u_{i,t+1})}{E_t(m_{t+1})}, \tag{8}$$

is the pricing error of the i^{th} security. Note that by construction the theory consistent factor model in (6) satisfies the equilibrium pricing condition (3) for all i. Also using the identity $Cov_t(m_{t+1}, \mathbf{f}_{t+1}) = E_t\left(m_{t+1}\mathbf{f}_{t+1}\right) - E_t\left(m_{t+1}\right) E_t\left(\mathbf{f}_{t+1}\right)$ in (7) the theory consistent factors, $\mathbf{g}_t$, can be written equivalently as:

$$\mathbf{g}_{t+1} = \mathbf{f}_{t+1} - E_t\left(\mathbf{f}_{t+1}\right) - \frac{Cov_t(m_{t+1}, \mathbf{f}_{t+1})}{E_t(m_{t+1})}. \tag{9}$$

This representation provides a transparent link between risk premia and (conditional) covariance of $\mathbf{f}_{t+1}$ and m_{t+1}. This follows since $E_t\left[\mathbf{f}_{t+1} - E_t\left(\mathbf{f}_{t+1}\right)\right] = 0$, and taking conditional expectations of (6) yields (note that $E_t\left(\eta_{it}\right) = \eta_{it}$)

$$E_t\left(r_{i,t+1} - r_t^f\right) = \boldsymbol{\beta}_i' E_t\left(\mathbf{g}_{t+1}\right) + \eta_{it}, \tag{10}$$

where $E_t\left(\mathbf{g}_{t+1}\right) = -\frac{Cov_t(m_{t+1}, \mathbf{f}_{t+1})}{E_t(m_{t+1})}$, is the vector of risk premia, with $E_t\left(g_{k,t+1}\right) = -\frac{Cov_t(m_{t+1}, f_{k,t+1})}{E_t(m_{t+1})}$ being the conditional risk premium of factor $g_{k,t+1}$. Therefore, the statistical factor $f_{k,t+1}$ has a non-zero conditional risk premium if it is correlated with the discount factor. A simple example of such a factor is consumption growth discussed in the example below.

The above results readily generalize to (unconditional) risk premia if $\mathbf{g}_{t+1}$ and η_{it} are stationary processes. In this case taking unconditional expectations of (10) we obtain:

$$E\left(r_{i,t+1} - r_t^f\right) = \boldsymbol{\beta}_i' E\left(\mathbf{g}_{t+1}\right) + E(\eta_{it}),$$

This corresponds to the APT cross section return regression

$$E\left(r_{i,t+1} - r_t^f\right) = \boldsymbol{\beta}_i' \boldsymbol{\lambda} + \eta_i, \tag{11}$$

where (in terms of the theory consistent factor model) we have $\boldsymbol{\lambda} = E\left(\mathbf{g}_{t+1}\right)$, and $\eta_i = E(\eta_{it})$. Here $\boldsymbol{\lambda}$ is the vector of risk premia, and η_i is the pricing error of the i^{th} security.

Our derivation allows us to relate $\boldsymbol{\lambda}$ to the theory consistent factor model. In particular:

$$\boldsymbol{\lambda} = E\left(\mathbf{g}_{t+1}\right) = -E\left[\frac{Cov_t(m_{t+1}, \mathbf{f}_{t+1})}{E_t(m_{t+1})}\right] = -E\left[(1 + r_t^f) Cov_t(m_{t+1}, \mathbf{f}_{t+1})\right], \tag{12}$$

or for an individual factor, $f_{k,t+1}$,

$$\lambda_k = -E\left[(1 + r_t^f) Cov_t(m_{t+1}, f_{k,t+1})\right], \text{ for } k = 1, 2, ..., K. \tag{13}$$

The above results are summarized in the following proposition:

Proposition 1 *Suppose that returns, $r_{i,t+1}$, on security $i = 1, 2..., n$ are generated according to the multi-factor model (1) subject to the inter-temporal equilibrium conditions in (3). Then the risk premium of factor $f_{k,t+1}$ is given by (13) and only factors that are conditionally correlated with the stochastic discount factor, m_{t+1}, have non-zero risk premia.*

Example 1 *To illustrate the derivation of the theory consistent factor model, consider the case of the stochastic discount factor which comes from the familiar consumption based asset pricing. Assume that utility is based on current and discounted expected future consumption, ρ is the subjective discount rate and there is power utility $U(C_t) = C_t^{(1-\kappa)}/(1-\kappa)$, with $\kappa > 0$. Then we have $m_{t+1} = e^{-\kappa \Delta c_{t+1}}/(1+\rho)$, where $c_t = \log(C_t)$. For this specification*

$$\mathbf{g}_{t+1} = \mathbf{f}_{t+1} - \frac{E_t\left(e^{-\kappa \Delta c_{t+1}} \mathbf{f}_{t+1}\right)}{E_t(e^{-\kappa \Delta c_{t+1}})},$$

and using

$$\begin{aligned} Cov_t(e^{-\kappa \Delta c_{t+1}}, \mathbf{f}_{t+1}) &= E_t\left(e^{-\kappa \Delta c_{t+1}} \mathbf{f}_{t+1}\right) - E_t\left(e^{-\kappa \Delta c_{t+1}}\right) E_t(\mathbf{f}_{t+1}) \\ &= \mathbf{f}_{t+1} - \frac{Cov_t(e^{-\kappa \Delta c_{t+1}}, \mathbf{f}_{t+1}) + E_t\left(e^{-\kappa \Delta c_{t+1}}\right) E_t(\mathbf{f}_{t+1})}{E_t(e^{-\kappa \Delta c_{t+1}})} \\ &= [\mathbf{f}_{t+1} - E_t(\mathbf{f}_{t+1})] - \frac{Cov_t(e^{-\kappa \Delta c_{t+1}}, \mathbf{f}_{t+1})}{E_t(e^{-\kappa \Delta c_{t+1}})}. \end{aligned}$$

Thus corresponding to (13) $\lambda_k = -E\left[(1+r_t^f) Cov_t(e^{-\kappa \Delta c_{t+1}}, f_{k,t+1})\right]$, for factor $f_{k,t+1}$ to have a non-zero risk premium we must have $Cov_t(e^{-\kappa \Delta c_{t+1}}, f_{k,t+1}) \neq 0$.

