TIME-VARYING RISK PREMIUM IN LARGE CROSS-SECTIONAL EQUITY DATASETS

Patrick Gagliardini[a], Elisa Ossola[b] and Olivier Scaillet[c]*

preliminary draft

November 30, 2010

Abstract

We develop an econometric methodology to infer the path of risk premia from large unbalanced panel of individual stock returns. We estimate the time-varying risk premia implied by conditional linear asset pricing models through simple two-pass cross-sectional regressions, and show consistency and asymptotic normality under increasing cross-sectional and time series dimensions. We address consistent estimation of the asymptotic variance, and testing for asset pricing restrictions. Our approach also delivers inference for a time-varying cost of equity. The empirical illustration on over 12,500 US stock returns from January 1960 to December 2009 shows that conditional risk premia and cost of equities are large and volatile in crisis periods. They exhibit large positive and negative strays from standard unconditional estimates and follow the macroeconomic cycles. The asset pricing restrictions are rejected for the usual unconditional four-factor model capturing market, size, value and momentum effects but not for its conditional version using scaled factors.

[a]University of Lugano and Swiss Finance Institute, [b]University of Lugano, [c]University of Genève and Swiss Finance Institute.

*Acknowledgements: We gratefully acknowledge the financial support of the Swiss National Science Foundation (Prodoc project PDFM11-114533).

1 Introduction

Risk premia measure financial compensation asked by investors for bearing risk. Risk is influenced by financial and macroeconomic variables. Conditional linear factor models aim at capturing their time-varying influence in a simple setting (see e.g. Shanken (1990), Cochrane (1996), Ferson and Schadt (1996), Ferson and Harvey (1991, 1999), Lettau and Ludvigson (2001), Petkova and Zhang (2005)). Time variation in risk is known to bias unconditional estimates of alphas and betas, and therefore asset pricing test conclusions (Jagannathan and Wang (1996), Lewellen and Nagel (2006), Boguth, Carlson, Fisher and Simutin (2010)). Ghysels (1998) discusses the pros and cons of modeling time-varying betas.

The workhorse to estimate equity risk premia in a linear multi-factor setting is the two-pass cross-sectional regression method developed by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973). Its large and finite sample properties for unconditional linear factor models have been addressed in a series of papers, see e.g. Shanken (1985, 1992), Jagannathan and Wang (1998), Shanken and Zhou (2007), Kan, Robotti and Shanken (2009) and the review paper of Jagannathan, Skoulakis and Wang (2009). Statistical inference for equity risk premia in conditional linear factor model has not yet been formally addressed in the literature despite its empirical relevance.

In this paper our goal is to study how we can infer the time-varying behaviour of equity risk premia from large stock return databases by using conditional linear factor models. Our approach is inspired by the recent trend in macro-econometrics and forecasting methods trying to extract cross-sectional and time-series information simultaneously from large panels (see e.g. Stock and Watson (2002a,b), Bai (2003, 2009), Bai and Ng (2002, 2006), Forni, Hallin, Lippi and Reichlin (2000, 2004, 2005), Pesaran (2006)). Ludvigson and Ng (2010) show that it is a promising route to follow to study bond risk premia. It is also inspired by the potential loss of information and bias induced by grouping stocks to build portfolios in asset pricing tests (Litzenberger and Ramaswamy (1979), Lo and MacKinlay (1990), Berk (2000), Conrad, Cooper and Kaul (2003), Phalippou (2007)). Avramov and Chordia (2006) have already shown that empirical findings given by conditional factor models about anomalies differ a lot when considering single securities instead of portfolios. Ang, Liu and Schwarz (2008) argue that a lot of efficiency may be lost when only considering portfolios as base assets, instead of individual stocks, to estimate equity risk premia in unconditional models. In our approach the large cross-section of stock returns also helps to get accurate estimation of the equity risk premia even if we get noisy time-series estimates of the factor loadings (the betas). Besides, when running asset-pricing tests, Lewellen, Nagel and Shanken (2010) advocate working with a large number of assets instead of working with a small number of portfolios exhibiting a tight factor structure. The former gives us a higher hurdle to meet in judging model explanation based on cross-sectional R^2.

Our theoretical contributions are twofold. First we provide a new two-pass cross-sectional estimator of the path over time of the risk premia from large unbalanced panels of excess returns. We derive the large sample properties of such an estimator in conditional linear factor models. The factor modeling permits conditional heteroskedasticity and cross-sectional dependence in the error terms (see Petersen (2008) for stressing the importance of residual dependence when computing standard errors in finance panel data). We derive consistency and asymptotic normality of our estimates by letting the time dimension T and the cross-section dimension n grow to infinity simultaneously, and not sequentially. We derive all properties for unbalanced panels to avoid the survivorship bias inherent to studies restricted to balanced subsets of available stock return databases (Brown, Goetzmann, Ross (1995)). The two-pass regression approach is simple and particularly easy to implement in an unbalanced setting. This explains our choice over more efficient, but numerically intractable, one-pass ML/GMM estimators or generalized least-squares estimators. When n is of the order of a couple of thousands assets, numerical optimization on a large parameter set or numerical inversion of a large weighting matrix is too challenging and unstable to benefit in practice from the theoretical efficiency gains. A sensible solution to this problem is to impose some strong ad hoc structural restrictions as for example in estimation of multivariate GARCH models. Second we provide a goodness-of-fit test for the conditional factor model underlying the estimation. The test exploits the asymptotic distribution of the sum of squared residuals of the second-pass cross-sectional regression (see Lewellen, Nagel and Shanken (2009), Kan, Robotti and Shanken (2009) for a related approach in unconditional models and asymptotics with fixed n). As a by-product, our approach permits inference for the cost of equity on individual stocks, in a time-varying setting (Fama and French (1997)). As known from standard textbooks in corporate finance, the cost of equity is such that cost of equity = risk free rate + factor loadings $\times$ factor risk premia. It is part of the cost of capital and is a central piece for evaluating investment projects by company managers.

For our empirical contributions we take the entire Center for Research in Security Prices (CRSP) dataset for over 12,500 stocks with monthly returns from January 1960 to December 2009. We look at factor models popular in the empirical finance literature to explain monthly equity returns. They differ by the choice of the factors. The first model is the CAPM (Sharpe (1964), Lintner (1965)) using market return as the single factor. Then, we consider the three-factor model of Fama and French (1993) based on two additional factors capturing the book-to-market and size effects, and a four-factor extension including a momentum factor (Jegadeesh and Titman (1993), Carhart (1997)). We study both unconditional and conditional factor models (Ferson and Schadt (1996), and Ferson and Harvey (1999)). The estimated path shows that the risk premia are large and volatile in crisis periods, e.g., the oil crisis in 1973-1974, the market crash in October 1987, and the crisis of the recent years. Furthermore, the conditional estimates exhibit large positive and negative strays from standard unconditional estimates and follow the macroeconomic cycles. The asset pricing restrictions are rejected for the usual unconditional four-factor model capturing market, size, value and momentum effects but not for its conditional version using scaled factors.

