Arbitrage pricing theory, the stochastic discount factor and estimation of risk premia from portfolios*

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

Ron P. Smith
Birkbeck, University of London

October 6, 2021

Abstract

The arbitrage pricing theory (APT) attributes differences in expected returns to exposure to systematic risk factors. Two aspects of the APT are considered. Firstly, the factors in the statistical asset pricing model are related to a theoretically consistent set of factors defined by their conditional covariation with the stochastic discount factor (SDF) used to price securities within inter-temporal asset pricing models. It is shown that risk premia arise from non-zero correlation of observed factors with SDF and the pricing errors arise from the correlation of the errors in the statistical model with SDF. Secondly, the estimates of factor risk premia using portfolios are compared to those obtained using individual securities. It is shown that in the presence of pricing errors consistent estimation of risk premia requires a large number of not fully diversified portfolios. Also, in general, it is not possible to rank estimators using individual securities and portfolios in terms of their small sample bias.

JEL Classifications: C38, G12

Key Words: Arbitrage Pricing Theory, Stochastic Discount Factor, portfolios, factor strength, identification of risk premia, two-pass regressions, Fama-MacBeth.

*We are grateful to the editor and two anonymous reviewers for very useful comments on an earlier version. 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 addresses two related issues that arise in the analyses of arbitrage pricing theory (APT) in finance. Firstly, APT is generally regarded as an empirical model and there is a question as to how the risk factors in the statistical APT model relate to the theoretical model of inter-temporal asset pricing in which a stochastic discount factor (SDF) represents the fundamentals. Secondly, estimation of risk premia is typically done using portfolios, and there is a question as to whether using portfolio returns, rather than individual security returns, is likely to result in more precise estimates of risk premia.

The APT, as 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 risk premia are usually estimated using a two-pass procedure suggested by Fama and MacBeth (1973, FM). In the first step the linear statistical factor model is estimated by running regressions of returns on each security or portfolio on K observed risk factors, f_{kt}, $k = 1, 2, ..., K$, $t = 1, 2, ..., T$. There is no shortage of suggested factors. The best known are the Fama-French three, market, value and size, or the five given in Fama and French (2015), but many more have been suggested. Harvey and Liu (2019) document a "factor zoo" of over 400 potential factors. The first pass regression is used to estimate the factor loadings, β_{ik}, which are assumed to be stable over the given sample period. The second pass is a cross section regression of average returns on the estimated factor loadings, $\widehat{\beta}_{ik}$, the coefficients of which are the risk premia, λ_k, for factor f_{kt}, which are then used to price the factors.[2] The properties of the two-pass estimators are investigated typically assuming zero pricing errors.

In this paper we provide a link between pricing errors and the estimation of risk premia, both when individual or portfolio returns are used. We show that pricing errors are still present even if portfolio returns are used. To this end we first consider how the factors in the statistical factor model are related to a theoretically consistent set of factors defined by their conditional covariation with the SDF used to price securities within inter-temporal asset pricing models. We show that a risk factor is priced only if its conditional covariance with the SDF is non-zero. In contrast, pricing errors arise when there is non-zero correlations between the idiosyncratic errors of asset returns and the SDF. Pricing errors correspond to persistent anomalies, unlike the mean zero random errors. The APT theory places bounds on the pricing errors, requiring them to be square summable.

Secondly, we compare the FM estimators of risk premia based on individual security and portfolio returns. Following the pioneering contribution of Fama and MacBeth, it is conventional in this literature to use mean returns and loadings for a relatively small number of portfolios P ($P < n$) formed from the underlying securities in the second pass regression,

[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.

rather than the individual securities themselves. It is argued that the sampling errors in the estimates of the first stage β_{ik} using individual security returns 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. We consider a wide variety of portfolio weights, both fully diversified and non-diversified ones. We begin by assuming known factor loadings. This allows us to focus on the estimation of risk premia in the second pass regression, without the complications arising from the first pass estimation of factor loadings, β_{ik}. We show that to eliminate the effects of pricing errors on estimation of risk premia a large number of securities must be considered, irrespective of whether individual security or portfolio returns are used. We also derive conditions on portfolio weights for consistent estimation of risk premia, and by means of a simple example we show that these conditions need not be satisfied when fully-diversified portfolios are used.

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. The small T case is relevant since factor loadings do not appear to be constant over time, hence the common practice of using rolling regressions on a relative short time period to estimate them. We obtain an expression for the small T bias of risk premia using portfolios corresponding to a similar result obtained by Shanken (1992) for the individual security returns. But for portfolios the small T bias depends on the portfolio weights as well as the error covariances, rather than the simple average of the error variances as in the Shanken case. A comparison of an estimator of the risk premia based on portfolios with the one based on individual securities shows that in general no clear cut ranking of the bias 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.

In this paper we follow the literature in assuming that observed factors are strong, which is necessary for $\sqrt{n}$-consistent estimation of the risk premia. The effect of having factors that are not strong is considered in a companion paper, Pesaran and Smith (2021), which uses a measure of factor strength proposed in Bailey, Kapetanios and Pesaran (2021).

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 in order to derive the APT risk premia and pricing errors. Section 3 considers how portfolios are formed. Section 4 sets out the theory consistent model for portfolios and considers the estimation of the risk premia for the factors from a cross section when the factor loadings are known. Section 5 analyses the effect of using portfolios when the factor loadings are unknown and provides a Shanken type bias correction formula. 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 $\sup_{n \geq n_0} (f_n/g_n) \leq C < \infty$.