2.2 Pricing errors

From (8), the pricing errors in the theory consistent factor model is

$$\eta_i = E(\eta_{it}) = -E\left[(1+r_t^f) E_t(m_{t+1} u_{i,t+1})\right], \tag{14}$$

which Ross (1976, condition 18) assumes to be bounded such that

$$\sum_{i=1}^{n} \eta_i^2 < C. \tag{15}$$

To further investigate the pricing error, decompose the errors in the statistical factor model, $u_{i,t+1}$, into a part correlated with m_{t+1} and a remaining idiosyncratic part uncorrelated with m_{t+1}, namely

$$u_{i,t+1} = \phi_i m_{t+1} + \varepsilon_{i,t+1}, \tag{16}$$

thus using (16) in (8)

$$\eta_i = -\phi_i E\left[(1+r_t^f) m_{t+1}^2\right]. \tag{17}$$

Then in terms of Ross's condition (15) we must have

$$\sum_{i=1}^{n} \eta_i^2 = \theta^2 \left(\sum_{i=1}^{n} \phi_i^2\right), \tag{18}$$

where $\theta = E\left[(1+r_t^f) m_{t+1}^2\right] > 0$. The strength of the pricing errors depends on their degree of pervasiveness, namely the rate at which $\sum_{i=1}^{n} \phi_i^2$ rises with n. The APT condition requires that $\sum_{i=1}^{n} \phi_i^2 < C$.

2.3 Portfolios

As noted in the introduction, the analysis of factor models is usually done in terms of portfolios. Consider a return portfolio $\overline{r}_{pt} = \sum_{i=1}^{n} w_{ip} r_{it}$, represented by the vector of weights, $\mathbf{w}_p = (w_{1p}, w_{2p}, ..., w_{np})'$, where $\sum_{i=1}^{n} w_{ip} = 1$. Then aggregating (11) it follows that

$$E_t\left(\overline{r}_{p,t+1} - r_t^f\right) = \bar{\boldsymbol{\beta}}_p' \boldsymbol{\lambda}_t + \bar{\eta}_{pt},$$
$$p = 1, 2, ..., P \geq K,$$

for $p = 1, 2, ..., P$, where

$$\bar{\boldsymbol{\beta}}_p = \sum_{i=1}^{n} w_{ip} \boldsymbol{\beta}_i, \text{ and } \bar{\eta}_{pt} = \sum_{i=1}^{n} w_{ip} \eta_{it}.$$

If the weights are granular such that $\mathbf{w}_p' \mathbf{w}_p = \ominus (n^{-1})$, and the pricing errors are bounded such that $\lambda_{max}(\mathbf{V}_\eta) < C$

$$Var\left(\sum_{i=1}^{n} w_{ip} \eta_{it}\right) = \mathbf{w}_p' \mathbf{V}_\eta \mathbf{w}_p \leq \left(\mathbf{w}_p' \mathbf{w}_p\right) \lambda_{max}(\mathbf{V}_\eta) \to 0.$$

It is possible to identify the risk premia, $\boldsymbol{\lambda}_t$, using portfolio returns $\overline{r}_{pt}$, for $p = 1, 2, ..., P$ so long as $\bar{\boldsymbol{\beta}}_p$ does not tend to zero.

In this regard, following the literature, the equilibrium condition, (3), can be written equivalently as

$$E_t\left(r_{i,t+1} - r_t^f\right) = -\frac{Cov_t(r_{i,t+1} - r_t^f, m_{t+1})}{E_t(m_{t+1})} = \beta_i' E_t(\mathbf{g}_{t+1}) + \eta_{it},$$

$$E_t\left(\sum_{i=1}^{n} w_{ip} r_{i,t+1} - \sum_{i=1}^{n} w_{ip} r_t^f\right) = -\frac{\sum_{i=1}^{n} w_{ip} Cov_t(r_{i,t+1} - r_t^f, m_{t+1})}{E_t(m_{t+1})}$$
$$= -\frac{Cov_t\left[\sum_{i=1}^{n} w_{ip}\left(r_{i,t+1} - r_t^f\right), m_{t+1}\right]}{E_t(m_{t+1})},$$

or more compactly

$$E_t\left(\bar{r}_{p,t+1} - r_t^f\right) = -\frac{Cov_t(\bar{r}_{p,t+1} - r_t^f, m_{t+1})}{E_t(m_{t+1})}.$$

The above analysis highlights the importance of distinguishing between the 'statistical factor model' given by (2), and the 'theory consistent factor model' given by (6). The focus of theoretical and empirical analysis should be on the theory consistent factor model, where it clearly shows that only factors that are known (or expected) to be correlated with the stochastic discount factor should be considered for inclusion in return regressions.

We now conduct a more formal presentation of the issues that surround identification of risk premia using portfolios.

3 Identification of risk premia

The debate over whether it is better to use portfolios or individual securities to estimate risk premia is an old one, and dates back to Fama and MacBeth (1973) who recommend using mean returns on a relatively small number of portfolios ($P < n$). Ang, Liu and Schwarz (2020) provide a survey of the issues involved, and consider the efficiency of estimation of risk premia using portfolios compared to individual securities. They comment "The majority of modern asset pricing papers testing expected return relations in the cross section now use portfolios." (p.710). Fama and MacBeth argued that more reliable estimates of $\boldsymbol{\lambda}$ can be obtained by using portfolios because the sampling errors in the estimates of the first stage β_{ik} from individual securities can be substantially reduced by using the $\bar{\beta}_{pk}$ of portfolios. 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 $\hat{\beta}_{ik}$ estimated over a training sample of four years. Portfolios are also often used to construct factors, like the five Fama and French (2015) factors.

