

Testing Beta-Pricing Models Using Large Cross-Sections

VALENTINA RAPONI, CESARE ROBOTTI, and PAOLO ZAFFARONI¹

4 January 2016

JEL # G12, C12, C52

¹Raponi, Robotti, and Zaffaroni are from Imperial College London. We thank Adelina Barbalau, Fernando Chague, Lubos Pastor, Krishna Ramaswamy, Jay Shanken, Raman Uppal, Pietro Veronesi, seminar participants at CORE, Imperial College London, Luxembourg School of Finance, University of Georgia, University of Southampton, Toulouse School of Economics, Tinbergen Institute, as well as conference participants at the 2015 Meetings of the Brazilian Finance Society, CFE 2015 and the NBER-NSF 2016 for helpful discussions and comments. This paper was previously circulated with the title “Ex-Post Risk Premia and Tests of Multi-Beta Models in Large Cross-Sections”.

Testing Beta-Pricing Models Using Large Cross-Sections

Abstract

This paper presents a limiting theory for estimating and testing linear beta-pricing models when a very large number of assets N is available together with a fixed, possibly very small, time-series dimension, applicable to both traded and non-traded factors. We focus on Shanken's (1992) estimator, and show that it exhibits many desirable properties. We first demonstrate that it is an OLS-based estimator that, unlike others, does not require preliminary estimation of the bias-adjustment. Under fairly standard assumptions, we prove its \sqrt{N} consistency and derive its asymptotic distribution, showing how its limiting covariance matrix can be consistently estimated. We then propose a new test of the no-arbitrage asset pricing restriction, and establish its asymptotic distribution (under the assumption of correct model specification) that only requires the number of assets N to diverge. Finally, we show how our theoretical results, in terms of estimation and testing, can be extended to deal with the more realistic case of unbalanced panels. The practical relevance of our findings is demonstrated using Monte Carlo simulations and an empirical application to asset-pricing models with traded risk factors. Our analysis suggests that the proposed large N framework can change, even substantially, the common empirical findings in terms of significance of estimated risk premia and validity of linear beta-pricing models.

Keywords: Beta-pricing models; Ex-post risk premia; Two-pass cross-sectional regression; Large N asymptotics; Specification test; Unbalanced panel.

JEL classification numbers: C12; C13; G12.

1 Introduction

Tens of thousands of stocks are traded every day in financial markets, providing an extremely rich information set to validate and estimate asset pricing models. For example, one can obtain the returns for 18,474 US stocks on December 2013 (source: CRSP). On the other hand, both academics and practitioners are reluctant to use time series spanning long time periods to avoid the risk of including structural breaks or, alternatively, to avoid the additional difficulty of parameterizing time variation of betas and risk premia. This implies that, for example, if one wishes to use only post-2009 observations because of the presumed change of the parameters of the equity returns' distribution due to the financial crisis, less than one hundred monthly time series observations or equivalently eight yearly time series observations are available, adversely affecting the finite-sample properties of estimates of any asset pricing model. The alternative approach of increasing the time series frequency of the sampled equity returns data, say to daily or even infra-daily observations, leads to a substantial increase of noisiness in the data. It also leads to additional complications related to unequal spacing or market micro-structure effects, which greatly complicate the statistical theory of any adopted estimator.

Therefore, it appears important to have a methodology that permits to carry out statistically correct inference on estimated risk premia and to test the validity of the underlying beta-pricing model, just by exploring such large available cross-sectional variation of returns across N individual securities and, at the same time, relying on a limited number of time series observations T . This is the contribution of the paper.

The traditional empirical methodology for exploring asset pricing models is instead asymptotically valid when T is large and N is fixed: it entails estimation of asset betas, which represent systematic risk measures, by means of time-series factor model regressions, followed by estimation of risk premia via a cross-sectional regression (CSR) of observed average returns on the estimated betas. For example, in the empirical strategy to analyse the Capital Asset Pricing model (CAPM) coined by Fama and MacBeth (1973), a CSR is run each month, with inference ultimately based on the time-series mean and standard error of the monthly risk premia estimates. The formal econometric analysis of such two-pass methodology was first provided by Shanken (1992) who shows how the asymptotic standard errors of the second-pass ordinary least squares (OLS) and generalized

least squares (GLS) risk premia estimators are influenced by estimation error in the first-pass betas, requiring an adjustment to the traditional Fama and MacBeth (1973) standard errors.¹

Unfortunately, estimation of risk premia in a large N environment rules out the possibility of obtaining estimators, such as the OLS CSR estimator, with the usual desirable statistical properties, for example consistency, when T is not allowed to grow.² However, one could estimate a related quantity denominated the “ex-post” risk premia by Shanken (1992): this equals the conventional risk premia plus the unexpected factors’ outcomes, namely the difference between the expected value and the sample average of the factors of the beta-pricing model.³ The “ex-post” risk premia is a meaningful object: it is an unbiased estimator of the (ex-ante) risk premia, and the beta-pricing model is still linear in the “ex-post” risk premia under the assumption of correct model specification, as exemplified by Shanken (1992). Moreover, the associated “ex-post” pricing errors, permit to construct valid specification tests of any beta-pricing model. Building on the arguments of Litzenberger and Ramaswamy (1979), Shanken (1992) proposes an estimator of the ex-post risk premia and shows that it is asymptotically unbiased when only N diverges. Noticeably, this result holds despite not imposing a strict structure on the $N \times N$ covariance matrix of the equity returns’ residuals, such as constant variances or null cross correlations, as assumed by Litzenberger and Ramaswamy (1979).⁴ This suggests that the Shanken estimator could have other desirable statistical properties, especially in large samples, which need nevertheless to be established under suitable regularity conditions. Moreover, unlike other approaches such as instrumental variable estimation, the Shanken estimator does not require any additional information besides a panel of asset returns and a sample of factors’ realizations. Similarly, its derivation does not hinge on the existence of a preliminary (consistent) estimator, which is often used to quantify the bias affecting

¹ See also the related paper by Black, Jensen, and Scholes (1972). Jagannathan and Wang (1998) relax the conditional homoskedasticity assumption in Shanken (1992) and derive expressions for the asymptotic variances of the OLS and GLS estimators that are valid under fairly general distributional assumptions. Hou and Kimmel (2006), Shanken and Zhou (2007), and Kan, Robotti, and Shanken (2013) provide a unifying treatment of the two-pass methodology in the presence of global (or fixed) model misspecification. We refer the readers to Jagannathan, Skoulakis, and Wang (2010), Kan and Robotti (2012), and Gospodinov and Robotti (2012) for a synthesis of the two-pass methodology. All these papers require large T and fixed N .

²When T is fixed, increasing the number of assets affects the residual variation but does not eliminate the uncertainty about the unanticipated factor realizations; see Shanken (1992) for details.

³Obviously as T gets large any discrepancy between the conventional (ex-ante) risk premia and the ex-post risk premia dissipates because the factors’ sample mean converge to their expected value.

⁴Somewhat surprisingly, the Shanken estimator does not suffer from the so-called curse of dimensionality, despite involving potentially an infinite number of parameters, viz. the elements of the residuals covariance matrix, as N diverges. We explain below the reasons for this extremely convenient feature.

the OLS CSR estimator. Finally, the Shanken's estimator can be used when either traded and non-traded factors are used in the beta-pricing model.

For these reasons, in this paper we provide a rigorous statistical analysis of the Shanken estimator. To provide further motivations to our analysis, we start by showing that the Shanken estimator of the ex-post risk premia is an element, with special properties, of a class of OLS bias-adjusted estimators. In particular we demonstrate mathematically that it is the only element of this class not requiring a preliminary estimation of the bias-adjustment, making it particularly convenient since it avoids, for example, any pre-testing biases and, at the same time, it does not require to sacrifice data for preliminary estimation. We then focus on the asymptotic properties of the Shanken estimator for large N and fixed T : under mild assumptions, in particular permitting a degree of cross-correlation among returns akin to the one typically postulated in the Arbitrage Pricing Theory (APT) of Ross (1976), we establish \sqrt{N} -consistency and asymptotic normality. Moreover, we derive an explicit, simple and easy to interpret, expression for its asymptotic covariance matrix, showing how it can be consistently estimated and used to conduct inference on the risk premia estimates, such as constructing correctly-sized confidence intervals. Besides estimation, we provide a new test for the validity of the asset-pricing restrictions and characterize its distribution when $N \rightarrow \infty$ and fixed T , under the null hypothesis that exact pricing holds and that the model is correctly specified. Noticeably, our test has power, i.e. is able to discriminate whether the beta-pricing model is correctly specified, despite that our test statistics is built on the ex-post pricing errors which are, necessarily, contaminated by the unexpected factor outcomes. Finally, thanks to the large N feature, our test appears to mitigate the concerns raised by Harvey et al (2015) and Barillas and Shanken (2016) regarding the conventional way in which beta-model specification tests have been used with the data, as discussed in the next section.

Although our theory is initially presented for the case when the same number of assets N is observed in every time period, we show how our results, both in terms of estimation and testing, can be modified when the panel is in fact unbalanced. This extension is useful particularly when only a limited number of time series observations is available, making extremely costly to eliminate observations just to achieve a balanced panel because this would lead to imprecise estimates and unnecessarily large confidence intervals. Finally, our analysis is applicable to models with traded factors, non-traded factors, or both. Our limiting results should be viewed as a theoretical

complement to the detailed simulation analysis in Chordia, Goyal, and Shanken (2015).

Besides Shanken (1992), there are only few other papers that deal with estimation of risk premia with a large cross section N . Gagliardini, Ossola, and Scaillet (2015) derive the limiting properties of a class of bias-adjusted estimators of the (ex-ante) risk premia under rather weak assumptions, when T , in addition to N , is allowed to go to infinity.⁵ The asymptotic theory in Gagliardini, Ossola, and Scaillet (2015) requires that that $N/T^\delta \rightarrow 0$, for some $0 < \delta < 3$, as $N \rightarrow \infty$. Like ours, their estimator requires a bias-adjustment precisely because N is also allowed to diverge. Bai and Zhou (2015) also investigate the joint asymptotics in N and T of a modified OLS CSR estimator of the (ex-ante) risk premia. Kim and Skoulakis (2014) propose a \sqrt{N} -consistent estimator of the ex post risk premia in a two-pass CSR setting, employing the so-called regression calibration approach used in errors-in-variables models, building upon Jagannathan, Skoulakis, and Wang (2010).⁶ Finally, Jegadeesh and Noh (2014) and Pukthuanthong, Roll, and Wang (2014) propose instrumental variable estimators of the ex-post risk premia, exploiting the assumed independence across time of the returns data. In terms of tests of the asset-pricing no-arbitrage condition, our test is correctly sized and has power despite only N is allowed to get arbitrarily large, unlike all other tests proposed so far which require T to diverge. For instance, extending the classical test of Gibbons, Ross, and Shanken (1989), Pesaran and Yamagata (2012) propose a number of tests of the asset-pricing no-arbitrage restriction: their setup accommodates traded factors and the feasible versions of their tests require joint asymptotics, such that $N/T^3 \rightarrow 0$ as N diverges. Gagliardini, Ossola, and Scaillet (2015) derive the asymptotic distribution of an asset-pricing model test, which also requires joint asymptotics but, like us, it allows for factors that are not necessarily traded portfolios.

We also explore the small N properties of the Shanken estimator and of our test statistic via Monte Carlo simulations. In particular, in terms of estimation of risk premia, we show both the effective improvement in terms of bias and mean square error of the Shanken estimator, in comparison to the CSR OLS estimator. We calibrate the model parameters using real data on

⁵Noticeably, Gagliardini, Ossola, and Scaillet (2015) derive the asymptotic distribution of their estimator both when the asset-pricing restriction is imposed and when it is not imposed. Time-variation in risk premia is allowed for by conditioning on observed state variables.

⁶Kim and Skoulakis (2014) estimator can be interpreted as providing a nice interpretation of the Shanken estimator: their estimator is almost identical to the Shanken estimator, the only difference being that the first- and second-pass regressions are evaluated on non-overlapping time periods. However, splitting the available sample of data unavoidably affects adversely the size of the standard errors of the estimated risk premia because it sacrifices some observations.

individual stocks from the Center for Research in Security Prices (CRSP). Overall, our simulation results suggest that the tests are reliable for the sample sizes often encountered in empirical finance. Finally, we apply our large N methodology to empirically investigate the performance of some prominent asset-pricing specifications using individual stock return data, namely a monthly dataset from January 1966 until December 2013 for about 3,500 stocks. We consider three linear beta-pricing models: the single-factor CAPM, the three-factor model of Fama and French (1993), and the five-factor model recently proposed by Fama and French (2015). The results are rather striking: for all the beta-pricing models under consideration, even for a short time window (three years), our methodology shows clearly that all the estimated risk premia are statistically significant across time, with few exceptions. This is completely at odds with the results obtained using the traditional approach, that relies on the Shanken's (1992) correct standard errors when T is large (the so-called Shanken's correction), where the estimated risk premia turn out to be non significantly different from zero for most time periods. Technically speaking, our methodology only requires T to be just larger than the number of factors corresponding to the beta-pricing model under consideration. For example $T > 1$ suffices when estimating the CAPM. Likewise, in terms of testing the validity of some specific beta-pricing model, our large N test is able to reject the CAPM at conventional significance levels even for the short time window of three years ($T = 36$ months) whereas, using the same data suitably grouped, we are unable to reject the CAPM using the Gibbons et al. (1989) methodology.

The rest of the paper is organised as follows. Section I provides a brief review of the two-pass CSR methodology and introduces the main notation and assumptions. It also formally shows how the Shanken's (1992) bias-adjusted estimator is the only member of a large class of CSR OLS modified estimators that does not require preliminary estimation of the bias-adjustment. Section II illustrates the usefulness of our results with an empirical exercise. Section III presents the asymptotic analysis of when $N \rightarrow \infty$ and T is fixed. Moreover, under the assumption that the asset pricing restriction holds, i.e. when exact pricing is imposed, we derive the limiting distribution of the specification test described above when N is large and T is fixed. Finally, we show how the main analysis can be extended to accommodate unbalanced panels. We explore the small- N properties of the Shanken estimator and of the associated test in Section IV. Section V summarises our conclusions. The proofs of the main results are collected in three final appendixes.

2 The Two-Pass Methodology

A beta-pricing model seeks to explain cross-sectional differences in expected asset returns in terms of asset betas computed relative to the model's systematic economic factors. Let $f_t = [f_{1t}, \dots, f_{Kt}]'$ be a K -vector of observed factors at time t and $R_t = [R_{1t}, \dots, R_{Nt}]'$ be an N -vector of test asset returns at time t .

Assume that asset returns are governed by the following factor model:

$$R_{it} = \alpha_i + \beta_{i1}f_{1t} + \dots + \beta_{iK}f_{Kt} + \epsilon_{it} = \alpha_i + \beta_i'f_t + \epsilon_{it}, \quad (1)$$

where i indicates the i th stock and t refers to time, α_i is a scalar representing the asset specific intercept, $\beta_i = [\beta_{i1}, \dots, \beta_{iK}]'$ is a vector of multiple regression betas of asset i with respect to the K factors, and the ϵ_{it} 's are the model residuals, which satisfy Assumption 2 below. In matrix notation, we can write the above model as

$$R_t = \alpha + Bf_t + \epsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where $\alpha = [\alpha_1, \dots, \alpha_N]'$, $B = [\beta_1, \dots, \beta_N]'$, and $\epsilon_t = [\epsilon_{1t}, \dots, \epsilon_{Nt}]'$.

For a better understanding of the notion of ex-post risk premia, let $\bar{R}_i = \frac{1}{T} \sum_{t=1}^T R_{it}$, $\bar{R} = [\bar{R}_1, \dots, \bar{R}_N]'$, and $\bar{\epsilon} = \frac{1}{T} \sum_{t=1}^T \epsilon_t$. Averaging (2) over time, imposing the exact pricing condition (19), and noting that $E[R_t] = \alpha + BE[f]$ by (2), yield

$$\bar{R} = X\Gamma^P + \bar{\epsilon}, \quad (3)$$

where we define $\Gamma^P = [\gamma_0, \gamma_1^P]'$ with

$$\gamma_1^P = \gamma_1 + \bar{f} - E[f]. \quad (4)$$

Definition (4) applies regardless of whether traded or non-traded factors are considered, although for traded factors one obtains $\gamma_1^P = \bar{f} - 1_K\gamma_0$, where 1_K is a $K \times 1$ vector of ones (see Shanken (1992), Section 1.1, for derivation of this result). Therefore, as explained below, a bias adjustment is not required in this case but, for simplicity, we will not exploit this restriction in the estimation of risk premia Γ^P . By (3), expected returns are still linear in the asset betas conditional on the factor outcomes through the quantity γ_1^P which, in turn, depends on the sample mean realization $\bar{f} - E[f]$. For this reason, the random coefficient vector γ_1^P is referred to, accordingly, as the vector

of ex-post risk premia. Obviously, γ_1^P is random but is unbiased for the (ex-ante) risk premia γ_1 . The notion of ex-post risk premia γ_1^P is meaningful only conditional on the factors' realizations: when $T \rightarrow \infty$, \bar{f} will converge to $E(f)$ and thus the ex-post and ex-ante risk premia will coincide but Γ and Γ^P will differ, in general, for any finite T . However, Γ^P remains unbiased for Γ when one considers any finite T .⁷ Therefore, the importance of the ex post risk premia naturally emerges when studying estimation of beta factor models with large N and fixed, especially small, T .

Notice that (3) cannot be used to estimate the ex-post risk premia Γ^P since X is unobserved. For this reason, the popular two-pass CSR method first obtains estimates of the betas of each asset i , $\hat{\beta}_i$, by running the following multivariate regression for every $i = 1, \dots, N$:

$$R_i = \alpha_i 1_T + F\beta_i + \epsilon_i, \quad (5)$$

where $R_i = [R_{i1}, \dots, R_{iT}]'$ is a time series of returns on asset i , $\epsilon_i = [\epsilon_{i1}, \dots, \epsilon_{iT}]'$ is the associated vector of idiosyncratic residuals and 1_T is a $T \times 1$ vector of ones. This is the first pass. It follows that

$$\hat{\beta}_i = \beta_i + (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\epsilon_i, \quad (6)$$

or, in matrix form,

$$\hat{B} = R'\tilde{F}(\tilde{F}'\tilde{F})^{-1} = B + \epsilon'\mathcal{P}, \quad (7)$$

where $\hat{B} = [\hat{\beta}_1, \dots, \hat{\beta}_N]'$, $R = [R_1, \dots, R_N]$ and $\epsilon = [\epsilon_1, \dots, \epsilon_N]$ are $N \times K$ (the first) and $T \times N$ (the second and third) matrices, respectively, and $\mathcal{P} = \tilde{F}(\tilde{F}'\tilde{F})^{-1}$. The corresponding $T \times N$ matrix of OLS residuals is denoted by $\hat{\epsilon} = [\hat{\epsilon}_1, \dots, \hat{\epsilon}_N] = R' - 1_T\bar{R}' - \hat{B}\tilde{F}'$. We then run a single CSR of the sample mean vector \bar{R} on $\hat{X} = [1_N, \hat{B}]$ to estimate the risk premia. This is the second pass. However, notice that we have two different feasible representations of (3), namely

$$\bar{R} = \hat{X}\Gamma + \eta, \quad (8)$$

with residuals $\eta = \left(\bar{\epsilon} + B(\bar{f} - E[f]) - (\hat{X} - X)\Gamma\right)$ and

$$\bar{R} = \hat{X}\Gamma^P + \eta^P, \quad (9)$$

⁷To gauge an idea of the magnitude, the estimated difference between the ex-post and the ex-ante risk premium is about 2% of the ex-ante risk premium when considering the market return factor, with its standard deviation ranging from 170%, when $T = 36$, to 80%, when $T = 120$.

with residuals $\eta^P = (\bar{\epsilon} - (\hat{X} - X)\Gamma^P)$. The OLS estimator applied to either (8) or (9) yields obviously the same result:

$$\hat{\Gamma} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{bmatrix} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R}. \quad (10)$$

However, as Shanken (1992) points out, $\hat{\Gamma}$ cannot be used as a consistent estimator of the ex ante risk premia Γ in (8) for a fixed T . The reason is that \bar{f} does not converge in probability to $E[f]$ unless $T \rightarrow \infty$. Although Bai and Zhou (2015) conjecture that the impact of the term $\bar{f} - E[f]$ is small in practice, we follow Shanken (1992) and conduct our analysis based on the representation (9) where the ex-post risk premia Γ^P represent the parameter of interest. Since the innovation η^P in (9) does not contain the term $B(\bar{f} - E[f])$ as opposed to the innovation η in (8), the bias term is less severe now. However Shanken (1992) and Bai and Zhou (2015) among others show that the CSR OLS $\hat{\Gamma}$ is still biased and inconsistent for Γ^P when T is fixed. Nevertheless, Shanken (1992) shows that this bias of the CSR OLS estimator can be corrected as follows. Denote the trace operator by $\text{tr}(\cdot)$ and a K -dimensional vector of zeros by 0_K . In addition, let $\hat{\sigma}^2 = \frac{1}{N(T-K-1)}\text{tr}(\hat{\epsilon}'\hat{\epsilon})$. Then, the bias-adjusted estimator of Shanken (1992) is given by⁸

$$\hat{\Gamma}^* = \begin{bmatrix} \hat{\gamma}_0^* \\ \hat{\gamma}_1^* \end{bmatrix} = (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \frac{\hat{X}'\bar{R}}{N}, \quad (11)$$

where

$$\hat{\Sigma}_X = \frac{\hat{X}'\hat{X}}{N} \quad (12)$$

and

$$\hat{\Lambda} = \begin{bmatrix} 0 & 0'_K \\ 0_K & \hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix}. \quad (13)$$

The formula for the Shanken's estimator $\hat{\Gamma}^*$ shows a *multiplicative* bias-adjustment through the term $(\hat{\Sigma}_X - \hat{\Lambda})^{-1}$. This prompts us to understand the analogies of the Shanken estimator $\hat{\Gamma}^*$ with the more conventional class of *additive* bias-adjusted CSR OLS estimators. To this end, it is useful to consider the following property satisfied by the CSR OLS estimator, obtained by a suitably re-writing Bai and Zhou, Theorem 1, (2015)):

$$\hat{\Gamma} = \Gamma^P + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \begin{bmatrix} 0 & 0'_K \\ 0_K & -\hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix} \Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right) = \Gamma^P - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda} \Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right).$$

⁸To keep the notation somewhat manageable, we do not distinguish between generic factors and factors that are portfolio returns. Therefore, we do not incorporate the additional pricing restriction that is implied when a given factor is a portfolio return.

This formula immediately suggests an easy way to construct an additive bias-adjusted estimator for Γ^P such as:

$$\hat{\Gamma}^{bias-adj} = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\hat{\Lambda}\hat{\Gamma}^{prelim},$$

where $\hat{\Gamma}^{prelim}$ defines an arbitrary preliminary estimator for Γ^P . Although this class of estimators $\hat{\Gamma}^{bias-adj}$ appears appealing, it requires the availability of a preliminary estimator, here denoted by $\hat{\Gamma}^{prelim}$.⁹ A possible approach is to impose that the preliminary estimator $\hat{\Gamma}^{prelim}$ and the resultant bias-adjusted estimator $\hat{\Gamma}^{bias-adj}$ coincide exactly, and seek the conditions required for uniqueness of the corresponding solution. It turns out that such approach is not only meaningful but its (unique) solution is precisely the Shanken estimator $\hat{\Gamma}^*$ in (11), which therefore is the unique additive bias-adjusted CSR OLS estimator that does not require preliminary estimation of the model. This is formalized in the following proposition, the proof of which is reported in Appendix B.

Proposition 1 *Assume that $\hat{\Sigma}_X - \hat{\Lambda}$ is non-singular. Then, the Shanken estimator $\hat{\Gamma}^*$ in (11) is the unique solution of the linear system of equations:*

$$\hat{\Gamma}^* = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\hat{\Lambda}\hat{\Gamma}^*.$$

Although computationally appealing, it remains to verify whether the Shanken estimator $\hat{\Gamma}^*$ exhibits desirable (asymptotic) statistical properties. This is studied in the next section, where we provide a formal asymptotic analysis of $\hat{\Gamma}^*$. As a final precaution, note that the errors-in-variables (EIV) correction in (11) entails subtracting the estimated covariance matrix of the beta estimation errors from $\hat{B}'\hat{B}$ leading to the bias-correction. However, it is possible that this EIV correction will overshoot making the matrix $(\hat{\Sigma}_X - \hat{\Lambda})$ nearly singular or even not positive definite for a given N . To deal with the possibility that the estimator will occasionally produce extreme results, in the simulation and empirical sections of the paper we multiply the matrix $\hat{\Lambda}$ by a scalar k ($0 \leq k \leq 1$), effectively implementing a shrinkage estimator.¹⁰ If k is zero, we get the CSR OLS estimator $\hat{\Gamma}$ back, whereas if k is one we obtain the modified Shanken's estimator $\hat{\Gamma}^*$. The choice of the shrinkage parameter k should be based on the eigenvalues of the matrix $(\hat{\Sigma}_X - k\hat{\Lambda})$. Starting from $k = 1$, if the minimum eigenvalue of this matrix is negative and/or the condition number (i.e. the ratio of

⁹Bai and Zhou (2015) propose to use the CSR OLS $\hat{\Gamma}$ itself as the preliminary estimator plugging it into the above formula in place of $\hat{\Gamma}^{prelim}$. This is justified only when $T \rightarrow \infty$, unlike the framework considered here.

¹⁰Our asymptotic theory would require $k = k_N$ to converge to unity at a suitably slow rate as N increases. We leave out this development to simplify the exposition.

the maximum and minimum eigenvalues) of this matrix is bigger than 20, then we lower k by an arbitrarily small amount. We iterate this procedure until the minimum eigenvalue is positive and the condition number becomes smaller than 20.¹¹ In our simulation experiments, we find that this shrinkage estimator is virtually unbiased leading to $k = 1$. This is mainly due to the fact that in our simulations we encounter very rare instances in which $(\hat{\Sigma}_X - \hat{\Lambda})$ is not positive definite.

We also investigate the limiting statistical properties of a new specification test based on the ex-post sample pricing errors

$$\hat{e}^P = (e_1^P, \dots, e_N^P)' = \bar{R} - \hat{X}\hat{\Gamma}^*, \quad (14)$$

by looking at the limiting behaviour of the (plain) average of the squared ex-post pricing errors, $N^{-1} \sum_{i=1}^N (\hat{e}_i^P)^2$, suitably normalized to converge to a normal distribution as $N \rightarrow \infty$.

Two recent advances challenged the conventional way in which beta-model specification tests have been used with the data. First, Harvey et al (2015) point out that, because of extensive data mining, with hundreds of research papers analyzing the same cross-sectional data of (portfolio) returns, the usual criteria to assess whether one factor is priced in a cross section of asset returns, namely a t-statistic greater than 2, is misleading. Based on multiple testing arguments, they show that instead a cut-off greater than 3 is required, with this cut-off bound to increase over time, as more empirical investigations are carried out. Our approach could mitigate this severe problem in three dimensions: first, it requires to use a panel of data on individual returns, in particular when N is much larger than T . Given that the usual methodologies cannot be used for this type of data set, namely large N and small T , as described in the Introduction, necessarily the risk of data mining is drastically reduced. Moreover, unlike the Fama and French portfolios datasets, a specific dataset of individual asset returns repeatedly examined in empirical asset pricing literature simply does not exist. Finally, the typical size of the standard error for our risk premia estimates, based on the available cross-sections made by thousands of individual returns, implies highly significant t-ratios much larger than 2.