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

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

2.1 Statistical and theory consistent factor models

Suppose that at time t there are n individual securities with returns, $r_{i,t+1}$, generated by a linear factor pricing model (LFPM) of the form:

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

where r_t^f is the risk free rate; a_{it}, for $i = 1, 2, ..., n_t$ are the intercepts in the factor model; $f_{k,t+1}$, $k = 1, 2, ..., K$ are the observed common factors with associated factor loadings, $\beta_{ik,t}$. The error $u_{i,t+1}$ is a mean zero serially uncorrelated idiosyncratic component of returns.[3] The model can be written more compactly as

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

where $\boldsymbol{\beta}_{it} = (\beta_{i1,t}, \beta_{i2,t}, ..., \beta_{iK,t})'$, 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 factor model. Under standard inter-temporal equilibrium asset pricing the equilibrium price for security i, P_{it}, is equal to the expected discounted value of the payoff, future price plus dividends, namely

$$P_{it} = E_t \left[m_{t+1}(P_{i,t+1} + D_{i,t+1}) \right], \tag{3}$$

where m_{t+1} is the stochastic discount factor (SDF) used to price all assets in the market, and $E_t(\circ)$ stands for a conditional expectations with respect to the information set, $\mathcal{I}_t = \left\{ r_\tau^f, \boldsymbol{\beta}_{i\tau}, \mathbf{f}_\tau, m_\tau, \text{ for } \tau = t, t-1, ... \right\}$, assumed to be common across all traders. A common information set is a strong assumption but is needed for the existence of equilibrium across security markets, at each moment of time, so that all securities can be priced in terms of the same SDF, m_{t+1}.

Denoting the holding period return by $r_{i,t+1} = \left(\Delta P_{i,t+1} + D_{i,t+1} \right) / P_{it}$, (3) can be written

$$1 = E_t \left[m_{t+1}(1 + r_{i,t+1}) \right]. \tag{4}$$

[3] 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.

Writing a similar condition for the risk free rate r_t^f we have

$$1 = E_t\left[m_{t+1}\left(1 + r_t^f\right)\right], \tag{5}$$

and subtracting (5) from (4) gives the inter-temporal equilibrium pricing condition

$$E_t\left[m_{t+1}(r_{i,t+1} - r_t^f)\right] = 0, \tag{6}$$

where

$$E_t(m_{t+1}) = 1/(1 + r_t^f) > 0. \tag{7}$$

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 (6), to give

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

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

$$\mathrm{a}_{it} = -\frac{\boldsymbol{\beta}_{it}' E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})} - \frac{E_t(m_{t+1}u_{i,t+1})}{E_t(m_{t+1})}. \tag{8}$$

Imposing this restriction by substituting (8) back into (1) yields the following theory consistent factor model

$$\textit{Theory consistent factor model: } r_{i,t+1} - r_t^f = \boldsymbol{\beta}_{it}'\mathbf{g}_{t+1} + \eta_{it} + u_{i,t+1}, \tag{9}$$

where $\mathbf{g}_{t+1}$ is the $K \times 1$ vector of theory-consistent factors defined as

$$\mathbf{g}_{t+1} = \mathbf{f}_{t+1} - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})}, \tag{10}$$

with associated pricing errors given by

$$\eta_{it} = -\frac{E_t(m_{t+1}u_{i,t+1})}{E_t(m_{t+1})} = -\frac{Cov_t(m_{t+1}, u_{i,t+1})}{E_t(m_{t+1})}. \tag{11}$$

To relate $\mathbf{g}_{t+1}$and η_{it} to the APT conditions we note that under APT (See Ross (1976))

$$E_t\left(r_{i,t+1} - r_t^f\right) = \boldsymbol{\beta}_{it}'\boldsymbol{\lambda}_t + \eta_{it}, \tag{12}$$

where $\boldsymbol{\lambda}_t$ is the $K \times 1$ vector of risk premia. Now taking conditional expectations of (9) and comparing the results with the APT condition we have[4]

$$\boldsymbol{\lambda}_t = E_t\left(\mathbf{g}_{t+1}\right) = E_t\left(\mathbf{f}_{t+1}\right) - \frac{E_t\left(m_{t+1}\mathbf{f}_{t+1}\right)}{E_t(m_{t+1})} = \boldsymbol{\mu}_t + \boldsymbol{\phi}_t, \tag{13}$$

where

$$\boldsymbol{\phi}_t = -\frac{E_t(m_{t+1}\mathbf{f}_{t+1})}{E_t(m_{t+1})}. \tag{14}$$

[4]Ross (1976) assumes that all factors are measured as deviations from their population means, and therefore implicitly assumes that $\boldsymbol{\lambda}_t = \boldsymbol{\phi}_t$.

It is worth noting that unlike $\mathbf{f}_{t+1}$, the theory-consistent factors, $\mathbf{g}_{t+1}$, are unobserved and can be identified only in terms of a given SDF, m_{t+1}. Further, the risk premia, $\boldsymbol{\lambda}_t$, is composed of $\boldsymbol{\mu}_t$ which is the conditional mean of the observed factors and does not depend on the SDF, m_{t+1}, and a second component, $\boldsymbol{\phi}_t$, that does depend on m_{t+1}. Furthermore, $\boldsymbol{\phi}_t$ can also be identified from cross section regressions of a_{it} on $\boldsymbol{\beta}_i$. This follows by noting that (8) can be written equivalently as

$$\mathrm{a}_{it} = \boldsymbol{\beta}_{it}' \boldsymbol{\phi}_t + \eta_{it}. \tag{15}$$

Seen from this perspective $\boldsymbol{\phi}_t$ can be regarded as the vector of alpha risk premia.

Our derivation also shows how pricing errors, η_{it}, in the APT condition arise from possible correlations between the SDF, m_{t+1}, and the idiosyncratic errors, $u_{i,t+1}$, across the individual securities, $i = 1, 2, ..., n$. In the context of misspecified factor models, η_{it} could also capture missing factors that are correlated with m_{t+1}.

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 linear factor pricing model (1) subject to the inter-temporal equilibrium conditions in (6). Then the vector of risk premia, $\boldsymbol{\lambda}_t$, is composed of $\boldsymbol{\mu}_t$, representing the returns on observed factors, and a vector of alpha risk premia given by (14). Namely $\boldsymbol{\lambda}_t = \boldsymbol{\mu}_t + \boldsymbol{\phi}_t$.*

By construction the theory consistent factor model in (9) satisfies the equilibrium pricing condition (6) 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 (10) the theory consistent factors, $\mathbf{g}_t$, can be written equivalently as:

$$\mathbf{g}_{t+1} = \mathbf{f}_{t+1} - E_t\left(\mathbf{f}_{t+1}\right) - \frac{Cov_t(m_{t+1}, \mathbf{f}_{t+1})}{E_t(m_{t+1})}. \tag{16}$$

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

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

Therefore, the statistical factor $f_{k,t+1}$ has a non-zero conditional risk premium if it is correlated with the SDF. A simple example of such a factor is consumption growth illustrated briefly in what follows:

Example 1 *To illustrate the derivation of the theory consistent factor model, consider the case of the SDF which comes from the familiar consumption based asset pricing model. In this model investor's utility is defined over current and discounted expected future consumption as*

$$U(C_t, C_{t+1}) = u(C_t) + \rho E_t\left[u(C_{t+1})\right],$$

where ρ is the subjective discount rate. The investor can buy or sell a security at price P_t with payoff $X_{t+1} = P_{t+1} + D_{t+1}$, where as before D_t is dividend. The consumer maximizes $E_t\left[U(C_t, C_{t+1})\right]$ subject to a budget constraint. The first order condition for this optimization problem is given by

$$P_t u'(C_t) = \rho E_t\left[u'(C_{t+1}) X_{t+1}\right],$$

which, corresponding to (4), and can be re-written in terms of the SDF, $m_{t+1} = \rho u'(C_{t+1})/u'(C_t)$, *as*

$$1 = E_t\left[m_{t+1}\left(1 + r_{t+1}\right)\right],$$

where $r_{t+1} = (\Delta P_{t+1} + D_{t+1})/P_t$. *Assuming the investor has a power utility,* $u(C_t) = \left(C_t^{1-\kappa} - 1\right)/(1-\kappa)$, *where* $\kappa > 0$ *is the coefficient of relative risk aversion 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 hence $\boldsymbol{\lambda}_t = \boldsymbol{\mu}_t + \boldsymbol{\phi}_t$, *where* $\boldsymbol{\mu}_t = E_t\left(\mathbf{f}_{t+1}\right)$, *and* $\boldsymbol{\phi}_t = -E_t\left(e^{-\kappa \Delta c_{t+1}} \mathbf{f}_{t+1}\right)/E_t(e^{-\kappa \Delta c_{t+1}})$. *The vector of risk premia,* $\boldsymbol{\lambda}_t$ *can be written equivalently as*

$$\boldsymbol{\lambda}_t = -\frac{Cov_t(e^{-\kappa \Delta c_{t+1}}, \mathbf{f}_{t+1})}{E_t(e^{-\kappa \Delta c_{t+1}})}.$$

In this application the risk premia does not depend on the subjective discount rate, ρ, *and is non-zero only if the risk factor is correlated with consumption growth.*

2.2 Pricing errors

From (11), the pricing errors in the theory consistent factor model is

$$\eta_{it} = -\frac{E_t(m_{t+1} u_{i,t+1})}{E_t(m_{t+1})}. \tag{17}$$

Ross (1976) did not allow for time variation, but applying his condition (18) that requires the pricing errors to be bounded gives

$$\sum_{i=1}^{n} \eta_{it}^2 < C. \tag{18}$$

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

$$u_{i,t+1} = \psi_i m_{t+1} + \varepsilon_{i,t+1}. \tag{19}$$

Thus using (19) in (11)

$$\eta_{it} = -\psi_i \theta_t, \tag{20}$$

where $\theta_t = E_t\left(m_{t+1}^2\right)/E_t(m_{t+1}) > 0$. The pricing errors are given anomalies. These appear where the price of an individual security is inconsistent with that implied by the asset pricing model, creating a return predictor that may persist for some time. This is different from the random errors, which have conditional expectations of zero.

Pricing errors only arise if the $u_{i,t+1}$ and the stochastic discount factor m_{t+1} are conditionally correlated and $\psi_i \neq 0$ for some i. Thus (18) becomes

$$\sum_{i=1}^{n} \eta_{it}^2 = \theta_t^2 \left(\sum_{i=1}^{n} \psi_i^2\right). \tag{21}$$

The strength of the pricing errors depends on their degree of pervasiveness, namely the rate at which $\sum_{i=1}^{n}\psi_i^2$ rises with n. The APT condition requires that $\sum_{i=1}^{n}\psi_i^2 < C$. The idiosyncratic errors, $\varepsilon_{i,t+1}$, although uncorrelated with m_{t+1}, could be cross-sectionally correlated, due to non-fundamental common factors uncorrelated with the stochastic discount factor that arise from herding behaviour or correlated beliefs, for instance at times of financial crisis. Alternatively it might arise from weak spatial correlations arising not from common factors but from local network effects for instance between firms in the same industry.

The above analysis highlights the importance of distinguishing between the 'statistical factor model' given by (2), and the 'theory consistent factor model' given by (9). 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 will aggregate the theory consistent factor model, (9), for individual securities to a corresponding theory consistent relation for portfolios. But first we examine how the portfolios are constructed from the individual securities before returning to the issue of whether one should use individual securities or portfolios to identify and estimate the vector of risk premia $\boldsymbol{\lambda}_t$ or $\boldsymbol{\phi}_t$.

3 The formation of return portfolios

The debate over whether it is better to use returns on portfolios or on individual securities to estimate risk premia is an old one. If one uses portfolios there is also the additional issue of how such portfolios are to be formed. In the case of a single factor model (CAPM), Fama and MacBeth (1973, p615) propose forming a small number of portfolios (say 20) based on ranked beta estimates for individual securities, and to minimize any biases arising from such a procedure, they suggest using betas estimated on an initial training sample to form return portfolios in a subsequent estimation sample over which the risk premia are estimated using their two-step method. The construction of portfolios becomes more complex when the return regressions contain more than one factor.[5]

Fama and MacBeth argued that betas of portfolio returns can be more precisely estimated as compared to the estimates obtained using individual security returns. However, even if this is true it does not necessarily follow that risk premia based on portfolio betas will be more precisely estimated. Furthermore, pricing errors continue to be an issue for portfolio returns which also need to be taken into account. A formal statistical analysis is clearly required to establish conditions under which risk premia are better estimated with portfolio returns. Ang,

[5] For instance, Fama and French (2015) for their Table 1, Panel A, sort the individual securities into five size groups and five book to market (B/M) groups, giving 25 separate, mutually exclusive, portfolios. However, with four characteristics - Size, B/M, operating profitability (OP) and investment - they comment that even $3 \times 3 \times 3 \times 3$ sorts, produce 81 poorly diversified portfolios that have low power in tests of asset pricing models. They compromise with sorts on size and pairs of the other three variables. They form two Size groups (small and big), using the median market cap for NYSE stocks as the breakpoint, and use NYSE quartiles to form four groups for each of the other two sort variables. For each combination of variables there are $2 \times 4 \times 4 = 32$ portfolios, but correlations between characteristics cause an uneven allocation of stocks. For example, B/M and OP are negatively correlated, especially among big stocks, so portfolios of stocks with high B/M and high OP can be poorly diversified. In fact, when they sort stocks independently on Size, B/M, and OP, the portfolio of big stocks in the highest quartiles of B/M and OP is often empty before July 1974. They then discuss how they spread stocks more evenly in the $2 \times 4 \times 4$ sorts.