In this section, we provide a theoretical framework for comparing the use of portfolios and individual securities for identification and estimation of risk premia. We consider two types of portfolio, a small number of fully diversified portfolios, and a large number of portfolios formed from a small number of securities. To clarify the central issues in identification, we first assume factor loadings are known and do not need to be estimated in the first-pass. This avoids the complications associated with the small T bias that comes from using estimated factor loadings. Unlike much of the literature on estimation of risk premia, we explicitly allow for pricing errors and establish the restrictions on the pricing errors needed for the identification of the risk premia.

In the next section, we relax the assumption of known factor loadings and compare the small T bias when one uses portfolios rather than individual securities.

3.1 Preliminaries

First we introduce the following definition of factor strength discussed in detail in Pesaran and Smith (2021) and make a number of assumptions to be used throughout the paper.

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

$$\sum_{i=1}^{n} \beta_{ik}^2 = \ominus(n^{\alpha_k}), \text{ for } k = 1, 2, ..., K, \tag{19}$$

and $0 \leq \alpha_k \leq 1$. We refer to $\{\alpha_k,\ j = 1, 2, ..., K\}$ as factor strengths. Factor f_{kt} is said to be strong if $\alpha_k = 1$, semi-strong if $1 > \alpha_k > 0.5$, and weak if $\alpha_k \leq 0.5$. We also require sufficient heterogeneity across the such that

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

where $\bar{\beta}_k = n^{-1} \sum_{i=1}^{n} \beta_{ik}$.

Remark 1 *The above definition of factor strength allows for the possibility of non-zero pricing errors ($\eta_i \neq 0$) in the theory consistent factor model (6), and in the related APT equilibrium condition (11).*

Assumption 1 *(Idiosyncratic errors) The errors $\{u_{it},\ i = 1, 2, ..., n; t = 1, 2, ..., T\}$ are serially independent over 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

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

Assumption 2 *(Pricing errors) The pricing errors, η_i, defined by (14) have zero means and satisfy the approximate bound*

$$\sum_{i=1}^{n} |\eta_i| = O_p\left(n^{\alpha_\eta}\right). \tag{20}$$

Assumption 3 *(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. $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F} \to_p \mathbf{\Sigma}_f = E\left[(\mathbf{f}_{t+1} - \mu_f)(\mathbf{f}_{t+1} - \mu_f)'\right] > 0$, where $\mu_f = E_t(\mathbf{f}_{t+1})$, K is a fixed number, $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}_T'$, and $\boldsymbol{\tau}_T$ is a $T \times 1$ vector on ones.*

Assumption 4 *(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 $\mathbf{\Sigma}_{\beta\beta}$, defined by*

$$\lim_{n\to\infty}\left(n^{-1}\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\right) = \mathbf{\Sigma}_{\beta\beta}, \tag{21}$$

is positive definite, where $\mathbf{M}_n = \mathbf{I}_n - n^{-1}\boldsymbol{\tau}_n\boldsymbol{\tau}_n'$, and $\boldsymbol{\tau}_n$ is an $n \times 1$ vector of ones.

Part (a) of Assumption 1 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 2 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 $\alpha_\eta = 0$.[6] Note also that the above assumptions do allow for correlations between pricing errors and the factor loadings.

Assumptions 3 and 4 are also standard in the literature.

Remark 2 *Assumption 4 can be relaxed if we were willing to settle for a slower rate of convergence, and the factor strengths, α_k for factors f_{tk}, $k = 1, 2, ..., K$ are known, then condition (21) can be further relaxed by requiring that $\lim_{n\to\infty}(\mathbf{D}_n\mathbf{B}_n'\mathbf{M}_n\mathbf{B}_n\mathbf{D}_n)$ is positive definite where $\mathbf{D}_n$ is a $K \times K$ diagonal matrix with elements $n^{-\alpha_k/2}$, for $k = 1, 2, ..., K$. See Pesaran and Smith (2021).*

Remark 3 *The asymptotic covariance matrix of factor loadings, $\mathbf{\Sigma}_{\beta\beta}$, defined by (21) is positive definite only if all the factors are strong, namely if $\alpha_k = 1$ for all $k = 1, 2, ..., K$.*

[6] This assumption is discussed further in remarks 5 and 6 of Pesaran and Smith (2021).

3.2 Identification using portfolios with known loadings

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 $\mathbf{w}_p = (w_{1p}, w_{2p}, ..., w_{np})'$, and consider P return portfolios, $\overline{r}_{pt}$, defined by

$$\overline{r}_{pt} = \sum_{i=1}^{n} w_{ip} r_{it} = \mathbf{w}_p' \mathbf{r}_{nt}, \text{ for } p = 1, 2, ..., P. \tag{22}$$

Collecting all the portfolio weights in the $n \times P$ portfolio weights matrix $\mathbf{W}_P = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P)$, we also have $\overline{\boldsymbol{\eta}}_P = \mathbf{W}_P' \boldsymbol{\eta}_n$, and

$$\overline{\mathbf{r}}_{Pt} = \mathbf{W}_P' \mathbf{r}_{nt}, \tag{23}$$

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

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

Assumption 5 *(Portfolio weights) The portfolio weights, w_{ip}, for $i = 1, 2, .., n; p = 1, 2, ..., P$ satisfy the following conditions*

$$\textit{(a):} \ \sum_{i=1}^{n} w_{ip} = 1, \ \textit{(b):} \ \sup_{p,n} \sum_{i=1}^{n} |w_{ip}| < C, \ \textit{and (c):} \ \sup_{i,P} \sum_{p=1}^{P} |w_{ip}| < C. \tag{24}$$