The outline of the paper is as follows. In Section 2 we present our approach in an unconditional linear factor setting. In Section 3 we provide the results for a conditional linear factor model. Section 4 contains the empirical results. Section 5 contains the simulation results. Finally, Section 6 concludes.

2 Unconditional factor model

In this section we consider an unconditional linear factor model in order to illustrate the main contributions of the article in a simple setting. This covers the CAPM where the single factor is the excess market return.

2.1 Model definition and asset pricing restriction

Let $R_{i,t}$ denote the excess return of asset i at date t, where $i = 1, ..., n$, and $t = 1, ..., T$. We assume that asset excess returns satisfy the unconditional linear factor model:

$$R_{i,t} = a_i + b_i' f_t + \varepsilon_{i,t}, \tag{1}$$

where vector f_t gathers the values of the K observable factors at date t, while the parameters a_i and b_i are the intercept and the factor sensitivities of asset i. Let $x_t = (1, f_t')'$ and $\beta_i = (a_i, b_i')'$, which yields the compact formulation:

$$R_{i,t} = \beta_i' x_t + \varepsilon_{i,t}. \tag{2}$$

Assumption A.1 below allows for a martingale difference sequence for the error terms (White (2001)) including potential conditional heteroskedasticity as well as weak cross-sectional dependence (Bai and Ng (2002)).

Assumption A.1 *There exists a positive constant M such that for all n, T:*
a) $E\left[\varepsilon_{i,t}|\varepsilon_{i,\underline{t-1}}, i = 1, ..., n, x_{\underline{t}}\right] = 0$, *with* $\varepsilon_{i,\underline{t-1}} = \{\varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \cdots\}$ *and* $x_{\underline{t}} = \{x_t, x_{t-1}, \cdots\}$;
b) $E\left[\varepsilon_{i,t}^2|x_{\underline{t}}\right] \leq M, \quad i = 1, ..., n;$ *c)* $\frac{1}{n}\sum_{i,j}\left|E\left[\varepsilon_{i,t}\varepsilon_{j,t}|x_{\underline{t}}\right]\right| \leq M, \quad t = 1, ..., T.$

More general error structures are possible but complicate consistent estimation of the asymptotic variances of the estimators (see Section 2.3).

We follow the literature on random coefficients panel models (e.g. Wooldridge (2002)). We assume that $(a_i, b_i')'$, for $i = 1, ..., n$, are drawn randomly according to some probability distribution.

Assumption A.2 *For all n, T, parameters $\beta_i, i = 1, ..., n$, are i.i.d. draws, independent of $\{f_t, \varepsilon_{i,t} : i = 1, ..., n, t = 1, ..., T\}$, with some distribution G on the compact support $\mathbb{B} \subset \mathbb{R}^{K+1}$.*

Assumption A.2 is compatible with i.i.d. draws $(a_i, b_i', w_i')'$, where $(a_i, b_i')'$ and some asset characteristics w_i, such as industry sector, are dependent. Then the distribution G is the marginal distribution of $(a_i, b_i')'$. Below we make use of the moments $\mu_b = E_G[b_i], \mu_b^{(2)} = E_G[b_i b_i']$ and $\Sigma_b = \mu_b^{(2)} - \mu_b\mu_b'$, where $E_G[\cdot]$ denotes expectation under distribution G.

The asset pricing restriction underlying the factor model (1) can be written either as

$$E[R_{i,t}|\beta_i = \beta] = b'\lambda, \tag{3}$$

where λ is the vector of the risk premia, or

$$a = b'\nu, \tag{4}$$

where $\nu = \lambda - E[f_t]$, for almost all $\beta = (a, b')' \in \mathbb{B}$ (Ross (1976), Chamberlain and Rothschild (1983), Connor (1984)). In the CAPM, we have $K = 1$ and $\nu = 0$. Equation (4) shows explicitly that the asset pricing restriction is a constraint on distribution G, namely its support is a plane through the origin. We use later the randomness of the coefficients to get well-defined probability limits of the second pass estimators for large n and T whether (4) holds or not.

2.2 Asymptotic properties of risk premium estimation

In available databases asset returns are not observed for all firms at all dates. We account for the unbalanced nature of the panel through the indicator $\mathbf{1}_{i,t}$ that is equal to 1 if the return of asset i is observed at date t, and 0 otherwise (Connor and Korajczyk (1987)). To ease exposition and to keep the factor structure linear, we treat the indicator sequences $(\mathbf{1}_{i,t})$ as deterministic. This is equivalent to assume a missing-at-random design (Rubin (1976), Heckman (1979)), that is, independence between unobservability and the factor structure. Another design would require an explicit firm-by-firm modeling of the unobservability mechanism and would yield a nonlinear factor structure.

We consider a two-pass approach (Fama and MacBeth (1973), Black, Jensen and Scholes (1972)) building on Equations (1) and (4). The first pass consists in computing time-series OLS estimators $\hat{\beta}_i = \left(\sum_t \mathbf{1}_{i,t} x_t x_t' \right)^{-1} \sum_t \mathbf{1}_{i,t} x_t R_{i,t}$, for $i = 1, ..., n$. The second pass consists in computing the cross-sectional OLS estimator $\hat{\nu} = \left(\sum_i \hat{b}_i \hat{b}_i' \right)^{-1} \left(\sum_i \hat{b}_i \hat{a}_i \right)$. The estimator of the risk premia is $\hat{\lambda} = \hat{\nu} + \frac{1}{T} \sum_t f_t$.

Starting from the asset pricing restriction (3) another estimator of λ is $\bar{\lambda} = \left(\sum_i \hat{b}_i \hat{b}_i' \right)^{-1} \left(\sum_i \hat{b}_i \bar{R}_i \right)$, where $\bar{R}_i = \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} R_{i,t}$ and $T_i = \sum_t \mathbf{1}_{i,t}$. This estimator is numerically equivalent to $\hat{\lambda}$ in the balanced case, while, in the general unbalanced case, it is equal to $\bar{\lambda} = \hat{\nu} + \left(\sum_i \hat{b}_i \hat{b}_i' \right)^{-1} \sum_i \hat{b}_i \hat{b}_i' \bar{f}_i$, where $\bar{f}_i = \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} f_t$. This second estimator is often studied by the literature (see, e.g., Shanken (1992), Kandel and Stambaugh (1995), Jagannathan and Wang (1998)). Estimating $E[f_t]$ with a simple average of the observed factor instead of a weighted average based on estimated betas simplifies the form of the asymptotic distribution in the unbalanced case (see below and Section 2.3). This explains our preference for $\hat{\lambda}$ over $\bar{\lambda}$ even if both estimators are consistent.

Proposition 1 summarizes consistency of estimators $\hat{\beta}_i, \hat{\nu}$ and $\hat{\lambda}$ under the double asymptotics $n, T \to \infty$. For sequences x_n and y_n, we denote $x_n \asymp y_n$ when x_n / y_n is bounded and bounded away from zero from below as $n \to \infty$.