Liu and Schwarz (2020) provide a survey of the issues involved.

For a formal analysis a central issue is the choice of portfolio weights, w_{ip}. Here we consider two types of portfolios: (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, $\bar{r}_{pt}$, defined by

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

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 have

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

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

Denote by m the maximum number of securities included in a single portfolio. 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, when n and $P \to \infty$, such that $n/P \to m \geq 1$. When $m = 1$ the number of portfolio returns and individual security returns coincide ($P = n$).

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 $\|\mathbf{W}_P\| = \lambda_{max}^{1/2}(\mathbf{W}_P' \mathbf{W}_P) = \ominus(n^{-1/2})$. 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 1 *(Portfolio weights) The portfolio weights, w_{ip}, for $i = 1, 2, .., n; p = 1, 2, ..., P$ satisfy the following conditions*

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

and

$$\text{(d): } \lambda_{min}(\mathbf{W}_P \mathbf{M}_P \mathbf{W}_P') > c > 0. \tag{25}$$

Remark 1 *The normalization restriction, $\sum_{i=1}^{n} w_{ip} = 1$, is made for convenience and is not necessary and other choices such as $\sum_{i=1}^{n} w_{ip} = 0$, can also be entertained. Short sales ($w_{ip} < 0$) are allowed, and it is easily verified that the Assumption 1 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 $w_{ip} \geq 0$. 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 $\|\mathbf{W}_P\|_\infty < C$. Condition (d) is required for identification of risk premia using portfolio returns, and, for example, rules out the use of linearly dependent portfolio weights when forming portfolios.*

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 noted above, 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.

4 Equilibrium conditions and estimation of risk premia for portfolios

This section, aggregates the theory consistent factor model, (9), for individual securities to a corresponding theory consistent relation for portfolios, and then considers the issue of whether one should use individual securities or portfolios to estimate the vector of risk premia $\boldsymbol{\lambda}_t$ or $\boldsymbol{\phi}_t$. To clarify the central issues, it is first assumed that 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 estimation of the risk premia.

4.1 Equilibrium conditions for portfolios

We now link the earlier discussion of equilibrium conditions to the use of portfolios. Consider a return portfolio $\bar{r}_{pt} = \sum_{i=1}^{n} w_{ip} r_{it}$, represented by the vector of weights, $\mathbf{w}_p = (w_{1p}, w_{2p}, ..., w_{np})'$, where $\sum_{i=1}^{n} w_{ip} = 1$.

Using (9), we first note that

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

Then aggregating to portfolios

$$E_t\left(\sum_{i=1}^{n} w_{ip} r_{i,t+1} - \sum_{i=1}^{n} w_{ip} r_t^f\right) = \bar{\boldsymbol{\beta}}_{pt}' E_t(\mathbf{g}_{t+1}) + \bar{\eta}_{pt},$$

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

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

Hence, noting that $E_t(\mathbf{g}_{t+1}) = \boldsymbol{\lambda}_t$ we have

$$E_t\left(\bar{r}_{p,t+1} - r_t^f\right) = \bar{\boldsymbol{\beta}}_{pt}' \boldsymbol{\lambda}_t + \bar{\eta}_{pt}, \tag{26}$$

As in the case of individual securities the equilibrium condition for portfolios can be also written equivalently as

$$E_t\left(\bar{r}_{p,t+1} - r_t^f\right) = -\frac{Cov_t(\bar{r}_{p,t+1} - r_t^f, m_{t+1})}{E_t(m_{t+1})}. \tag{27}$$

4.2 Estimation of risk premia using portfolios with known loadings

For estimation of risk premia, from now on, we assume that T is sufficiently short such that $\beta_{ik,t}$ λ_{kt} and η_{it} can be treated as fixed constants, namely $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_i$, $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}$, and $\eta_{it} = \eta_i$, for $t = 1, 2, ..., T$, and assume that $\bar{r}_{p,t+1} - r_t^f$ is stationary such that $E\left(\bar{r}_{p,t+1} - r_t^f\right) = \bar{\mu}_p - r^f$. Under these assumptions and using (26) we have

$$\bar{\mu}_p = r^f + \bar{\boldsymbol{\beta}}_p' \boldsymbol{\lambda} + \bar{\eta}_p, \; p = 1, 2, ..., P, \tag{28}$$

where

$$\overline{\boldsymbol{\beta}}_p = \sum_{i=1}^n w_{ip} \boldsymbol{\beta}_i = \boldsymbol{B}_n' \mathbf{w}_p, \quad \bar{\eta}_p = \sum_{i=1}^n w_{ip} \eta_i = \mathbf{w}_p' \boldsymbol{\eta}_n, \tag{29}$$

$\mathbf{B}_n = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_n)'$, and $\boldsymbol{\eta}_n = (\eta_1, \eta_2, ..., \eta_n)'$. This equilibrium condition holds when $\bar{\mu}_p$ and portfolio betas, $\bar{\boldsymbol{\beta}}_p$, are known or estimated. To highlight the conditions needed for consistent estimation of $\boldsymbol{\lambda}$ here we assume $\bar{\mu}_p$ and $\bar{\boldsymbol{\beta}}_p$ are known.