Remark 4 *The normalization restriction, $\sum_{i=1}^{n} w_{ip} = 1$, is made for convenience and is not necessary and other choices such as $\sum_{i=1}^{n} w_i^p = 0$, can also be entertained. Short sales ($w_i^p < 0$) are allowed, and it is easily verified that the Assumption 5 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 assumption follows from the normalization condition if $w_i \geq 0$. 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 $\mathbf{W}_P$, namely $\|\mathbf{W}_P\|_1 < C$ and $\|\mathbf{W}_P\|_\infty < C$.*

Remark 5 *For the purpose of 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, $\|\mathbf{W}_P\|$, varies with the number of securities included in each portfolio. In the case of fully diversified portfolios we require that $\|\mathbf{W}_P\| = \ominus \left(n^{-1/2}\right)$, and for non-diversified portfolios we will assume that $\|\mathbf{W}_P\| = \ominus \left(m^{-1/2}\right)$ 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 $\mathbf{w}_p' \mathbf{w}_{p'} = 0$ for all $p \neq p'$, and $\mathbf{w}_p' \mathbf{w}_p = 1/m$, where m is the integer part of n/P, and $\|\mathbf{W}_P\| = m^{-1/2}$. 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 and individual securities coincide.*

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

$$\overline{r}_{p,t+1} - r_t^f = \overline{\mathrm{a}}_p + \overline{\boldsymbol{\beta}}_p' \mathbf{f}_{t+1} + \overline{u}_{p,t+1}, \text{ for } p = 1, 2, ..., P, \tag{25}$$

where

$$\overline{\mathrm{a}}_p = \sum_{i=1}^{n} w_{ip}\mathrm{a}_i,\ \overline{\boldsymbol{\beta}}_p = \sum_{i=1}^{n} w_{pi}\boldsymbol{\beta}_i, \text{ and } \overline{u}_{p,t+1} = \sum_{i=1}^{n} w_{ip}u_{i,t+1}. \tag{26}$$

The APT equilibrium condition for portfolios, corresponding to (11), is given by

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

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

$$\bar{\boldsymbol{\mu}}_P = r^f\boldsymbol{\tau}_P + \overline{\boldsymbol{B}}_P\boldsymbol{\lambda} + \bar{\boldsymbol{\eta}}_P,$$

where $\bar{\boldsymbol{\mu}}_P = (\bar{\mu}_1, \bar{\mu}_2, ..., \bar{\mu}_P)'$, $\overline{\boldsymbol{B}}_p' = \left(\overline{\boldsymbol{\beta}}_1, \overline{\boldsymbol{\beta}}_2, ..., \overline{\boldsymbol{\beta}}_P\right)$, $\bar{\boldsymbol{\eta}}_P = (\bar{\eta}_1, \bar{\eta}_2, ..., \bar{\eta}_P)'$. To identify $\boldsymbol{\lambda}$ using the portfolio return equations it is now required that

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

where $\mathbf{M}_P = \mathbf{I}_P - P^{-1}\boldsymbol{\tau}_P\boldsymbol{\tau}_P'$, and $\boldsymbol{\tau}_P$ is a $P \times 1$ vector of ones. Note that

$$\overline{\boldsymbol{\beta}}_p = \sum_{i=1}^{n} w_{ip}\boldsymbol{\beta}_i = \boldsymbol{B}_n'\mathbf{w}_p, \quad \bar{\eta}_p = \sum_{i=1}^{n} w_{ip}\eta_i = \mathbf{w}_p'\boldsymbol{\eta}_n,$$

and

$$\begin{aligned} \overline{\boldsymbol{B}}_P' &= \boldsymbol{B}_n'\left(\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P\right) = \boldsymbol{B}_n'\mathbf{W}_P, \\ \bar{\boldsymbol{\eta}}_P &= \mathbf{W}_P'\boldsymbol{\eta}_n. \end{aligned} \tag{27}$$

Now write the identification conditions when portfolio returns are used as

$$P^{-1}\left(\overline{\boldsymbol{B}}_P'\mathbf{M}_P\overline{\boldsymbol{B}}_P\right) = P^{-1}\left(\boldsymbol{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}_n\right) > \mathbf{0}, \tag{28}$$

and

$$P^{-1}\left(\boldsymbol{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\eta_n\right) \to_p \mathbf{0}. \tag{29}$$

Using portfolios does not relax the identification condition but requires that the portfolio weights are such that $\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'$ is a full rank matrix. Factors must be strong 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 $\boldsymbol{\lambda}$, suppose that $n^{-1}\boldsymbol{B}_n'\mathbf{B}_n \to \mathbf{0}$, as $n \to \infty$, and hence $\boldsymbol{\lambda}$ cannot be identified at the standard $\sqrt{n}$ rate using individual securities. Consider the limiting properties of $P^{-1}\left(\overline{\boldsymbol{B}}_P'\mathbf{M}_P\overline{\boldsymbol{B}}_P\right)$ given by (28), and note that[7]

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

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

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

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

and $P^{-1}\left\|\overline{\boldsymbol{B}}_P'\mathbf{M}_P\overline{\boldsymbol{B}}_P\right\| \to 0$ if $n^{-1}\boldsymbol{B}_n'\mathbf{B}_n \to \mathbf{0}$. The same result follows in the case of fully diversified portfolios where P is fixed and $\|\mathbf{W}_P\|^2 = \ominus\left(\frac{1}{n}\right)$. Condition (29) matches an associated condition for individual securities. Using (29) note that