Proposition 1 *Under Assumptions A.1, C.1-C.2 and* $\frac{1}{T}\sum_t (f_t - E[f_t]) = o_p(1)$, *we get*

$$a)\ \sup_i \left\|\hat{\beta}_i - \beta_i\right\| = o_p(1), \qquad b)\ \|\hat{\nu} - \nu\| = o_p(1), \qquad c)\ \left\|\hat{\lambda} - \lambda\right\| = o_p(1),$$

when $n, T \to \infty$ *such that* $n \asymp T^{\gamma}$ *for* $\gamma > 0$, *and* $T_i \geq CT$ *for all* i *with* $C > 0$.

The conditions in Proposition 1 allow for n large w.r.t. T (short panel asymptotics) when $\gamma > 1$. The condition on T_i in Proposition 1 ensures the uniform consistency of $\hat{\beta}_i$ across assets. Shanken (1992) shows consistency of $\hat{\nu}$ and $\hat{\lambda}$ for a fixed n and $T \to \infty$. This consistency does not imply Proposition 1. Shanken (1992) (see also Litzenberger and Ramaswamy (1979)) further shows that we can estimate ν consistently in the second pass with a modified cross-sectional estimator for a fixed T and $n \to \infty$. Since $\lambda = \nu + E[f_t]$, consistent estimation of the risk premia themselves is impossible for a fixed T.

Proposition 2 below gives the large-sample distributions under the double asymptotics $n, T \to \infty$. Let us define the limiting sums contributing to the asymptotic variances: $S_1 = \lim_{n\to\infty} \frac{1}{n}\sum_i \tau_i^2 S_{ii}$, $S_2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i,j} \tau_i \tau_j S_{ij}$, and $Q_x = \lim_{T\to\infty} \frac{1}{T}\sum_t E[x_t x_t']$, where $S_{ij} = \operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_t \sigma_{ij,t} x_t x_t'$, $\sigma_{ij,t} = E[\varepsilon_{i,t}\varepsilon_{j,t}|\underline{x_t}]$, and $\tau_i = \sqrt{T/T_i}$, $i = 1, ..., n$. The following assumption describes the CLTs underlying the proof of the distributional properties. These CLTs hold under weak serial and cross-sectional dependencies such as temporal mixing and block dependence.

Assumption A.3 *As* $n, T \to \infty$ *such that* $n \asymp T^{\gamma}$ *for* $\gamma \in \Gamma_1 \subset \mathbb{R}^+$ *and* $T_i \geq CT$ *for all* i *with* $C > 0$, *a) for all* i, $Y_{i,T} = \frac{1}{\sqrt{T_i}}\sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \Rightarrow N(0, S_{ii})$; *b)* $\frac{1}{\sqrt{n}}\sum_i \tau_i (Y_{i,T} \otimes b_i) \Rightarrow N(0, S_b)$, *where* $S_b = S_1 \otimes \Sigma_b + S_2 \otimes \mu_b \mu_b'$; *c)* $\frac{1}{\sqrt{T}}\sum_t (f_t - E[f_t]) \Rightarrow N(0, \Sigma_f)$, *where* $\Sigma_f = \lim_{T\to\infty} \frac{1}{T}\sum_{t,s} Cov(f_t, f_s)$.

Proposition 2 *Under Assumptions A.1-A.3, and C.1-C.8, we have: a)* $\sqrt{T_i}\left(\hat{\beta}_i - \beta_i\right) \Rightarrow N\left(0, Q_x^{-1} S_{ii} Q_x^{-1}\right)$, *for any* i, *conditionally on* β_i; *b)* $\sqrt{nT}\left(\hat{\nu} - \nu - \frac{1}{T}B_\nu\right) \Rightarrow N(0, \Sigma_\nu)$, *where* $B_\nu = \left(\mu_b^{(2)}\right)^{-1} E_2' Q_x^{-1} S_1 Q_x^{-1} c_\nu$, *and* $\Sigma_\nu = \left(\mu_b^{(2)}\right)^{-1}\left[c_\nu' Q_x^{-1} S_1 Q_x^{-1} c_\nu \Sigma_b + c_\nu' Q_x^{-1} S_2 Q_x^{-1} c_\nu \mu_b \mu_b'\right]\left(\mu_b^{(2)}\right)^{-1}$, *with* $E_2 = (0 : Id_K)'$ *and* $c_\nu = (1, -\nu')'$; *c)* $\sqrt{T}\left(\hat{\lambda} - \lambda\right) \Rightarrow N(0, \Sigma_f)$, *when* $n, T \to \infty$ *such that* $n \asymp T^{\gamma}$ *for* $\gamma \in \Gamma_1 \subset \mathbb{R}^+$ *as in Assumption A.3, and* $T_i \geq CT$ *for all* i *and* $C > 0$.

Proposition 2 shows that the estimator $\hat{\nu}$ has a faster convergence rate than the estimator $\hat{\lambda}$. The estimator $\hat{\nu}$ involves $\sum_i \hat{b}_i \hat{a}_i$. Both $\hat{a}_i$ and $\hat{b}_i$ contain an estimation error. The cross-sectional sum of the products of these estimation errors explains the bias term B_ν/T which centers the asymptotic distribution. Since $c_\nu' Q_x^{-1} c_\nu = 1 + \lambda' \Sigma_f^{-1} \lambda$, the asymptotic covariance of $\hat{\nu}$ in the homoskedastic case $\sigma_{ij,t} = \sigma_{ij}$, is $\Sigma_\nu = \left(1 + \lambda'\Sigma_f^{-1}\lambda\right) \left(\mu_b^{(2)}\right)^{-1} \left[\left(\lim_{n\to\infty} \frac{1}{n}\sum_i \tau_i^2 \sigma_{ij}\right)\Sigma_b + \left(\lim_{n\to\infty} \frac{1}{n}\sum_{i,j} \tau_i\tau_j\sigma_{ij}\right)\mu_b\mu_b'\right]\left(\mu_b^{(2)}\right)^{-1}$. The asymptotic distribution of $\hat{\lambda}$ is only driven by the variability of the factor since the convergence rate of the sample average $\frac{1}{T}\sum_t f_t$ dominates the convergence rate of $\hat{\nu}$. This result is an oracle property for $\hat{\lambda}$, namely that its asymptotic distribution is the same irrespective of the knowledge of ν. This property is in sharp difference with the single asymptotics with a fixed n and $T \to \infty$. In the balanced case and with homoskedastic errors, Theorem 1 of Shanken (1992) shows that the rate of convergence of $\hat{\lambda}$ is $\sqrt{T}$ and that its asymptotic variance is $\Sigma_{\lambda,n} = \Sigma_f + \left(1 + \lambda'\Sigma_f^{-1}\lambda\right)\left(\mu_{b,n}^{(2)}\right)^{-1}\left(\frac{1}{n^2}\sum_{i,j} b_i b_j' \sigma_{ij}\right)\left(\mu_{b,n}^{(2)}\right)^{-1}$, with $\mu_{b,n}^{(2)} = \frac{1}{n}\sum_i b_i b_i'$, for fixed n and $T \to \infty$. A direct extension to our heteroskedastic and unbalanced setting yields $\Sigma_{\lambda,n} = \Sigma_f + \Sigma_{\nu,n}$, where $\Sigma_{\nu,n} = \left(\mu_{b,n}^{(2)}\right)^{-1}\left[\frac{1}{n^2}\sum_{i,j}\tau_i\tau_j b_i b_j' \left(c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)\right]\left(\mu_{b,n}^{(2)}\right)^{-1}$. Letting $n \to \infty$ gives Σ_f under weak cross-sectional dependence. Thus, exploiting the full cross-section of assets improves efficiency asymptotically, and the positive definite matrix $\Sigma_{\lambda,n} - \Sigma_f$ corresponds to the efficiency gain.