For estimation 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 in (28) to obtain

$$\bar{\boldsymbol{\mu}}_P = r^f \boldsymbol{\tau}_P + \overline{\boldsymbol{B}}_P \boldsymbol{\lambda} + \bar{\boldsymbol{\eta}}_P, \tag{30}$$

where $\bar{\boldsymbol{\mu}}_P = (\bar{\mu}_1, \bar{\mu}_2, ..., \bar{\mu}_P)'$, $\overline{\boldsymbol{B}}_p' = \left(\overline{\boldsymbol{\beta}}_1, \overline{\boldsymbol{\beta}}_2, ..., \overline{\boldsymbol{\beta}}_P\right)$, $\bar{\boldsymbol{\eta}}_P = (\bar{\eta}_1, \bar{\eta}_2, ..., \bar{\eta}_P)'$, and $\boldsymbol{\tau}_P$ is a $P \times 1$ vector of ones. The least squares estimator of $\boldsymbol{\lambda}$ is now given by

$$\hat{\boldsymbol{\lambda}}_p = \left(\overline{\boldsymbol{B}}_P' \mathbf{M}_P \overline{\boldsymbol{B}}_P\right)^{-1} \overline{\boldsymbol{B}}_P' \mathbf{M}_P \bar{\boldsymbol{\mu}}_P, \tag{31}$$

where $\mathbf{M}_P = \mathbf{I}_P - P^{-1} \boldsymbol{\tau}_P \boldsymbol{\tau}_P'$. Using (30) we now have

$$\hat{\boldsymbol{\lambda}}_p - \boldsymbol{\lambda}^0 = \left(\overline{\boldsymbol{B}}_P' \mathbf{M}_P \overline{\boldsymbol{B}}_P\right)^{-1} \overline{\boldsymbol{B}}_P' \mathbf{M}_P \bar{\boldsymbol{\eta}}_P,$$

where $\boldsymbol{\lambda}^0$ is the true value of $\boldsymbol{\lambda}$. It is clear that $\sqrt{P}$ consistent estimation of $\boldsymbol{\lambda}^0$ requires that

$$\lambda_{min}\left(P^{-1} \overline{\boldsymbol{B}}_P' \mathbf{M}_P \overline{\boldsymbol{B}}_P\right) > c > 0, \text{ and } P^{-1}\left(\overline{\boldsymbol{B}}_P' \mathbf{M}_P \bar{\boldsymbol{\eta}}_P\right) \to_p \mathbf{0}. \tag{32}$$

To investigate whether the above two conditions hold when portfolio returns are used, we need to make assumptions about the factor loadings and pricing errors of the underlying individual returns. With this in mind we introduce the following assumptions in addition to Assumption 1 already made for the portfolio weights, $\mathbf{W}_P$.

Assumption 2 *(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 such that for all nby*

$$\text{(a): } \lambda_{min}\left(n^{-1} \boldsymbol{B}_n' \mathbf{B}_n\right) > c > 0, \text{ (b): } \lim_{n \to \infty}\left(n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{B}_n\right) = \boldsymbol{\Sigma}_{\beta\beta} > 0, \tag{33}$$

and $\boldsymbol{\Sigma}_{\beta\beta}$, 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.

We also make the following assumption on the pricing errors, η_i.

Assumption 3 *(Pricing errors) The pricing errors, η_i satisfy the approximate bound condition*

$$\sum_{i=1}^{n} \eta_i^2 = O_p\left(n^{\alpha_\eta}\right). \tag{34}$$

where $\alpha_\eta \geq 0$.

Assumption 3 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 $\alpha_\eta = 0$, as in the APT condition given by (18)

Consider now the two conditions in (32), required for $\sqrt{P}$ consistent estimation of the risk premia, noting that

$$\bar{\boldsymbol{\eta}}_P = \mathbf{W}'_P \boldsymbol{\eta}_n, \text{ and } \overline{\boldsymbol{B}}'_P = \boldsymbol{B}'_n\left(\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P\right) = \boldsymbol{B}'_n \mathbf{W}_P, \tag{35}$$

it follows that $P^{-1}\overline{\boldsymbol{B}}'_P \mathbf{M}_P \overline{\boldsymbol{B}}_P = P^{-1}\boldsymbol{B}'_n \mathbf{W}_P \mathbf{M}_P \mathbf{W}'_P \mathbf{B}_n$, and the first condition in (32) is met if $\lambda_{min}\left(P^{-1}\boldsymbol{B}'_n \mathbf{W}_P \mathbf{M}_P \mathbf{W}'_P \mathbf{B}_n\right) > c > 0$. Using Corollary 4.5.1 in Horn and Johnson (1985), we note that

$$\lambda_{min}\left(P^{-1}\boldsymbol{B}'_n \mathbf{W}_P \mathbf{M}_P \mathbf{W}'_P \mathbf{B}_n\right) \geq \lambda_{min}\left(P^{-1}\boldsymbol{B}'_n \mathbf{B}_n\right) \lambda_{min}\left(\mathbf{W}_P \mathbf{M}_P \mathbf{W}'_P\right),$$

and since $P = n/m$, with m fixed, then under Assumptions (1) and (2) we have

$$\lambda_{min}\left(P^{-1}\boldsymbol{B}'_n \mathbf{W}_P \mathbf{M}_P \mathbf{W}'_P \mathbf{B}_n\right) \geq c > 0,$$

as required. However, there is no guarantee that the above condition will be met in the case of fully diversified portfolios where P is fixed and $\left\|\mathbf{W}_P\right\|^2 = \ominus\left(n^{-1}\right)$. This is illustrated by the following simple example.

Example 2 *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, $\sigma_\xi^2 > 0$, 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 $\bar{\xi}_p = \sum_{i=1}^n w_{ip}\xi_i = \mathbf{w}'_p \boldsymbol{\xi}$, 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*

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

which does not depend on β. 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

$$P^{-1}\sum_{p=1}^{P} \bar{\boldsymbol{\xi}}'_p \mathbf{M}_P \bar{\boldsymbol{\xi}}_p \leq P^{-1}\sum_{p=1}^{P} \bar{\boldsymbol{\xi}}'_p \bar{\boldsymbol{\xi}}_p = O_p\left(m^{-1}\right).$$

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, $P \to \infty$.

To establish the second condition in (32) we note that[6]

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

In the case of non-diversified portfolios $\left\|\mathbf{W}_P\right\|^2 = \ominus\left(m^{-1}\right)$, $n = mP$, and by Assumptions (2) and 3 we have $\left\|n^{-1/2}\mathbf{B}_n\right\| = \lambda_{max}^{1/2}\left(n^{-1}\mathbf{B}_n'\mathbf{B}_n\right) < C < \infty$, $\left\|n^{-1/2}\boldsymbol{\eta}_n\right\| = \left(n^{-1}\boldsymbol{\eta}_n'\boldsymbol{\eta}_n\right)^{1/2} = O(n^{(\alpha_\eta-1)/2})$. Then $P^{-1}\left\|\boldsymbol{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\boldsymbol{\eta}_n\right\| = O(n^{\frac{\alpha_\eta-1}{2}}) \to 0$, if $\alpha_\eta < 1$, as n (or P) $\to \infty$. Recall that m is fixed as n and $P \to \infty$.