Second, Barillas and Shanken (2016) show that, when comparing beta-pricing models with traded factors only, the choice of the test assets is completely irrelevant with respect to many

¹¹Following Greene (2008), Gagliardini, Ossola, and Scaillet (2015) rely on similar methods to implement their trimming conditions. Alternatively, one can determine k by cross-validation.

popular criteria.¹² Although we do not consider model comparison explicitly, our test statistic can be used for that purpose. However, the important irrelevance result of Barillas and Shanken (2016) will not apply to our framework, regardless of whether the beta-pricing model under consideration includes trading or non-trading factors only. This is due to two main reasons, both consequences of the large N approach; first, the sum of the (plain) squared pricing errors, employed in our test statistic, does not satisfy the irrelevance result, as opposed to the sum of the GLS-weighted squared pricing error as established in Section 4.2 of Barillas and Shanken (2016), building upon the insights of Lewellen, Nagel and Shanken (2010). The GLS-weighting is not feasible in our context because when N is larger than T , the inverse of the residual covariance matrix Σ cannot be estimated unless excessively strong parametric restrictions are imposed. For instance, the sample covariance matrix has rank T , smaller than N , and thus lacks invertibility. Second, our test statistic uses the ex-post pricing errors e_i^P , which are shown below to be approximately equal to $e_i + Q'\epsilon_i$ for large N , with $Q = 1_T/T - \tilde{F}(\tilde{F}'\tilde{F})^{-1}$, where e_i and ϵ_i are the ex-ante pricing errors and the first-pass true residuals, respectively. It follows that the test assets returns, or equivalently, their residuals ϵ_i , appear coupled with the factors' returns through Q , and are not in insulation, ruling out the irrelevance result. In contrast, in the conventional large T framework, the matrix Q vanishes and the irrelevance result applies.

3 Empirical Analysis

In this section, we empirically estimate the risk premia associated with some prominent beta-pricing models using individual stock return data, and investigate their performance. This demonstrates how the empirical results obtained using our large N methodology can differ, even dramatically, from the results obtained with the more traditional large T methodologies.

We consider three linear beta-pricing models: (i) the single-factor CAPM, (ii) the three-factor model of Fama and French (1993, FF3), and (iii) the five-factor model recently proposed by Fama and French (2015, FF5). The five factors entering these empirical specifications are the market excess return (mkt), the return difference between portfolios of stocks with small and large market capitalizations (smb), the return difference between portfolios of stocks with high and low book-

¹²In particular, Barillas and Shanken (2016) establish this irrelevance result with respect to three criteria for model comparison: the GRS test statistics, the likelihood ratio with certain elliptical distributions and the GLS CSR R^2 .

to-market ratios (hml), the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios (rmw), and the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios (cma). The data on the above factors is available from from Kenneth French’s website. We use monthly data on individual stocks from the CRSP database, available from January 1966 until December 2013. We carry out the empirical analysis using balanced panel with three different time windows of, respectively, three-, six- and ten-year (i.e. $T=36, 72$ and 120 , respectively). For each of these time windows, we estimate each of the above beta-pricing model by rolling the window one month at the time. In this way, we obtain time series of estimated risk premia and of the test statistics based on overlapping time windows of fixed length T . After filtering the data, we obtain an average number of approximately 2,800 stocks for the three-year periods, 1,900 for the six-year period and 1,200 for the ten-year period.¹³

We choose to conduct our analysis over relatively short time spans in order to assess the performance of our large- N approach. To this end, we compare our results with the ones obtained using the conventional large- T approach. This is done for both the risk premia estimates and the specification test statistic. In particular, in terms of estimation, we compare the risk premia estimates based on the Shanken estimator $\hat{\Gamma}^*$ with the OLS estimates $\hat{\Gamma}$. For both estimators, we also derive their corresponding 95% confidence intervals. However, whereas for the Shanken estimator we use the large- N standard errors derived in Theorem 1 below, for the OLS estimator we adopt the standard errors of Theorem 1(ii) in Shanken (1992), which are valid for large T only. Hereafter, we will refer to the former confidence intervals as the large- N intervals, while the latter OLS intervals will be denoted as the large- T confidence intervals.

The results are reported in a series of figures.¹⁴ Using a time window of three years ($T = 36$ monthly observations), the top panel of Figure I shows the rolling-window estimates for the risk premium of the market excess return (γ_{mkt}), together with the corresponding 95% confidence intervals, derived using both the large- T and the large- N approach. Using a sample of approximately

¹³Specifically, we download monthly stock data from January 2008 to December 2013 from the CRSP database and apply two filters in the selection of stocks. First, we require that a stock has a Standard Industry Classification (SIC) code (we adopt the 49 industry classifications listed on Kenneth French’s website). Second, we keep a stock in our sample only for the months in which its price is at least 3 dollars. The resulting dataset consists of 3435 individual stocks and we randomly select 3000 of them.

¹⁴Tables with estimates and confidence intervals are available upon request.

$N = 2,800$ stocks, we obtain 541 time series estimates of the market risk premium, from January 1968 to January 2013. The blue line represents the Shanken estimates, $\hat{\gamma}_{mkt}^*$, while the OLS estimates, $\hat{\gamma}_{mkt}$, are represented by the red line. The light blue band represents the 95% confidence interval using the large- N approach whereas the orange band refers to the large- T intervals.

Figure I about here

A striking feature emerges. The large- T confidence intervals include the zero value, implying that we cannot reject (at 95% confidence level) the hypothesis of zero risk premium of the market excess return for most dates in the sample. The only few exceptions are observed between 1975-1981 and 1999-2001. In contrast, our large- N intervals seldom include zero, except for the period around 1975 as well as between 1982 and 1983. More in general, the large- T approach tends to favour the hypothesis of constant risk premia whereas our large- N results points toward a significant time-variation in risk premia. The second and third panels of Figure I report the results for rolling windows of $T=72$ (six years) and $T=120$ (ten years), respectively. As expected, the large- T confidence intervals get narrower, although still sizeably wider than the large- N intervals, yielding risk premia estimates which are now often significant. Finally, the large- N estimates appear systematically larger than the corresponding large- T estimates for most dates, especially when the rolling window is longer. This is the result of the systematic (negative) bias affecting the OLS estimator. At the same time, for long rolling window, the time-variation in estimated risk premia appear much less prominent, as expected.

Figures II.a, II.b, II.c report the rolling window estimates for the risk premia of the market excess return (γ_{mkt}), the *smb* return (γ_{smb}) and the *hml* return (γ_{hml}), respectively, obtained by estimating the FF3 model. The results are aligned with what observed for the CAPM: whereas the large- N results tend to show estimated risk premia statistically different from zero, the large- T approach gives the opposite indications. This divergence is attenuated when the estimation (rolling) window is longer across time. Interestingly, for the *smb* risk premia, even the large- N often leads to estimated risk premia non statistically different from zero especially for the largest estimation window of ten years. For the other two factors, *mkt* and *hml*, the corresponding large- N risk premia estimates appear non-zero in most cases, regardless of the size of the estimation window.

Figure II about here

Finally, Figures III.a, III.b, III.c, III.d and III.e report the estimation results for the FF5 model. The results are analogous to the previous cases for the first three common factors (*mkt*, *smb*, *hml*). For the additional two factors (*rmw* and *cma*), the corresponding large- T risk premia estimates are non-significant for most periods regardless of the estimation window. In contrast, the large- N procedure shows that the estimates of these risk premia are in fact significantly different from zero for about half of the cases, the evidence being somewhat stronger for the *rmw* risk premium.

Figure III about here

We finally consider the performance of our specification test, described in detail in Section 4.4 below. We report only the results for the CAPM, in Figure IV and Figure V.¹⁵ We compare the p-values of our test, based on the asymptotic distribution for large N (derived in Theorem 3 below), with the p-values associated with the Gibbons, Ross, and Shanken (1989) test (hereafter GRS), which is a common testing procedure valid for large T and fixed (moderate) N . For both ours and the GRS test, the null hypothesis is $H_0 : e_i = 0$ for every asset i , namely that the beta-pricing model is correctly specified.

The black line in Figure IV denotes the time series of p-values associated with our test statistic for the time windows of three years (top panel), six years (middle panel) and ten years (bottom panel), respectively. A black line below the dotted line (that indicates the value of 0.05) means that the CAPM, estimated over that particular sample period, is rejected at 5% level. Figure IV clearly shows that our test rejects the validity of the CAPM for the large majority of the data periods, even for the shortest time window of $T = 36$. At the same time, as expected, rejection of the CAPM happens more frequently as the time window increases from $T=36$ to $T=120$.

Figure IV about here

Figure V reports, instead, the p-value corresponding to the GRS test applied to 25 portfolios, as opposed to the thousands of individual stocks used by our test. Indeed, in order to facilitate the comparison, based on our data of individual asset returns, we have constructed 25 time series of portfolios, by taking simple average of the appropriate number of stocks.¹⁶ The result is, again,

¹⁵The results are available for the FF3 and FF5 model upon request.

¹⁶For instance, when one considers $T = 36$ each portfolio is made by approximately 110 stocks.

rather striking: in contrast to our large- N test, the GRS test is almost always unable to reject the CAPM at 5% when considering the shortest time window of three years of data ($T = 36$). The CAPM will be rejected about half of the times for the six years window and will almost always be rejected for the long time window of ten years.

Figures V about here

4 Theoretical Analysis

The analysis in this section assumes that $N \rightarrow \infty$ and T is fixed. We first establish the limiting distribution of Shanken’s bias-adjusted estimator and explain how its asymptotic variance can be consistently estimated. We then characterize the limiting behavior of a test of the asset-pricing restriction, viz. Assumption 4 below, deploying the estimated ex-post pricing errors \hat{e}^P associated with the Shanken estimator. Finally, we show how our analysis can be extended to deal with an unbalanced panel.

4.1 Assumptions

The following are our main regularity assumptions. Further technical assumptions are relegated to Appendix A.

Assumption 1. (*loadings*) As $N \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N \beta_i \rightarrow \mu_\beta, \tag{15}$$

$$\frac{1}{N} \sum_{i=1}^N \beta_i \beta_i' \rightarrow \Sigma_\beta, \tag{16}$$

where Σ_β is a finite symmetric and positive-definite matrix.

The first part of Assumption 1 states that the limiting cross-sectional average of the betas exists, while the second part states that the limiting cross-sectional average of squared betas exists and is a symmetric and positive-definite matrix. To simplify the exposition of our theory, we are not assuming the β_i to be random (see Gagliardini et al (2015) for a beta-pricing model with random betas). By construction this implies that the β_i are cross-sectionally unrelated to any other

characteristics of the returns' distribution, in particular to the returns' idiosyncratic innovations ϵ_{it} . In Section 3.2 we will discuss the consequences of relaxing this assumption.

Assumption 2. (*residuals; see Shanken (1992)*) Assume that the vector ϵ_t is independently and identically distributed (*i.i.d.*) over time with

$$E[\epsilon_t|F] = 0_N \quad (17)$$

and

$$\text{Var}[\epsilon_t|F] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{pmatrix} = \Sigma \quad \text{is of rank } N, \quad (18)$$

where $F = [f_1, \dots, f_T]'$ is the $T \times K$ matrix of factors, 0_N is a N -vector of zeros and σ_{ij} denotes the (i, j) th element of Σ , with $i, j = 1, \dots, N$. Hereafter we denote $\sigma_i^2 = \sigma_{ii}$.

The *i.i.d.* assumption over time is common to many studies, including Shanken (1992). This assumption could be relaxed at the cost of more cumbersome derivations. Conditions (17) and (18) are implied if the factors f_t and the innovations ϵ_s are mutually independent for any s, t . Noticeably, (18) is not imposing any structure on the elements of Σ . In particular, we are not imposing the returns R_{it} to be uncorrelated across assets or to exhibit the same variance. Instead we allow for a substantial degree of heterogeneity in the cross-section of stock returns. Although the expression for Σ is here left unspecified, obviously our asymptotic theory limits the degree of cross-correlation between the residuals ϵ_{it} (see Assumption 5 in Appendix A). Essentially, we will require that the sum of the $|\sigma_{ij}|$ across every row (or column) of Σ is bounded, that is $\sup_{1 \leq j \leq N} \sum_{i=1}^N |\sigma_{ij}| \leq C < \infty$. This condition is slightly stronger than (and thus implies) the corner-stone assumption of the APT which requires Σ to have bounded maximum eigenvalue (see Chamberlain and Rothschild, 1983).

Regarding the observed factors, we require very minimal assumptions because our asymptotic analysis holds conditional on the factors realization F , with a fixed T . The following is the only assumption we make.

Assumption 3. (*factors*) Assume that $E[f_t] = E[f]$ does not vary over time. Moreover, for every $T \geq K$,

$$\tilde{F}'\tilde{F} \quad \text{is of full rank,}$$

where $\tilde{F} = (I_T - \frac{1_T 1_T'}{T})F = F - 1_T \bar{f}'$ and $I_T, 1_T$ are the $T \times T$ identity matrix and the T -dimensional

column vector of ones, respectively, implying that \bar{f} is the sample mean of the f_t .

Finally, to close the asset-pricing model, one must postulate the form of no-arbitrage required. All the results developed in this paper assume that exact pricing holds which, in view of the constant mean assumption for f_t , can be expressed as follows:

Assumption 4. (*exact pricing*)

$$E[R_t] = X\Gamma, \quad (19)$$

where $X = [1_N, B]$ is assumed to be of full column rank for every N , 1_N is the N -dimensional column vector of ones, and $\Gamma = [\gamma_0, \gamma_1]'$ is a vector consisting of the zero-beta rate (γ_0) and ex-ante risk premia associated to the K factors (γ_1).

When the model is misspecified, Assumption 4 is not satisfied and the N -vector of pricing errors, $e = E[R_t] - X\Gamma$, will be different from the N -dimensional zero vector. Throughout this paper, we will only explore non-zero pricing errors when providing a sufficient condition for power of our asset pricing test.

4.2 Asymptotic Distribution of the Bias-Adjusted OLS Estimator

In this subsection, we study the asymptotic distribution of $\hat{\Gamma}^*$ under the assumption that the model is correctly specified, namely that exact no-arbitrage holds (Assumption 4). Let $\Sigma_X = \begin{bmatrix} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta \end{bmatrix}$, $\sigma^2 = \lim \frac{1}{N} \sum_{i=1}^N \sigma_i^2$, $U_\epsilon = \lim \frac{1}{N} \sum_{i,j=1}^N E \left[\text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \text{vec}(\epsilon_j \epsilon_j' - \sigma_j^2 I_T)' \right]$, $M = I_T - D(D'D)^{-1}D'$, where I_T is a $T \times T$ identity matrix, $D = [1_T, F]$, $Q = \frac{1_T}{T} - \mathcal{P}\gamma_1^P$, and $Z = (Q \otimes \mathcal{P}) + \frac{\text{vec}(M)}{T-K-1} \gamma_1^{P'} \mathcal{P}' \mathcal{P}$, where all the limits are finite by our assumptions, as $N \rightarrow \infty$. In the following theorem, we provide the rate of convergence and the limiting distribution of $\hat{\Gamma}^*$.

Theorem 1

(i) Under Assumptions 1–4 and 5 (listed in Appendix A),

$$\hat{\Gamma}^* - \Gamma^P = O_p \left(\frac{1}{\sqrt{N}} \right). \quad (20)$$

(ii) Under Assumptions 1–4 and 5–6 in (listed in Appendix A),

$$\sqrt{N} \left(\hat{\Gamma}^* - \Gamma^P \right) \xrightarrow{d} \mathcal{N} \left(0_{K+1}, V + \Sigma_X^{-1} W \Sigma_X^{-1} \right), \quad (21)$$

where

$$V = \frac{\sigma^2}{T} \left[1 + \gamma_1^{P'} \left(\tilde{F}' \tilde{F} / T \right)^{-1} \gamma_1^P \right] \Sigma_X^{-1} \quad (22)$$

and

$$W = \begin{bmatrix} 0 & 0'_K \\ 0_K & Z' U_\epsilon Z \end{bmatrix}. \quad (23)$$

Proof: See Appendix B and Lemmas 1 to 5 in Appendix A.

Note that the expression in (21) for the asymptotic covariance matrix is very simple and, moreover, it has a very neat interpretation. The first term of this asymptotic variance, V , accounts for the estimation error in the betas, and it is essentially identical to the large T expression of the correct asymptotic covariance matrix associated with the CSR OLS estimator, derived by Shanken [Theorem 1(ii)](1992): part $\frac{\sigma^2}{T} \Sigma_X^{-1}$ is the classical CSR OLS covariance matrix, which one would obtain if the betas were observed, whereas part $c = \gamma_1^{P'} \left(\tilde{F}' \tilde{F} / T \right)^{-1} \gamma_1^P$ is an asymptotic adjustment for EIV, with $c \frac{\sigma^2}{T} \Sigma_X^{-1}$ being the corresponding overall EIV contribution of the asymptotic covariance matrix. As Shanken (1992) points out, the EIV adjustment reflects the fact that the variability of the estimated betas is directly related to residual variance, σ^2 , and inversely related to factor variability $\left(\tilde{F}' \tilde{F} / T \right)^{-1}$. The last term of the asymptotic covariance, $\Sigma_X^{-1} W \Sigma_X^{-1}$, arises because of the bias adjustment that characterises exclusively $\hat{\Gamma}^*$ (but not $\hat{\Gamma}$) which also vanishes when $T \rightarrow \infty$. In addition, the W matrix also accounts for the cross-sectional variation in the residual variances of asset returns through U_ϵ .

To conduct statistical inference, we need a consistent estimator of the asymptotic covariance matrix $V + \Sigma_X^{-1} W \Sigma_X^{-1}$. Let $M^{(2)} = M \odot M$, where \odot denotes the Hadamard product operator. In addition, define

$$\hat{\sigma}_4 = \frac{\frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{\epsilon}_{it}^4}{3 \text{tr} (M^{(2)})}. \quad (24)$$

and let

$$\hat{Z} = (\hat{Q} \otimes \mathcal{P}) + \frac{\text{vec}(M)}{T - K - 1} \hat{\gamma}_1^{*'} \mathcal{P}' \mathcal{P} \quad (25)$$

with

$$\hat{Q} = \frac{1_T}{T} - \mathcal{P} \hat{\gamma}_1^*. \quad (26)$$

The following theorem provides a consistent estimator of the asymptotic covariance matrix of the $\hat{\Gamma}^*$ estimator.

Theorem 2 Under Assumptions 1 to 3 and 4–5 (listed in Appendix A), we have

$$\hat{V} + \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \hat{W} \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \xrightarrow{p} V + \Sigma_X^{-1} W \Sigma_X^{-1}, \quad (27)$$

where

$$\hat{V} = \frac{\hat{\sigma}^2}{T} \left[1 + \hat{\gamma}_1^{*'} \left(\tilde{F}' \tilde{F} / T \right)^{-1} \hat{\gamma}_1^* \right] \left(\hat{\Sigma}_X - \hat{\Lambda} \right)^{-1}, \quad (28)$$

$$\hat{W} = \begin{bmatrix} 0 & 0'_K \\ 0_K & \hat{Z}' \hat{U}_\epsilon \hat{Z} \end{bmatrix}, \quad (29)$$

and \hat{U}_ϵ is a consistent plug-in estimator of U_ϵ described in Appendix C.

Proof: See Appendix B and Lemmas 1 to 6 in Appendix A.

A remarkable feature of the above result is that a consistent estimate of the asymptotic covariance matrix of $\hat{\Gamma}^*$ could be obtained, whilst leaving the residual covariance matrix Σ unspecified. In fact, with Σ having in general $N(N+1)/2$ distinct elements and our asymptotic theory only allowing $N \rightarrow \infty$, it follows that consistent estimation of Σ is completely unfeasible, a phenomenon known in econometrics as the curse-of-dimensionality. However, a key feature of the Shanken's estimator is that it depends on Σ only through $\sum_{i=1}^N \sigma_i^2 / N$. Moreover, its asymptotic covariance matrix depends on the average $\sum_{i,j=1}^N \sigma_{ij} / N$. Our large N asymptotic theory shows how both quantities can be estimated consistently, unlike for the individual covariances σ_{ij} . This crucial feature of the Shanken's estimator is in sharp contrast to the GLS estimator which requires to estimate consistently every element of Σ .

4.3 Consequences of random loadings β_i

We now discuss the consequences of allowing for random β_i . Consider at first the case when these, although random, are mutually independent to any other cross-sectional characteristic of the individual asset returns. Then, in this case, no consequences arise in terms of the asymptotic properties of the Shanken estimator $\hat{\Gamma}^*$. The only, marginal, changes involve Assumptions 1 and 4. In particular, equations (15) and (16) in Assumption 1 must be stated in terms of convergence in probability, instead of the conventional convergence valid of non random sequences; equation (19) in Assumption 4 must be replaced by $E[R_t|X] = X\Gamma$. All the other assumptions are unchanged

except that now (A.11) involves random β_i . Then, by easy calculations,

$$\lim \text{Var}\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(C'_T \otimes \begin{pmatrix} 1 \\ \beta_i \end{pmatrix}\right) \epsilon_i\right) \quad (30)$$

$$= \lim \frac{1}{N} \sum_{i,j=1}^N E\left(C'_T \otimes \begin{pmatrix} 1 \\ \beta_i \end{pmatrix}\right) \epsilon_i \epsilon'_j \left(C_T \otimes (1 \ \beta'_j)\right) \quad (31)$$

$$= (C'_T C_T) \lim \frac{1}{N} \sum_{i=1}^N \sigma_i^2 E\left(\begin{pmatrix} 1 \\ \beta_i \end{pmatrix} (1 \ \beta'_i)\right) + (C'_T C_T) \lim \frac{1}{N} \sum_{\substack{i \neq j \\ i=1}}^N \sigma_{ij} E\left(\begin{pmatrix} 1 \\ \beta_i \end{pmatrix} (1 \ \beta'_j)\right) \quad (32)$$

$$= (C'_T C_T) \sigma^2 \Sigma_X. \quad (33)$$

In fact the second term on the right hand side of (32) converges to zero under our assumptions, given that $E\|\beta_i \beta'_j\| \leq E(\beta'_i \beta_i)^{\frac{1}{2}} E(\beta'_j \beta_j)^{\frac{1}{2}} \leq C < \infty$ and $\sum_{\substack{i \neq j \\ i=1}}^N |\sigma_{ij}| = o(N)$. Expression (33) coincides exactly with the asymptotic covariance matrix of Theorem 1, which holds for non-random β_i .

Consider now the case when the β_i are potentially cross-sectionally correlated with the ϵ_i . We argue that such covariance structure could not be identified based on the OLS estimators $\hat{\beta}_i$ and $\hat{\epsilon}_i$, either for finite or arbitrarily large N , implying that the possibility of cross-correlation should be ruled out. In fact, inspecting the proof of Theorem 1, it turns out that the asymptotic distribution of $\sqrt{N}(\hat{\Gamma}^* - \Gamma^P)$ depends, among others, on $N^{-\frac{1}{2}} \sum_{i=1}^N \beta_i \epsilon'_i Q$ where we recall that $Q = \frac{1}{T} - \mathcal{P}\gamma_1^P$.

Let $\delta_i \equiv E\beta_i \epsilon_{it} = \text{Cov}(\beta_i, \epsilon_{it})$ implying

$$E\beta_i \epsilon'_i \equiv \Delta_i = \delta_i 1'_T, \quad (34)$$

where the second equality follows by the identically-distributed assumption of the ϵ_{it} across time.

Then

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \beta_i \epsilon'_i Q = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\beta_i \epsilon'_i - \Delta_i) Q + \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta_i Q.$$

The first term on the right hand side of the last equation will converge to a normal distribution, by a simple generalization of Assumption 6-(iii). Given (34), the bias term can be re-written as $\sqrt{N}^{-1} \sum_{i=1}^N \Delta_i Q = \sqrt{N}^{-1} \sum_{i=1}^N \delta_i 1'_T Q = \sqrt{N}^{-1} \sum_{i=1}^N \delta_i$ because $1'_T Q = 1$. It is evident that in order to avoid an asymptotic bias for $\hat{\Gamma}^*$, although each δ_i could be non-zero, their average must satisfy $\sqrt{N}^{-1} \sum_{i=1}^N \delta_i = o(1)$, including the case $\delta_i = 0_K$. We now illustrate how this restriction is naturally asked for when considering the OLS estimator of δ_i .

In particular, the OLS estimator of $N^{-1} \sum_{i=1}^N \Delta_i = N^{-1} \sum_{i=1}^N E(\beta_i \epsilon'_i)$ will be $N^{-1} \sum_{i=1}^N \hat{\Delta}_i$ with $\hat{\Delta}_i \equiv \hat{\beta}_i \hat{\epsilon}'_i$. However, by easy calculations, one can show that $\hat{\epsilon}_i$ and Q are orthogonal for any

finite T (and N). Set $M_A = I - A(A'A)^{-1}A'$ for any generic, full row-rank, A , implying that M_A is the projection matrix that projects onto the space orthogonal to the space spanned by the columns of A . Then

$$\hat{\epsilon}'_i Q = \epsilon'_i M Q = \epsilon'_i M_{\tilde{F}} M_{1_T} Q = -\epsilon'_i M_{\tilde{F}} M_{1_T} \mathcal{P} \gamma_1^P = \epsilon'_i M_{\tilde{F}} \mathcal{P} \gamma_1^P = 0,$$

since $M_{1_T} 1_T = 0$ and $M_{\tilde{F}} \tilde{F} = 0$, where we use the identity $M = M_{\tilde{F}} M_{1_T} = M_{1_T} M_{\tilde{F}}$, where using our notation $M = M_D$ with $D = (1_T, F)$. Therefore, the estimated bias term $N^{-1} \sum_{i=1}^N \hat{\Delta}_i Q$ is identically zero for any finite N . On the other hand, even without the post-multiplication by Q , as N diverges

$$\frac{1}{N} \sum_{i=1}^N \hat{\beta}_i \hat{\epsilon}'_i = \frac{1}{N} \sum_{i=1}^N (\beta_i + \mathcal{P}' \epsilon_i) \epsilon'_i M \quad (35)$$

$$= \left(\frac{1}{N} \sum_{i=1}^N \beta_i \epsilon'_i \right) M + \frac{\mathcal{P}'}{N} \left(\sum_{i=1}^N \epsilon_i \epsilon'_i \right) M \quad (36)$$

$$\rightarrow_p \delta 1'_T M + \sigma^2 \mathcal{P}' M = 0, \quad (37)$$

regardless of whether $\delta \equiv \lim N^{-1} \sum_{i=1}^N \delta_i$ is zero or not, because both $1'_T M = 0$ and $\mathcal{P}' M = 0$.

Therefore, both the finite N and the large N arguments strongly suggest that the assumption:

$$\delta_i = \text{Cov}(\beta_i, \epsilon_{it}) = 0_K,$$

is plausible in our large N environment, or alternatively the slightly more general case $\sqrt{N}^{-1} \sum_{i=1}^N \delta_i = o(1)$.