To conclude this section let us recall the definition of the cost of equity CE_i for firm i: that is $CE_{i,t} = r_{f,t} + b_i'\lambda$, where $r_{f,t}$ denotes the risk-free rate. Then we can deduce from Proposition 2 and the asymptotic independence of estimators $\hat{\lambda}$ and $\hat{b}_i$ that

$$\sqrt{T}\left(\widehat{CE}_i - CE_i\right) \Rightarrow N\left(0, \tau_i^2 \lambda' E_2' Q_x^{-1} S_{ii} Q_x^{-1} E_2 \lambda + b_i' \Sigma_f b_i\right),$$

where $\widehat{CE}_{i,t} = r_{f,t} + \hat{b}_i'\hat{\lambda}$. This extends the standard error results for CE estimates of Fama and French (1997) to large unbalanced panels.

2.3 Confidence intervals

We can use Proposition 2 to build confidence intervals by means of consistent estimation of the asymptotic variances and B_ν. We can check with these intervals whether the risk of a given factor $f_{k,t}$ is not remunerated, i.e., $\lambda_k = 0$, or the restriction $\nu_k = 0$ holds when the factor is a portfolio excess return. We estimate Q_x with $\hat{Q}_x = \frac{1}{T}\sum_t x_t x_t'$ and S_{ii} with $\hat{S}_{ii} = \frac{1}{T_i}\sum_t \mathbf{1}_{i,t}\hat{\varepsilon}_{i,t}^2 x_t x_t'$, where $\hat{\varepsilon}_{i,t} = R_{i,t} - \hat{\beta}_i' x_t$. We estimate Σ_f by a standard HAC estimator $\hat{\Sigma}_f$ such as in Newey and West (1994) or Andrews and Monahan (1992). Hence, the construction of confidence intervals with valid asymptotic coverage for components of $\hat{\beta}_i$ and $\hat{\lambda}$ is straightforward. On the contrary, getting a HAC estimator for $\bar{\Sigma}_f$ appearing in the asymptotic distribution of $\bar{\lambda}$ is not obvious in the unbalanced case.

The construction of confidence intervals for the components of $\hat{\nu}$ is more difficult. Indeed, S_2 involves a limiting double sum over S_{ij} scaled by n and not n^2. A naive approach consists in replacing S_{ij} by any consistent estimator such as $\hat{S}_{ij} = \frac{1}{T_{ij}}\sum_t \mathbf{1}_{ij,t}\hat{\varepsilon}_{i,t}\hat{\varepsilon}_{j,t} x_t x_t'$, where $T_{ij} = \sum_t \mathbf{1}_{ij,t}$ and $\mathbf{1}_{ij,t} = \mathbf{1}_{i,t}\mathbf{1}_{j,t}$, but this does not work here. To handle this, we rely on recent proposals in the statistical literature on consistent estimation of large-dimensional sparse covariance matrices by thresholding (Bickel and Levina (2008), El Karoui (2008)). The idea is to assume sparse contributions of the S_{ij}'s to the double sum. Then we only have to account for sufficiently large contributions in the estimation, i.e., contributions larger than a threshold vanishing asymptotically. Thresholding permits an estimation invariant to asset permutations; this choice of estimator is motivated by the absence of any natural cross-sectional ordering among the matrices S_{ij}.

Assumption A.4 describes the sparsity structure.

Assumption A.4 *There exists constants $M > 0$, and $q, \delta \in [0, 1)$, such that $\|S_{ii}\| \leq M$, for $i = 1, ..., n$ and $\max_i \sum_j \|S_{ij}\|^q = O\left(n^\delta\right)$.*

In Assumption A.4 the individual contribution $\|S_{ii}\|$ of each asset is bounded, and most cross-asset contributions $\|S_{ij}\|$ can be neglected. Assumption A.4 does not impose sparsity of the covariance matrix of the returns themselves. It is satisfied under weak cross-sectional dependence between the error terms, for instance, under a block dependence structure. As sparsity increases, we can choose coefficients q and δ closer to zero in Assumption A.4.

As in Bickel and Levina (2008), let us introduce the thresholded estimator $\tilde{S}_{ij} = \hat{S}_{ij}\mathbf{1}_{\{\|\hat{S}_{ij}\|\geq\kappa\}}$ of S_{ij}, which we refer to as $\hat{S}_{ij}$ thresholded at $\kappa = \kappa_{n,T}$. We set $\hat{S}_1 = \frac{1}{n}\sum_i \tau_i^2 \hat{S}_{ii}$, $\tilde{S}_2 = \frac{1}{n}\sum_{i,j}\tau_i\tau_j\tilde{S}_{ij}$, $\hat{\mu}_b = \frac{1}{n}\sum_i \hat{b}_i$, $\hat{\mu}_b^{(2)} = \frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'$ and $\hat{\Sigma}_b = \hat{\mu}_b^{(2)} - \hat{\mu}_b\hat{\mu}_b'$. We also take $\hat{c}_\nu = (1, -\hat{\nu}')'$.

We can derive an asymptotically valid confidence interval for the components of $\hat{\nu}$ from the next proposition giving a feasible asymptotic normality result.

Proposition 3 *Under Assumptions A.1-A.4, and C.1-C.8, we have* $\tilde{\Sigma}_\nu^{-1/2}\sqrt{nT}\left(\hat{\nu} - \frac{1}{T}\hat{B}_\nu - \nu\right) \Rightarrow N\left(0, Id_K\right)$ *where* $\hat{B}_\nu = \left(\hat{\mu}_b^{(2)}\right)^{-1}\left(\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}\hat{S}_{ii}\hat{Q}_{x,i}^{-1}\hat{c}_\nu\right)$ *and* $\tilde{\Sigma}_\nu = \left(\hat{\mu}_b^{(2)}\right)^{-1}\left[\left(\hat{c}_\nu'\hat{Q}_x^{-1}\hat{S}_1\hat{Q}_x^{-1}\hat{c}_\nu\right)\hat{\Sigma}_b + \left(\hat{c}_\nu'\hat{Q}_x^{-1}\tilde{S}_2\hat{Q}_x^{-1}\hat{c}_\nu\right)\hat{\mu}_b\hat{\mu}_b'\right]\left(\hat{\mu}_b^{(2)}\right)^{-1}$*, when* $n, T \to \infty$ *such that* $n \asymp T^\gamma$ *for* $\gamma \in \Gamma_1 \cap \left(0, \min\left\{2, \eta\frac{1-q}{2\delta}\right\}\right)$, *and* $\kappa = M\sqrt{\frac{\log n}{T^\eta}}$ *for a constant* M *and* $\eta \in (0,1]$ *as in Assumption C.7.*

Constant $\eta \in (0,1]$ is related with the time series dependence of processes $(\varepsilon_{i,t})$ and (x_t); we have $\eta = 1$ when $(\varepsilon_{i,t})$ and (x_t) are i.i.d..