It is clear that whether one uses individual securities or portfolios to eliminate the effects of the pricing errors on estimation of $\boldsymbol{\lambda}_0$, it is required that $n^{(\alpha_\eta-1)/2} \to 0$, as $n \to \infty$. If α_η is large, convergence of $\hat{\boldsymbol{\lambda}}_p$ to its true value, $\boldsymbol{\lambda}^0$, can be very slow even with known factor loadings. For $\alpha_\eta = 1/2$, which corresponds to about 20 out of 400 securities having non-zero pricing errors, convergence is not at the usual $n^{-1/2}$ rate but at $n^{-1/4}$ rate. This is very slow and requires a much larger sample size to obtain comparable precision to the case where there are no pricing errors.

Also $\sqrt{n}$ consistent estimation of $\boldsymbol{\lambda}_0$ requires that all factors are strong in the sense that $\lambda_{min}\left(n^{-1}\boldsymbol{B}_n'\mathbf{B}_n\right)$ tends to a strictly positive number as $n \to \infty$. Condition (33) of Assumption 2 can be relaxed by replacing it with $lim_{n\to\infty}\lambda_{min}\left(\mathbf{D}_n\mathbf{B}_n'\mathbf{B}_n\mathbf{D}_n\right) > c > 0$, where $\mathbf{D}_n$ is a $K \times K$ diagonal matrix with elements $n^{-\alpha_k/2}$, for $k = 1, 2, ..., K$, and α_k measures the strength of factor f_{kt}. In this case the estimator of the risk premia associated with factor f_{kt} converges at the slower rate of $\alpha_k/2$, instead of the standard rate of $\sqrt{n}$, rate when $\alpha_k < 1$. Using portfolio returns does not relax the requirement that factors must be strong for $\sqrt{n}$ consistent estimation. If the factor under consideration is not strong and there are pricing errors, the convergence will be even slower. For further details see Pesaran and Smith (2021).

5 Estimation of risk premia using individual or portfolios returns: unknown factor loadings

The above analysis considered the case when the true factor loadings, β_{ik}, are known and showed that for consistent estimation of risk premia n needs to be large. In practice the factor loadings must be estimated. Since the loadings tend to vary over time, short samples are typically used to estimate them, often with rolling regressions. Thus it is the finite T, large n, case that is relevant in practice.

One argument for the use of portfolios is that they reduce the small T bias in estimators of the risk premia. Thus it is of some interest to compare the finite T, large n, bias of two-pass estimators of $\boldsymbol{\lambda}$ when using either the returns on individual securities or on portfolios. To focus on the small T bias we consider the usual case examined in the literature where all factors are strong, $\alpha_k = 1$, but we allow for pricing errors. We begin with individual securities.

5.1 Using individual security returns

To allow for sampling errors when factor loadings are estimated we need the following additional assumptions:

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

Assumption 4 *(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. $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F} \rightarrow_p \boldsymbol{\Sigma}_f = E\left[(\mathbf{f}_{t+1} - \mu)(\mathbf{f}_{t+1} - \mu)'\right] > 0$, where $E_t(\mathbf{f}_{t+1}) = \mu$, 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 of ones.*

Assumption 5 *(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 $0 < c < \sigma_{ii} < C < \infty$,*

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

and

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

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.

Now stacking the returns on the n individual securities by time we have

$$\mathbf{r}_{nt} = \mathbf{a}_n + \boldsymbol{B}_n\mathbf{f}_t + \mathbf{u}_{nt}, \text{ for } t = 1, 2, ..., T, \tag{36}$$

where $\mathbf{r}_{nt} = (r_{1t}, r_{2t}, ..., r_{nt})'$ is an $n \times 1$ vector of 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

$$\mathbf{r}_{i\circ} = \mathrm{a}_i\boldsymbol{\tau}_T + \mathbf{F}\boldsymbol{\beta}_i + \mathbf{u}_{i\circ}, \tag{37}$$

where $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, and $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, ..., u_{iT})'$. As noted above, the true value of the vector of risk premia, $\boldsymbol{\lambda}$, is defined by the cross section regressions (CSR)

$$E(r_{it}) = \lambda_0 + \boldsymbol{\beta}_i'\boldsymbol{\lambda} + \eta_i, \text{ for } i = 1, 2, ..., n, \tag{38}$$

where η_i is the pricing error and λ_0 is the zero beta return.

The two-pass estimator of risk premia, $\boldsymbol{\lambda}$, based on individual returns is given by[7]

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

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

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

$\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}_T'$, and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under (37), $\hat{\boldsymbol{\beta}}_{i,T} = \boldsymbol{\beta}_i + (\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{u}_{i\circ}$, and hence

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

[7] 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 $\mathbf{U}_n = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, ..., \mathbf{u}_{n\circ})'$, and $\mathbf{G}_T = \mathbf{M}_T\mathbf{F}\left(\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)^{-1}$. Also, averaging the return equations (37) over t for each i, we have

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

where $\bar{\mathbf{f}}_T = T^{-1}\sum_{t=1}^T \mathbf{f}_t$, and $\bar{u}_{i\circ} = T^{-1}\sum_{t=1}^T u_{it}$. Hence, using the above results together with the APT condition given by (38), we have

$$\bar{\mathbf{r}}_n = \lambda_0\boldsymbol{\tau}_n + \boldsymbol{B}_n\boldsymbol{\lambda}_T^* + \bar{\mathbf{u}} + \boldsymbol{\eta}, \tag{43}$$

where $\boldsymbol{\lambda}_T^*$ is what is called the "ex post" pricing error

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda}^0 + \mathbf{d}_T, \tag{44}$$

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

$\bar{\mathbf{u}} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, ..., \bar{u}_{n\circ})'$, and $\boldsymbol{\eta}$ is the $n \times 1$ vector of pricing errors. As before $\boldsymbol{\lambda}^0$ denotes the true value of $\boldsymbol{\lambda}$.