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

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

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

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

$$\left\|n^{-1/2}\boldsymbol{\eta}_n\right\|^2 = n^{-1}\boldsymbol{\eta}_n'\boldsymbol{\eta}_n = n^{-1}\sum_{i=1}^{n}\eta_i^2 = O(n^{\alpha_\eta - 1}),$$

and hence $P^{-1}\left\|\boldsymbol{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\eta_n\right\| \to 0$, if $\alpha_\eta < 1$, 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, β_{ik}, are known the factor risk premia could only be identified at the standard $\sqrt{n}$ if the factors are strong, $\alpha_k = 1$ such that $\sum_{i=1}^{n}(\beta_{ik} - \overline{\beta}_k)^2 = \ominus(n)$. In practice the factor loadings must be estimated and then additional restrictions are required. In what follows we derive the finite T, large n, bias of two-pass estimators of risk premia when portfolio returns are used. But for the purpose of comparison of risk premia estimates based on individual securities and portfolio returns we first consider the estimation based on individual securities discussed in the companion paper by Pesaran and Smith (2021).

4.1 Using individual security returns

Stacking the returns on individual securities by time we have

$$\mathbf{r}_{nt} = \mathbf{a}_n + \boldsymbol{B}_n\mathbf{f}_t + \mathbf{u}_{nt},\ \text{for } t = 1, 2, ..., T, \tag{30}$$

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 $\mathbf{u}_{nt} = (u_{1t}, u_{2t}, ..., u_{nt})'$. Stacking the return equations by individual securities we have

$$\mathbf{r}_{i\circ} = \mathrm{a}_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{31}$$

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 risk premia, $\boldsymbol{\lambda}$, are defined by the cross section regressions (CSR)

$$E(r_{it}) = \lambda_0 + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \eta_i, \text{ for } i = 1, 2, ..., n, \tag{32}$$

where η_i is the pricing error.

The two-pass estimator of risk premia, $\boldsymbol{\lambda}$, based on individual returns is given by[8]

$$\hat{\boldsymbol{\lambda}}_n = \left(\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\hat{\mathbf{B}}_{nT}\right)^{-1}\hat{\mathbf{B}}_{nT}'\mathbf{M}_n\bar{\mathbf{r}}_n, \tag{33}$$

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

$$\hat{\boldsymbol{\beta}}_{i,T} = \left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{34}$$

$\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}_T'$, and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under (31), $\hat{\boldsymbol{\beta}}_{i,T} = \boldsymbol{\beta}_i + \left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}$, and hence

$$\hat{\mathbf{B}}_{nT} = \boldsymbol{B}_n + \mathbf{U}_n\mathbf{G}_T, \tag{35}$$

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 (31) over t for each i, we have

$$\bar{r}_{i\circ} = \mathrm{a}_i + \boldsymbol{\beta}_i'\bar{\mathbf{f}}_T + \bar{u}_{i\circ}, \text{ and } E(\bar{r}_i) = \mathrm{a}_i + \boldsymbol{\beta}_i'E\left(\bar{\mathbf{f}}_T\right), \tag{36}$$

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 (32), we have

$$\bar{\mathbf{r}}_n = \lambda_0\boldsymbol{\tau}_n + \boldsymbol{B}_n\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}} + \boldsymbol{\eta}, \tag{37}$$

where

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda} + \mathbf{d}_T, \tag{38}$$

$$\mathbf{d}_T = \bar{\mathbf{f}}_T - E\left(\bar{\mathbf{f}}_T\right) = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right], \tag{39}$$

$\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, and $\boldsymbol{\eta}$ is the $n \times 1$ vector of pricing errors.

As established in Pesaran and Smith (2021), for any fixed $T > k$ we have (as $n \to \infty$)

$$\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda} \to_p \left[\boldsymbol{\Sigma}_{\beta\beta} + \frac{\overline{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Sigma}_{\beta\beta}\mathbf{d}_T - \frac{\overline{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}\right). \tag{40}$$

where $\hat{\boldsymbol{\lambda}}_n$ is defined by (33) and

$$\overline{\sigma}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 > 0. \tag{41}$$

[8] 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.

4.2 Using portfolio returns

Consider now the estimates of $\boldsymbol{\lambda}$ based on portfolios discussed in Section 3.2. Using portfolio returns defined by (23), we assume the portfolio weights, w_{ip}, are fixed and do not depend on the factor loadings or the errors. The risk premia can be estimated either forming portfolio betas, as in (26), 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 $p = 1, 2, ..., P$. The resultant estimates will be identical. Denoting the portfolio estimate of $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\lambda}}_P$ we have

$$\hat{\lambda}_P = \left(\overline{\hat{\mathbf{B}}}_{PT}' \mathbf{M}_P \overline{\hat{\mathbf{B}}}_{PT} \right)^{-1} \left(\overline{\hat{\mathbf{B}}}_{PT}' \mathbf{M}_P \bar{\mathbf{r}}_P \right), \tag{42}$$

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 above. Using these results $\boldsymbol{\lambda}_P$ can now be written equivalently as

$$\hat{\boldsymbol{\lambda}}_P = \left(\hat{\boldsymbol{B}}_{nT}' \mathbf{W}_P \mathbf{M}_P \mathbf{W}_P' \hat{\boldsymbol{B}}_{nT} \right)^{-1} \left(\hat{\boldsymbol{B}}_{nT}' \mathbf{W}_P \mathbf{M}_P \mathbf{W}_P' \bar{\mathbf{r}}_n \right). \tag{43}$$

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 5. The asymptotic properties of $\hat{\boldsymbol{\lambda}}_P$ can now be derived using (35) and (37) in (43) under the following identification assumption:

Assumption 6 *(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*

$$\lim_{P \to \infty} \left(P^{-1} \boldsymbol{B}_n' \mathbf{W}_P \mathbf{M}_P \mathbf{W}_P' \mathbf{B}_n \right) = \boldsymbol{\Sigma}_{\beta\beta,w} > 0, \tag{44}$$

is positive definite.