2.4 Global specification tests

The null hypothesis underlying the asset pricing restriction (4) is

$$\mathcal{H}_0: \text{ there exists } \nu \in \mathbb{R}^K \text{ such that } a = b'\nu, \quad \text{for almost all } \left(a, b'\right)' \in \mathbb{B}.$$

Under $\mathcal{H}_0$, we have $E_G\left[(a - b'\nu)^2\right] = 0$. Since ν is estimated via the cross-sectional regression of the estimates $\hat{a}_i$ on the estimates $\hat{b}_i$, we suggest a test based on the sum of squared residuals SSR of the cross-sectional regression. The SSR is $\hat{Q}_e = \frac{1}{n}\sum_i \hat{e}_i^2$, with $\hat{e}_i = \hat{c}_\nu'\hat{\beta}_i$, which is an empirical counterpart of $E_G\left[(a - b'\nu)^2\right]$.

Let us define $S_{ii,T} = \frac{1}{T_i}\sum_t \mathbf{1}_{i,t}\sigma_{ii,t}x_tx_t'$, and introduce the commutation matrix $W_{m,n}$ of order $mn \times mn$ such that $W_{m,n}vec\left[A\right] = vec\left[A'\right]$ for any $A \in \mathbb{R}^{m\times n}$, where the vector operator $vec\left[\cdot\right]$ stacks the elements of an $m \times n$ matrix as a $mn \times 1$ vector. If $m = n$, we write W_n instead $W_{n,n}$. For two $(K+1) \times (K+1)$ matrices A and B, $W_{(K+1)}\left(A \otimes B\right) = \left(B \otimes A\right)W_{(K+1)}$ also holds (see Chapter 3 of Magnus and Neudecker (2007) for other properties).

Assumption A.5 *For $n, T \to \infty$ such that $n \asymp T^\gamma$ for $\gamma \in \Gamma_2 \subset \Gamma_1$, and $T_i \geq CT$ with $C > 0$, we have* $\frac{1}{\sqrt{n}} \sum_i \tau_i^2 \left(Y_{i,T} \otimes Y_{i,T} - vec\left[S_{ii,T} \right] \right) \Rightarrow N(0, \Omega)$, *where*
$\Omega = \lim_{n \to \infty} \frac{1}{n} \sum_{i,j} \frac{\tau_i^4 \tau_j^4}{\tau_{ij}^4} \left[S_{ij} \otimes S_{ij} + \left(S_{ij} \otimes S_{ij} \right) W_{(K+1)} \right]$, *and* $\tau_{ij} = \sqrt{\frac{T}{T_{ij}}}$.

Assumption A.5 is a high-level CLT condition. This assumption can be proved under primitive conditions on the time series and cross-sectional dependence. For instance, we prove in the Supplementary Materials that Assumption A.5 holds under a cross-sectional block dependence structure for the errors. Intuitively, the expression of the variance-covariance matrix Ω is related to the fact that, for random $(K+1) \times 1$ vectors Y_1 and Y_2 which are jointly normal with covariance matrix S, we have $Cov\left(Y_1 \otimes Y_1, Y_2 \otimes Y_2 \right) = S \otimes S + (S \otimes S) W_{(K+1)}$.

Let us now introduce the following statistic $\hat{\xi}_{nT} = T\sqrt{n} \left(\hat{Q}_e - \frac{1}{T} \hat{B}_\xi \right)$, where $\hat{B}_\xi = \frac{1}{n} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_{x,i} \hat{S}_{ii} \hat{Q}_{x,i} \hat{c}_\nu$ is the estimated centering term. Under the null hypothesis $\mathcal{H}_0$, we prove $\hat{\xi}_{nT} = \left(vec\left[\hat{Q}_x^{-1} \hat{c}_\nu \hat{c}_\nu' \hat{Q}_x^{-1} \right] \right)' \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \left(Y_{iT} \otimes Y_{i,T} - vec\left[S_{ii,T} \right] \right) + o_p(1)$, which implies $\hat{\xi}_{nT} \Rightarrow N(0, \Sigma_\xi)$, where $\Sigma_\xi = 2 \lim_{n \to \infty} \frac{1}{n} \sum_{i,j} \frac{\tau_i^4 \tau_j^4}{\tau_{ij}^4} \left(c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu \right)^2$ as $n, T \to \infty$ (see Appendix A.4.3). Then a feasible testing procedure exploits the consistent estimator $\tilde{\Sigma}_\xi = 2 \frac{1}{n} \sum_{i,j} \frac{\tau_i^4 \tau_j^4}{\tau_{ij}^4} \left(\hat{c}_\nu' \hat{Q}_x^{-1} \tilde{S}_{ij} \hat{Q}_x^{-1} \hat{c}_\nu \right)^2$ of the asymptotic variance Σ_ξ.

Proposition 4 *Under $\mathcal{H}_0$, and Assumptions A.1-A.5 and C.1-C.9, we have* $\tilde{\Sigma}_\xi^{-1/2} \hat{\xi}_{nT} \Rightarrow N(0,1)$, *as $n, T \to \infty$ such that $n \asymp T^\gamma$ for $\gamma \in \Gamma_2 \cap \left(0, \min\left\{ 2, \eta \frac{1-q}{2\delta} \right\} \right)$*

In the homoskedastic case, the asymptotic variance of $\hat{\xi}_{nT}$ is $\Sigma_\xi = 2 \left(1 + \lambda' \Sigma_f^{-1} \lambda \right)^2 \lim_{n \to \infty} \frac{1}{n} \sum_i \frac{\tau_i^4 \tau_j^4}{\tau_{ij}^4} \sigma_{ij}^2$. In the CAPM we have $K = 1$ and $\nu = 0$ which implies that $\sqrt{\frac{\lambda^2}{\Sigma_f}}$ is equal to the slope of the Capital Market Line $\sqrt{\frac{E[f_t]^2}{\Sigma_f}}$, i.e., the Sharpe Ratio of the market portfolio.