4.4 Limiting Distribution of the Specification Test

In this section, we are interested in deriving an OLS-type test of the validity of the beta-pricing model, based on the Shanken's estimator. The null hypothesis underlying the asset-pricing restriction can be formulated as

$$H_0 : e_i = 0 \quad \text{for every } i = 1, 2, \dots, \quad (38)$$

where $e_i = E[R_{it}] - \gamma_0 - \beta'_i \gamma_1$ is the pricing error associated with asset i . The null hypothesis H_0 easily follows by simple re-writing of Assumption 4. Let $X_i = [1, \beta'_i]$, $\hat{X}_i = [1, \hat{\beta}'_i]$, and denote by

\hat{e}_i^P the ex-post sample pricing error for asset i . Then, we have

$$\hat{e}_i^P = \bar{R}_i - \hat{X}_i \hat{\Gamma}^* \quad (39)$$

$$= e_i + Q' \epsilon_i - \hat{X}_i (\hat{\Gamma}^* - \Gamma^P). \quad (40)$$

It follows that

$$\hat{e}_i^P \xrightarrow{P} e_i + Q' \epsilon_i \equiv e_i^P. \quad (41)$$

Equation (41) shows that even when the ex-ante pricing error, e_i , is zero, \hat{e}_i^P will not converge in probability to zero. This is a consequence of the fact that, when T is fixed, $Q' \epsilon_i = \bar{e}_i$ will not converge to zero even under the null of zero ex-ante pricing errors. This is the price that we have to pay when N is large and T is fixed. Nonetheless, a test of H_0 with good size and power properties can be developed. Since we estimate Γ^P via OLS cross-sectional regressions, we propose a test based on the sum of the squared ex-post sample pricing errors, that is,

$$\hat{Q} = \frac{1}{N} \sum_{i=1}^N (\hat{e}_i^P)^2. \quad (42)$$

Consider the centered statistic

$$\mathcal{S} = \sqrt{N} \left(\hat{Q} - \frac{\hat{\sigma}^2}{T} \left(1 + \hat{\gamma}_1^{*'} (\tilde{F}' \tilde{F} / T)^{-1} \hat{\gamma}_1^* \right) \right). \quad (43)$$

The following theorem provides the limiting distribution of \mathcal{S} under $H_0 : e_i = 0$ for all i .

Theorem 3 *Under Assumptions 1 to 4 and 5–6 (listed in Appendix A), where by Assumption 4 $H_0 : e_i = 0$ holds for all i , we have*

$$\mathcal{S} \xrightarrow{d} \mathcal{N}(0, \mathcal{V}), \quad (44)$$

where $\mathcal{V} = Z_Q' U_\epsilon Z_Q$ and $Z_Q = (Q \otimes Q) - \frac{\text{vec}(M)}{T-K-1} Q' Q$.

Proof: See Appendix B and Lemmas 1 to 5 in Appendix A.

The expression for the asymptotic variance of the test in (44) is rather simple. This variance can be consistently estimated by replacing Q with \hat{Q} and U_ϵ with \hat{U}_ϵ . Specifically, using Theorem 2 and Lemma 6 in Appendix A, we have

$$\hat{Z}_Q' \hat{U}_\epsilon \hat{Z}_Q \xrightarrow{P} Z_Q' U_\epsilon Z_Q, \quad (45)$$

where

$$\hat{Z}_Q = \left(\hat{Q} \otimes \hat{Q} \right) - \frac{\text{vec}(M)}{T - K - 1} \hat{Q}' \hat{Q}. \quad (46)$$

Then, under H_0 , it follows that

$$\frac{\mathcal{S}}{(\hat{Z}'_Q \hat{U}_\epsilon \hat{Z}_Q)^{\frac{1}{2}}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (47)$$

It turns out that our test will have power when e_i^2 is greater than zero for the majority of the assets.¹⁷ In the next section, we will undertake a Monte Carlo simulation experiment calibrated to real data in order to determine whether our test possesses desirable power properties.

Our test is not immune to repackaging, namely to arbitrary linear combinations of the test assets. Typically, to allow for re-packaging one should standardise the pricing errors by Σ^{-1} . See Lewellen, Nagel and Shanken (2010) for more formal arguments in favour of this approach.¹⁸ However, estimating Σ^{-1} is not possible unless strong restrictions are imposed such as, for example, when it is assumed diagonal or spherical. For instance, the $N \times N$ sample covariance matrix of the OLS residuals $\hat{e}'\hat{e}/T$ has rank $T < N$ implying singularity.

4.5 Unbalanced Panel

In this section, we extend our methodological results to the case of an unbalanced panel. Following Gagliardini, Ossola, and Scaillet (2015), we assume a missing at random design (see, for example, Rubin, 1976), that is independence between unobservability and return generating process. This allows us to keep the factor structure linear. In the following analysis, we explicitly account for the randomness of T_i , the time-series sample size for asset i . Define the following $T \times T$ matrix

$$J_i = \text{diag}(J_{i1} \dots J_{it} \dots J_{iT}) \quad i = 1, \dots, N, \quad (48)$$

where $J_{it} = 1$ if the return on asset i is observed by the econometrician at date t , and zero otherwise. We assume that J_{it} is *i.i.d.* across i and t . In addition, let $R_{i,u} = J_i R_i$, $F_{i,u} = J_i F$, and $\epsilon_{i,u} = J_i \epsilon_i$, and assume that asset returns are governed by the multifactor model

$$J_{it} R_{it} = J_{it} \alpha_i + J_{it} f'_t \beta_i + J_{it} \epsilon_{it}, \quad (49)$$

¹⁷To be precise, the pricing errors e_i can be zero for only a number N_0 of assets, such that $N_0/N \rightarrow 0$ as $N \rightarrow \infty$. This condition allows N_0 to diverge although not too fast.

¹⁸Consider a (non-singular) linear combinations of the test assets such as AR_t for an arbitrary non-singular matrix A , implying that the residuals satisfy $A\epsilon_t$. Then, the ex-post pricing errors would become Ae^p and their squared sum $e^{p'} A' A e^p$, depending on the choice of the matrix A . Instead the weighted sum, by the inverse of the residual covariance matrix, becomes $e^{p'} A' (A \Sigma A')^{-1} A e^p = e^{p'} A' (A')^{-1} \Sigma^{-1} A^{-1} A e^p = e^{p'} \Sigma^{-1} e^p$, regardless of the chosen A .

that is, the same data generating process of the previous section multiplied by J_{it} . Let $\bar{R}_{i,u} = \frac{1}{T_i} \sum_{t=1}^T J_{it} R_{it}$, $\bar{f}_{i,u} = \frac{1}{T_i} \sum_{t=1}^T J_{it} f_t$, and $\bar{\epsilon}_{i,u} = \frac{1}{T_i} \sum_{t=1}^T J_{it} \epsilon_{it}$. Averaging (49) over time, imposing the asset-pricing restriction, and noting that $E[R_{it}] = \alpha_i + \beta'_i E[f]$ yields

$$\bar{R}_{i,u} = \gamma_0 + \hat{\beta}'_{i,u} \gamma_{1i,u}^P + \eta_{i,u}^P, \quad (50)$$

where $\gamma_{1i,u}^P = \gamma_1 + \bar{f}_{i,u} - E[f]$, $\eta_{i,u}^P = \bar{\epsilon}_{i,u} - (\hat{\beta}_{i,u} - \beta_i)' \gamma_{1i,u}^P$, $\hat{\beta}_{i,u} = \beta_i + \mathcal{P}'_{i,u} \epsilon_i$, $\mathcal{P}_{i,u} = \tilde{F}'_{i,u} (\tilde{F}'_{i,u} \tilde{F}_{i,u})^{-1}$, and $\tilde{F}_{i,u} = F_{i,u} - J_i 1_T \bar{f}'_{i,u}$. Since the panel is unbalanced, there is now a sequence of ex-post risk premia, one for each asset i .

In matrix form, we have

$$\bar{R}_u = \gamma_0 1_N + \begin{bmatrix} \hat{\beta}'_{1,u} & & 0'_{K(N-1)} \\ \vdots & \ddots & \vdots \\ 0'_{K(N-1)} & & \hat{\beta}'_{N,u} \end{bmatrix} \begin{bmatrix} \gamma_{11,u}^P \\ \vdots \\ \gamma_{1N,u}^P \end{bmatrix} + \begin{bmatrix} \eta_{1,u}^P \\ \vdots \\ \eta_{N,u}^P \end{bmatrix}, \quad (51)$$

where $\bar{R}_u = (\bar{R}_{1,u}, \dots, \bar{R}_{N,u})'$. Define the $N \times K$ matrix $\hat{X}_u = [1_N, \hat{B}_u]$, where $\hat{B}_u = (\hat{\beta}_{1,u}, \dots, \hat{\beta}_{N,u})'$. Denote by $\hat{\epsilon}_{i,u}$ the T -vector of residuals from the first-pass (unbalanced) OLS regressions in

$$R_{i,u} = \alpha_i J_i 1_T + F_{i,u} \beta_i + \epsilon_{i,u}, \quad i = 1, \dots, N. \quad (52)$$

The proposed modified estimator of the ex-post risk premia in the unbalanced panel case is

$$\hat{\Gamma}_u^* = \begin{bmatrix} \hat{\gamma}_{0,u}^* \\ \hat{\gamma}_{1,u}^* \end{bmatrix} = \left(\hat{\Sigma}_{X,u} - \hat{\Lambda}_u \right)^{-1} \frac{\hat{X}'_u \bar{R}_u}{N}, \quad (53)$$

where

$$\hat{\Sigma}_{X,u} = \frac{\hat{X}'_u \hat{X}_u}{N}, \quad \hat{\Lambda}_u = \begin{bmatrix} 0 & 0'_K \\ 0_K & \hat{\sigma}_u^2 \hat{\mathcal{F}}_u \end{bmatrix}, \quad (54)$$

$$\hat{\sigma}_u^2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T_i - K - 1} \text{tr}(\hat{\epsilon}_{i,u} \hat{\epsilon}'_{i,u}) \right), \quad (55)$$

and

$$\hat{\mathcal{F}}_u = \frac{1}{N} \sum_{i=1}^N \left(\tilde{F}'_{i,u} \tilde{F}_{i,u} \right)^{-1}. \quad (56)$$

The estimator $\hat{\Gamma}_u^*$ in (53) generalizes the modified estimator of Shanken (1992) to the unbalanced panel case and coincides with the Shanken's estimator when the panel is balanced. Let $\tau = E[1/T_i]$,

$\theta = \text{Var} \left(\frac{J_{it}}{T_i} \right)$, and assume that these moments exist. In addition, define $\Sigma_{X,i} = \begin{bmatrix} 1 & \beta_i' \\ \beta_i & \beta_i \beta_i' \end{bmatrix}$ and let $\Sigma_{F\beta} = \text{plim} \frac{1}{N} \sum_{i=1}^N \beta_i' F' F \beta_i \Sigma_{X,i}$, $\mathcal{F}_u = \text{plim} \frac{1}{N} \sum_{i=1}^N \mathcal{P}'_{i,u} \mathcal{P}_{i,u}$, and $Q_{i,u} = \frac{J_{i1T}}{T_i} - \mathcal{P}_{i,u} \gamma_1^P$. Finally, define $Z_{i,u} = \left[\left(Q_{i,u} \otimes \mathcal{P}_{i,u} \right) + \frac{\text{vec}(M_{i,u})}{T_i - K - 1} \gamma_1^{P'} \mathcal{P}'_{i,u} \mathcal{P}_{i,u} \right]$ and $M_{i,u} = [I_T - J_i D (D' J_i D)^{-1} D' J_i] J_i$. The consistency and asymptotic normality of the proposed estimator are provided in the following theorem.

Theorem 4

(i) Under Assumptions 1–4 and 5 (listed in Appendix A),

$$\hat{\Gamma}_u^* - \Gamma^P = O_p \left(\frac{1}{\sqrt{N}} \right). \quad (57)$$

(ii) Under Assumptions 1–4 and 5–6 (listed in Appendix A),

$$\sqrt{N} \left(\hat{\Gamma}_u^* - \Gamma^P \right) \xrightarrow{d} \mathcal{N} \left(0_{K+1}, V_u + \Sigma_X^{-1} (W_u + \Theta) \Sigma_X^{-1} \right), \quad (58)$$

where

$$V_u = \sigma^2 \left(\tau + \gamma_1^{P'} \mathcal{F}_u \gamma_1^P \right) \Sigma_X^{-1}, \quad (59)$$

$$W_u = \begin{bmatrix} 0 & 0'_K \\ 0_K & \text{plim} \frac{1}{N} \sum_{i=1}^N Z'_{i,u} U_\epsilon Z_{i,u} \end{bmatrix}, \quad (60)$$

$$\Theta = \theta \Sigma_{F\beta} - \sigma^2 \Psi, \quad (61)$$

with

$$\Psi = \begin{bmatrix} 0 & \gamma_1^{P'} \mathcal{F}_\gamma \\ \mathcal{F}_\gamma \gamma_1^P & \mathcal{F}_{\gamma\beta} \end{bmatrix}, \quad (62)$$

$$\mathcal{F}_\gamma = \text{plim} \frac{1}{N} \sum_{i=1}^N \mathcal{P}'_{i,u} \mathcal{P}_{i,u} (\bar{f}_{i,u} - \bar{f})' \beta_i, \quad (63)$$

and

$$\begin{aligned} \mathcal{F}_{\gamma\beta} &= \text{plim} \frac{1}{N} \sum_{i=1}^N (\beta_i \beta_i' (\bar{f}_{i,u} - \bar{f}) \gamma_1^{P'} \mathcal{P}'_{i,u} \mathcal{P}_{i,u} + \mathcal{P}'_{i,u} \mathcal{P}_{i,u} \gamma_1^P (\bar{f}_{i,u} - \bar{f})' \beta_i \beta_i' \\ &\quad - (\bar{f}_{i,u} - \bar{f})' \beta_i \beta_i' (\bar{f}_{i,u} - \bar{f}) \mathcal{P}'_{i,u} \mathcal{P}_{i,u}). \end{aligned} \quad (64)$$

Proof: See Online Appendix.

It should be noted that the asymptotic covariance matrix in Theorem 4 is similar to the one for the balanced panel case provided in Theorem 1, and it still has a sandwich form. The additional terms in part (ii) of Theorem 4 account for the randomness of the sample size T_i . When the panel is balanced, Theorem 4 reduces to Theorem 1 since $T_i = T$, $J_{it} = 1$, $\bar{f}_{i,u} = \bar{f}$, which implies that $\tau = 1/T$, $\theta = 0$, $\Psi = \Theta = 0_{(K+1) \times (K+1)}$, and all the relevant quantities do not depend on i anymore.

In conducting statistical tests, we need a consistent estimator of the asymptotic covariance matrix in Theorem 4(ii). Let $\hat{\tau} = \frac{1}{N} \sum_{i=1}^N \frac{1}{T_i}$, $\hat{\Sigma}_{X,i}^a = \begin{bmatrix} 1 & \hat{\beta}'_{i,u} \\ \hat{\beta}_{i,u} & \hat{\Sigma}_{\hat{\beta}_{i,u}}^a \end{bmatrix}$, where $\hat{\Sigma}_{\hat{\beta}_{i,u}}^a = \hat{\beta}_{i,u} \hat{\beta}'_{i,u} - \hat{\sigma}_u^2 \mathcal{P}'_{i,u} \mathcal{P}_{i,u}$, $\hat{b}_i = \text{tr}(F' F \hat{\Sigma}_{X,i}^a)$, and $A_i = \mathcal{P}'_{i,u} \mathcal{P}_{i,u} F' F$. Also, let $\hat{U}_i = \sum_{t=1}^T (\mathcal{P}'_{i,u} \otimes f'_t \mathcal{P}'_{i,u}) \hat{U}_\epsilon (\mathcal{P}_{i,u} \otimes \mathcal{P}_{i,u} f_t)$, where \hat{U}_ϵ (as in the balanced panel case) is a plug-in estimator of U_ϵ that depends only on $\hat{\sigma}_{4,u} = \frac{\frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{\epsilon}_{it,u}^4}{3 \frac{1}{N} \sum_{i=1}^N \text{tr}(M_{i,u}^{(2)})}$, with $\hat{\epsilon}_{it,u}$ being the t -th element of $\hat{\epsilon}_{i,u}$ and $M_{i,u}^{(2)} = M_{i,u} \odot M_{i,u}$. Finally, let $\hat{\Sigma}_{F\beta} = \frac{1}{N} \sum_{i=1}^N \hat{b}_i \hat{\Sigma}_{X,i}^a - \hat{\Upsilon}$, where $\hat{\Upsilon} = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} 0 & 2\hat{\sigma}_u^2 \hat{\beta}'_{i,u} A'_i \\ 2\hat{\sigma}_u^2 A_i \hat{\beta}_{i,u} & 2\hat{\sigma}_u^2 (A_i \hat{\Sigma}_{\hat{\beta}_{i,u}}^a + \hat{\Sigma}_{\hat{\beta}_{i,u}}^a A'_i) + \hat{U}_i \end{bmatrix}$, $\hat{\theta} = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \frac{J_{it}}{T_i^2} - \frac{1}{T^2}$, and $\hat{Z}_{i,u} = \left[(\hat{Q}_{i,u} \otimes \mathcal{P}_{i,u}) + \frac{\text{vec}(M_{i,u})}{T_i - K - 1} \hat{\gamma}_{1,u}^* \mathcal{P}'_{i,u} \mathcal{P}_{i,u} \right]$, where $\hat{Q}_{i,u} = \frac{J_i 1_T}{T_i} - \mathcal{P}_{i,u} \hat{\gamma}_{1,u}^*$.

The following theorem provides a consistent estimator of the asymptotic covariance matrix of the estimates.

Theorem 5 *Under Assumptions 1 to 4 and 5-6 (listed in Appendix A), we have*

$$\hat{V}_u + \left(\hat{\Sigma}_{X,u} - \hat{\Lambda}_u \right)^{-1} (\hat{W}_u + \hat{\Theta}) \left(\hat{\Sigma}_{X,u} - \hat{\Lambda}_u \right)^{-1} \xrightarrow{p} V_u + \Sigma_X^{-1} (W_u + \Theta) \Sigma_X^{-1}, \quad (65)$$

where

$$\hat{V}_u = \left[\hat{\sigma}_u^2 \left(\hat{\tau} + \hat{\gamma}_{1,u}^* \hat{\mathcal{F}}_u \hat{\gamma}_{1,u}^* \right) \right] \left(\hat{\Sigma}_{X,u} - \hat{\Lambda}_u \right)^{-1}, \quad (66)$$

$$\hat{W}_u = \begin{bmatrix} 0 & 0'_K \\ 0_K & \frac{1}{N} \sum_{i=1}^N \hat{Z}'_{i,u} \hat{U}_\epsilon \hat{Z}_{i,u} \end{bmatrix}, \quad (67)$$

$$\hat{\Theta} = \hat{\theta} \hat{\Sigma}_{F\beta} - \hat{\sigma}_u^2 \hat{\Psi}, \quad (68)$$

with

$$\hat{\Psi} = \begin{bmatrix} 0 & \hat{\gamma}_{1,u}^* \hat{\mathcal{F}}_\gamma \\ \hat{\mathcal{F}}'_\gamma \hat{\gamma}_{1,u}^* & \hat{\mathcal{F}}_{\gamma\beta} \end{bmatrix}, \quad (69)$$

$$\hat{\mathcal{F}}_\gamma = \frac{1}{N} \sum_{i=1}^N \mathcal{P}'_{i,u} \mathcal{P}_{i,u} (\bar{f}_{i,u} - \bar{f})' \hat{\beta}_{i,u}, \quad (70)$$

$$\begin{aligned} \hat{\mathcal{F}}_{\gamma\beta} &= \frac{1}{N} \sum_{i=1}^N \hat{\Sigma}_{\hat{\beta}_{i,u}}^a (\bar{f}_{i,u} - \bar{f}) \hat{\gamma}_{1,u}^* \mathcal{P}'_{i,u} \mathcal{P}_{i,u} \\ &\quad + \frac{1}{N} \sum_{i=1}^N \mathcal{P}'_{i,u} \mathcal{P}_{i,u} \hat{\gamma}_{1,u}^* (\bar{f}_{i,u} - \bar{f})' \hat{\Sigma}_{\hat{\beta}_{i,u}}^a \\ &\quad - \frac{1}{N} \sum_{i=1}^N (\bar{f}_{i,u} - \bar{f})' \hat{\Sigma}_{\hat{\beta}_{i,u}}^a (\bar{f}_{i,u} - \bar{f}) \mathcal{P}'_{i,u} \mathcal{P}_{i,u}. \end{aligned} \quad (71)$$

Proof: See Online Appendix.

Turning to the specification test analysis, let

$$\hat{e}_u^P = \bar{R}_u - \hat{X}_u \hat{\Gamma}_u^* \quad (72)$$

be the N -vector of ex-post sample pricing errors. Define $\hat{Q}_u = \frac{\hat{e}_u^{P'} \hat{e}_u^P}{N}$ as the sum of squared ex-post sample pricing errors and denote by $\hat{\Sigma}_{\hat{\beta}_u}^a = \left(\frac{\hat{B}_u' \hat{B}_u}{N} - \hat{\sigma}_u^2 \hat{\mathcal{F}}_u \right)$, $\hat{b} = \text{tr}(F' F \hat{\Sigma}_{\hat{\beta}_u}^a)$, $\omega_N = \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \left(\frac{J_{it}}{T_i} - \frac{1}{T} \right)^2 \text{tr} \left(\mathcal{P}_{i,u} f_t f_t' \mathcal{P}'_{i,u} \right)$, and $Z_{Q_{i,u}} = \left[\left(Q'_{i,u} \otimes Q'_{i,u} \right) - \frac{Q'_{i,u} Q_{i,u} \text{vec}(M_{i,u})'}{T_i - K - 1} \right]'$. Finally, consider the centered statistic

$$\mathcal{S}_u = \sqrt{N} \left(\hat{Q}_u - \hat{\sigma}_u^2 (\hat{\tau} + \hat{\gamma}_{1,u}^* \hat{\mathcal{F}}_u \hat{\gamma}_{1,u}^*) - \hat{\theta} \hat{b} \right). \quad (73)$$

Theorem 6 *Under Assumptions 1 to 4 and 4–5 (listed in Appendix A) where by Assumption 4 $H_0 : e_i = 0$ holds for all i , we have*

$$\mathcal{S}_u \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_u + \mathcal{W}_u), \quad (74)$$

where

$$\mathcal{V}_u = \text{plim} \frac{1}{N} \sum_{i=1}^N \tilde{Z}'_{Q_{i,u}} U_\epsilon \tilde{Z}_{Q_{i,u}}, \quad (75)$$

$$\mathcal{W}_u = 4\sigma^2 \text{plim} \frac{1}{N} \sum_{i=1}^N W_i' W_i \quad (76)$$

with

$$\tilde{Z}_{Q_{i,u}} = Z_{Q_{i,u}} + \left(\omega_N \left(\frac{\text{vec}(M_{i,u})}{T_i - K - 1} \right) - \sum_{t=1}^T \left(\frac{J_{it}}{T_i} - \frac{1}{T} \right)^2 \text{vec} \left(\mathcal{P}_{i,u} f_t f_t' \mathcal{P}'_{i,u} \right) \right) \quad (77)$$

and

$$W_i = \left[(\gamma_{1i,u}^P - \gamma_1^P)' \beta_i Q'_{i,u} - \sum_{t=1}^T \left(\frac{J_{it}}{T_i} - \frac{1}{T} \right)^2 \beta_i' f_t f_t' \mathcal{P}'_{i,u} \right]'. \quad (78)$$

Note that when the panel is balanced, Theorem 6 reduces to Theorem 3 since $\frac{J_{it}}{T_i} = \frac{1}{T}$ and $\bar{f}_{i,u} = \bar{f}$, which implies that $\mathcal{W}_u = 0$, $Q_{i,u} = Q$, and $\tilde{Z}_{Q_{i,u}} = Z_{Q_{i,u}} = Z_Q$.

This variance can be consistently estimated. Let $\hat{Z}_{Q_{i,u}} = \left[\left(\hat{Q}'_{i,u} \otimes \hat{Q}'_{i,u} \right) - \frac{\hat{Q}'_{i,u} \hat{Q}_{i,u} \text{vec}(M_{i,u})'}{T_i - K - 1} \right]'$ and $\hat{\tilde{Z}}_{Q_{i,u}} = \hat{Z}_{Q_{i,u}} + \left(\omega_N \left(\frac{\text{vec}(M_{i,u})}{T_i - K - 1} \right) - \sum_{t=1}^T \left(\frac{J_{it}}{T_i} - \frac{1}{T} \right)^2 \text{vec} \left(\mathcal{P}_{i,u} f_t f_t' \mathcal{P}'_{i,u} \right) \right)$. Then,

$$\hat{V}_u = \frac{1}{N} \sum_{i=1}^N \hat{\tilde{Z}}'_{Q_{i,u}} \hat{U}_{\epsilon,u} \hat{\tilde{Z}}_{Q_{i,u}} \quad (79)$$

and

$$\begin{aligned} \hat{W}_u &= 4\hat{\sigma}_u^2 \frac{1}{N} \sum_{i=1}^N \left(\hat{Q}'_{i,u} \hat{Q}_{i,u} (\bar{f}_{i,u} - \bar{f})' \hat{\Sigma}_{\hat{\beta}_{i,u}}^a (\bar{f}_{i,u} - \bar{f}) \right. \\ &\quad + \sum_{t=1}^T \left(\frac{J_{it}}{T_i} - \frac{1}{T} \right)^4 \text{tr} \left(f_t f_t' \mathcal{P}'_{i,u} \mathcal{P}_{i,u} f_t f_t' \hat{\Sigma}_{\hat{\beta}_{i,u}}^a \right) \\ &\quad \left. - 2 \hat{Q}'_{i,u} \mathcal{P}_{i,u} \sum_{t=1}^T \left(\frac{J_{it}}{T_i} - \frac{1}{T} \right)^2 f_t f_t' \hat{\Sigma}_{\hat{\beta}_{i,u}}^a (\bar{f}_{i,u} - \bar{f}) \right). \quad (80) \end{aligned}$$

Monte Carlo simulations (not reported to conserve space) show that the parameter and specification tests based on Theorems 5 and 6 have excellent size and power properties even when the number of missing observations is 30-40% of the entire sample.

5 Simulation Evidence

In this section, we undertake a Monte Carlo simulation experiment to study the empirical rejection rates of the specification and t -tests for the OLS bias-adjusted estimator of Shanken (1992). The return generating process under the null of a correctly specified asset-pricing model is given by

$$R_t = \gamma_0 1_N + B(\gamma_1 + f_t - E[f]) + \epsilon_t, \quad (81)$$

where $\epsilon_t \sim \mathcal{N}(0, \Sigma)$. To study the power of the specification test, we generate the returns on the test assets as in (4), that is, we do not impose the asset-pricing restriction.

In all of our simulation experiments, we consider balanced panels with time-series dimension of $T = 36$ and $T = 72$ observations. Specifically, f_t in (81) and in (4) is the excess market return (from Kenneth French’s website) from January 2008 to December 2010 for $T = 36$, and the excess market return from January 2008 to December 2013 for $T = 72$. In our simulation designs, the factor realizations are taken as given and kept fixed throughout. This is consistent with the fact that our analysis of the ex-post risk premia is conditional on the realizations of the factors. In addition, $E[f]$ in (81) is set equal to the time-series mean of f_t over the 2008-2010 sample when performing the analysis for $T = 36$ and to the time-series mean of f_t over the 2008-2013 sample when performing the analysis for $T = 72$. To obtain representative values for the parameters γ_0 , γ_1 , B , and Σ in (81) and (4), we employ a sample of 3000 stocks from CRSP in addition to the excess market return. See Section 2 for a description of the data. Based on this balanced panel of 3000 stock returns and the excess market return, for each time-series sample size, we compute the OLS estimates of B , γ_0 , and γ_1 . Then, we set the B , γ_0 , and γ_1 parameters in (81) and in (4) equal to these OLS estimates. The calibration of Σ is a more delicate task and is described in the next subsection. In the simulations, we consider cross-sections of $N = 100, 500, 1000$, and 3000 stocks. All results are based on 10,000 Monte Carlo replications. Our econometric approach, designed for large N and fixed T , should be able to handle this large number of assets over relative short time spans. The rejection rates of the various tests are computed using our asymptotic results in Section 2.