Remark 6 *When portfolio weights, w_{ip}, satisfy the bounds in (24), then it is readily seen that part (b) of the above assumption follows from part (b) of Assumption 4, and it is therefore somewhat weaker. Similarly, part (a) of the above assumption follows from part (a) of Assumption 4. 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 4, and also imposes further restrictions on the portfolio weights.*

Remark 7 *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

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

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 the number of portfolios, $P \to \infty$.

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

Theorem 1 *(Small T bias of portfolio estimator of risk premia) Consider the multi-factor linear return model (30) and the associated risk premia, $\boldsymbol{\lambda}$, defined by (32), and suppose that Assumptions (1), (2), (3), and (6) hold, and $\alpha_\eta < 1$ where α_η is defined by (20). 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 5 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 $P \to \infty$)*

$$\hat{\boldsymbol{\lambda}}_P - \boldsymbol{\lambda} \to_p \left[\boldsymbol{\Sigma}_{\beta\beta,w} + \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left[\boldsymbol{\Sigma}_{\beta\beta,w}\mathbf{d}_T - \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\lambda\right]. \tag{45}$$

where $\hat{\boldsymbol{\lambda}}_n$ is defined by (33), $\mathbf{d}_T = T^{-1}\sum_{t=1}^T\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]$,

$$\boldsymbol{\Sigma}_{\beta\beta,w} = \lim_{P\to\infty}\left(\frac{\mathbf{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}_n}{P}\right), \text{ and } \bar{\omega}^2 = \lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^P\left(\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p\right) > 0, \tag{46}$$

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

A proof is provided in sub-section A.3 of the Appendix.

It is clear from the above theorem that the small T bias continues to be present when portfolio returns are used to estimate $\boldsymbol{\lambda}$. Following Shanken (1992) it is possible to construct a bias-correced version of $\hat{\boldsymbol{\lambda}}_P$. Suppose that $\bar{\omega}^2$ is known then Shanken type bias-corrected estimator of $\boldsymbol{\lambda}$ is given by

$$\hat{\boldsymbol{\lambda}}_P = \left[\frac{\hat{\mathbf{B}}_{nT}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\hat{\mathbf{B}}_{nT}}{P} - \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\hat{\boldsymbol{B}}_{nT}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\bar{\mathbf{r}}_n}{P}\right). \tag{47a}$$

Now using (A.28) and (A.29) in the Appendix we have

$$\frac{\hat{\mathbf{B}}_{nT}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\hat{\mathbf{B}}_{nT}}{P}-\frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\rightarrow_p\boldsymbol{\Sigma}_{\beta\beta,\omega},$$

$$\frac{\hat{\boldsymbol{B}}_{nT}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\bar{\mathbf{r}}_n}{P}\rightarrow_p\boldsymbol{\Sigma}_{\beta\beta,\omega}\boldsymbol{\lambda}_T^*.$$

Using these results in (47a), and assuming that all factors are strong (so that $\boldsymbol{\Sigma}_{\beta\beta,\omega}$ is full rank), we obtain

$$\tilde{\boldsymbol{\lambda}}_P\rightarrow_p\boldsymbol{\lambda}_T^*=\boldsymbol{\lambda}+\mathbf{d}_{fT},\tag{48}$$

where $\mathbf{d}_T$ is defined by (39), and $\boldsymbol{\lambda}_T^*$ is Shanken's "ex-post" risk premia. However, to implement this correction a small T unbiased (as $n\rightarrow\infty$) estimator of $\bar{\omega}^2=\lim_{P\rightarrow\infty}\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p\right)$. Since portfolio weights, $\mathbf{w}_p$, are given, small T unbiased estimator of $\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p$, requires small T unbiased estimation of $\boldsymbol{\Sigma}_u$ which does not seem possible unless u_{it} are cross-sectionally independent. In general estimation of $\boldsymbol{\Sigma}_u$ requires both n and T large even when u_{it} are weakly cross correlated. In contrast, when individual security returns are used the Shanken correction requires small T unbiased estimation of $\overline{\sigma}^2=\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$ which does not involve covariances, σ_{ij}, and can be estimated by $\widehat{\overline{\sigma}}_{nT}^2=\frac{1}{n(T-k-1)}\sum_{t=1}^{T}\sum_{i=1}^{n}\hat{u}_{it}^2$ which is shown to converge to $\overline{\sigma}^2$ for a fixed $T>k+1$ and as $n\rightarrow\infty$.

Also when the focus of analysis is $\boldsymbol{\lambda}$, whether the bias of estimating $\boldsymbol{\lambda}$ can be reduced using portfolio returns instead of individual security returns is unclear and depends in a complicated way 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 $\boldsymbol{\Sigma}_{\beta\beta,w}$. The issue is illustrated in the following example.

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 (40) and (45))*

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

Further, 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 $(n,P\rightarrow\infty)$ of the following inequality

$$\frac{\beta_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\beta_n}{P}>\left(\frac{\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p\right)}{\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2}\right)\frac{\beta_n'\mathbf{M}_n\beta_n}{n}.\tag{49}$$

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*

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

Similarly

$$\bar{\beta}_1 = \mathbf{w}'_1\boldsymbol{\beta} = m^{-1}\sum_{i=1}^{m}\beta_i,\ \bar{\beta}_2 = \mathbf{w}'_2\boldsymbol{\beta} = m^{-1}\sum_{i=m+1}^{2m}\beta_i, ..., \bar{\beta}_P = \mathbf{w}'_P\boldsymbol{\beta} = m^{-1}\sum_{i=(P-1)m+1}^{n}\beta_i, \quad (50)$$

with the sample average of $\bar{\beta}_p$ *across* p *given by*

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

Using these results we now have

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

Similarly, for the estimate of λ *based on individual securities 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* $\mathbf{w}'_p\boldsymbol{\Sigma}_u\mathbf{w}_p = m^{-2}\boldsymbol{\tau}'_m\boldsymbol{\Sigma}_{p,u}\tau_m$*, where* $\boldsymbol{\Sigma}_{p,u}$ *is the* $m \times m$ *covariance matrix of the errors of the returns included in the* p^{th} *portfolio, and*