We can use $\hat{V}_a = \frac{1}{n} \sum_i \left(\hat{a}_i - \bar{\hat{a}} \right)^2$, with $\bar{\hat{a}} = \frac{1}{n} \sum_i \hat{a}_i$, to get $\tilde{\Sigma}_\xi^{-1/2} \hat{\xi}_{nT} = \hat{V}_a \tilde{\Sigma}_\xi^{-1/2} (1 - \hat{\rho}^2 - \frac{1}{T} \hat{V}_a^{-1} \hat{B}_\xi)$, with $\hat{\rho}^2 = 1 - \frac{\hat{Q}_e}{\hat{V}_a}$. This yields an interpretation of the test statistic in terms of an estimate $\hat{\rho}^2$ of the cross-sectional R^2. However, the population R^2 is not well-defined in the CAPM under increasing cross-sectional dimension since $V_a = \lim_{n\to\infty} \frac{1}{n}\sum_i (a_i - \bar{a})^2 = 0$, and this explains why we prefer the writing in terms of SSR. For fixed n we can rely on the test statistic $T\hat{V}_a\left(1-\hat{\rho}^2\right) = T\hat{Q}_e$, which is asymptotically distributed as $\frac{1}{n}\sum_j w_j \chi_j^2$ for $j = 1, \ldots, (n-K)$, where the χ_j^2 are i.i.d. chi-square variables with 1 degree of freedom, and the weights w_j are the non-zero eigenvalues of matrix $\Omega^{1/2}(I_n - \bar{B}(\bar{B}'\bar{B})^{-1}\bar{B}')\Omega^{1/2}$ with elements $\Omega_{ij} = \tau_i\tau_j c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu$ for Ω and rows $\bar{B}_i = b_i'$ (see Kan et al. (2009) for similar asymptotic results on $T(1-\hat{\rho}^2)$ with fixed n and a generalized least squares estimator in the second pass). By letting n grow, the sum of chi-square variables converges to a Gaussian variable after recentering and rescaling, which yields the result of Proposition 4.

The alternative hypothesis is

$$\mathcal{H}_1 : \quad \inf_{\nu\in\mathbb{R}^K} E_G\left[\left(a - b'\nu\right)^2\right] > 0.$$

Let us define the pseudo-true value $\nu_\infty = \arg\inf_{\nu\in\mathbb{R}^K} E_G\left[\left(a - b'\nu\right)^2\right] = \left(E_G\left[b_i b_i'\right]\right)^{-1} E_G\left[b_i a_i\right]$ (White (1982), Gourieroux, Monfort and Trognon (1984)) and population errors $e_i = a_i - b_i'\nu_\infty = c_\infty'\beta_i$, $i = 1, ..., n$, for all n. In the next proposition, we prove consistency of the test, namely that the statistic $\hat{\xi}_{nT}$ diverges to $+\infty$ under the alternative hypothesis $\mathcal{H}_1$ for large n and T. We also give the asymptotic distribution of estimators $\hat{\nu}$ and $\hat{\lambda}$ under $\mathcal{H}_1$.

Proposition 5 *Under $\mathcal{H}_1$ and Assumptions A.1-A.5 and C.1-C.9, we have $\hat{\xi}_{nT} \overset{p}{\to} +\infty$, and $\sqrt{n}\left(\hat{\nu} - \nu_\infty\right) \Rightarrow N\left(0, \Sigma_{\nu_\infty}\right)$, where $\Sigma_{\nu_\infty} = \left(\mu_b^{(2)}\right)^{-1} E_G\left[b_i b_i' e_i^2\right]\left(\mu_b^{(2)}\right)^{-1}$ and $\sqrt{T}\left(\hat{\lambda} - \lambda_\infty\right) \Rightarrow N\left(0, \Sigma_f\right)$, and $\lambda_\infty = \nu_\infty - E\left[f_t\right]$, as $n, T \to \infty$ such that $n \asymp T^\gamma$ for $\gamma \in \Gamma_2 \cap \left(0, \min\left\{2, \eta\frac{1-q}{2\delta}\right\}\right)$.*

Under the alternative hypothesis $\mathcal{H}_1$, the rate of convergence of $\hat{\nu}$ is slower than under $\mathcal{H}_0$, while the rate of convergence of $\hat{\lambda}$ remains the same. The asymptotic distribution of $\hat{\nu}$ is the same as the one got from a cross-sectional regression of a_i on b_i; pre-estimation of a_i and b_i has no impact since the induced error-in-variable vanishes asymptotically.

Along the same lines we can derive a test for the null hypothesis when the factors come from tradable assets, i.e., are portfolio excess returns:

$$\mathcal{H}_0 : a = 0 \text{ for almost all } \beta = \left(a, b'\right)' \in \mathbb{B},$$

against the alternative hypothesis

$$\mathcal{H}_1 : E_G \left[a^2\right] > 0.$$

We only have to substitute $\hat{Q}_a = \frac{1}{n} \sum_i \hat{a}_i^2$ for $\hat{Q}_e$, and $E_1 = (1, 0')'$ for $\hat{c}_\nu$ in Propositions 4 and 5. A notable difference with the classical approach of Gibbons, Ross and Shanken (1989) for a balanced panel with fixed n and $T \to \infty$ is that our testing procedure does not require inverting an estimate of the asymptotic covariance matrix of $\sqrt{T} \left(\hat{a}_1 - a_1, ..., \hat{a}_n - a_n\right)'$.

3 Conditional factor model

In this section we extend the setting of Section 2 to conditional models with scaled factors in order to model possibly time-varying risk premia. We do not follow rolling short-window regression approaches to account for time-variation (Fama and French (1997), Lewellen and Nagel (2006)) since we favor a structural econometric framework to conduct formal inference in large cross-sectional equity datasets. A five-year window of monthly data yields a very short time-series panel for which asymptotics with fixed (small) T and large n are better suited, but keeping T fixed impedes consistent estimation of the risk premia as already mentioned in the previous section.

3.1 Model definition and asset pricing restriction

We assume that asset excess returns satisfy the conditional linear factor model:

$$R_{i,t} = a_{i,t} + b_{i,t}' f_t + \varepsilon_{i,t}, \tag{5}$$

where the time-varying intercept and factor sensitivities depend on lagged instrumental variables $Z_{t-1} \in \mathbb{R}^d$, that is, $a_{i,t} = a_i \left(Z_{t-1}\right)$ and $b_{i,t} = b_i \left(Z_{t-1}\right)$ (Ferson and Harvey (1991)). The instruments $Z_t = (1, f_t', Z_t^{*\prime})'$ include the constant, the observable factors f_t, and additional observable variables $Z_t^* \in \mathbb{R}^q$, with $d = 1 + K + q$. The latter may include powers to account for possible nonlinearities.

The next assumption is similar to Assumption A.1.