5.1 Percentage Errors and Root Mean Squared Errors of the Estimates

We start from the case in which Σ is a scalar matrix, that is, $\Sigma = \sigma^2 I_T$. In the simulations, we set σ^2 equal to the cross-sectional average (over the 3000 stocks) of the σ_i^2 ’s estimated from the data. Table I reports the percentage error (bias) and root mean squared error (RMSE), all in percent, of the OLS estimator and of the OLS bias-adjusted estimator of Shanken (1992). Panels A and B are for $T = 36$ and $T = 72$, respectively.

Table I about here

Panel A clearly shows that the bias of the OLS estimator is substantial. For $\hat{\gamma}_0$, the bias ranges from 28.8% for $N = 100$ to 22.9% for $N = 3000$, while for $\hat{\gamma}_1$ the bias ranges from -24.8% for

$N = 100$ to -17.8% for $N = 3000$. For $\hat{\Gamma}^*$, the bias is small for $N = 100$ (-2.3% for $\hat{\gamma}_0^*$ and 1.8% for $\hat{\gamma}_1^*$) and becomes negligible for $N \geq 500$. As for the RMSE, the typical bias-variance trade-off emerges up to $N = 500$, with the OLS estimator exhibiting a smaller RMSE than the OLS bias-adjusted estimator. When $N > 500$, the RMSE of the OLS bias-adjusted estimator becomes substantially smaller than the one of the OLS estimator. Panel B for $T = 72$ conveys a similar message. As expected from the theoretical analysis, the larger time-series dimension helps in reducing the bias and RMSE associated with the OLS estimator. However, the bias for the OLS estimator is still substantial and ranges from -18.5% for $N = 100$ to -11.7% for $N = 3000$. For the bias-adjusted estimator, the bias becomes negligible even for $N = 100$ when $T = 72$.

Next, we consider the case in which the Σ matrix is either diagonal or full. As emphasized above, our theoretical results hinge upon the assumption that the model disturbances are weakly cross-sectionally correlated. In order to generate shocks under a weak factor structure, we consider the following data generating process (DGP). Define

$$\epsilon^{(1)} = \eta \left(\frac{\sqrt{\theta}}{N^\delta} \right) c' + \sqrt{1 - \theta} Z, \quad (82)$$

where η and c are T and N -vectors of i.i.d. standard normal random variables, respectively, Z is a $(T \times N)$ matrix of i.i.d. standard normal random variables, $0 \leq \theta \leq 1$ is a shrinkage parameter that controls the weight assigned to the diagonal and extra-diagonal elements of Σ , and δ is a parameter that controls the strength of the cross-sectional dependence of the shocks (the bigger is δ , the weaker is the dependence). Our $T \times N$ matrix of shocks is then generated as

$$\epsilon = \epsilon^{(1)} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_N^2 \end{bmatrix}^{0.5} \begin{bmatrix} \frac{\theta}{N^{2\delta}} c_1^2 + (1 - \theta) & & & \\ & \frac{\theta}{N^{2\delta}} c_2^2 + (1 - \theta) & & \\ & & \ddots & \\ & & & \frac{\theta}{N^{2\delta}} c_N^2 + (1 - \theta) \end{bmatrix}^{-0.5}, \quad (83)$$

where c_i is the i -th element of c . Given this specification for the shocks, for our theoretical results to hold we require $\delta > 0.25$.

In Table II, we report results for the diagonal case, that is, we set $\theta = 0$ in the above DGP. To obtain representative values of the shock variances, while accounting for the fact that $\hat{\Sigma}$ is ill-conditioned when T is small and N is large, we first estimate the residual variances from the historical data. Then, at each Monte Carlo iteration, we generate a string of Beta(p, q)-distributed

random variables with the p and q parameters calibrated to the cross-sectional mean and variance of the $\hat{\sigma}_i^2$'s. This resampling procedure is used to minimize the impact of an ill-conditioned $\hat{\Sigma}$ on the simulation results.

Table II about here

Overall, we find that the OLS estimator exhibits a slightly higher bias compared to the scalar Σ case. The OLS bias-adjusted estimator continues to perform very well in terms of bias for all the time-series and cross-sectional dimensions considered. The RMSEs of both estimators are now a bit higher than in the scalar Σ case, and the OLS bias-adjusted estimator still outperforms the OLS estimator for $N \geq 500$.

Finally, in Tables III and IV, we allow for weak cross-sectional dependence of the model disturbances by setting $\theta = 0.5$ in the above DGP.

Tables III and IV about here

In Table III, we consider the situation in which δ , the parameter that regulates the strength of the cross-sectional dependence, is equal to 0.5. Consistent with our theoretical results, the bias-adjusted estimator continues to perform very well in this scenario. Setting $\delta = 0.25$ in Table IV has only a modest effect on the bias and RMSEs of the two estimators. Overall, the first 4 tables reveal a superiority of the bias-adjusted estimator of Shanken (1992) over the OLS estimator, not only in terms of bias, but also in terms of RMSE when $N > 500$. Furthermore, the bias-adjusted estimator shows little sensitivity to changes in the length of the time series, consistent with the idea that this estimator should perform well for any given T .

5.2 Rejection Rates of the t -tests

In Tables V through VIII, we consider the empirical rejection rates of centered t -tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of time-series and cross-sectional observations using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with covariance

matrix calibrated as in Tables I through IV. The t -statistics are compared with the critical values from a standard normal distribution. We consider three t -statistics. For the OLS estimator of the ex-post risk premia, the first t -statistic is the one that uses the traditional Fama-MacBeth standard error (t_{FM}), while the second t -statistic is the one that uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992, t_{EIV}). Both of these t -statistics were developed in a large T and fixed N framework. We report them here to determine how misleading inference can be when using these t -statistics in a large N and fixed T setup. Finally, the third t -statistic is the one associated with the OLS bias-adjusted estimator and is based on the asymptotic distribution in part (ii) of our Theorem 1.

Table V about here

Starting from the scalar Σ case, Table V shows that the t -statistics associated with the OLS estimator only slightly overreject the null hypothesis for $N = 100$. However, as N increases, the performance of these t -statistics substantially deteriorates. For example, when $N = 3000$, the rejections rate of the Fama-MacBeth t -statistic associated with $\hat{\gamma}_1$ is either 41.6% for $T = 36$ or 33.3% for $T = 72$ at the 5% nominal level. The strong size distortions of the Fama-MacBeth t -test don't show any improvement when accounting for the EIV bias due to the estimation of the betas in the first stage. In contrast our proposed t -statistic, based on Theorems 1 and 2, performs extremely well for all T and N . A similar picture emerges in the Σ full case (Tables VI and VII), with the rejection rates of our proposed t -test being always aligned with the critical values from a standard normal distribution.

Tables VI and VII about here

In Table VIII, we increase the strength of the cross-sectional dependence of the residuals by setting $\delta = 0.25$.

Table VIII about here

In this situation, we start to notice some slight over-rejections for the t -test associated with the OLS bias-adjusted estimator. For example, when $T = 36$ and $N = 3000$, the rejection rate for the

t -test associated with $\hat{\gamma}_1^*$ is 6.8% at the 5% level, and when $T = 72$ and $N = 3000$, the rejection rate for the t -test associated with $\hat{\gamma}_1^*$ is 5.8% at the 5% level. Overall, these results suggest that our proposed t -test is relatively well behaved even when moving towards a fairly strong factor structure in the residuals. Furthermore, using the standard tools that were developed in a large T and fixed N framework can lead to strong over-rejections of the null hypothesis, with the likely consequence that a factor will be found to be priced even when it does not help explain the cross-sectional variation in individual stock returns.

5.3 Rejection Rates of the Specification Test

In Tables IX and X, we investigate the size and power properties of our specification test based on the results in Theorem 3. Table IX is for $T = 36$, while Table X is for $T = 72$.

Tables IX and X about here

Since the specification test has a standard normal distribution, we consider two-sided p -values in the computation of the rejection rates. The results in the two tables suggest that the rejection rates of our test under the null that the model is correctly specified are excellent for the scalar and diagonal cases. When simulating with Σ full, the specification test is very well sized when $\delta = 0.5$ but it over-rejects a bit too much when $\delta = 0.25$. The power properties of our specification test are fairly good when $N = 100$ and excellent when $N \geq 500$. As expected, power increases when the number of assets becomes large and the rejection rates are similar across time-series sample sizes. Overall, these simulation results suggest that the tests should be fairly reliable for the time-series and cross-sectional dimensions encountered in our empirical work.

6 Conclusion

This paper is concerned with estimation of risk premia when data is available for a large cross-section of securities N but only for a limited number of time periods. Since in this context the cross-sectional OLS estimator of the conventional ex-ante risk premia is asymptotically biased and inconsistent, the focus of the paper is on the modified, bias-adjusted, estimator of the ex-post risk premia proposed by Shanken (1992). From a methodological point of view, we demonstrate that

the Shanken estimator: (i) is an OLS-based estimator that, however, does not require preliminary estimation of the bias-adjustment; (ii) it converges to the true ex-post risk premia at usual rate \sqrt{N} ; (iii) it has an asymptotically normal distribution; (iv) its limiting covariance matrix can be consistently estimated. Besides estimation, building upon the pricing errors stemming from the Shanken's estimator, we propose a new test of the no-arbitrage asset pricing restriction, and establish its asymptotic distribution (assuming that the restriction holds) as N diverges. Finally, we show how our results can be extended to deal with the more realistic case of unbalanced panels, allowing to take advantage of the large cross-sections of stocks existing only for certain time periods. Montecarlo simulations corroborate our theoretical finding, both in terms of estimation and in terms of testing for the asset pricing restriction.

Our empirical analysis employs individual monthly stock returns from the CRSP database over overlapping three-, six- and ten-year periods from 1966 until 2013. The three prominent asset-pricing specifications that we consider are the CAPM, the three-factor model of Fama and French (1993), and the newly proposed five-factor model of Fama and French (2015). We find some convincing pricing ability for all the factors for each of the three models, even when using a relatively short time window of three years, for most periods. In contrast, the same risk premia appear almost always statistically different from zero when using the traditional approach of estimating risk premia based on large-T asymptotics. Similarly, in terms of asset pricing test, our methodology will tend to reject the CAPM even when using a short time window, unlike the traditional large-T approach of Gibbons, Ross and Shanken (1989). Only when the time window increases, six-years and above, the CAPM will also appear to be rejected by the large-T approach.

Appendix A: Additional Assumptions and Lemmas

All the limits are taken for $N \rightarrow \infty$. In addition, the expectation operator used throughout this appendix has to be understood as conditional of F .

Assumption 5. (Idiosyncratic component) We require

(i)

$$\frac{1}{N} \sum_{i=1}^N (\sigma_i^2 - \sigma^2) = o\left(\frac{1}{\sqrt{N}}\right) \quad (\text{A.1})$$

with $0 < \sigma^2 < \infty$.

(ii)

$$\sum_{i,j=1}^N |\sigma_{ij}| \mathbb{1}_{\{i \neq j\}} = o(N), \quad (\text{A.2})$$

where $\sigma_{ij} = E[\epsilon_{it}\epsilon_{jt}]$ and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

(iii)

$$\frac{1}{N} \sum_{i=1}^N \mu_{4i} \rightarrow \mu_4 \quad (\text{A.3})$$

with $0 < \mu_4 < \infty$ and $\mu_{4i} = E[\epsilon_{it}^4]$.

(iv)

$$\frac{1}{N} \sum_{i=1}^N \sigma_i^4 \rightarrow \sigma_4 \quad (\text{A.4})$$

with $0 < \sigma_4 < \infty$.

(v)

$$\sup_i \mu_{4i} \leq C < \infty \quad (\text{A.5})$$

for a generic constant C .

(vi)

$$E[\epsilon_{it}^3] = 0. \quad (\text{A.6})$$

(vii)

$$\frac{1}{N} \sum_{i=1}^N \kappa_{4,iiii} \rightarrow \kappa_4 \quad (\text{A.7})$$

with $0 \leq |\kappa_4| < \infty$, where $\kappa_{4,iiii} = \kappa_4(\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it})$ denotes the fourth-order cumulant of the random variables $\{\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it}\}$.

(viii) For every $3 \leq h \leq 8$, all the mixed cumulants of order h are such that

$$\sup_{i_1} \sum_{i_2, \dots, i_h=1}^N |\kappa_{h, i_1 i_2 \dots i_h}| = o(N) \quad (\text{A.8})$$

for at least one i_j ($2 \leq j \leq h$) different than i_1 .

Assumption 5 essentially describes the cross-sectional behavior of the model disturbances. Assumption 5(i) limits the cross-sectional heterogeneity of the return conditional variances. Assumption 5(ii) implies that the conditional correlation among asset returns is sufficiently weak. In particular, Assumptions 5(ii) and 5(v) imply that $\sup_i \sum_{j=1}^N |\sigma_{ij}| \leq C < \infty$, which in turn implies that the maximum eigenvalue of the conditional covariance of asset returns is bounded. The latter is the most common assumption in factor pricing models such as the Arbitrage Pricing Theory (see, for example, Chamberlain and Rothschild, 1983). In Assumption 5(iii), we simply assume the existence of the limit of the conditional fourth moment average across assets. In Assumption 5(iv), the magnitude of σ_4 reflects the degree of cross-sectional heterogeneity of the conditional variance of asset returns. Assumption 5(v) is a bounded fourth moment condition uniform across assets, which implies $\sup_i \sigma_i^2 \leq C < \infty$. Assumption 5(vi) is a convenient symmetry assumption but it is not strictly necessary for our results. It could be relaxed at the cost of a more cumbersome notation. Assumption 5(vii) allows for non Gaussianity of asset returns when $|\kappa_4| > 0$. For example, this assumption is satisfied when the marginal distribution of asset returns is a Student t with degrees of freedom greater than four. However, when estimating the asymptotic covariance matrix of the Shanken's estimator one needs to set $\kappa_4 = 0$ merely for identification purposes, as indicated below (cf Lemma 6).

Assumption 6.

(i)

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i \xrightarrow{d} \mathcal{N}(0_T, \sigma^2 I_T). \quad (\text{A.9})$$

(ii)

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \xrightarrow{d} \mathcal{N}(0_{T^2}, U_\epsilon). \quad (\text{A.10})$$

(iii) For a generic T -dimensional column vector C_T ,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(C_T' \otimes \begin{pmatrix} 1 \\ \beta_i \end{pmatrix} \right) \epsilon_i \xrightarrow{d} \mathcal{N}(0_{K+1}, V_c), \quad (\text{A.11})$$

where $V_c \equiv c\sigma^2\Sigma_X$ and $c = C_T' C_T$. In particular, $\frac{1}{\sqrt{N}} \sum_{i=1}^N (C_T' \otimes \beta_i) \epsilon_i \xrightarrow{d} \mathcal{N}(0_K, V_c^\dagger)$, where $V_c^\dagger \equiv c\sigma^2\Sigma_\beta$.

Lemma 1

(i) Under Assumptions 2 to 5, we have

$$\hat{\sigma}^2 - \sigma^2 = O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{A.12})$$

(ii) In addition, under Assumption 6, we have

$$\sqrt{N}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, u_{\sigma^2}). \quad (\text{A.13})$$

Proof

(i) Rewrite $\hat{\sigma}^2 - \sigma^2$ as

$$\begin{aligned} \hat{\sigma}^2 - \sigma^2 &= \left(\hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \right) + \left(\frac{1}{N} \sum_{i=1}^N \sigma_i^2 - \sigma^2 \right) \\ &= \left(\hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \right) + o\left(\frac{1}{\sqrt{N}}\right) \end{aligned} \quad (\text{A.14})$$

by Assumption 5(i). Moreover,

$$\begin{aligned} \hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^N \sigma_i^2 &= \frac{\text{tr}(M\epsilon\epsilon')}{N(T-K-1)} - \frac{\text{tr}(M)}{T-K-1} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \\ &= \frac{\text{tr}\left(P\left(\sum_{i=1}^N \sigma_i^2 I_T - \epsilon\epsilon'\right)\right)}{N(T-K-1)} + \frac{\text{tr}(\epsilon\epsilon') - T \sum_{i=1}^N \sigma_i^2}{N(T-K-1)}. \end{aligned} \quad (\text{A.15})$$

As for the second term on the right-hand side of (A.15), we have

$$\begin{aligned} \frac{\text{tr}(\epsilon\epsilon') - T \sum_{i=1}^N \sigma_i^2}{N(T-K-1)} &= \frac{\sum_{i=1}^N \sum_{t=1}^T (\epsilon_{it}^2 - \sigma_i^2)}{N(T-K-1)} \\ &= O_p\left(\frac{1}{\sqrt{N}} \frac{\sqrt{T}}{(T-K-1)}\right) = O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \quad (\text{A.16})$$

As for the first term on the right-hand side of (A.15), we have

$$\begin{aligned} \frac{\text{tr} \left(P \left(\sum_{i=1}^N \sigma_i^2 I_T - \epsilon \epsilon' \right) \right)}{N(T-K-1)} &= \frac{\sum_{t=1}^T d_t (D'D)^{-1} D' \left(\sum_{i=1}^N \sigma_i^2 \iota_t - \sum_{i=1}^N \epsilon_i \epsilon_{it} \right)}{N(T-K-1)} \\ &= \frac{\sum_{t=1}^T p_t \left(\sum_{i=1}^N \sigma_i^2 \iota_t - \sum_{i=1}^N \epsilon_i \epsilon_{it} \right)}{N(T-K-1)}, \end{aligned} \quad (\text{A.17})$$

where ι_t is a T -vector with 1 in the t -th position and zeros elsewhere, d_t is the t -th row of D , and $p_t = d_t (D'D)^{-1} D'$. Since (A.17) has zero mean, we only need to consider its variance to determine the rate of convergence. We have

$$\begin{aligned} &\text{Var} \left(\frac{\sum_{t=1}^T p_t \left(\sum_{i=1}^N \sigma_i^2 \iota_t - \sum_{i=1}^N \epsilon_i \epsilon_{it} \right)}{N(T-K-1)} \right) \\ &= \frac{1}{N^2(T-K-1)^2} E \left[\sum_{i,j=1}^N \sum_{t,s=1}^T p_t (\sigma_i^2 \iota_t - \epsilon_i \epsilon_{it}) (\sigma_j^2 \iota_s - \epsilon_j \epsilon_{js})' p_s' \right] \\ &= \frac{1}{N^2(T-K-1)^2} \sum_{i,j=1}^N \sum_{t,s=1}^T p_t E \left[(\sigma_i^2 \iota_t - \epsilon_i \epsilon_{it}) (\sigma_j^2 \iota_s - \epsilon_j \epsilon_{js})' \right] p_s'. \end{aligned} \quad (\text{A.18})$$

Moreover, we have

$$\begin{aligned} &E \left[(\sigma_i^2 \iota_t - \epsilon_i \epsilon_{it}) (\sigma_j^2 \iota_s - \epsilon_j \epsilon_{js})' \right] = E \left[\sigma_i^2 \sigma_j^2 \iota_t \iota_s' + \epsilon_i \epsilon_j' \epsilon_{it} \epsilon_{js} - \sigma_i^2 \iota_t \epsilon_j' \epsilon_{js} - \sigma_j^2 \epsilon_{it} \epsilon_i \iota_s' \right] \\ &= \begin{cases} \mu_{4i} \iota_t \iota_t' + \sigma_i^4 (I_T - 2 \iota_t \iota_t') & \text{if } i = j, t = s \\ \kappa_{4,ij} \iota_t \iota_t' + \sigma_{ij}^2 (I_T + \iota_t \iota_t') & \text{if } i \neq j, t = s \\ \sigma_i^4 \iota_s \iota_t' & \text{if } i = j, t \neq s \\ \sigma_{ij}^2 \iota_s \iota_t' & \text{if } i \neq j, t \neq s. \end{cases} \end{aligned} \quad (\text{A.19})$$

It follows that

$$\begin{aligned}
& \text{Var} \left(\frac{\sum_{t=1}^T p_t \left(\sum_{i=1}^N \sigma_i^2 \iota_t - \sum_{i=1}^N \epsilon_i \epsilon_{it} \right)}{N(T-K-1)} \right) \\
&= \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^T \sum_{i=1}^N p_t (\mu_{4i} \iota_t \iota_t' + \sigma_i^4 (I_T - 2\iota_t \iota_t')) p_t' \\
&+ \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^T \sum_{i \neq j} p_t (\kappa_{4,ij} \iota_t \iota_t' + \sigma_{ij}^2 (I_T + \iota_t \iota_t')) p_t' \\
&+ \frac{1}{N^2(T-K-1)^2} \sum_{i=1}^N \sigma_i^4 \sum_{t \neq s} p_t \iota_s \iota_t' p_s' \\
&+ \frac{1}{N^2(T-K-1)^2} \sum_{i \neq j} \sigma_{ij}^2 \sum_{t \neq s} p_t \iota_s \iota_t' p_s' \\
&= O\left(\frac{1}{N}\right) \tag{A.20}
\end{aligned}$$

by Assumptions 5(ii), 5(iii), 5(iv), and 5(viii), which implies that the first term on the right-hand side of (A.15) is $O_p\left(\frac{1}{\sqrt{N}}\right)$. Putting the pieces together concludes the proof of part (i).

(ii) Using Assumption 5(i) and the properties of the vec operator, we can write $\sqrt{N}(\hat{\sigma}^2 - \sigma^2)$ as

$$\sqrt{N}(\hat{\sigma}^2 - \sigma^2) = \frac{1}{T-K-1} \text{vec}(M)' \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + o(1). \tag{A.21}$$

The desired result then follows from using Assumption 6(ii). This concludes the proof of part (ii).

Lemma 2 *Let*

$$\Lambda = \begin{bmatrix} 0 & 0_K' \\ 0_K & \sigma^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix}. \tag{A.22}$$

(i) *Under Assumptions 1 to 5, we have*

$$\hat{X}'\hat{X} = O_p(N). \tag{A.23}$$

In addition, under Assumption 6, we have

(ii)

$$\hat{\Sigma}_X \xrightarrow{P} \Sigma_X + \Lambda, \tag{A.24}$$

and

(iii)

$$\frac{(\hat{X} - X)'(\hat{X} - X)}{N} \xrightarrow{p} \Lambda. \quad (\text{A.25})$$

Proof

(i) Consider

$$\hat{X}'\hat{X} = \begin{bmatrix} N & 1'_N\hat{B} \\ \hat{B}'1_N & \hat{B}'\hat{B} \end{bmatrix}. \quad (\text{A.26})$$

Then, we have

$$\hat{B}'1_N = \sum_{i=1}^N \hat{\beta}_i = \sum_{i=1}^N \beta_i + \mathcal{P}' \sum_{i=1}^N \epsilon_i. \quad (\text{A.27})$$

Under Assumptions 2 to 5,

$$\begin{aligned} \text{Var} \left(\sum_{t=1}^T \sum_{i=1}^N \epsilon_{it}(f_t - \bar{f}) \right) &= \sum_{t,s=1}^T \sum_{i,j=1}^N (f_t - \bar{f})(f_s - \bar{f})' E[\epsilon_{it}\epsilon_{js}] \\ &\leq \sum_{t=1}^T \sum_{i,j=1}^N (f_t - \bar{f})(f_t - \bar{f})' |\sigma_{ij}| \\ &= O \left(N\sigma^2 \sum_{t=1}^T (f_t - \bar{f})(f_t - \bar{f})' \right) = O(NT). \end{aligned} \quad (\text{A.28})$$

Using Assumption 1, we have

$$\hat{B}'1_N = O_p \left(N + \left(\frac{N}{T} \right)^{\frac{1}{2}} \right) = O_p(N). \quad (\text{A.29})$$

Next, consider

$$\begin{aligned} \hat{B}'\hat{B} &= \sum_{i=1}^N \hat{\beta}_i \hat{\beta}_i' \\ &= \sum_{i=1}^N (\beta_i + \mathcal{P}'\epsilon_i) (\beta_i' + \epsilon_i'\mathcal{P}) \\ &= \sum_{i=1}^N \beta_i \beta_i' + \mathcal{P}' \left(\sum_{i=1}^N \epsilon_i \epsilon_i' \right) \mathcal{P} \\ &\quad + \mathcal{P}' \left(\sum_{i=1}^N \epsilon_i \beta_i' \right) + \left(\sum_{i=1}^N \beta_i \epsilon_i' \right) \mathcal{P}. \end{aligned} \quad (\text{A.30})$$

By Assumption 1,

$$\sum_{i=1}^N \beta_i \beta_i' = O(N). \quad (\text{A.31})$$

Using similar arguments as for (A.28),

$$\mathcal{P}' \left(\sum_{i=1}^N \epsilon_i \beta_i' \right) = O_p \left(\left(\frac{N}{T} \right)^{\frac{1}{2}} \right) \quad (\text{A.32})$$

and

$$\left(\sum_{i=1}^N \beta_i \epsilon_i' \right) \mathcal{P} = O_p \left(\left(\frac{N}{T} \right)^{\frac{1}{2}} \right). \quad (\text{A.33})$$

For $\mathcal{P}' \left(\sum_{i=1}^N \epsilon_i \epsilon_i' \right) \mathcal{P}$, consider its central part and take the norm of its expectation. Using Assumptions 2 to 5,

$$\begin{aligned} & \left\| E \left[\tilde{F}' \left(\sum_{i=1}^N \epsilon_i \epsilon_i' \right) \tilde{F} \right] \right\| \\ &= \left\| E \left[\sum_{t,s=1}^T \sum_{i=1}^N (f_t - \bar{f})(f_s - \bar{f})' \epsilon_{it} \epsilon_{is}' \right] \right\| \\ &\leq \sum_{t,s=1}^T \sum_{i=1}^N \|(f_t - \bar{f})(f_s - \bar{f})'\| |E[\epsilon_{it} \epsilon_{is}']| \\ &= \sum_{t=1}^T \sum_{i=1}^N \|(f_t - \bar{f})(f_t - \bar{f})'\| \sigma_i^2 \\ &= O \left(N \sigma^2 \sum_{t=1}^T \|(f_t - \bar{f})(f_t - \bar{f})'\| \right) = O(NT). \end{aligned} \quad (\text{A.34})$$

Then, we have

$$\mathcal{P}' \left(\sum_{i=1}^N \epsilon_i \epsilon_i' \right) \mathcal{P} = O_p \left(\frac{N}{T} \right) \quad (\text{A.35})$$

and

$$\hat{B}' \hat{B} = O_p \left(N + \left(\frac{N}{T} \right)^{\frac{1}{2}} + \frac{N}{T} \right) = O_p(N). \quad (\text{A.36})$$

This concludes the proof of part (i).