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

It is now easily seen that $\bar{\omega}_n^2 = m^{-1}\bar{\sigma}_n^2$*, when* $\boldsymbol{\Sigma}_{p,u}$ *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 (49), 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 7)*

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

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*. Furthermore, the number of portfolios,* P*, still needs to be sufficiently large.*

5 Concluding remarks

This paper examines two questions associated with the APT. The first question is the relationship between the statistical factor model determining returns and the theoretically consistent factor model which takes account of the restrictions implied by the inter-temporal equilibrium pricing conditions. The factors included in the statistical model are priced only if they have non-zero conditional correlation with the stochastic discount factor. Pricing errors arise from

non-zero correlations between the errors in the statistical factor model and the stochastic discount factor. From a theoretical perspective, the factors used in the return regressions should be the ones that are thought to be correlated with fundamentals, represented by the stochastic discount factor.

The second question addressed in this paper is the *pros* and *cons* of using portfolios, rather than individual securities, in the estimation of risk premia. We show that conditions for identification of the risk premia are unaffected whether one uses portfolios or individual securities. In both cases, even when the true factor loadings, β_{ik}, are known, the factor risk premia can only be identified at the standard $\sqrt{n}$ rate if the factors are strong, $\alpha_k = 1$, and the pricing errors are sufficiently weak, namely if $\alpha_\eta = 0$. For general values of $\alpha_k \leq 1$ and $\alpha_\eta \geq 0$, Pesaran and Smith (2021) show the risk premia of factor f_{tk} with strength α_k can be estimated at the rate of $n^{(\alpha_k - \alpha_\eta)/2}$. Their proof is based on individual security returns, and can be shown to extend to portfolio returns that we consider in this paper.

Portfolios are used in an attempt to reduce the generated regressor bias that results from the effect of the sampling error of the estimated first stage loadings. However, as shown in this paper, the small T bias continues to be present when portfolio returns are used to estimate risk premia. Whether the bias can be reduced using portfolio returns instead of individual security returns is unclear and depends in a complicated way on the covariances of the individual securities within the portfolio. We derive the bias correction for portfolios, but whereas with individual securities the bias correction is operational when the factors are strong, this does not seem to be the case for portfolios. Again this is because the correction will depend on the covariances of the individual securities within the portfolio. In any event, if portfolios are used, the number of portfolios, P, must still be sufficiently large, which presents the investigator with a fine balance between the number of individual securities to be allocated to individual portfolios for estimation of the loadings, and the number of portfolios to be used in the second pass to estimate the risk premia.

A Mathematical Appendix

A.1 Introduction

We first state a number of lemmas that we shall then use to prove Theorem 1.

A.2 Statement and proofs of lemmas

Lemma A.1 *Consider the errors* $\{u_{it}, i = 1, 2, ..., n; t = 1, 2, ..., T\}$ *in the factor model defined by (30), and suppose that Assumption 1 holds. Then for any* t *and* t' *(as* $n \to \infty$*)*

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

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

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. See Pesaran and Smith (2021) section A.2. ■

Lemma A.2 *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 suppose that assumptions 1, 2 and part (b) of 4 hold, and* $\alpha < 1$.[9] *Then*

$$\frac{\mathbf{B}'\mathbf{M}_n\mathbf{U}}{n} \to_p \mathbf{0}, \tag{A.4}$$

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

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

$$\frac{\mathbf{U}'\mathbf{M}_n\mathbf{U}}{n} \to_p \overline{\sigma}^2\mathbf{I}_T, \tag{A.7}$$

$$\frac{\mathbf{U}'\mathbf{M}_n\overline{\mathbf{u}}}{n} \to_p \frac{\overline{\sigma}^2}{T}\boldsymbol{\tau}_T, \tag{A.8}$$

$$\frac{\mathbf{U}'\mathbf{M}_n\boldsymbol{\eta}}{n} \to_p \mathbf{0} \tag{A.9}$$

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*

[9] As compared to the notation in the body of the paper, we have dropped the subscript n from $\mathbf{B}_n$ as defined by (30).

Proof. See Pesaran and Smith (2021) section A.2. ■

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 5, 1, 2 and 4 hold, $\alpha_\eta<1$, and $\|\mathbf{W}_P\|=\ominus\left(m^{-1/2}\right)$. Then for a fixed m, k and T, and as $P\to\infty$, such that $P/n\to\pi$, $(0<\pi<1)$, then we have*

$$\frac{\mathbf{U}'\mathbf{W}_P\tau_P}{P}\to_p\mathbf{0}, \tag{A.10}$$

$$\frac{\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{U}}{P}\to_p\mathbf{0}, \tag{A.11}$$

$$\frac{\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\overline{\mathbf{u}}}{P}\to_p\mathbf{0}, \tag{A.12}$$

$$\frac{\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\eta}}{P}\to_p\mathbf{0}, \tag{A.13}$$

$$\frac{\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{\eta}}{P}\to_p\mathbf{0}, \tag{A.14}$$

$$\frac{\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{U}}{P}\to_p\bar{\omega}^2\mathbf{I}_T, \tag{A.15}$$

$$\frac{\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\overline{\mathbf{u}}}{P}\to_p\frac{\bar{\omega}^2}{T}\boldsymbol{\tau}_T, \tag{A.16}$$

where $\mathbf{M}_P=\mathbf{I}_P-\frac{1}{P}\boldsymbol{\tau}_P\boldsymbol{\tau}'_P$, $\overline{\mathbf{u}}=(\overline{u}_{1\circ},\overline{u}_{2\circ},...,\overline{u}_{n\circ})'$, $\overline{u}_{i\circ}=T^{-1}\sum_{t=1}^T u_{it}$, $\bar{\omega}^2=\lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^P\left(\mathbf{w}'_p\boldsymbol{\Sigma}_u\mathbf{w}_p\right)$, 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.10) 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