Assumption B.1 *There exists a positive constant M such that for all n, T:*
a) $E\left[\varepsilon_{i,t}|\varepsilon_{i,\underline{t-1}}, i=1,...,n, Z_{\underline{t}}\right]=0$, *with* $Z_{\underline{t}}=\{Z_{t-1}, Z_{t-2},\cdots\}$;
b) $E\left[\varepsilon_{i,t}^2|Z_{\underline{t}}\right]\leq M, \quad i=1,...,n;$ *c)* $\frac{1}{n}\sum_{i,j}\left|E\left[\varepsilon_{i,t}\varepsilon_{j,t}|Z_{\underline{t}}\right]\right|\leq M, \quad t=1,...,T.$

To get a scaled factor model we specify that the vector of factor sensitivities $b_{i,t}$ is a linear function of lagged instruments Z_{t-1} (Ferson and Schadt (1996)): $b_{i,t}=B_iZ_{t-1}$, where $B_i\in\mathbb{R}^{K\times d}$, for any i,t. We also specify that the vector of risk premia is a linear function of lagged instruments Z_{t-1} (Cochrane (1996), Jagannathan and Wang (1996)): $\lambda_t=\Lambda Z_{t-1}$, where $\Lambda\in\mathbb{R}^{K\times d}$, for any t. Furthermore, we assume that the conditional expectation of Z_t given the past $Z_{\underline{t-1}}$ depends on Z_{t-1} only and is linear, as, for instance, in a Vector Autoregressive (VAR) model of order 1. Then $E\left[f_t|Z_{t-1}\right]=FZ_{t-1}$, where $F\in\mathbb{R}^{K\times d}$, for any t. Under these functional specifications the asset pricing restriction linking the expected excess returns and the factor sensitivities

$$E\left[R_{i,t}|R_{i,\underline{t-1}}, i=1,...,n, Z_{\underline{t-1}}\right]=b_{i,t}'\lambda_t, \tag{6}$$

implies that the intercept $a_{i,t}$ is a quadratic function of lagged instruments Z_{t-1}, that is, $a_{i,t}=Z_{t-1}'B_i'\left(\Lambda-F\right)Z_{t-1}$. This shows that assuming a priori linearity of $a_{i,t}$ in the lagged instruments Z_{t-1} is in general not compatible with linearity of $b_{i,t}$ and $E\left[f_t|Z_{t-1}\right]$.

The conditional factor model (5) then becomes

$$R_{i,t}=Z_{t-1}'B_i'\left(\Lambda-F\right)Z_{t-1}+Z_{t-1}'B_i'f_t+\varepsilon_{i,t}, \tag{7}$$

which is nonlinear in the parameters Λ, F, and B_i. In order to implement the two-pass methodology in a conditional context it is useful to rewrite model (7) as a model that is linear in transformed parameters and new regressors. The regressors include the scaled factors $x_{2,t}=f_t\otimes Z_{t-1}\in\mathbb{R}^{Kd}$ and the predetermined variables $x_{1,t}=vech\left[X_t\right]\in\mathbb{R}^p$, where $p=d(d+1)/2$ and symmetric matrix $X_t=\left[X_{t,i,j}\right]\in\mathbb{R}^{d\times d}$ is such that $X_{t,i,j}=Z_{i,t-1}^2$, if $i=j$, and $X_{t,i,j}=2Z_{i,t-1}Z_{j,t-1}$, otherwise. The vector-half operator $vech\left[\cdot\right]$ stacks the lower elements of a $m\times m$ matrix as a $m\left(m+1\right)/2\times 1$ vector (see Chapter 2 in Magnus and Neudecker (2007) for properties of this matrix tool). To parallel the analysis of the unconditional case, we can express model (7) as in (2) through appropriate redefinitions of x_t and β_i (see Appendix):

$$R_{i,t}=\beta_i'x_t+\varepsilon_{i,t}, \tag{8}$$

where $x_t = \left(x'_{1,t}, x'_{2,t}\right)'$ and $\beta_i = \left(\beta'_{1,i}, \beta'_{2,i}\right)'$ is such that

$$\beta_{1,i} = \frac{1}{2} D_d^+ \left[(\Lambda - F)' \otimes I_d + I_d \otimes (\Lambda - F)' W_{K,d} \right] \beta_{2,i}, \quad \beta_{2,i} = vec\left[B'_i\right]. \tag{9}$$

The matrix D_d^+ is the $p \times d^2$ Moore-Penrose inverse of the duplication matrix D_d, such that $vech\,[A] = D_d^+ vec\,[A]$ for any $A \in \mathbb{R}^{d\times d}$ (see Chapter 3 in Magnus and Neudecker (2007)). When $Z_t = 1$, we have $d = p = 1$, and model (8) reduces to model (2).

For inference purposes we introduce the next random coefficient assumption which is similar to Assumption A.2.

Assumption B.2 *For all* n, T, *vectors* β_i, *for* $i = 1, ..., n$, *are i.i.d. draws, independent of* $\{Z_t, \varepsilon_{i,t} : i = 1, ..., n, t = 1, ..., T\}$, *with some distribution* G *on the compact support* $\mathbb{B} \subset \mathbb{R}^{p+Kd}$.

In (9) the $p \times 1$ vector $\beta_{1,i}$ is a linear transformation of the $dK \times 1$ vector $\beta_{2,i}$. This clarifies that the asset pricing restriction (6) implies a constraint on the distribution G via the support $\mathbb{B}$. The coefficients of this linear transformation depend on matrix $\Lambda - F$. For the purpose of estimating the loading coefficients of the risk premium in matrix Λ, the parameter restrictions can be written as (see Appendix):

$$\beta_{1,i} = \beta_{3,i}\nu, \tag{10}$$

where $\nu = vec\left[\Lambda' - F'\right]$ and $\beta_{3,i} = D_d^+ \left(B'_i \otimes I_d\right)$. Furthermore, we can relate the $p \times dK$ matrix $\beta_{3,i}$ to the vector $\beta_{2,i}$ (see Appendix):

$$vec\left[\beta_{3,i}\right] = P\beta_{2,i}, \tag{11}$$

where $P = I_K \otimes \left[\left(I_d \otimes D_d^+\right)(W_d \otimes I_d)(I_d \otimes vec\left[I_d\right])\right]$. The link (11) is instrumental in deriving the asymptotic results. The parameters $\beta_{1,i}$ and $\beta_{2,i}$ correspond to the parameters a_i and b_i of the unconditional case, where the matrix P is equal to I_K. Equations (10) and (11) in the conditional setting are the counterparts of restriction (4) in the static setting.

3.2 Asymptotic properties of time-varying risk premium estimation

We consider a two-pass approach building on Equations (8) and (10). The first pass consists in computing time-series OLS estimators $\hat{\beta}_i = \left(\sum_t \mathbf{1}_{i,t} x_t x'_t\right)^{-1} \sum_t \mathbf{1}_{i,t} x_t R_{i,t}$, for $i = 1, \ldots, n$. The second pass consists in computing the cross-sectional OLS estimator $\hat{\nu} = \left(\sum_i \hat{\beta}'_{3,i} \hat{\beta}_{3,i} \right)^{-1} \sum_i \hat{\beta}'_{3,i} \hat{\beta}_{1,i}$. The estimator of the time-varying risk premia is $\hat{\lambda}_t = \hat{\Lambda} Z_{t-1}$ where we deduce $\hat{\Lambda}$ from the relationship $vec\left[\hat{\Lambda}'\right] = \hat{\nu} + vec\left[\hat{F}'\right]$ with the estimator $\hat{F}$ obtained by a SUR regression of factors f_t on lagged instruments Z_{t-1}: $\hat{F} = \sum_t f_t Z'_{t-1} \left(\sum_t Z_{t-1} Z'_{t-1} \right)^{-1}$.