(ii) Using part (i) and under Assumption 2 to 6, we have

$$N^{-1} \hat{B}' 1_N = \frac{1}{N} \sum_{i=1}^N \beta_i + O_p \left(\frac{1}{\sqrt{N}} \right) \quad (\text{A.37})$$

and

$$\begin{aligned}
N^{-1}\hat{B}'\hat{B} &= \frac{1}{N} \sum_{i=1}^N \beta_i \beta_i' + \mathcal{P}' \left(\frac{1}{N} \sum_{i=1}^N \epsilon_i \epsilon_i' \right) \mathcal{P} + \mathcal{P}' \left(\frac{1}{N} \sum_{i=1}^N \epsilon_i \beta_i' \right) + \left(\frac{1}{N} \sum_{i=1}^N \beta_i \epsilon_i' \right) \mathcal{P} \\
&= \frac{1}{N} \sum_{i=1}^N \beta_i \beta_i' + \mathcal{P}' \left(\frac{1}{N} \sum_{i=1}^N \epsilon_i \epsilon_i' - \frac{1}{N} \sum_{i=1}^N \sigma_i^2 I_T + \frac{1}{N} \sum_{i=1}^N \sigma_i^2 I_T - \sigma^2 I_T + \sigma^2 I_T \right) \mathcal{P} \\
&\quad + \mathcal{P}' \left(\frac{1}{N} \sum_{i=1}^N \epsilon_i \beta_i' \right) + \left(\frac{1}{N} \sum_{i=1}^N \beta_i \epsilon_i' \right) \mathcal{P} \\
&= \frac{1}{N} \sum_{i=1}^N \beta_i \beta_i' + \mathcal{P}' \left(\frac{1}{N} \sum_{i=1}^N (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \right) \mathcal{P} + \frac{1}{N} \sum_{i=1}^N (\sigma_i^2 - \sigma^2) \mathcal{P}' \mathcal{P} + \sigma^2 \mathcal{P}' \mathcal{P} \\
&\quad + \mathcal{P}' \left(\frac{1}{N} \sum_{i=1}^N \epsilon_i \beta_i' \right) + \left(\frac{1}{N} \sum_{i=1}^N \beta_i \epsilon_i' \right) \mathcal{P} \\
&= \frac{1}{N} \sum_{i=1}^N \beta_i \beta_i' + \sigma^2 \mathcal{P}' \mathcal{P} + O_p \left(\frac{1}{\sqrt{N}} \right) + o \left(\frac{1}{\sqrt{N}} \right) + O_p \left(\frac{1}{\sqrt{N}} \right) + O_p \left(\frac{1}{\sqrt{N}} \right).
\end{aligned} \tag{A.38}$$

Assumption 1 concludes the proof of part (ii).

(iii) Note that

$$\begin{aligned}
\frac{(\hat{X} - X)'(\hat{X} - X)}{N} &= \frac{1}{N} \begin{bmatrix} 0'_N \\ (\hat{B} - B)' \end{bmatrix} [0_N, (\hat{B} - B)] \\
&= \begin{bmatrix} 0 & 0'_K \\ 0_K & \mathcal{P}' \frac{\epsilon \epsilon'}{N} \mathcal{P} \end{bmatrix},
\end{aligned} \tag{A.39}$$

where 0_N is an N -vector of zeros. As in part (ii) we can write

$$\frac{\epsilon \epsilon'}{N} = \frac{1}{N} \sum_{i=1}^N (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + \left(\frac{1}{N} \sum_{i=1}^N (\sigma_i^2 - \sigma^2) \right) I_T + \sigma^2 I_T. \tag{A.40}$$

Assumptions 5(i) and 6(ii) conclude the proof since

$$\mathcal{P}' \frac{\epsilon \epsilon'}{N} \mathcal{P} = \sigma^2 \mathcal{P}' \mathcal{P} + O_p \left(\frac{1}{\sqrt{N}} \right) + o \left(\frac{1}{\sqrt{N}} \right). \tag{A.41}$$

Lemma 3

(i) Under Assumptions 1 to 5, we have

$$X' \bar{\epsilon} = O_p \left(\sqrt{N} \right). \tag{A.42}$$

(ii) In addition, under Assumption 6, we have

$$\frac{1}{\sqrt{N}} X' \bar{\epsilon} \xrightarrow{d} \mathcal{N}(0_{K+1}, V). \quad (\text{A.43})$$

Proof

(i) We have

$$X' \bar{\epsilon} = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} 1'_N \\ B' \end{bmatrix} \epsilon_t \quad (\text{A.44})$$

and

$$\begin{aligned} \text{Var} \left(\frac{1}{T} \sum_{t=1}^T 1'_N \epsilon_t \right) &= \frac{1}{T^2} \sum_{t,s=1}^T \sum_{i,j=1}^N E[\epsilon_{it} \epsilon_{js}] \\ &\leq \frac{1}{T^2} \sum_{t=1}^T \sum_{i,j=1}^N |\sigma_{ij}| \\ &= O \left(\frac{NT}{T^2} \sigma^2 \right) = O(N). \end{aligned} \quad (\text{A.45})$$

Moreover, using Assumptions 1 and 5(ii),

$$\begin{aligned} \text{Var} \left(\frac{1}{T} \sum_{t=1}^T B' \epsilon_t \right) &= \frac{1}{T^2} \sum_{t,s=1}^T \sum_{i,j=1}^N E[\epsilon_{it} \epsilon_{js}] \beta_i \beta'_j \\ &\leq \frac{1}{T^2} \sum_{t=1}^T \sum_{i,j=1}^N |\beta_i \beta'_j| |\sigma_{ij}| \\ &= O \left(\frac{NT}{T^2} \sigma^2 \right) = O(N). \end{aligned} \quad (\text{A.46})$$

Putting the pieces together, $X' \bar{\epsilon} = O_p(\sqrt{N})$. This concludes the proof of part (i).

(ii) We have

$$\begin{aligned} \frac{1}{\sqrt{N}} X' \bar{\epsilon} &= \frac{1}{\sqrt{N}} X' \epsilon' \frac{1_T}{T} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{1'_T}{T} \otimes \begin{bmatrix} 1 \\ \beta_i \end{bmatrix} \right) \epsilon_i. \end{aligned} \quad (\text{A.47})$$

Assumption 6(iii) concludes the proof of part (ii).

Lemma 4

(i) Under Assumptions 2 to 5, we have

$$(\hat{X} - X)'X\Gamma^P = O_p\left(\sqrt{N}\right). \quad (\text{A.48})$$

(ii) In addition, under Assumption 6, we have

$$\frac{1}{\sqrt{N}}(\hat{X} - X)'X\Gamma^P \xrightarrow{d} \mathcal{N}(0_{K+1}, \mathcal{K}), \quad (\text{A.49})$$

where

$$\mathcal{K} \equiv \sigma^2 \begin{bmatrix} 0 & 0'_K \\ 0_K & \Gamma^{P'}\Sigma_X\Gamma^P(\tilde{F}'\tilde{F})^{-1} \end{bmatrix}. \quad (\text{A.50})$$

Proof

(i) We have

$$(\hat{X} - X)'X\Gamma^P = \begin{bmatrix} 0'_N \\ \mathcal{P}'\epsilon \end{bmatrix} X\Gamma^P. \quad (\text{A.51})$$

Using similar arguments as for (A.28) concludes the proof of part (i).

(ii) Using the properties of the vec operator

$$\begin{aligned} \frac{1}{\sqrt{N}}(\hat{X} - X)'X\Gamma^P &= \frac{1}{\sqrt{N}} \begin{bmatrix} 0 & 0'_K \\ \mathcal{P}'\epsilon_{1N} & \mathcal{P}'\epsilon_B \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1^P \end{bmatrix} \\ &= \frac{1}{\sqrt{N}} \begin{bmatrix} 0 \\ \mathcal{P}'\epsilon X\Gamma^P \end{bmatrix} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{bmatrix} 0'_T \\ \Gamma^{P'} \begin{bmatrix} 1 \\ \beta_i \end{bmatrix} \otimes \mathcal{P}' \end{bmatrix} \epsilon_i. \end{aligned} \quad (\text{A.52})$$

Using Assumption 6(iii) concludes the proof of part (ii).

Lemma 5

(i) Under Assumptions 2 to 5, we have

$$(\hat{X} - X)'\bar{\epsilon} = O_p\left(\sqrt{N}\right). \quad (\text{A.53})$$

(ii) In addition, under Assumption 6, we have

$$\frac{1}{\sqrt{N}}(\hat{X} - X)'\bar{\epsilon} \xrightarrow{d} \mathcal{N}(0_{K+1}, \mathcal{W}). \quad (\text{A.54})$$

Proof

(i)

$$\begin{aligned}
(\hat{X} - X)' \bar{\epsilon} &= \begin{bmatrix} 0 \\ \mathcal{P}' \epsilon \bar{\epsilon} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathcal{P}' \epsilon \epsilon' \frac{1}{T} \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ \mathcal{P}' \left[\left(\epsilon \epsilon' - \sum_{i=1}^N \sigma_i^2 I_T \right) + \left(\sum_{i=1}^N \sigma_i^2 - N \sigma^2 \right) I_T \right] \frac{1}{T} \end{bmatrix} = O_p(\sqrt{N})
\end{aligned} \tag{A.55}$$

by Assumption 5.

(ii)

$$\begin{aligned}
\frac{1}{\sqrt{N}} (\hat{X} - X)' \bar{\epsilon} &= \frac{1}{\sqrt{N}} \begin{bmatrix} 0 \\ \mathcal{P}' \left[\left(\epsilon \epsilon' - \sum_{i=1}^N \sigma_i^2 I_T \right) + \left(\sum_{i=1}^N \sigma_i^2 - N \sigma^2 \right) I_T \right] \frac{1}{T} \end{bmatrix} \\
&= \begin{bmatrix} 0'_{T^2} \\ \left(\frac{1}{T} \otimes \mathcal{P}' \right) \end{bmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + o(1).
\end{aligned} \tag{A.56}$$

The $o(1)$ term in (A.56) is due to Assumption 5(i). Using Assumption 6(ii) concludes the proof of part (ii).

Lemma 6 *Under Assumption 5 and the identification assumption $\kappa_4 = 0$, we have*

$$\hat{\sigma}_4 \xrightarrow{p} \sigma_4. \tag{A.57}$$

Proof

We need to show that (i) $E(\hat{\sigma}_4) \rightarrow \sigma_4$ and (ii) $\text{Var}(\hat{\sigma}_4) = O\left(\frac{1}{N}\right)$.

(i) By Assumptions 5(iv), 5(vi), and 5(vii), we have

$$\begin{aligned}
E \left[\frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{\epsilon}_{it}^4 \right] &= \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N E [\hat{\epsilon}_{it}^4] \\
&= \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \sum_{s_1, s_2, s_3, s_4=1}^T m_{ts_1} m_{ts_2} m_{ts_3} m_{ts_4} E [\epsilon_{is_1} \epsilon_{is_2} \epsilon_{is_3} \epsilon_{is_4}] \\
&= \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \kappa_{4,iiii} \sum_{s=1}^T m_{ts}^4 + 3 \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \sigma_i^4 \left(\sum_{s=1}^T m_{ts}^2 \right)^2 \\
&\rightarrow \kappa_4 \sum_{t=1}^T \sum_{s=1}^T m_{ts}^4 + 3 \sigma_4 \sum_{t=1}^T \left(\sum_{s=1}^T m_{ts}^2 \right)^2,
\end{aligned} \tag{A.58}$$

where $\hat{\epsilon}_{it} = i_t' M \epsilon_i$ and $M = [m_{ts}]$ for $t, s = 1, \dots, T$. Note that

$$\begin{aligned}
\sum_{s=1}^T m_{ts}^2 &= \|m_t\|^2 \\
&= i_t' M i_t \\
&= i_t' (I_T - D(D'D)^{-1}D') i_t \\
&= 1 - \text{tr} (D(D'D)^{-1}D' i_t i_t') \\
&= 1 - \text{tr} (P i_t i_t') \\
&= 1 - p_{tt} \\
&= m_{tt},
\end{aligned} \tag{A.59}$$

where p_{tt} is the (t, t) -element of P . Then, we have

$$\sum_{t=1}^T \left(\sum_{s=1}^T m_{ts}^2 \right)^2 = \sum_{t=1}^T m_{tt}^2 = \text{tr} (M^{(2)}). \tag{A.60}$$

By setting $\kappa_4 = 0$, it follows that

$$E [\hat{\sigma}_4] \rightarrow \sigma_4. \tag{A.61}$$

This concludes the proof of part (i).

(ii) As for the variance of $\hat{\sigma}_4$, we have

$$\begin{aligned}
\text{Var} \left(\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \hat{\epsilon}_{it}^4 \right) &= \frac{1}{N^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \text{Cov} (\hat{\epsilon}_{it}^4, \hat{\epsilon}_{js}^4) \\
&= \frac{1}{N^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \sum_{\substack{u_1, u_2, \\ u_3, u_4=1}}^T \sum_{\substack{v_1, v_2, \\ v_3, v_4=1}}^T m_{tu_1} m_{tu_2} m_{tu_3} m_{tu_4} m_{sv_1} m_{sv_2} m_{sv_3} m_{sv_4} \\
&\quad \times \text{Cov} (\epsilon_{iu_1} \epsilon_{iu_2} \epsilon_{iu_3} \epsilon_{iu_4}, \epsilon_{jv_1} \epsilon_{jv_2} \epsilon_{jv_3} \epsilon_{jv_4}) \\
&= \frac{1}{N^2} \sum_{i,j=1}^N \sum_{t,s=1}^T \sum_{\substack{u_1, u_2, \\ u_3, u_4=1}}^T \sum_{\substack{v_1, v_2, \\ v_3, v_4=1}}^T m_{tu_1} m_{tu_2} m_{tu_3} m_{tu_4} m_{sv_1} m_{sv_2} m_{sv_3} m_{sv_4} \\
&\quad \times \left(\kappa_8 (\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4}) \right. \\
&\quad \quad (6,2) \\
&\quad + \sum \kappa_6 (\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}) \text{Cov} (\epsilon_{jv_3}, \epsilon_{jv_4}) \\
&\quad \quad (4,4) \\
&\quad + \sum \kappa_4 (\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{jv_1}, \epsilon_{jv_2}) \kappa_4 (\epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_3}, \epsilon_{jv_4}) \\
&\quad \quad (4,2,2) \\
&\quad + \sum \kappa_4 (\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{jv_1}, \epsilon_{jv_2}) \text{Cov} (\epsilon_{iu_3}, \epsilon_{iu_4}) \text{Cov} (\epsilon_{jv_3}, \epsilon_{jv_4}) \\
&\quad \quad (2,2,2,2) \\
&\quad \left. + \sum \text{Cov} (\epsilon_{iu_1}, \epsilon_{iu_2}) \text{Cov} (\epsilon_{iu_3}, \epsilon_{jv_1}) \text{Cov} (\epsilon_{iu_4}, \epsilon_{jv_2}) \text{Cov} (\epsilon_{jv_3}, \epsilon_{jv_4}) \right), \tag{A.62}
\end{aligned}$$

where $\kappa_4(\cdot)$, $\kappa_6(\cdot)$, and $\kappa_8(\cdot)$ denote the fourth, sixth, and eighth-order mixed cumulants, respectively. By $\sum^{(\nu_1, \nu_2, \dots, \nu_k)}$ we denote the sum over all possible partitions of a group of K random variables into k subgroups of size $\nu_1, \nu_2, \dots, \nu_k$, respectively. As an example, consider $\sum^{(6,2)}$. $\sum^{(6,2)}$ defines the sum over all possible partitions of the group of eight random variables $\{\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4}\}$ into two subgroups of size six and two, respectively. Moreover, since $E[\epsilon_{it}] = E[\epsilon_{it}^3] = 0$, we do not need to consider further partitions in the above relation.¹⁹ Then, under Assumptions 5(i), 5(ii), 5(v), and 5(viii), it follows that

$$\text{Var} \left(\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \hat{\epsilon}_{it}^4 \right) = O \left(\frac{1}{N} \right) \tag{A.63}$$

and $\text{Var}(\hat{\sigma}_4) = O\left(\frac{1}{N}\right)$. This concludes the proof of part (ii).

¹⁹According to the theory on cumulants (Brillinger, 1975), evaluation of $\text{Cov}(\epsilon_{iu_1} \epsilon_{iu_2} \epsilon_{iu_3} \epsilon_{iu_4}, \epsilon_{jv_1} \epsilon_{jv_2} \epsilon_{jv_3} \epsilon_{jv_4})$ requires to consider the indecomposable partitions of the two sets $\{\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}\}$, $\{\epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4}\}$, meaning that there must be at least one subset that includes an element of both sets.

Appendix B: Main proofs

Proof of Proposition 1

Consider the class of additive bias-adjusted estimators $\hat{\Gamma}^{bias-adj}$ for Γ^P :

$$\hat{\Gamma}^{bias-adj} = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\hat{\Lambda}\hat{\Gamma}^{prelim} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\hat{\Lambda}\hat{\Gamma}^{prelim}.$$

where $\hat{\Gamma}^{prelim}$ denotes any preliminary \sqrt{N} -consistent estimator of Γ^P . Simply imposing the restriction that $\hat{\Gamma}^{bias-adj} = \hat{\Gamma}^{prelim}$, and re-arranging, one gets:

$$\left[I_{K+1} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \begin{bmatrix} 0 & 0'_K \\ 0_K & \hat{\sigma}^2(\hat{F}'\hat{F})^{-1} \end{bmatrix} \right] \hat{\Gamma}^{bias-adj} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R},$$

which implies

$$\hat{\Gamma}^{bias-adj} = \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}'\bar{R}}{N} = \hat{\Gamma}^*, \quad (\text{B.1})$$

that is, one obtains the modified estimator of Shanken (1992).

Proof of Theorem 1

(i) Starting from (11), the modified estimator of Shanken (1992) can be written as

$$\begin{aligned} \hat{\Gamma}^* &= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}'\bar{R}}{N} \\ &= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}'}{N} \left[\hat{X}\Gamma^P + \bar{\epsilon} - (\hat{X} - X)\Gamma^P \right] \\ &= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'\hat{X}}{N}\Gamma^P + \frac{\hat{X}'}{N}\bar{\epsilon} - \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P \right] \\ &= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left(\frac{\hat{X}'\hat{X}}{N}\right) \left[\Gamma^P + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}\bar{\epsilon} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P \right] \\ &= \left[I_{K+1} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda} \right]^{-1} \left[\Gamma^P + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}\bar{\epsilon} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P \right]. \end{aligned} \quad (\text{B.2})$$

Hence:

$$\begin{aligned}
\hat{\Gamma}^* - \Gamma^P &= \left(\frac{\hat{X}'\hat{X}}{N} - \hat{\Lambda} \right)^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{N} - \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P + \hat{\Lambda}\Gamma^P \right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{N} - \left(\frac{\hat{X}'}{N}(\hat{X} - X) - \hat{\Lambda} \right) \Gamma^P \right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{N} - \left[\frac{B'\epsilon'\mathcal{P}\gamma_1^P}{N} + \mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^P - \hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \right] \right]. \quad (\text{B.3})
\end{aligned}$$

By Lemmas 1(i) and 2(i), $(\hat{\Sigma}_X - \hat{\Lambda}) = O_p(1)$. In addition, Lemmas 3(i) and 5(i) imply that

$$\begin{aligned}
\frac{\hat{X}'\bar{\epsilon}}{N} &= \frac{1}{N}(\hat{X} - X)'\bar{\epsilon} + \frac{1}{N}X'\bar{\epsilon} \\
&= O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned} \quad (\text{B.4})$$

and Assumption 6(i) implies that

$$\mathcal{P}' \sum_{i=1}^N \epsilon_i = O_p\left(\sqrt{N}\right). \quad (\text{B.5})$$

Note that

$$\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^P - \hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \quad (\text{B.6})$$

can be rewritten as

$$\mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \frac{1}{N} \sum_{i=1}^N \sigma_i^2 I_T \right) \mathcal{P}\gamma_1^P - \left[(\hat{\sigma}^2 - \sigma^2) - \left(\frac{1}{N} \sum_{i=1}^N \sigma_i^2 - \sigma^2 \right) \right] (\tilde{F}'\tilde{F})^{-1}\gamma_1^P. \quad (\text{B.7})$$

Assumption 6(ii) implies that

$$\mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \frac{\sum_{i=1}^N \sigma_i^2}{N} I_T \right) \mathcal{P}\gamma_1^P = O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{B.8})$$

Using Lemma 1(i) and Assumption 5(i) concludes the proof of part (i) since $\hat{\sigma}^2 - \sigma^2 = O_p\left(\frac{1}{\sqrt{N}}\right)$ and $\frac{1}{N} \sum_{i=1}^N \sigma_i^2 - \sigma^2 = o\left(\frac{1}{\sqrt{N}}\right)$.

(ii) Starting from (B.3), we have

$$\begin{aligned}
\sqrt{N}(\hat{\Gamma}^* - \Gamma^P) &= (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{\sqrt{N}} - \left(\frac{\hat{X}'}{\sqrt{N}}(\hat{X} - X)\Gamma^P \right) + \sqrt{N}\hat{\Lambda}\Gamma^P \right] \\
&= (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{\sqrt{N}} - \begin{bmatrix} 1'_N \\ \hat{B}' \end{bmatrix} \begin{bmatrix} 0_N, & \frac{\epsilon' \mathcal{P}}{\sqrt{N}} \end{bmatrix} \Gamma^P + \sqrt{N}\hat{\Lambda}\Gamma^P \right] \\
&= (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \left[\frac{X'\bar{\epsilon}}{\sqrt{N}} + \frac{1}{\sqrt{N}} \begin{bmatrix} 0'_N \\ \mathcal{P}'\epsilon \end{bmatrix} \frac{\epsilon' 1_T}{T} - \frac{1}{\sqrt{N}} \begin{bmatrix} 1'_N \epsilon' \mathcal{P} \\ \hat{B}' \epsilon' \mathcal{P} \end{bmatrix} \gamma_1^P + \sqrt{N}\hat{\Lambda}\Gamma^P \right] \\
&= (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \left[\begin{bmatrix} 1'_N \\ B' \end{bmatrix} \frac{\epsilon' 1_T}{T\sqrt{N}} + \begin{bmatrix} -1'_N \frac{\epsilon' \mathcal{P}}{\sqrt{N}} \gamma_1^P \\ \mathcal{P}' \frac{\epsilon \epsilon'}{\sqrt{N}} \frac{1_T}{T} - B' \frac{\epsilon' \mathcal{P}}{\sqrt{N}} \gamma_1^P - \mathcal{P}' \frac{\epsilon \epsilon'}{\sqrt{N}} \mathcal{P} \gamma_1^P \end{bmatrix} \right. \\
&\quad \left. + \sqrt{N}\hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \right] \\
&= (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \left[\begin{array}{c} \frac{1'_N \epsilon'}{\sqrt{N}} \left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P \right) \\ \frac{B' \epsilon'}{\sqrt{N}} \left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P \right) + \mathcal{P}' \frac{\epsilon \epsilon'}{\sqrt{N}} \left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P \right) + \frac{\text{tr}(M\epsilon\epsilon')}{\sqrt{N}(T-K-1)} \mathcal{P}' \mathcal{P} \gamma_1^P \end{array} \right] \\
&= (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \left[\begin{bmatrix} \frac{1'_N \epsilon'}{\sqrt{N}} Q \\ \frac{B' \epsilon'}{\sqrt{N}} Q \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{P}' \frac{\epsilon \epsilon'}{\sqrt{N}} Q + \frac{\text{tr}(M\epsilon\epsilon')}{\sqrt{N}(T-K-1)} \mathcal{P}' \mathcal{P} \gamma_1^P \end{bmatrix} \right] \\
&\equiv (\hat{\Sigma}_X - \hat{\Lambda})^{-1} (I_1 + I_2). \tag{B.9}
\end{aligned}$$

Using Lemmas 1(i) and 2(ii), we have

$$(\hat{\Sigma}_X - \hat{\Lambda}) \xrightarrow{p} \left(\begin{bmatrix} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta + \sigma^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix} - \begin{bmatrix} 0 & 0'_K \\ 0_K & \sigma^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix} \right) = \Sigma_X. \tag{B.10}$$

Consider now the terms I_1 and I_2 . Both terms have mean zero and, under Assumption 5(vi), they are asymptotically uncorrelated. Assumptions 1, 5(i), 6(i), and 6(iii) imply that

$$\begin{aligned}
\text{Var}(I_1) &= E \left[\begin{array}{cc} Q' \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon'_j Q & Q' \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon'_j (Q \otimes \beta'_j) \\ \frac{1}{\sqrt{N}} \sum_{i=1}^N (Q' \otimes \beta_i) \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon'_j Q & \frac{1}{\sqrt{N}} \sum_{i=1}^N (Q' \otimes \beta_i) \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^N \epsilon'_j (Q \otimes \beta'_j) \end{array} \right] \\
&= \begin{bmatrix} Q' \frac{1}{N} \sum_{i=1}^N E[\epsilon_i \epsilon'_i] Q & Q' \frac{1}{N} \sum_{i=1}^N E[\epsilon_i \epsilon'_i] (Q \otimes \beta'_i) \\ \frac{1}{N} \sum_{i=1}^N (Q' \otimes \beta_i) E[\epsilon_i \epsilon'_i] Q & \frac{1}{N} \sum_{i=1}^N (Q' \otimes \beta_i) E[\epsilon_i \epsilon'_i] (Q \otimes \beta'_i) \end{bmatrix} + o(1) \\
&\rightarrow \begin{bmatrix} \sigma^2 Q' Q & \sigma^2 Q' (Q \otimes \mu'_\beta) \\ \sigma^2 (Q' \otimes \mu_\beta) Q & \sigma^2 (Q' Q \otimes \Sigma_\beta) \end{bmatrix} \\
&= \sigma^2 Q' Q \Sigma_X = \frac{\sigma^2}{T} \left[1 + \gamma_1^{P'} (\tilde{F}'\tilde{F}/T)^{-1} \gamma_1^P \right] \Sigma_X. \tag{B.11}
\end{aligned}$$

Next, consider I_2 . Since $\mathcal{P}' \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 Q + \frac{1}{T-K-1} \text{tr} \left(M \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2 \right) \mathcal{P}' \mathcal{P} \gamma_1^P = 0_K$, we have

$$\begin{aligned} I_2 &= \begin{bmatrix} 0 \\ (Q' \otimes P') \text{vec} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \right) + \frac{1}{T-K-1} \text{tr} \left(M \frac{1}{\sqrt{N}} \sum_{i=1}^N (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \right) \mathcal{P}' \mathcal{P} \gamma_1^P \end{bmatrix} \\ &\equiv \begin{bmatrix} 0 \\ I_{22} \end{bmatrix}. \end{aligned} \quad (\text{B.12})$$

Therefore, $\text{Var}(I_2)$ has the following form:

$$\text{Var}(I_2) = \begin{bmatrix} 0 & 0'_K \\ 0_K & E[I_{22} I_{22}'] \end{bmatrix}. \quad (\text{B.13})$$

Under Assumptions 5(i) and 6(ii), we have

$$\begin{aligned} E[I_{22} I_{22}'] &= E \left[(Q' \otimes P') \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^N \text{vec}(\epsilon_j \epsilon_j' - \sigma_j^2 I_T)' (Q \otimes P) \right] \\ &\quad + E \left[(Q' \otimes P') \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^N \text{vec}(\epsilon_j \epsilon_j' - \sigma_j^2 I_T)' \frac{\text{vec}(M)}{T-K-1} \gamma_1^{P'} \mathcal{P}' \mathcal{P} \right] \\ &\quad + E \left[\mathcal{P}' \mathcal{P} \gamma_1^P \frac{\text{vec}(M)'}{T-K-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^N \text{vec}(\epsilon_j \epsilon_j' - \sigma_j^2 I_T)' (Q \otimes P) \right] \\ &\quad + E \left[\mathcal{P}' \mathcal{P} \gamma_1^P \frac{\text{vec}(M)'}{T-K-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \text{vec}(\epsilon_i \epsilon_i' - \sigma_i^2 I_T) \frac{1}{\sqrt{N}} \sum_{j=1}^N \text{vec}(\epsilon_j \epsilon_j' - \sigma_j^2 I_T)' \frac{\text{vec}(M)}{T-K-1} \right. \\ &\quad \left. \times \gamma_1^{P'} \mathcal{P}' \mathcal{P} \right] \\ &\rightarrow \left[(Q' \otimes P') + \mathcal{P}' \mathcal{P} \gamma_1^P \frac{\text{vec}(M)'}{T-K-1} \right] U_\epsilon \left[(Q \otimes P) + \frac{\text{vec}(M)}{T-K-1} \gamma_1^{P'} \mathcal{P}' \mathcal{P} \right]. \end{aligned} \quad (\text{B.14})$$

Defining $Z = \left[(Q \otimes P) + \frac{\text{vec}(M)}{T-K-1} \gamma_1^{P'} \mathcal{P}' \mathcal{P} \right]$ concludes the proof of part (ii).