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

$$\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}^0 \to_p \left[\boldsymbol{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Sigma}_{\beta\beta}\mathbf{d}_T - \frac{\bar{\sigma}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}^0\right). \tag{46}$$

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

$$\bar{\sigma}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sigma_i^2 > 0. \tag{47}$$

The bias of $\hat{\boldsymbol{\lambda}}_n$ is due to terms that involve $\mathbf{d}_T$ and $\bar{\sigma}^2$. Following Shanken (1992), $\bar{\sigma}^2$ can be consistently estimated (for a fixed $T > k+1$) by

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

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

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

5.2 Using portfolio returns

Consider now the estimator of $\boldsymbol{\lambda}^0$ based on portfolios discussed in Section 4.2. Using portfolio returns defined by (29), 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 (29), 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 resulting estimates will be identical. Denoting the portfolio estimator of $\boldsymbol{\lambda}^0$ by $\hat{\boldsymbol{\lambda}}_P$ we have

$$\hat{\lambda}_P = \left(\bar{\hat{\mathbf{B}}}'_{PT}\mathbf{M}_P\bar{\hat{\mathbf{B}}}_{PT}\right)^{-1}\left(\bar{\hat{\mathbf{B}}}'_{PT}\mathbf{M}_P\bar{\mathbf{r}}_P\right), \tag{50}$$

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

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

To relate $\hat{\boldsymbol{\lambda}}_P$ to the estimator, $\hat{\boldsymbol{\lambda}}_n$, based on the individual securities, we note that $\bar{\hat{\mathbf{B}}}_{PT} = \mathbf{W}'_P\hat{\boldsymbol{B}}_{nT}$, 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 $\hat{\lambda}_P$ can now be written equivalently as

$$\hat{\boldsymbol{\lambda}}_P = \left(\hat{\boldsymbol{B}}'_{nT}\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\hat{\boldsymbol{B}}_{nT}\right)^{-1}\left(\hat{\boldsymbol{B}}'_{nT}\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{r}}_n\right). \tag{51}$$

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 1. To establish the asymptotic properties of $\hat{\boldsymbol{\lambda}}_P$ we also need the following 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) $\boldsymbol{\Sigma}_{\beta\beta,w}$ defined by*

$$\lim_{P\to\infty}\left(P^{-1}\boldsymbol{B}'_n\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\mathbf{B}_n\right) = \boldsymbol{\Sigma}_{\beta\beta,w}, \tag{52}$$

is positive definite.

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 2, and it is therefore somewhat weaker. Similarly, part (a) of the above assumption follows from part (a) of Assumption 2. 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 2, and also requires condition (d) of Assumption 1.

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 (36) and the associated risk premia, $\boldsymbol{\lambda}$, defined by (38), and suppose that Assumptions (3), (4), (5), and (6) hold, and $\alpha_\eta < 1$ where α_η is defined by (34). 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 (1) 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 fixed, then for any fixed* $T > K$ *we have (as* $P \to \infty$*)*

$$\hat{\boldsymbol{\lambda}}_P - \boldsymbol{\lambda}^0 \to_p \left[\boldsymbol{\Sigma}_{\beta\beta,w} + \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left[\boldsymbol{\Sigma}_{\beta\beta,w}\mathbf{d}_T - \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}^0\right]. \tag{53}$$

where $\hat{\lambda}_P$ *is defined by (50),* $\boldsymbol{\lambda}^0$ *is the true value of* $\boldsymbol{\lambda}$*,* $\mathbf{d}_T = T^{-1}\sum_{t=1}^{T}\left[\mathbf{f}_t - E(\mathbf{f}_t)\right]$,

$$\boldsymbol{\Sigma}_{\beta\beta,w} = \lim_{P\to\infty}\left(\frac{\mathbf{B}_n'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}_n}{P}\right), \text{ and } \bar{\omega}^2 = \lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p\right) > 0, \tag{54}$$

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

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-corrected version of $\hat{\boldsymbol{\lambda}}_P$, corresponding to (49). Suppose that $\bar{\omega}^2$ is known then a Shanken type bias-corrected portfolio estimator of $\boldsymbol{\lambda}$ is given by

$$\hat{\boldsymbol{\lambda}}_P^{BC} = \left[\frac{\hat{\mathbf{B}}_{nT}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\hat{\mathbf{B}}_{nT}}{P} - \frac{\bar{\omega}^2}{T}\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\right]^{-1}\left(\frac{\hat{\boldsymbol{B}}_{nT}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\bar{\mathbf{r}}_n}{P}\right). \tag{55}$$

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

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

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

Using these results in (55), and assuming that $\boldsymbol{\Sigma}_{\beta\beta,\omega}$ is full rank (see part (c) of Assumption 6) we obtain

$$\hat{\boldsymbol{\lambda}}_P^{BC} \to_p \boldsymbol{\lambda}_T^* = \boldsymbol{\lambda}^0 + \mathbf{d}_{fT}, \tag{56}$$

where $\mathbf{d}_T$ is defined by (45), and $\boldsymbol{\lambda}_T^*$ is Shanken's "ex-post" risk premia.

However, to implement this correction requires a small T unbiased (as $n \to \infty$) estimator of $\bar{\omega}^2 = \lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p\right)$ which depend on the error covariances, σ_{ij}. Recall that

$$\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ip}w_{jp}\sigma_{ij}.$$

This contrasts the case when individual security returns are used, where the Shanken correction requires small T unbiased estimation of $\overline{\sigma}^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$ which does not involve the error covariances, σ_{ij}, and can be estimated consistently 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 $n \to \infty$. An estimator of $\bar{\omega}^2$ can be obtained by replacing σ_{ij} by its sample estimator, $\hat{\sigma}_{ij,T} = T^{-1}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}$, to obtain

$$\widehat{\bar{\omega}}_{nT}^2 = \frac{1}{TP}\sum_{p=1}^{P}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ip}w_{jp}\hat{u}_{it}\hat{u}_{jt}. \tag{57}$$

But $\hat{u}_{it}$, being based on the estimates of $\boldsymbol{\beta}_i$, is subject to additional sampling errors, and in terms of u_{it} is given by

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

Substituting the above expression in (57) now yields

$$\begin{aligned}\widehat{\bar{\omega}}^2_{nT} &= \frac{1}{TP}\sum_{p=1}^{P}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ip}w_{jp}\left(u_{it} - \bar{u}_i\right)\left(u_{jt} - \bar{u}_j\right) + P^{-1}\sum_{p=1}^{P}\mathbf{q}'_{p,nT}\left(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}\right)\mathbf{q}_{p,nT} \\ &- 2\frac{1}{TP}\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{ip}\left(u_{it} - \bar{u}_i\right)\left(\mathbf{f}_t - \bar{\mathbf{f}}_T\right)'\mathbf{q}_{p,nT},\end{aligned}$$

where $\mathbf{q}_{p,nT} = \sum_{i=1}^{n} w_{ip}\left(\hat{\boldsymbol{\beta}}_{i,T} - \boldsymbol{\beta}_i\right)$. It is now easily seen that for T fixed the above estimator does not converge to $\bar{\omega}^2$, and a bias correction version does not seem to be attainable, either.