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

which tends to zero as $P\to\infty$, since under Assumptions 5 and 1, $\sup_{i,P}|\bar{w}_{iP}|<C$, 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.11 and note that

$$P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\boldsymbol{U}=P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{W}'_P\,\boldsymbol{U}-\left(P^{-1}\mathbf{B}'\mathbf{W}_P\tau_P\right)\left(P^{-1}\tau'_P\mathbf{W}_P\boldsymbol{U}\right), \tag{A.17}$$

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 $\bar{\boldsymbol{\beta}}_p = \sum_{i=1}^{n} \bar{w}_{ip}\boldsymbol{\beta}_i$, and by Assumption 6 $\sup_p \left\|\bar{\boldsymbol{\beta}}_p\right\| < C$. 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.10), it follows that

$$P^{-2}\mathbf{B}'\mathbf{W}_P\tau_P\tau_P'\mathbf{W}_P\ \boldsymbol{U} \to_p \mathbf{0}. \tag{A.18}$$

The first term of (A.17) 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

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

Also $\left|\phi_{is,P}\right| \leq \sup_{s,p}\left|\bar{\beta}_{sp}\right|\sum_{p=1}^{P}\left|w_{ip}\right| < C$ and $\sup_i\sum_{j=1}^{n}\left|\sigma_{ij}\right| < C$, by Assumptions 5, 1,and 6. 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.18) in (A.17) now establishes (A.11). To prove (A.12) we first note that since $\overline{\mathbf{u}} = \left(\overline{u}_{1\circ}, \overline{u}_{2\circ}, ..., \overline{u}_{n\circ}\right)' = T^{-1}\boldsymbol{U}\tau_T$, where $\overline{u}_{i\circ} = T^{-1}\sum_{t=1}^{T}u_{it}$, and hence $\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\overline{\mathbf{u}} = T^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\boldsymbol{U}\tau_T$, and

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

and tends to zero in probability by virtue of result (A.11). To prove (A.13) we note that

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

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 6, and

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

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.14) follows similarly. To establish (A.15), in view of (A.10) 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

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

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

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

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.16) follows from (A.15), noting that $\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\overline{\mathbf{u}} = T^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}_P'\boldsymbol{U}\tau_T$. ■

A.3 Proof of theorem 1

We first present some definitions for the case using individual securities. Consider the two-pass estimator of $\boldsymbol{\lambda}$ defined by (33), and to simplify notations, write it as

$$\hat{\boldsymbol{\lambda}}_n = \left(\frac{\hat{\mathbf{B}}'\mathbf{M}_n\hat{\mathbf{B}}}{n}\right)^{-1}\left(\frac{\hat{\mathbf{B}}'\mathbf{M}_n\overline{\mathbf{r}}}{n}\right), \tag{A.19}$$

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

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{r}_{i\circ}, \tag{A.20}$$

and $\mathbf{r}_{i\circ} = (r_{i1},r_{i2},...,r_{iT})'$. Under the factor model (30)

$$\mathbf{r}_{i\circ} = \alpha_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{A.21}$$

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

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}. \tag{A.22}$$

Stacking these results over i yields:

$$\hat{\mathbf{B}} = \mathbf{B} + \mathbf{U}\mathbf{G}_T \tag{A.23}$$

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

$$\mathbf{G}_T = \mathbf{M}_T\mathbf{F}(\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} \tag{A.24}$$

Also using result (37) in the paper we have (in terms of the simplified notations used here)

$$\bar{\mathbf{r}} = \lambda_0 \boldsymbol{\tau}_n + \mathbf{B}\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}} + \boldsymbol{\eta} \tag{A.25}$$

where

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda} + \mathbf{d}_T, \text{ and } \mathbf{d}_T = \bar{\mathbf{f}}_T - E\left(\bar{\mathbf{f}}_T\right). \tag{A.26}$$

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

Consider the portfolio estimator $\boldsymbol{\lambda}$ given by (43) and write it simply as

$$\hat{\boldsymbol{\lambda}}_P = \left(P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\hat{\boldsymbol{B}}\right)^{-1}\left(P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\bar{\mathbf{r}}\right), \tag{A.27}$$

Substituting $\hat{\boldsymbol{B}}$ and $\bar{\mathbf{r}}$ using (A.23) and (A.25) 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

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

and recall that $\boldsymbol{\lambda}_T^*$ is defined by (A.26). 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 1, 2 and 6, and using the results of Lemma A.3, we have (as $P \to \infty$, for a fixed m, T and k):

$$P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\hat{\boldsymbol{B}} \to_p \boldsymbol{\Sigma}_{\beta\beta,\omega} + \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}, \tag{A.28}$$

$$P^{-1}\hat{\boldsymbol{B}}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\bar{\mathbf{r}} \to_p \boldsymbol{\Sigma}_{\beta\beta,\omega}\boldsymbol{\lambda}_T^*, \tag{A.29}$$

where $\bar{\omega}^2$ and $\boldsymbol{\Sigma}_{\beta\beta,\omega}$ are defined by (46). Result (45) then follows by using the above in (A.27), and writing the outcome in terms of $\hat{\boldsymbol{\lambda}}_P - \boldsymbol{\lambda}$.

REFERENCES

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

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

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

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

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

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

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

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

Pesaran, M. H. and R. P. Smith (2021) Factor Strengths, Pricing Errors, and Estimation of Risk Premia, CESifo Working Paper, 8947.

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

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

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

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