Proof of Theorem 2

By Theorem 1(i), $\hat{\gamma}_1^* \xrightarrow{P} \gamma_1^P$. Lemma 1(i) implies that $\hat{\Lambda}$ is a consistent estimator of Λ . Hence, using Lemma 2(ii), we have that $(\hat{\Sigma}_X - \hat{\Lambda}) \xrightarrow{P} \Sigma_X$, which implies that $\hat{V} \xrightarrow{P} V$. A consistent estimator of W requires a consistent estimate of the matrix U_ϵ , which can be obtained using Lemma 6. This concludes the proof of Theorem 2.

Proof of Theorem 3

We first establish a simpler, asymptotically equivalent, expression for $\sqrt{N} \left(\frac{\hat{e}^{P'} \hat{e}^P}{N} - \hat{\sigma}^2 \hat{Q}' \hat{Q} \right)$. Then, we derive the asymptotic distribution of this approximation. Consider the sample ex-post pricing

errors

$$\hat{e}^P = \bar{R} - \hat{X}\hat{\Gamma}^*. \quad (\text{B.15})$$

Starting from $\bar{R} = \hat{X}\Gamma^P + \eta^P$ with $\eta^P = \bar{\epsilon} - (\hat{X} - X)\Gamma^P$, we have

$$\begin{aligned} \hat{e}^P &= \hat{X}\Gamma^P + \bar{\epsilon} - (\hat{X} - X)\Gamma^P - \hat{X}\hat{\Gamma}^* \\ &= \bar{\epsilon} - \hat{X}(\hat{\Gamma}^* - \Gamma^P) - (\hat{X} - X)\Gamma^P. \end{aligned} \quad (\text{B.16})$$

Then,

$$\begin{aligned} \hat{e}^{P'}\hat{e}^P &= \epsilon'\bar{\epsilon} + \Gamma^{P'}(\hat{X} - X)'(\hat{X} - X)\Gamma^P \\ &\quad - 2(\hat{\Gamma}^* - \Gamma^P)'\hat{X}'\bar{\epsilon} - 2\Gamma^{P'}(\hat{X} - X)'\bar{\epsilon} \\ &\quad + 2\Gamma^{P'}(\hat{X} - X)'\hat{X}(\hat{\Gamma}^* - \Gamma^P) \\ &\quad + (\hat{\Gamma}^* - \Gamma^P)'\hat{X}'\hat{X}(\hat{\Gamma}^* - \Gamma^P). \end{aligned} \quad (\text{B.17})$$

Note that

$$\frac{\epsilon'\bar{\epsilon}}{N} = \frac{1}{T^2}1_T'\frac{\epsilon\epsilon'}{N}1_T \xrightarrow{p} \frac{\sigma^2}{T}, \quad (\text{B.18})$$

and, by Lemma 2(iii),

$$\Gamma^{P'}\frac{(\hat{X} - X)'(\hat{X} - X)}{N}\Gamma^P = \gamma_1^{P'}\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^P \xrightarrow{p} \sigma^2\gamma_1^{P'}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P. \quad (\text{B.19})$$

Using Lemmas 3(i) and 5(i) and Theorem 1, we have

$$\frac{(\hat{\Gamma}^* - \Gamma^P)'\hat{X}'\bar{\epsilon}}{N} = \frac{(\hat{\Gamma}^* - \Gamma^P)'(\hat{X} - X)'\bar{\epsilon}}{N} + \frac{(\hat{\Gamma}^* - \Gamma^P)'X'\bar{\epsilon}}{N} = O_p\left(\frac{1}{N}\right) \quad (\text{B.20})$$

and

$$\frac{\Gamma^{P'}(\hat{X} - X)'\bar{\epsilon}}{N} = O_p\left(\frac{1}{\sqrt{N}}\right). \quad (\text{B.21})$$

In addition, using Lemmas 2(i), 2(iii), 4(i), and Theorem 1, we have

$$\begin{aligned} \frac{\Gamma^{P'}(\hat{X} - X)'\hat{X}(\hat{\Gamma}^* - \Gamma^P)}{N} &= \frac{\Gamma^{P'}(\hat{X} - X)'(\hat{X} - X)(\hat{\Gamma}^* - \Gamma^P)}{N} + \frac{\Gamma^{P'}(\hat{X} - X)'X(\hat{\Gamma}^* - \Gamma^P)}{N} \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{N}\right) \end{aligned} \quad (\text{B.22})$$

and

$$\frac{(\hat{\Gamma}^* - \Gamma^P)' \hat{X}' \hat{X} (\hat{\Gamma}^* - \Gamma^P)}{N} = O_p\left(\frac{1}{N}\right). \quad (\text{B.23})$$

It follows that

$$\frac{\hat{e}^{P'} \hat{e}^P}{N} \xrightarrow{p} \frac{\sigma^2}{T} + \sigma^2 \gamma_1^{P'} (\tilde{F}' \tilde{F})^{-1} \gamma_1^P \equiv \sigma^2 Q' Q. \quad (\text{B.24})$$

Collecting terms and rewriting explicitly only the ones that are $O_p\left(\frac{1}{\sqrt{N}}\right)$, we have

$$\frac{\hat{e}^{P'} \hat{e}^P}{N} = \frac{\epsilon' \bar{\epsilon}}{N} \quad (\text{B.25})$$

$$+ \frac{\Gamma^{P'} (\hat{X} - X)' (\hat{X} - X) \Gamma^P}{N} \quad (\text{B.26})$$

$$- 2 \frac{\Gamma^{P'} (\hat{X} - X)' \bar{\epsilon}}{N} \quad (\text{B.27})$$

$$+ 2 \frac{\Gamma^{P'} (\hat{X} - X)' (\hat{X} - X) (\hat{\Gamma}^* - \Gamma^P)}{N} \quad (\text{B.28})$$

$$+ O_p\left(\frac{1}{N}\right). \quad (\text{B.29})$$

Consider the sum of the three terms in (B.25)-(B.27). Under Assumption 5(i), we have

$$\begin{aligned} & \frac{\epsilon' \bar{\epsilon}}{N} + \frac{\Gamma^{P'} (\hat{X} - X)' (\hat{X} - X) \Gamma^P}{N} - 2 \frac{\Gamma^{P'} (\hat{X} - X)' \bar{\epsilon}}{N} \\ &= \frac{1'_T \epsilon \epsilon' 1_T}{T} \frac{1_T}{N} + \gamma_1^{P'} \mathcal{P}' \frac{\epsilon \epsilon'}{N} \mathcal{P} \gamma_1^P - 2 \frac{1'_T \epsilon \epsilon'}{T} \frac{1_T}{N} \mathcal{P} \gamma_1^P \\ &= \frac{1'_T \epsilon \epsilon'}{T} \frac{1_T}{N} \left(\frac{1_T}{T} - \mathcal{P} \gamma_1^P \right) - \frac{1'_T \epsilon \epsilon'}{T} \frac{1_T}{N} \mathcal{P} \gamma_1^P + \gamma_1^{P'} \mathcal{P}' \frac{\epsilon \epsilon'}{N} \mathcal{P} \gamma_1^P \\ &= \frac{1'_T \epsilon \epsilon'}{T} \frac{1_T}{N} Q - Q' \frac{\epsilon \epsilon'}{N} \mathcal{P} \gamma_1^P \\ &= Q' \frac{\epsilon \epsilon'}{N} \frac{1_T}{T} - Q' \frac{\epsilon \epsilon'}{N} \mathcal{P} \gamma_1^P \\ &= Q' \frac{\epsilon \epsilon'}{N} Q = Q' \left(\frac{\epsilon \epsilon'}{N} - \bar{\sigma}^2 I_T \right) Q + \sigma^2 Q' Q + o\left(\frac{1}{\sqrt{N}}\right), \end{aligned} \quad (\text{B.30})$$

where the $o\left(\frac{1}{\sqrt{N}}\right)$ term comes from $(\bar{\sigma}^2 - \sigma^2)Q'Q$. As for the term in (B.28), define

$$\left(\hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}, \quad (\text{B.31})$$

where every block of $\left(\hat{\Sigma}_X - \hat{\Lambda} \right)^{-1}$ is $O_p(1)$ by the nonsingularity of Σ_X and Slutsky's theorem.

Using the same arguments as for Theorem 2, we have

$$\begin{aligned}
& \frac{2\Gamma^{P'}(\hat{X} - X)'(\hat{X} - X)(\hat{\Gamma}^* - \Gamma^P)}{N} \\
&= 2 \left[\gamma_1^{P'} \mathcal{P}' \frac{\epsilon\epsilon'}{N} \mathcal{P} \hat{\Sigma}_{21}, \gamma_1^{P'} \mathcal{P}' \frac{\epsilon\epsilon'}{N} \mathcal{P} \hat{\Sigma}_{22} \right] \left[\frac{B'\epsilon'Q}{N} + Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \right] \\
&= 2\gamma_1^{P'} \mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \mathcal{P} \hat{\Sigma}_{21} \frac{1'_N \epsilon' Q}{N} + 2\gamma_1^{P'} \mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \mathcal{P} \hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
&\quad + 2\gamma_1^{P'} \mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \mathcal{P} \hat{\Sigma}_{22} Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \\
&\quad + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{21} \frac{1'_N \epsilon' Q}{N} + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
&\quad + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + o_p \left(\frac{1}{N} \right) \\
&= 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{21} \frac{1'_N \epsilon' Q}{N} + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
&\quad + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + o_p \left(\frac{1}{N} \right) + O_p \left(\frac{1}{N} \right), \tag{B.32}
\end{aligned}$$

where the two approximations on the right-hand side of the previous expression refer to

$$\begin{aligned}
& 2(\bar{\sigma}^2 - \sigma^2) \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{21} \frac{1'_N \epsilon' Q}{N} + 2(\bar{\sigma}^2 - \sigma^2) \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
& + 2(\bar{\sigma}^2 - \sigma^2) \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) = o_p \left(\frac{1}{N} \right) \tag{B.33}
\end{aligned}$$

and

$$\begin{aligned}
& 2\gamma_1^{P'} \mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \mathcal{P} \hat{\Sigma}_{21} \frac{1'_N \epsilon' Q}{N} + 2\gamma_1^{P'} \mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \mathcal{P} \hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
& + 2\gamma_1^{P'} \mathcal{P}' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) \mathcal{P} \hat{\Sigma}_{22} Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) = O_p \left(\frac{1}{N} \right), \tag{B.34}
\end{aligned}$$

respectively. Therefore, we have

$$\begin{aligned}
\frac{\hat{e}^{P'} \hat{e}^P}{N} &= Q' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) Q + \sigma^2 Q' Q \\
&\quad + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{21} \frac{1'_N \epsilon' Q}{N} + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
&\quad + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P} \hat{\Sigma}_{22} Z' \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + O_p \left(\frac{1}{N} \right) + o_p \left(\frac{1}{N} \right) + o \left(\frac{1}{\sqrt{N}} \right). \tag{B.35}
\end{aligned}$$

It follows that

$$\begin{aligned}
\frac{\hat{e}^{P'}\hat{e}^P}{N} - \hat{\sigma}^2\hat{Q}'\hat{Q} &= Q' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) Q - \left(\hat{\sigma}^2\hat{Q}'\hat{Q} - \sigma^2 Q'Q \right) \\
&\quad + 2\sigma^2\gamma_1^{P'}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21} \frac{1'_N\epsilon'Q}{N} + 2\sigma^2\gamma_1^{P'}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
&\quad + 2\sigma^2\gamma_1^{P'}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + O_p \left(\frac{1}{N} \right) + o_p \left(\frac{1}{N} \right) + o \left(\frac{1}{\sqrt{N}} \right).
\end{aligned} \tag{B.36}$$

Note that

$$\begin{aligned}
&\hat{\sigma}^2\hat{Q}'\hat{Q} - \sigma^2Q'Q \\
&= \frac{1}{T}(\hat{\sigma}^2 - \sigma^2) + \hat{\sigma}^2\hat{\gamma}_1^{*'}(\tilde{F}'\tilde{F})^{-1}\hat{\gamma}_1^* - \sigma^2\gamma_1^{P'}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \\
&= \frac{1}{T}(\hat{\sigma}^2 - \sigma^2) + (\hat{\sigma}^2 - \sigma^2)\gamma_1^{P'}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P + 2\sigma^2(\hat{\gamma}_1^* - \gamma_1^P)'(\tilde{F}'\tilde{F})^{-1}\gamma_1^P + O_p \left(\frac{1}{N} \right) \\
&= (\hat{\sigma}^2 - \sigma^2) \left(\frac{1}{T} + \gamma_1^{P'}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \right) + 2\sigma^2(\hat{\gamma}_1^* - \gamma_1^P)'(\tilde{F}'\tilde{F})^{-1}\gamma_1^P + O_p \left(\frac{1}{N} \right) \\
&= (\hat{\sigma}^2 - \sigma^2) \left(\frac{1}{T} + \gamma_1^{P'}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \right) + 2\sigma^2\gamma_1^{P'}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21} \frac{1'_N\epsilon'Q}{N} + 2\sigma^2\gamma_1^{P'}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22} \frac{B'\epsilon'Q}{N} \\
&\quad + 2\sigma^2\gamma_1^{P'}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + O_p \left(\frac{1}{N} \right) + O_p \left(\frac{1}{N\sqrt{N}} \right),
\end{aligned} \tag{B.37}$$

where $\sigma^2(\hat{\gamma}_1^* - \gamma_1^P)'(\tilde{F}'\tilde{F})^{-1}(\hat{\gamma}_1^* - \gamma_1^P) + 2(\hat{\sigma}^2 - \sigma^2)(\hat{\gamma}_1^* - \gamma_1^P)'(\tilde{F}'\tilde{F})^{-1}\gamma_1^P = O_p \left(\frac{1}{N} \right)$ and $(\hat{\sigma}^2 - \sigma^2)(\hat{\gamma}_1^* - \gamma_1^P)'(\tilde{F}'\tilde{F})^{-1}(\hat{\gamma}_1^* - \gamma_1^P) = O_p \left(\frac{1}{N\sqrt{N}} \right)$. It follows that

$$\begin{aligned}
&\frac{\hat{e}'\hat{e}}{N} - \hat{\sigma}^2\hat{Q}'\hat{Q} \\
&= Q' \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) Q - (\hat{\sigma}^2 - \sigma^2) \left(\frac{1}{T} + \gamma_1^{P'}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \right) + O_p \left(\frac{1}{N\sqrt{N}} \right) + O_p \left(\frac{1}{N} \right) + o \left(\frac{1}{\sqrt{N}} \right) + o_p \left(\frac{1}{\sqrt{N}} \right) \\
&= \left[(Q' \otimes Q') - \frac{Q'Q}{T-K-1} \text{vec}(M)' \right] \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + o_p \left(\frac{1}{\sqrt{N}} \right) \\
&= Z'_Q \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + o_p \left(\frac{1}{\sqrt{N}} \right),
\end{aligned} \tag{B.38}$$

where we have condensed $O_p \left(\frac{1}{N\sqrt{N}} \right) + O_p \left(\frac{1}{N} \right) + o \left(\frac{1}{\sqrt{N}} \right) + o_p \left(\frac{1}{\sqrt{N}} \right)$ into the single term $o_p \left(\frac{1}{\sqrt{N}} \right)$ for simplicity. Hence,

$$\sqrt{N} \left(\frac{\hat{e}'\hat{e}}{N} - \hat{\sigma}^2\hat{Q}'\hat{Q} \right) = \sqrt{N} Z'_Q \text{vec} \left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T \right) + o_p(1), \tag{B.39}$$

implying that the asymptotic distribution of $\sqrt{N}\left(\frac{\hat{\epsilon}'\hat{\epsilon}}{N} - \hat{\sigma}^2\hat{Q}'\hat{Q}\right)$ is equivalent to the asymptotic distribution of $\sqrt{N}Z'_Q\text{vec}\left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2I_T\right)$. Finally, by Assumption 6(ii), we have

$$\sqrt{N}Z'_Q\text{vec}\left(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2I_T\right) \xrightarrow{d} \mathcal{N}\left(0, Z'_QU_\epsilon Z_Q\right). \quad (\text{B.40})$$

This concludes the proof of Theorem 3.

Appendix C: Form of U_ϵ

Denote by U_ϵ the $T^2 \times T^2$ matrix

$$U_\epsilon = \begin{bmatrix} U_{11} & \cdots & U_{1t} & \cdots & U_{1T} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ U_{t1} & \cdots & U_{tt} & \cdots & U_{tT} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{T1} & \cdots & U_{Tt} & \cdots & U_{TT} \end{bmatrix}. \quad (\text{C.1})$$

Each block of U_ϵ is a $T \times T$ matrix. The blocks along the main diagonal, denoted by U_{tt} , $t = 1, 2, \dots, T$, are themselves diagonal matrices with $(\kappa_4 + 2\sigma_4)$ in the (t, t) -th position and σ_4 in the (s, s) position for every $s \neq t$, that is,

$$U_{tt} = \begin{matrix} \downarrow \\ t\text{-th column} \\ \begin{bmatrix} \sigma_4 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sigma_4 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & \sigma_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \sigma_4 \end{bmatrix} \end{matrix} \xrightarrow{t\text{-th row}}. \quad (\text{C.2})$$

The blocks outside the main diagonal, denoted by U_{ts} , $s, t = 1, 2, \dots, T$ with $s \neq t$, are all made of

zeros except for the (s, t) -th position that contains σ_4 , that is,

$$\begin{aligned}
 & \qquad \qquad \qquad \qquad \qquad \qquad \qquad \downarrow \\
 & \qquad \qquad \qquad \qquad \qquad \qquad \qquad t\text{-th column} \\
 U_{ts} &= \begin{matrix} \rightarrow \\ s\text{-th row} \end{matrix} \begin{bmatrix} 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_4 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix}. \tag{C.3}
 \end{aligned}$$

Under Assumption 5 and Lemma 6 in Appendix A, it is easy to show that \hat{U}_ϵ in Theorem 2 is a consistent plug-in estimator of U_ϵ that only depends on $\hat{\sigma}_4$.

References

- Bai, Jushan, and Guofu Zhou, 2015, Fama-MacBeth two-pass regressions: Improving risk premia estimates, *Finance Research Letters*, forthcoming.
- Barillas, F. and Shanken, J. 2016, Which Alpha?, *Review of Financial Studies*, forthcoming.
- Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).
- Brillinger, David A., 1975, *Time Series: Data Analysis and Theory*, 3rd ed., Holt, Rinehart & Winston, Toronto.
- Chamberlain, G., and M. Rothschild, 1983, Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets, *Econometrica* 51, 1281–1304.
- Chordia, Tarun, Amit Goyal, and Jay Shanken, 2015, Cross-sectional asset pricing with individual stocks: Betas versus characteristics, Working Paper, Emory University.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607–636.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.
- Gagliardini, Patrik, Elisa Ossola, and Olivier Scaillet, 2015, Time-varying risk premium in large cross-sectional equity datasets, *Econometrica*, forthcoming.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.
- Gospodinov, Nikolay, and Cesare Robotti, 2012, Asset pricing theories, models, and tests, in H. Kent Baker, Greg Filbeck, eds.: *Portfolio Theory and Management*, Chapter 3, 46–72, Oxford University Press.

- Greene, William, 2008, *Econometric Analysis*. Prentice Hall.
- Hou, Kewei, and Robert Kimmel, 2006, On the estimation of risk premia in linear factor models, Working Paper, Ohio State University.
- Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.
- Jagannathan, Ravi, and Zhenyu Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *Journal of Finance* 53, 1285–1309.
- Jagannathan, Ravi, Georgios Skoulakis, and Zhenyu Wang, 2010, The analysis of the cross-section of security returns, in Yacine Ait-Sahalia and Lars Hansen, eds.: *Handbook of Financial Econometrics* (Elsevier, Amsterdam).
- Jegadeesh, Narasimhan, and Joonki Noh, 2014, Empirical tests of asset pricing models with individual stocks, Working Paper, Emory University.
- Kan, Raymond, and Cesare Robotti, 2012, Evaluation of asset pricing models using two-pass cross-sectional regressions, in Jin-Chuan Duan, James E. Gentle, Wolfgang Haerdle, eds.: *Handbook of Computational Finance*, Chapter 9, 223–251, Springer.
- Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.
- Kim, Soohun, and Georgios Skoulakis, 2014, Estimating and testing linear factor models using large cross sections: The regression-calibration approach, Working Paper, Georgia Institute of Technology.
- Harvey, C.R. and Liu, Y. and Zhu, H. 2016, ... and the Cross-Section of Expected Returns. Working Paper, Duke.
- Lewellen, J. and Nagel, S. and Shanken, J. 2010, A Skeptical Appraisal of Asset Pricing Tests, *Journal of Financial Economics* 96, 175-194.
- Lintner, J, 1965, The Valuation of Risk Assets and The Selection of Risky Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics* 47, 13–37

- Litzenberger, Robert H., and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of Financial Economics* 7, 16–195.
- Pesaran, Hashem M., and Takashi Yamagata, 2012, Testing CAPM with a large number of assets, Working Paper, Cambridge University.
- Pukthuanthong, Kuntara, Richard Roll, and Junbo Wang, 2014, Resolving the errors-in-variables bias in risk premium estimation, Working Paper, California Institute of Technology.
- Ross, S. 1976, The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory* 13, 341–360.
- Rubin, Donald B., 1976, Inference and missing data, *Biometrika* 63, 581–592.
- Shanken, Jay, 1985, Multivariate tests of the zero-beta CAPM, *Journal of Financial Economics* 14, 327–348.
- Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.
- Shanken, Jay, and Guofu Zhou, 2007, Estimating and testing beta pricing models: Alternative methods and their performance in simulations, *Journal of Financial Economics* 84, 40–86.
- Sharpe, W. F. 1963, A Simplified Model for Portfolio Analysis, *Management Science* 9, 277–293.

Table I
Bias and RMSE of the OLS and OLS Bias-Adjusted Estimators in a One-Factor Model (Σ Scalar)

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.

Statistics	$N = 100$	$N = 500$	$N = 1000$	$N = 3000$
Panel A: $T = 36$				
Bias($\hat{\gamma}_0$)	28.8%	26.2%	24.6%	22.9%
Bias($\hat{\gamma}_0^*$)	-2.3%	-0.3%	0.3%	-0.2%
RMSE($\hat{\gamma}_0$)	0.3675	0.1875	0.1427	0.1066
RMSE($\hat{\gamma}_0^*$)	0.4509	0.1892	0.1255	0.0699
Bias($\hat{\gamma}_1$)	-24.8%	-20.0%	-18.8%	-17.8%
Bias($\hat{\gamma}_1^*$)	1.8%	0.1%	-0.2%	0.2%
RMSE($\hat{\gamma}_1$)	0.3539	0.1642	0.1277	0.1000
RMSE($\hat{\gamma}_1^*$)	0.4529	0.1655	0.1098	0.0609
Panel B: $T = 72$				
Bias($\hat{\gamma}_0$)	11.6%	9.8%	8.7%	7.9%
Bias($\hat{\gamma}_0^*$)	-0.8%	-0.0%	-0.0%	-0.1%
RMSE($\hat{\gamma}_0$)	0.2504	0.1198	0.0877	0.0628
RMSE($\hat{\gamma}_0^*$)	0.2881	0.1165	0.0766	0.0426
Bias($\hat{\gamma}_1$)	-18.5%	-14.1%	-12.4%	-11.7%
Bias($\hat{\gamma}_1^*$)	1.0%	-0.0%	0.2%	0.1%
RMSE($\hat{\gamma}_1$)	0.2437	0.1063	0.0787	0.0597
RMSE($\hat{\gamma}_1^*$)	0.2868	0.1026	0.0674	0.0379

Table II
Bias and RMSE of the OLS and OLS Bias-Adjusted Estimators in a One-Factor Model (Σ Diagonal)

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.

Statistics	$N = 100$	$N = 500$	$N = 1000$	$N = 3000$
Panel A: $T = 36$				
Bias($\hat{\gamma}_0$)	30.1%	25.8%	24.8%	23.0%
Bias($\hat{\gamma}_0^*$)	-0.7%	-0.8%	0.4%	-0.1%
RMSE($\hat{\gamma}_0$)	0.4047	0.1976	0.1495	0.1100
RMSE($\hat{\gamma}_0^*$)	0.5027	0.2054	0.1364	0.0763
Bias($\hat{\gamma}_1$)	-25.5%	-19.6%	-18.7%	-17.9%
Bias($\hat{\gamma}_1^*$)	0.9%	0.6%	-0.1%	0.1%
RMSE($\hat{\gamma}_1$)	0.3949	0.1733	0.1339	0.1033
RMSE($\hat{\gamma}_1^*$)	0.5104	0.1815	0.1208	0.0681
Panel B: $T = 72$				
Bias($\hat{\gamma}_0$)	11.2%	10.0%	8.6%	8.0%
Bias($\hat{\gamma}_0^*$)	-1.2%	0.2%	-0.1%	0.0%
RMSE($\hat{\gamma}_0$)	0.2673	0.1246	0.0899	0.0643
RMSE($\hat{\gamma}_0^*$)	0.3116	0.1223	0.0804	0.0446
Bias($\hat{\gamma}_1$)	-18.1%	-14.3%	-12.3%	-11.8%
Bias($\hat{\gamma}_1^*$)	1.5%	-0.3%	0.3%	-0.0%
RMSE($\hat{\gamma}_1$)	0.2621	0.1112	0.0809	0.0612
RMSE($\hat{\gamma}_1^*$)	0.3120	0.1087	0.0711	0.0400

Table III
Bias and RMSE of the OLS and OLS Bias-Adjusted Estimators in a One-Factor Model (Σ Full, $\delta = 0.5$)

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.

Statistics	$N = 100$	$N = 500$	$N = 1000$	$N = 3000$
Panel A: $T = 36$				
Bias($\hat{\gamma}_0$)	28.8%	26.0%	24.6%	22.7%
Bias($\hat{\gamma}_0^*$)	-2.6%	-0.6%	0.3%	-0.4%
RMSE($\hat{\gamma}_0$)	0.4065	0.1960	0.1506	0.1089
RMSE($\hat{\gamma}_0^*$)	0.5081	0.2031	0.1385	0.0760
Bias($\hat{\gamma}_1$)	-24.2%	-19.6%	-18.9%	-17.7%
Bias($\hat{\gamma}_1^*$)	2.7%	0.7%	-0.3%	0.3%
RMSE($\hat{\gamma}_1$)	0.3963	0.1727	0.1352	0.1028
RMSE($\hat{\gamma}_1^*$)	0.5159	0.1806	0.1220	0.0681
Panel B: $T = 72$				
Bias($\hat{\gamma}_0$)	11.8%	9.4%	8.6%	8.0%
Bias($\hat{\gamma}_0^*$)	-0.5%	-0.5%	-0.1%	-0.0%
RMSE($\hat{\gamma}_0$)	0.2671	0.1227	0.0910	0.0642
RMSE($\hat{\gamma}_0^*$)	0.3099	0.1225	0.0820	0.0447
Bias($\hat{\gamma}_1$)	-19.0%	-13.6%	-12.4%	-11.7%
Bias($\hat{\gamma}_1^*$)	0.5%	0.6%	0.1%	0.1%
RMSE($\hat{\gamma}_1$)	0.2614	0.1104	0.0819	0.0611
RMSE($\hat{\gamma}_1^*$)	0.3096	0.1100	0.0720	0.0405

Table IV
Bias and RMSE of the OLS and OLS Bias-Adjusted Estimators in a One-Factor Model (Σ Full, $\delta = 0.25$)

The table reports the percentage bias (Bias) and root mean squared error (RMSE), all in percent, over 10,000 simulated data sets. The model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12.