Also whether the bias in estimation of $\boldsymbol{\lambda}$ 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 characterized 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 3 *Suppose $K = 1$, 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 (46) and (53))*

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

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

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

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

$$\frac{\beta'_n\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\beta_n}{P} > \left(\frac{\frac{1}{P}\sum_{p=1}^{P}\left(\mathbf{w}'_p\boldsymbol{\Sigma}_u\mathbf{w}_p\right)}{\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2}\right)\frac{\beta'_n\mathbf{M}_n\beta_n}{n}. \tag{58}$$

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. 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

$$\bar{\beta}_1 = \mathbf{w}_1'\boldsymbol{\beta} = m^{-1}\sum_{i=1}^{m}\beta_i,\ \bar{\beta}_2 = \mathbf{w}_2'\boldsymbol{\beta} = m^{-1}\sum_{i=m+1}^{2m}\beta_i, ..., \bar{\beta}_P = \mathbf{w}_P'\boldsymbol{\beta} = m^{-1}\sum_{i=(P-1)m+1}^{n}\beta_i, \quad (59)$$

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 for the estimator of λ *based on portfolio returns we 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 estimator 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* $\mathbf{w}_p'\boldsymbol{\Sigma}_u\mathbf{w}_p = m^{-2}\boldsymbol{\tau}_m'\boldsymbol{\Sigma}_{p,u}\boldsymbol{\tau}_m$*, 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 (58), 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 Example 2)

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

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.

The above 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 for $\sqrt{n}$ consistent estimation of the risk premia.

6 Concluding remarks

This paper examines two questions associated with tests of the APT, both of which involve the role of pricing errors. 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. We show that factors included in the statistical model are priced only if they have non-zero conditional correlation with the stochastic discount factor, and the pricing errors arise from non-zero correlations between the idiosyncratic 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 as characterized by the stochastic discount factor.

The second question addressed in this paper is the *pros* and *cons* of using portfolio returns rather than individual security returns in estimation of risk premia. We show that when there are pricing errors it is crucial to have a large cross section dimension, whether the cross section is of individual securities or of portfolios. One argument given for using portfolios is 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. Similarly to Shanken bias-corrected estimator based using individual returns, we also derive a bias corrected estimator of the risk premia based on portfolio returns. But whereas with individual securities the bias correction is operational, 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 comprising the portfolios. 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 of Fama-MacBeth estimator of 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 (36), and suppose that Assumption 5 holds. Then for any* t *and* t' *(as* $n \to \infty$*)*

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

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

and

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

where

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

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 5, 3 and part (b) of 2 hold, and* $\alpha_\eta < 1$.[8] *Then*

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

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

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

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

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

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

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

[8] As compared to the notation in the body of the paper, we have dropped the subscript n from $\mathbf{B}_n$ as defined by (36).

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 1, 5, 3 and 2 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*

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

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

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

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

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

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

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

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

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

which tends to zero as $P \to \infty$, since under Assumptions 1 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.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,

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

and in view of (A.10), it follows that

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

The first term of (A.17) can be written as

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

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 1, 5,and 6. Hence, it follows that $Var\left(c_{st,P}\right) \rightarrow 0$, for all $s = 1, 2, .., K$ and $t = 1, 2, ..., T$, and hence $P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{W}_P'\ \boldsymbol{U} \rightarrow_p \mathbf{0}$. 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

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

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 $\lim_{P\rightarrow\infty}\left\|P^{-1/2}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\right\|^2 = \lim_{P\rightarrow\infty}\lambda_{\max}\left(P^{-1}\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\mathbf{B}\right) < C$, 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 $\left\|\mathbf{W}_P\right\|^2 = \ominus\left(m^{-1}\right)$, $P/n \rightarrow \pi$, then $P^{-1}\left\|\mathbf{B}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}_P'\boldsymbol{\eta}\right\| = \ominus\left(n^{-1}\sum_{i=1}^{n}\eta_i^2\right) = \ominus\left(n^{\alpha_\eta-1}\right)$, which tends to zero since $\alpha_\eta < 1$. 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 $\mathbf{C}_P = P^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}'_P\boldsymbol{U}$ is given by $c_{tt',P} = P^{-1}\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ip}w_{jp}u_{it}u_{jt'}$ and we have

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

and hence $E\left(\mathbf{C}_P\right) = \bar{\omega}_P^2\mathbf{I}_T$, 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 $E\left(c_{tt',P}^2\right)$ when $t\neq t'$ and $E\left(c_{tt',P}-\bar{\omega}_P^2\right)^2$ 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 $\mathbf{U}'\mathbf{W}_P\mathbf{M}_P\mathbf{W}'_P\bar{\mathbf{u}} = T^{-1}\mathbf{U}'\mathbf{W}_P\mathbf{W}'_P\boldsymbol{U}\tau_T$. ■

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

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

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

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

and $\mathbf{r}_{i\circ} = (r_{i1}, r_{i2}, ..., r_{iT})'$. Under the factor model (36)

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

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

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

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

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

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

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

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 (51) 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

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

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 5, 3 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 (54). Result (53) 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.

Bailey, N., G. Kapetanios and M. H. Pesaran (2021) Measurement of Factor Strength: Theory and Practice, *Journal of Applied Econometrics, 36, 587-613.*

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.

Horn, R. A. and C.R. Johnson (1985), *Matrix Algebra,* Cambridge University Press, Cambridge.

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*, Volume 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.