Statistics	$N = 100$	$N = 500$	$N = 1000$	$N = 3000$
Panel A: $T = 36$				
Bias($\hat{\gamma}_0$)	28.8%	26.6%	24.2%	23.5%
Bias($\hat{\gamma}_0^*$)	-2.5%	0.1%	-0.3%	0.5%
RMSE($\hat{\gamma}_0$)	0.4191	0.2053	0.1536	0.1135
RMSE($\hat{\gamma}_0^*$)	0.5254	0.2152	0.1450	0.0809
Bias($\hat{\gamma}_1$)	-24.8%	-19.9%	-18.5%	-18.3%
Bias($\hat{\gamma}_1^*$)	2.0%	0.2%	0.2%	-0.4%
RMSE($\hat{\gamma}_1$)	0.4116	0.1824	0.1380	0.1072
RMSE($\hat{\gamma}_1^*$)	0.5355	0.1935	0.1288	0.0731
Panel B: $T = 72$				
Bias($\hat{\gamma}_0$)	12.2%	9.7%	8.8%	7.9%
Bias($\hat{\gamma}_0^*$)	-0.1%	-0.2%	0.1%	-0.1%
RMSE($\hat{\gamma}_0$)	0.2795	0.1287	0.0939	0.0645
RMSE($\hat{\gamma}_0^*$)	0.3252	0.1292	0.0853	0.0459
Bias($\hat{\gamma}_1$)	-19.3%	-13.9%	-12.6%	-11.7%
Bias($\hat{\gamma}_1^*$)	0.0%	0.2%	-0.1%	0.2%
RMSE($\hat{\gamma}_1$)	0.2761	0.1155	0.0854	0.0615
RMSE($\hat{\gamma}_1^*$)	0.3279	0.1158	0.0763	0.0416

Table V
Size of t -tests in a One-Factor Model (Σ Scalar)

The table presents the size properties of t -tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. $t_{FM}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the traditional Fama-MacBeth standard error, while $t_{EIV}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992). Finally, the rejection rates for the t -test associated with the OLS bias-adjusted estimator are based on the asymptotic distribution in part (ii) of Theorem 1. The t -statistics are compared with the critical values from a standard normal distribution.

Panel A: $T = 36$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.128	0.074	0.021	0.141	0.078	0.022
500	0.186	0.113	0.040	0.213	0.132	0.047
1000	0.243	0.156	0.059	0.290	0.197	0.075
3000	0.438	0.324	0.153	0.538	0.416	0.219
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.127	0.073	0.020	0.140	0.077	0.022
500	0.185	0.113	0.039	0.211	0.132	0.047
1000	0.243	0.156	0.059	0.289	0.197	0.075
3000	0.437	0.323	0.152	0.537	0.415	0.218
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.097	0.051	0.010	0.100	0.048	0.010
500	0.105	0.053	0.011	0.107	0.055	0.012
1000	0.103	0.052	0.010	0.105	0.054	0.011
3000	0.098	0.051	0.011	0.100	0.049	0.010

Table V (Continued)
Size of t -tests in a One-Factor Model (Σ Scalar)

Panel B: $T = 72$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.123	0.063	0.016	0.124	0.066	0.016
500	0.167	0.099	0.030	0.181	0.109	0.033
1000	0.211	0.133	0.041	0.237	0.154	0.053
3000	0.378	0.263	0.109	0.449	0.333	0.150
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.122	0.063	0.015	0.123	0.065	0.016
500	0.166	0.099	0.030	0.181	0.108	0.033
1000	0.210	0.132	0.040	0.236	0.153	0.052
3000	0.377	0.261	0.108	0.448	0.331	0.149
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.096	0.047	0.009	0.100	0.048	0.009
500	0.097	0.049	0.010	0.098	0.049	0.010
1000	0.100	0.047	0.009	0.103	0.048	0.009
3000	0.103	0.054	0.010	0.106	0.054	0.010

Table VI
Size of t -tests in a One-Factor Model (Σ Diagonal)

The table presents the size properties of t -tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. $t_{FM}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the traditional Fama-MacBeth standard error, while $t_{EIV}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992). Finally, the rejection rates for the t -test associated with the OLS bias-adjusted estimator are based on the asymptotic distribution in part (ii) of Theorem 1. The t -statistics are compared with the critical values from a standard normal distribution.

Panel A: $T = 36$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.122	0.066	0.019	0.125	0.072	0.018
500	0.163	0.104	0.033	0.179	0.112	0.036
1000	0.226	0.141	0.050	0.248	0.166	0.060
3000	0.398	0.292	0.128	0.474	0.362	0.174
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.120	0.065	0.018	0.124	0.070	0.017
500	0.163	0.103	0.033	0.179	0.111	0.036
1000	0.225	0.141	0.050	0.247	0.165	0.060
3000	0.397	0.291	0.127	0.473	0.362	0.173
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.093	0.045	0.011	0.091	0.044	0.010
500	0.102	0.051	0.010	0.096	0.049	0.011
1000	0.099	0.048	0.009	0.101	0.051	0.009
3000	0.099	0.053	0.012	0.099	0.051	0.010

Table VI (Continued)
Size of t -tests in a One-Factor Model (Σ Diagonal)

Panel B: $T = 72$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.115	0.060	0.015	0.121	0.064	0.015
500	0.157	0.089	0.027	0.165	0.096	0.030
1000	0.199	0.121	0.036	0.219	0.137	0.044
3000	0.353	0.250	0.103	0.416	0.302	0.134
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.114	0.059	0.014	0.119	0.063	0.015
500	0.157	0.089	0.027	0.163	0.096	0.029
1000	0.198	0.120	0.036	0.218	0.136	0.044
3000	0.351	0.248	0.102	0.414	0.301	0.132
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.097	0.048	0.010	0.096	0.048	0.007
500	0.095	0.046	0.010	0.093	0.047	0.010
1000	0.097	0.049	0.011	0.095	0.049	0.010
3000	0.103	0.052	0.010	0.102	0.051	0.010

Table VII
Size of t -tests in a One-Factor Model (Σ Full, $\delta = 0.5$)

The table presents the size properties of t -tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. $t_{FM}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the traditional Fama-MacBeth standard error, while $t_{EIV}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992). Finally, the rejection rates for the t -test associated with the OLS bias-adjusted estimator are based on the asymptotic distribution in part (ii) of Theorem 1. The t -statistics are compared with the critical values from a standard normal distribution.

Panel A: $T = 36$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.126	0.069	0.020	0.125	0.070	0.021
500	0.166	0.097	0.030	0.181	0.109	0.034
1000	0.227	0.143	0.049	0.258	0.170	0.063
3000	0.393	0.282	0.123	0.472	0.354	0.168
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.124	0.068	0.019	0.123	0.068	0.021
500	0.166	0.096	0.030	0.180	0.109	0.034
1000	0.227	0.142	0.049	0.257	0.170	0.063
3000	0.392	0.281	0.122	0.470	0.353	0.167
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.097	0.045	0.012	0.094	0.046	0.011
500	0.094	0.045	0.009	0.095	0.045	0.010
1000	0.106	0.051	0.011	0.102	0.050	0.010
3000	0.100	0.051	0.011	0.100	0.053	0.011

Table VII (Continued)
Size of t -tests in a One-Factor Model (Σ Full, $\delta = 0.5$)

Panel B: $T = 72$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.113	0.062	0.014	0.119	0.061	0.014
500	0.150	0.086	0.025	0.165	0.096	0.029
1000	0.202	0.127	0.041	0.228	0.141	0.047
3000	0.353	0.246	0.102	0.417	0.302	0.137
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.112	0.062	0.014	0.117	0.060	0.014
500	0.149	0.085	0.025	0.164	0.096	0.029
1000	0.201	0.126	0.041	0.227	0.141	0.047
3000	0.352	0.244	0.100	0.415	0.301	0.136
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.094	0.046	0.010	0.091	0.044	0.009
500	0.095	0.047	0.010	0.094	0.050	0.011
1000	0.105	0.052	0.011	0.102	0.052	0.010
3000	0.102	0.052	0.012	0.102	0.053	0.013

Table VIII
Size of t -tests in a One-Factor Model (Σ Full, $\delta = 0.25$)

The table presents the size properties of t -tests of statistical significance. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12. $t_{FM}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the traditional Fama-MacBeth standard error, while $t_{EIV}(\cdot)$ denotes the t -statistic associated with the OLS estimator that uses the EIV-adjusted standard error in Theorem 1(ii) of Shanken (1992). Finally, the rejection rates for the t -test associated with the OLS bias-adjusted estimator are based on the asymptotic distribution in part (ii) of Theorem 1. The t -statistics are compared with the critical values from a standard normal distribution.

Panel A: $T = 36$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.125	0.068	0.017	0.124	0.068	0.018
500	0.163	0.095	0.034	0.174	0.109	0.039
1000	0.215	0.131	0.046	0.241	0.155	0.057
3000	0.389	0.280	0.125	0.459	0.343	0.164
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.123	0.067	0.017	0.123	0.067	0.017
500	0.162	0.095	0.033	0.174	0.109	0.039
1000	0.214	0.130	0.046	0.240	0.155	0.057
3000	0.388	0.278	0.124	0.458	0.341	0.163
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.109	0.060	0.015	0.112	0.059	0.015
500	0.115	0.062	0.018	0.117	0.064	0.019
1000	0.122	0.065	0.016	0.119	0.066	0.017
3000	0.121	0.069	0.018	0.124	0.068	0.018

Table VIII (Continued)
Size of t -tests in a One-Factor Model (Σ Full, $\delta = 0.25$)

Panel B: $T = 72$

N	0.10	0.05	0.01	0.10	0.05	0.01
	$t_{FM}(\hat{\gamma}_0)$			$t_{FM}(\hat{\gamma}_1)$		
100	0.119	0.060	0.014	0.123	0.066	0.015
500	0.155	0.091	0.025	0.163	0.098	0.030
1000	0.199	0.126	0.042	0.222	0.138	0.050
3000	0.334	0.229	0.092	0.390	0.280	0.124
	$t_{EIV}(\hat{\gamma}_0)$			$t_{EIV}(\hat{\gamma}_1)$		
100	0.117	0.059	0.014	0.122	0.065	0.015
500	0.155	0.090	0.025	0.162	0.098	0.030
1000	0.198	0.125	0.042	0.222	0.138	0.049
3000	0.333	0.228	0.091	0.388	0.278	0.123
	$t(\hat{\gamma}_0^*)$			$t(\hat{\gamma}_1^*)$		
100	0.108	0.057	0.012	0.110	0.059	0.015
500	0.114	0.062	0.015	0.119	0.065	0.015
1000	0.121	0.063	0.015	0.122	0.067	0.016
3000	0.111	0.057	0.012	0.114	0.058	0.014

Table IX
Rejection Rates of the Specification Test in a One-Factor Model

The table presents the size and power properties of the test of correct model specification. The null hypothesis is that the model is correctly specified. The alternative hypothesis is that the model is misspecified. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2010:12 ($T = 36$). Finally, the rejection rates for the specification test are based on the asymptotic distribution in Theorem 3. The rejection rates of the test are based on two-sided p -values.

N	Size			Power		
	10%	5%	1%	10%	5%	1%
Panel A: Σ Scalar						
100	0.103	0.049	0.009	0.882	0.823	0.675
500	0.098	0.050	0.009	1.000	1.000	0.998
1000	0.101	0.052	0.011	1.000	1.000	1.000
3000	0.101	0.050	0.009	1.000	1.000	1.000
Panel B: Σ Diagonal						
100	0.085	0.037	0.010	0.634	0.529	0.340
500	0.093	0.046	0.010	0.983	0.967	0.894
1000	0.099	0.050	0.009	1.000	1.000	0.996
3000	0.097	0.046	0.011	1.000	1.000	1.000
Panel C: Σ Full ($\delta = 0.5$)						
100	0.084	0.040	0.011	0.639	0.534	0.332
500	0.101	0.050	0.012	0.982	0.965	0.887
1000	0.095	0.049	0.011	1.000	1.000	0.997
3000	0.108	0.056	0.011	1.000	1.000	1.000
Panel D: Σ Full ($\delta = 0.25$)						
100	0.110	0.060	0.021	0.621	0.522	0.336
500	0.145	0.084	0.029	0.977	0.956	0.874
1000	0.145	0.088	0.029	1.000	0.999	0.993
3000	0.146	0.087	0.030	1.000	1.000	1.000

Table X
Rejection Rates of the Specification Test in a One-Factor Model

The table presents the size and power properties of the test of correct model specification. The null hypothesis is that the model is correctly specified. The alternative hypothesis is that the model is misspecified. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 10,000 simulations, assuming that the model disturbances are generated from a multivariate normal distribution with a covariance matrix calibrated to 3000 NYSE-AMEX-NASDAQ individual stock returns over the period 2008:1–2013:12 ($T = 72$). Finally, the rejection rates for the specification test are based on the asymptotic distribution in Theorem 3. The rejection rates of the test are based on two-sided p -values.

N	Size			Power		
	10%	5%	1%	10%	5%	1%
Panel A: Σ Scalar						
100	0.095	0.045	0.009	0.929	0.891	0.781
500	0.101	0.047	0.009	1.000	1.000	1.000
1000	0.104	0.055	0.010	1.000	1.000	1.000
3000	0.099	0.048	0.010	1.000	1.000	1.000
Panel B: Σ Diagonal						
100	0.085	0.041	0.010	0.771	0.676	0.480
500	0.098	0.046	0.010	1.000	1.000	0.997
1000	0.101	0.049	0.012	1.000	1.000	1.000
3000	0.102	0.051	0.011	1.000	1.000	1.000
Panel C: Σ Full ($\delta = 0.5$)						
100	0.085	0.039	0.011	0.770	0.681	0.482
500	0.092	0.046	0.009	1.000	0.999	0.996
1000	0.094	0.049	0.010	1.000	1.000	1.000
3000	0.097	0.047	0.010	1.000	1.000	1.000
Panel D: Σ Full ($\delta = 0.25$)						
100	0.120	0.063	0.023	0.749	0.660	0.470
500	0.140	0.083	0.029	1.000	0.999	0.994
1000	0.149	0.086	0.030	1.000	1.000	1.000
3000	0.153	0.093	0.034	1.000	1.000	1.000

Figure I
CAPM:

estimates and confidence intervals of market excess return (*mkt*) risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the market excess return obtained by estimating the CAPM. We report the time series of the Shanken estimator $\hat{\gamma}_{mkt}^*$ (blue line), obtained with a rolling time window of three- (top panel) , six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{mkt}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{mkt}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP database, monthly observations from January 1966 until December 2013.

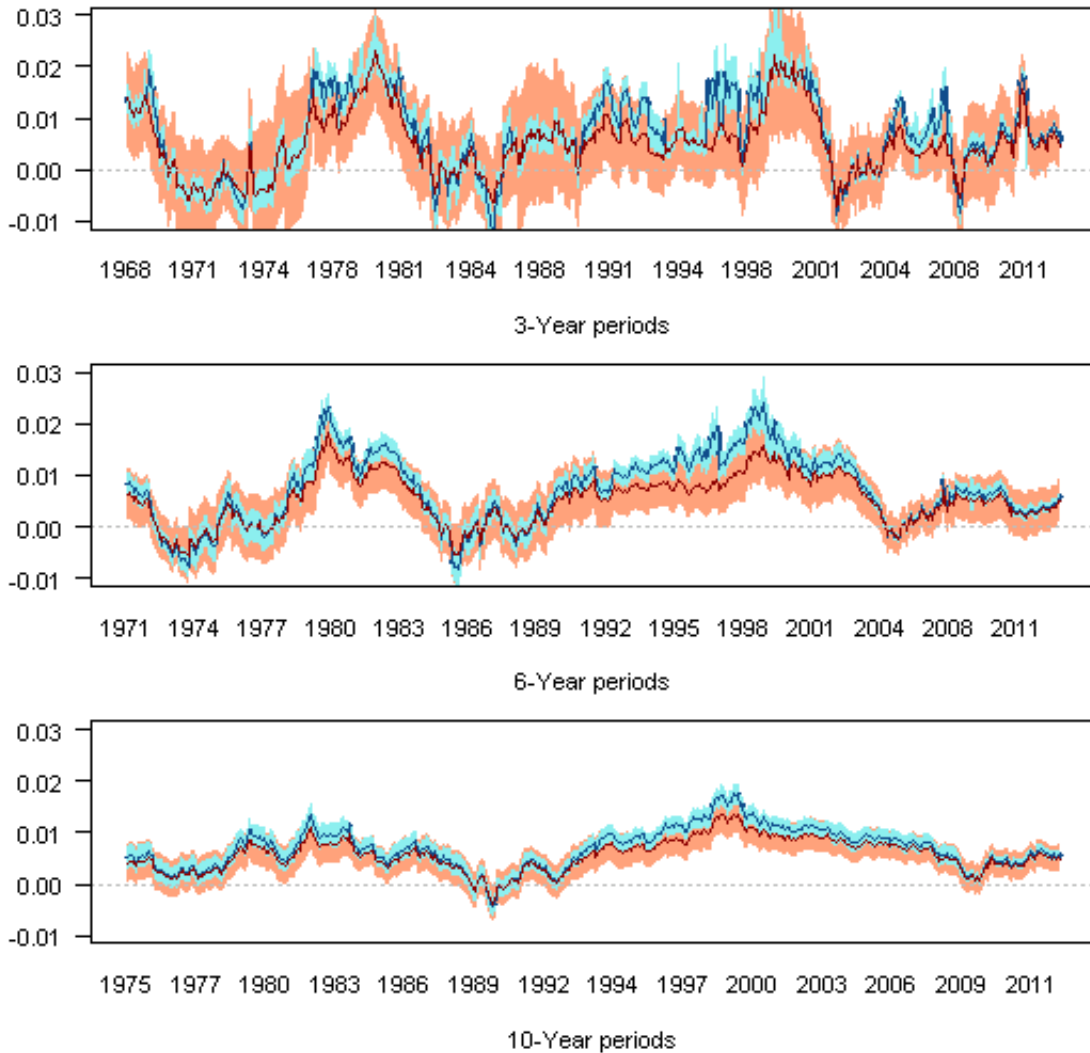


Figure II.a
Fama and French (1993) three-factor model:
estimates and confidence intervals of market excess return (*mkt*) risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the market excess return obtained by estimating the Fama and French (1993) three-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{mkt}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{mkt}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{mkt}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

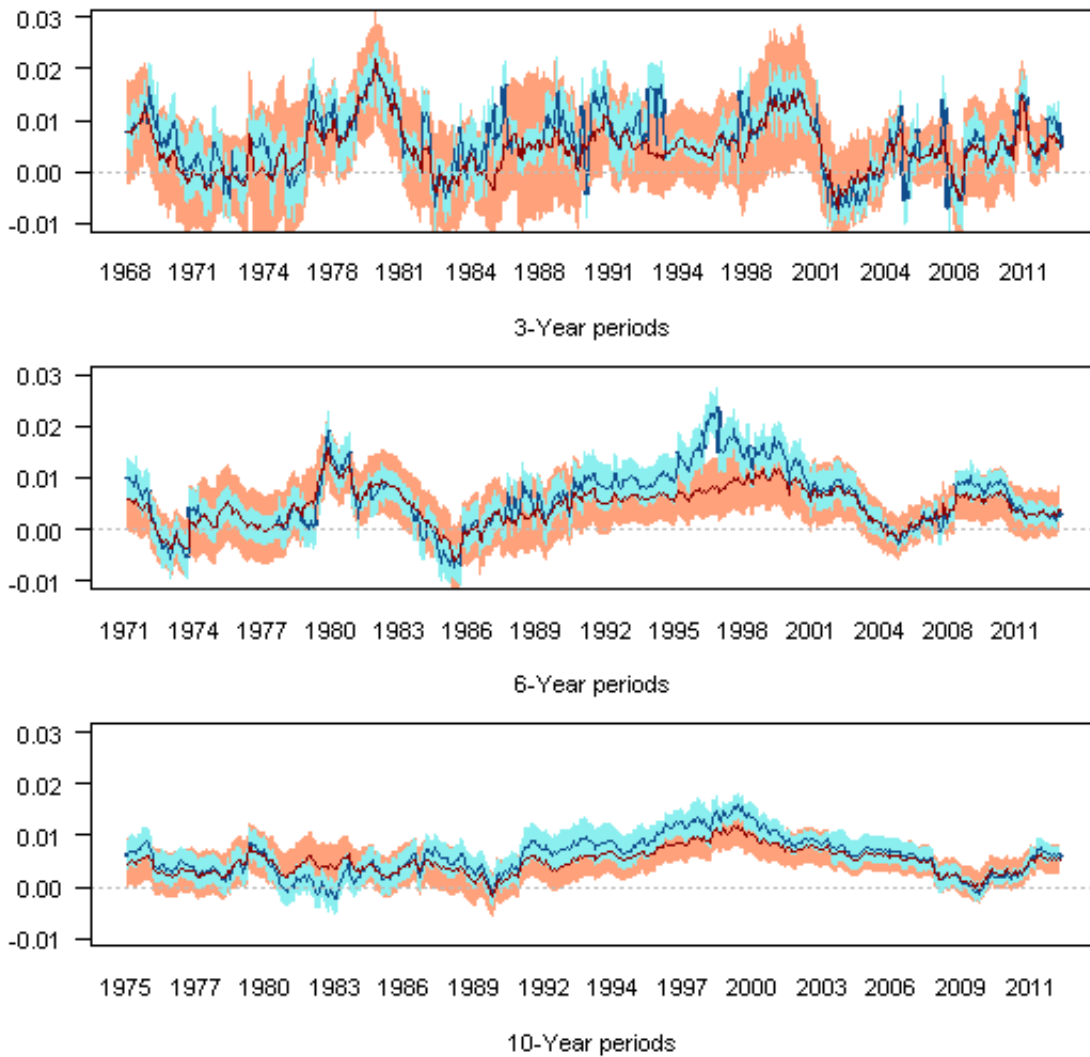


Figure II.b
Fama and French (1993) three-factor model:
estimates and confidence intervals of small-minus-big (*smb*) factor risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the *smb* factor (the return difference between portfolios of stocks with small and large market capitalizations) obtained by estimating the Fama and French (1993) three-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{smb}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{smb}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{smb}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

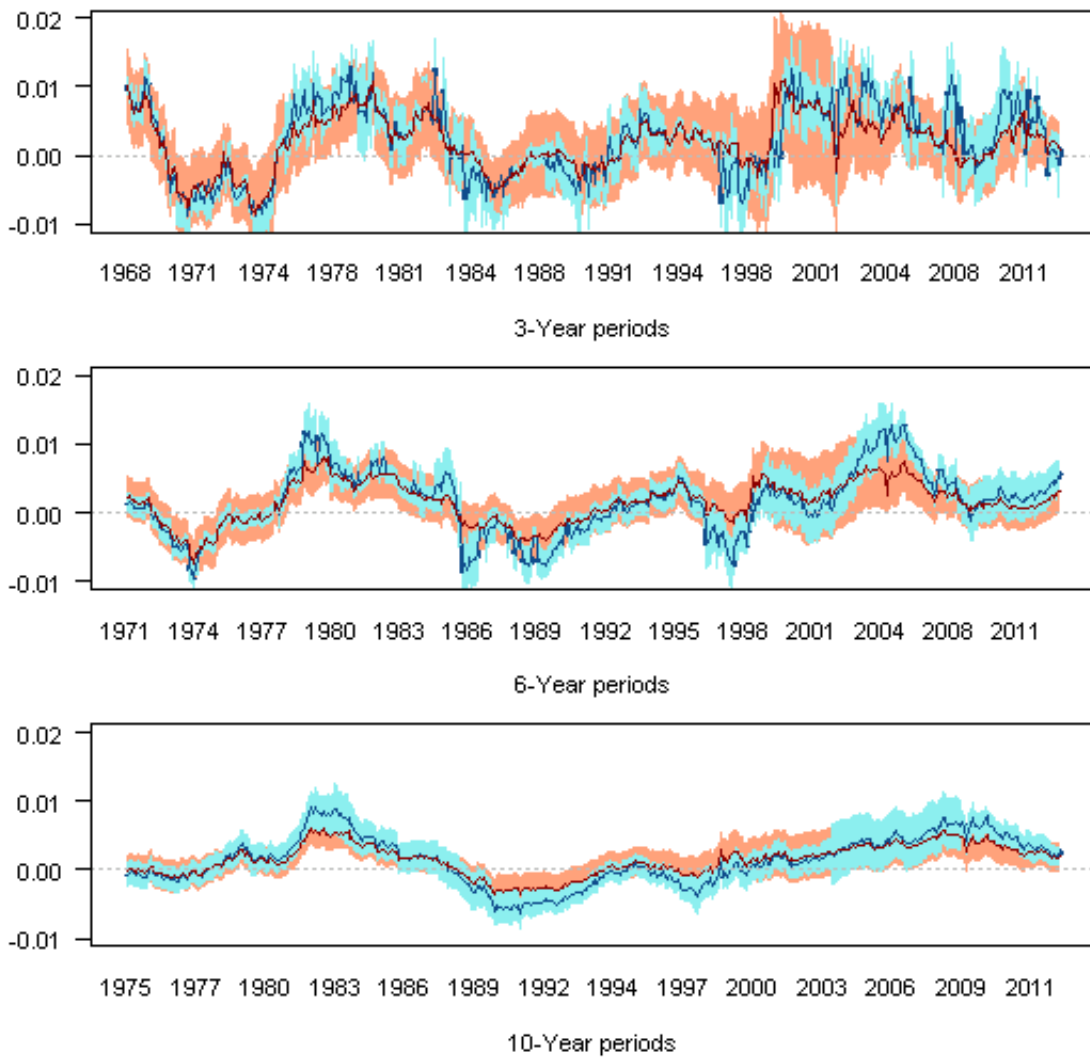


Figure II.c
Fama and French (1993) three-factor model:
estimates and confidence intervals of high-minus-low (*hml*) factor risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the *hml* factor (the return difference between portfolios of stocks with high and low book-to-market) obtained by estimating the Fama and French (1993) three-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{hml}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{hml}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{hml}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

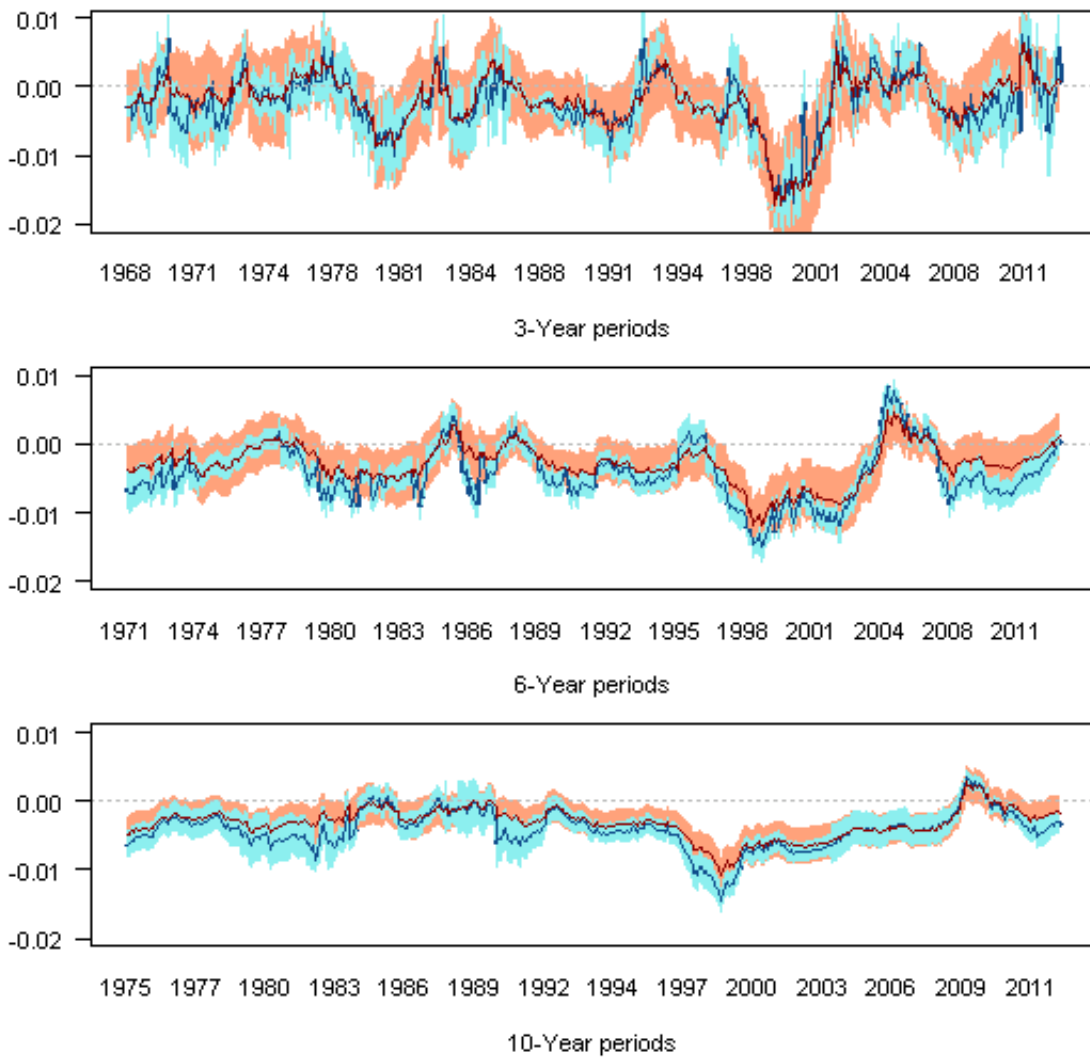


Figure III.a
Fama and French (2014) five-factor model:
estimates and confidence intervals of market excess return (*mkt*) risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the market excess return obtained by estimating the Fama and French (2014) five-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{mkt}^*$ (blue line), obtained with a rolling time window of three- (top panel) , six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{mkt}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{mkt}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

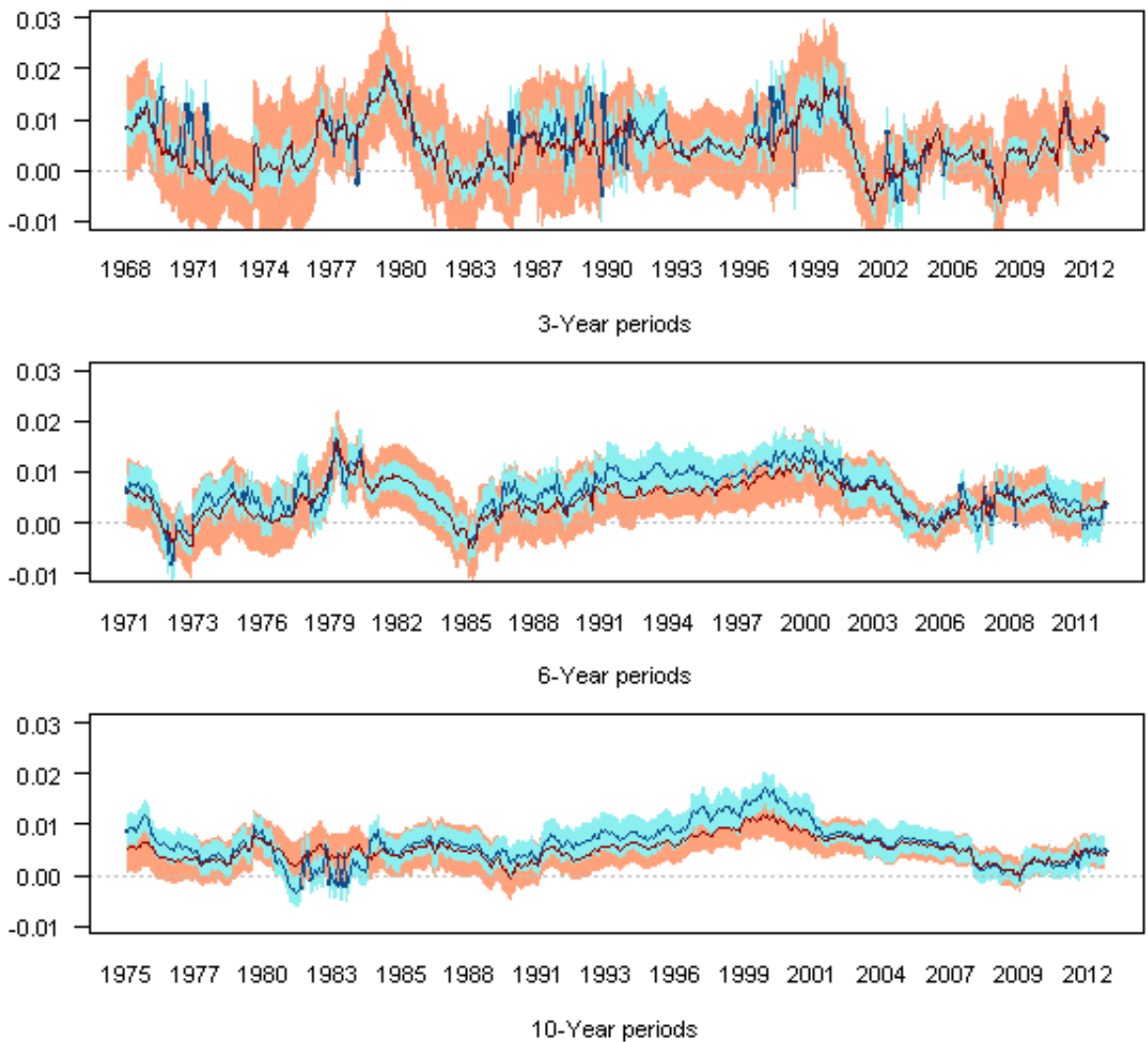


Figure III.b

Fama and French (2014) five-factor model:

estimates and confidence intervals of small-minus-big (*smb*) factor risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the *smb* factor (the return difference between portfolios of stocks with small and large market capitalizations) obtained by estimating the Fama and French (2014) five-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{smb}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{smb}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{smb}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

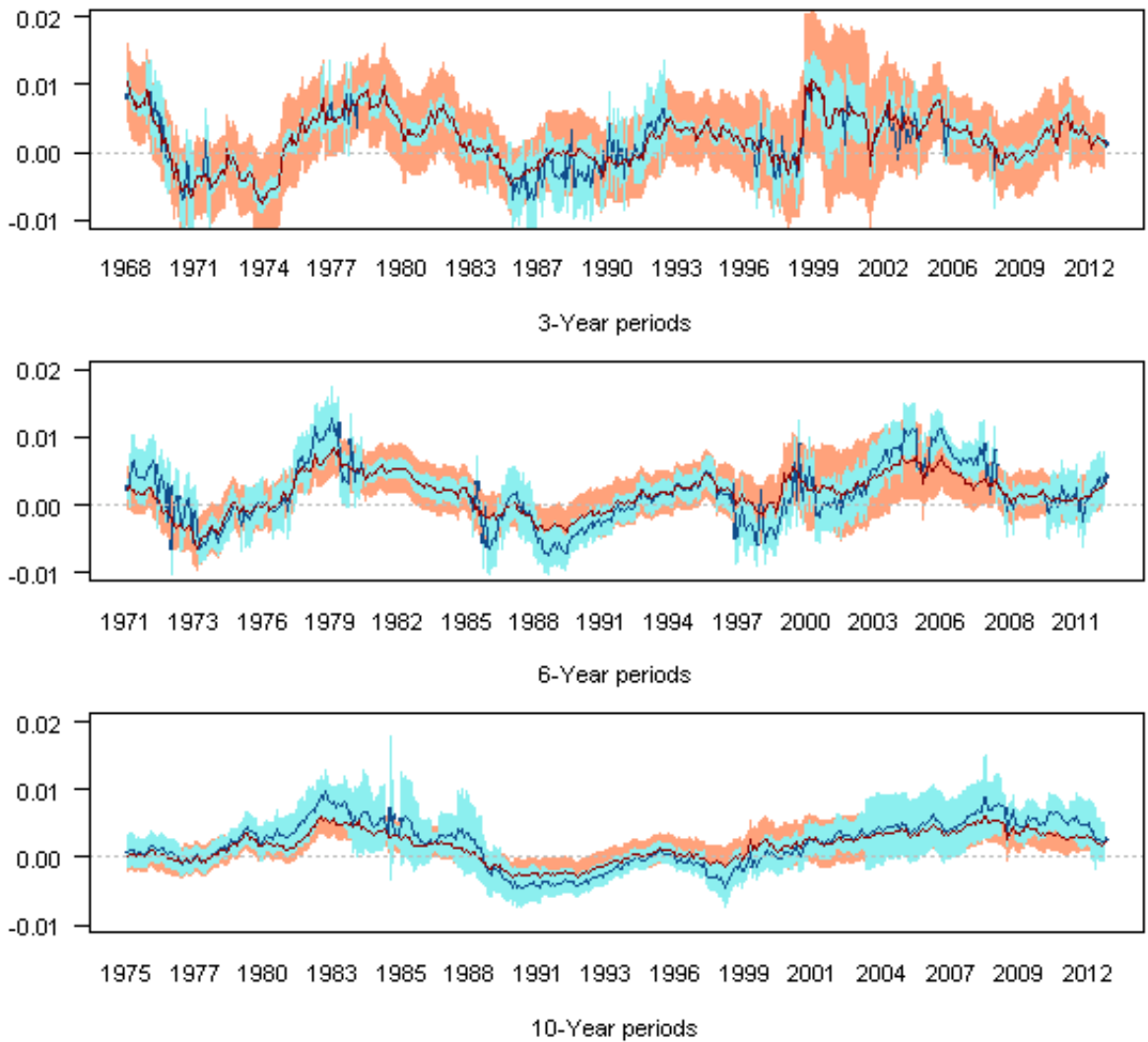


Figure III.c

Fama and French (2014) five-factor model:

estimates and confidence intervals of high-minus-low (*hml*) factor risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the *hml* factor (the return difference between portfolios of stocks with high and low book-to-market) obtained by estimating the Fama and French (2014) five-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{hml}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{hml}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{hml}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

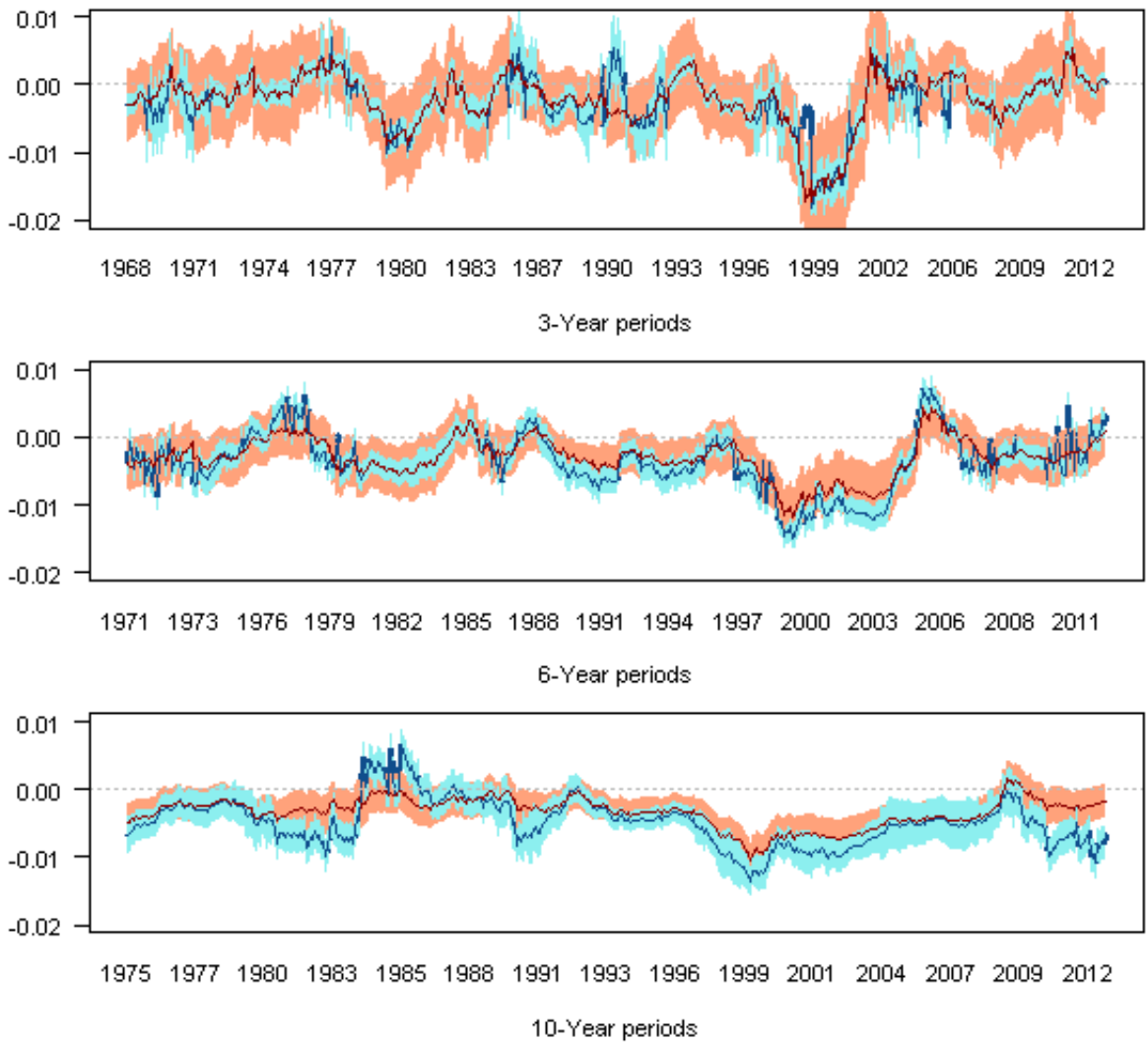


Figure III.d

**Fama and French (2014) five-factor model:
estimates and confidence intervals of profitability (*rmw*) factor risk premium.**

The figure presents the estimates and the associated confidence intervals of the risk premium on the *rmw* factor (the difference between the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios) obtained by estimating the Fama and French (2014) five-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{rmw}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{rmw}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{rmw}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

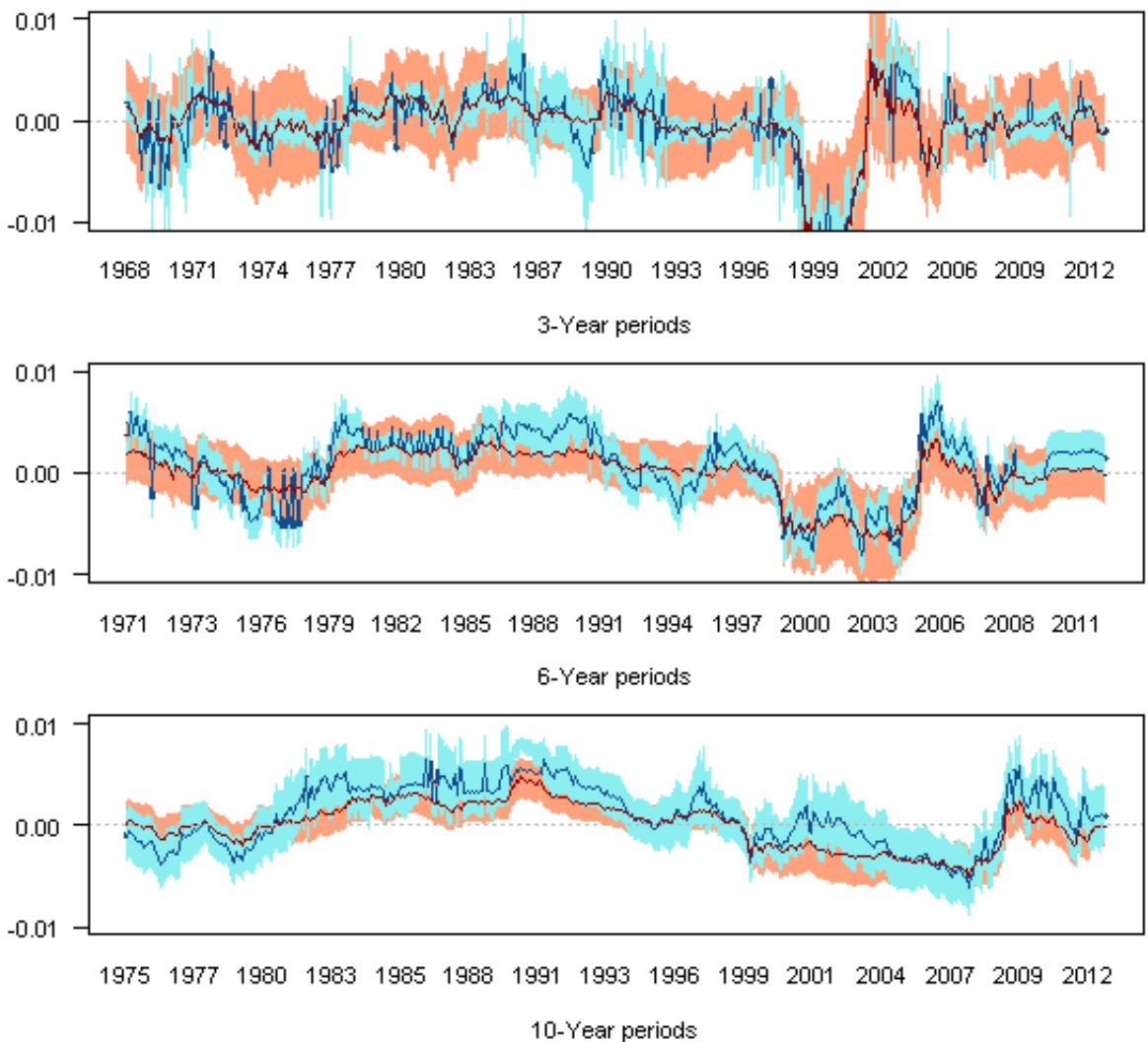


Figure III.e
Fama and French (2014) five-factor model:
estimates and confidence intervals of investment (*cma*) factor risk premium.

The figure presents the estimates and the associated confidence intervals of the risk premium on the *cma* factor (the difference between the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios) obtained by estimating the Fama and French (2014) five-factor model. We report the time series of the Shanken estimator $\hat{\gamma}_{cma}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{cma}^*$ based on the large-N standard errors of Theorem 1. We also report the time series of the OLS estimator $\hat{\gamma}_{cma}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

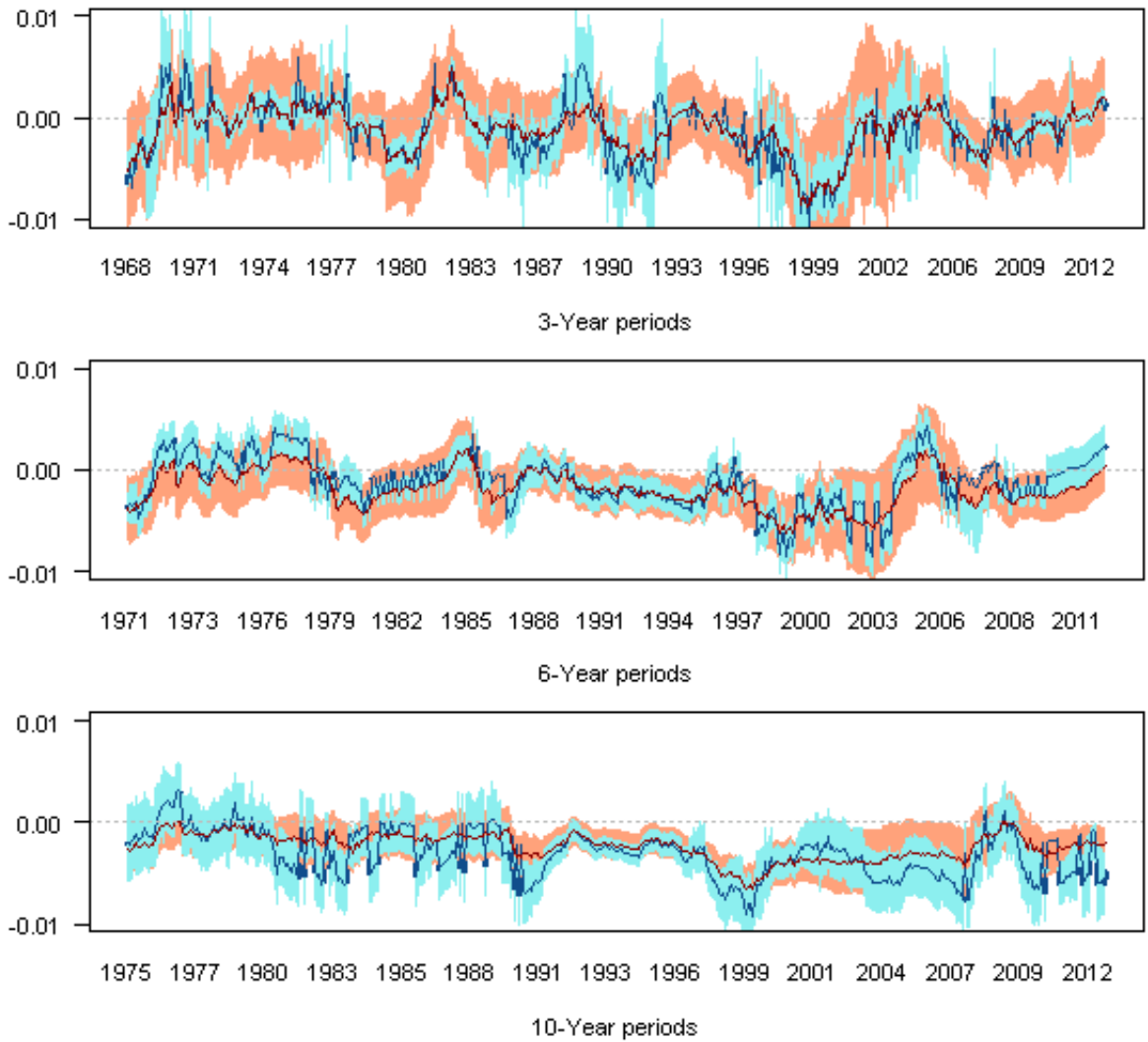


Figure IV
CAPM: specification test.

The figure reports the time series of p-values (black line) corresponding to the asset-pricing test of Theorem 3 corresponding to the null hypothesis that the CAPM holds, obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The dotted red line indicates the 5% level. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.

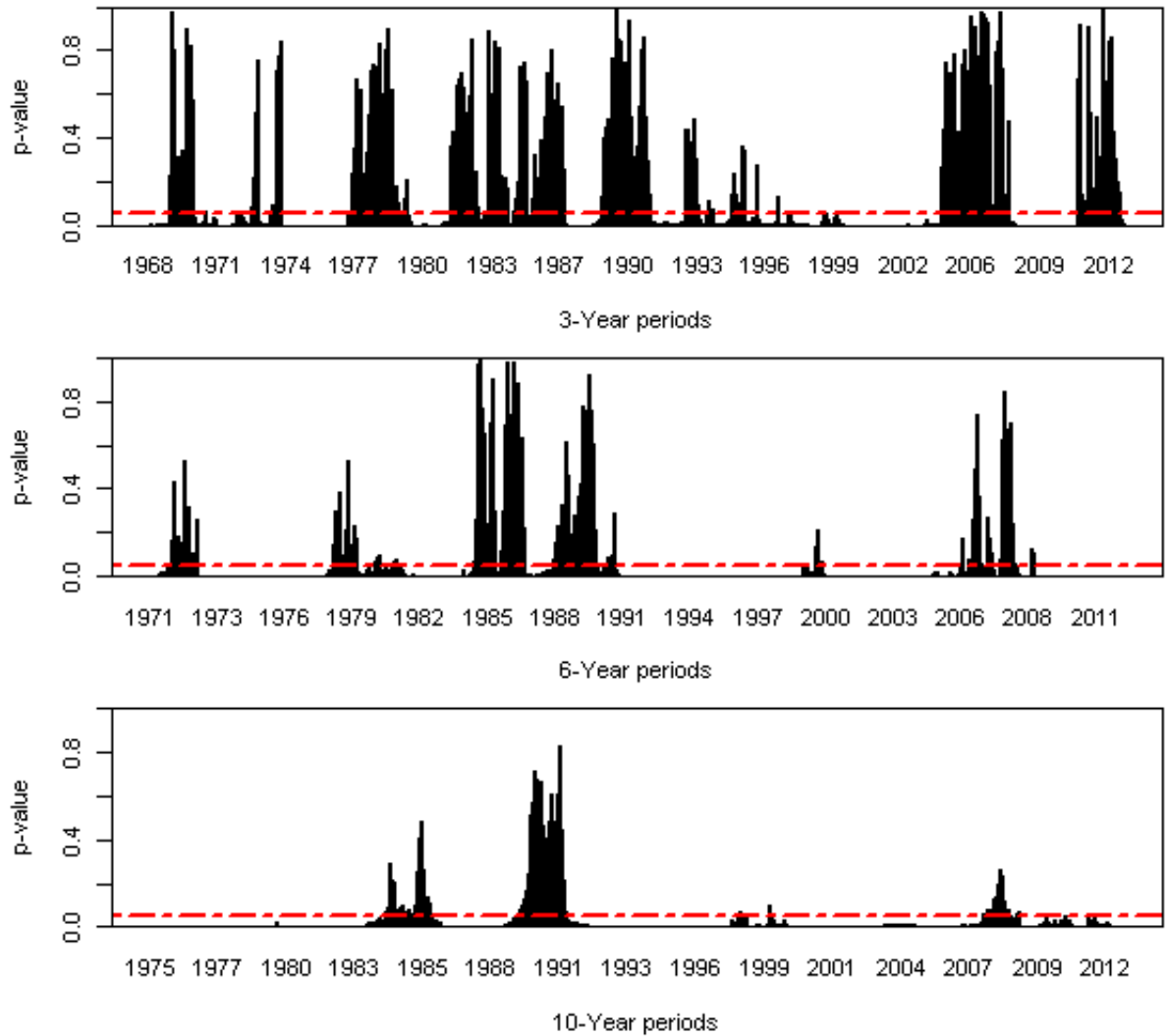


Figure V
CAPM: specification test of Gibbons et al (1989).

This figure reports the time series of p-values (black line) corresponding to the asset-pricing test of Gibbons et al. (1989) corresponding to the null hypothesis that the CAPM holds, obtained with a rolling time window of three- (top panel), six- (central panel) and ten-year (bottom panel). The dotted red line indicates the 5% level. We use individual stock data from the CRSP aggregated into 25 equally weighted portfolios, monthly observations from January 1966 until December 2013.