Proposition 6 summarizes consistency of estimators $\hat{\beta}_i$, $\hat{\nu}$, and $\hat{\Lambda}$ under the double asymptotics $n, T \to \infty$. It extends Proposition 1 to the conditional case.

Proposition 6 *Under Assumptions B.1 and C.1,C.2,C.10, we get*

$$a) \sup_i \left\| \hat{\beta}_i - \beta_i \right\| = o_p(1), \qquad b) \left\| \hat{\nu} - \nu \right\| = o_p(1), \qquad c) \left\| \hat{\Lambda} - \Lambda \right\| = o_p(1),$$

when $n, T \to \infty$ such that $n \asymp T^\gamma$ for $\gamma > 0$, and $T_i \geq CT$ for all i with $C > 0$.

Part c) implies $\sup_t \left\| \hat{\lambda}_t - \lambda_t \right\| = o_p(1)$ under for instance a boundeness assumption on process Z_t.

Proposition 7 below gives the large-sample distributions under the double asymptotics $n, T \to \infty$. It extends Proposition 2 to the conditional case through adequate use of selection matrices. The following assumption is similar to Assumption A.3. Below we make use of the moments $\mu_{\beta_3} = E_G\left[vec\left[\beta'_{3,i}\right]\right]$, $\mu^{(2)}_{\beta_3} = E_G\left[vec\left[\beta'_{3,i}\right] vec\left[\beta'_{3,i}\right]'\right]$, $\Sigma_{\beta_3} = \mu^{(2)}_{\beta_3} - \mu_{\beta_3}\mu'_{\beta_3}$ and $Q_{\beta_3} = E_G\left[\beta'_{3,i}\beta_{3,i}\right]$. We also need $Q_z = \lim_{T\to\infty} \frac{1}{T} \sum_t E\left[Z_t Z'_t\right]$, otherwise, we keep the same notations as in Section 2 for the limiting sums contributing to the asymptotic variances.

Assumption B.3 *As $n, T \to \infty$ such that $n \asymp T^\gamma$ for $\gamma \in \Gamma_1 \subset \mathbb{R}^+$ and $T_i \geq CT$ for all i with $C > 0$, a) for all i, $Y_{i,T} = \frac{1}{\sqrt{T_i}} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \Rightarrow N(0, S_{ii})$; b) $\frac{1}{\sqrt{n}} \sum_i \tau_i \left(Y_{i,T} \otimes vec\left[\beta'_{3,i}\right]\right) \Rightarrow N(0, S_{\beta_3})$, where $S_{\beta_3} = S_1 \otimes \Sigma_{\beta_3} + S_2 \otimes \mu_{\beta_3}\mu'_{\beta_3}$; c) $\frac{1}{\sqrt{T}} \sum_t u_t \otimes Z_{t-1} \Rightarrow N(0, \Sigma_u)$, where $\Sigma_u = E\left[u_t u'_t \otimes Z_{t-1} Z'_{t-1}\right]$ and $u_t = f_t - F Z_{t-1}$.*

Proposition 7 *Under Assumptions B.1-B.3 and C.1-C.8, we have a) $\sqrt{T_i}\left(\hat{\beta}_i - \beta_i\right) \Rightarrow N\left(0, Q_x^{-1} S_{ii} Q_x^{-1}\right)$, for any i, conditionally on β_i; b) $\sqrt{nT}\left(\hat{\nu} - \nu - \frac{1}{T} B_\nu\right) \Rightarrow N(0, \Sigma_\nu)$ where $B_\nu = \left(\mu^{(2)}_{\beta_3}\right)^{-1} J_1 vec\left[W_{p,dK} P E'_2 Q_x^{-1} S_1 Q_x^{-1} C_\nu\right]$, and*

$\Sigma_\nu = \left(vec\left[C'_\nu Q_x^{-1}\right]' \otimes Q_{\beta_3}^{-1}\right) S_{\beta_3} \left(vec\left[C'_\nu Q_x^{-1}\right] \otimes Q_{\beta_3}^{-1}\right)$, *with* $C_\nu = \left(E'_1 - \left(\nu' \otimes I_p\right) PE'_2\right)'$, $E_1 = (I_p : 0)'$ *and* $E_2 = (0 : I_{dK})'$, $J_1 = \left(vec\left[I_p\right]' \otimes I_{dK}\right)$*; c)*$\sqrt{T} vec\left[\hat{\Lambda}' - \Lambda'\right] \Rightarrow N\left(0, \Sigma_\Lambda\right)$ *where* $\Sigma_\Lambda = \left(I_K \otimes Q_z^{-1}\right) \Sigma_u \left(I_K \otimes Q_z^{-1}\right)$*, when* $n, T \to \infty$ *such that* $n \asymp T^\gamma$ *for* $\gamma \in \Gamma_1 \subset \mathbb{R}^+$ *as in Assumption A.3, and* $T_i \geq CT$ *for all* i *and* $C > 0$.

Since $\lambda_t = \Lambda Z_{t-1} = \left(Z'_{t-1} \otimes I_K\right) W_{d,K} vec\left[\Lambda'\right]$, part c) implies conditionally on Z_{t-1} : $\sqrt{T}\left(\hat{\lambda}_t - \lambda_t\right) \Rightarrow N\left(0, \left(Z'_{t-1} \otimes I_K\right) W_{d,K} \Sigma_\Lambda W_{K,d}\left(Z_{t-1} \otimes I_K\right)\right)$.

Paralleling with the unconditional case, we can estimate a time varying cost of equity $CE_{i,t} = r_{f,t} + b_{i,t}\lambda_t$ with $\widehat{CE}_{i,t} = r_{f,t} + \hat{b}'_{i,t}\hat{\lambda}_t$. From the properties of the tr operator, we deduce that $\hat{b}'_{i,t}\hat{\lambda}_t = tr[Z'_{t-1}\hat{B}'_i\hat{\Lambda}Z_{t-1}] = vec[Z_{t-1}Z'_{t-1}]'vec[\hat{B}'_i\hat{\Lambda}]$. Then, from the properties of the vec operator, we get

$$\begin{aligned} \sqrt{T}\left(\widehat{CE}_{i,t} - CE_{i,t}\right) &= \tau_i\left(\lambda'_t \otimes Z'_{t-1}\right) E'_2 \sqrt{T_i}\left(\hat{\beta}_i - \beta_i\right) \\ &\quad + \left(Z'_{t-1} \otimes b'_{i,t}\right) W_{d,K} \sqrt{T} vec\left(\hat{\Lambda}' - \Lambda'\right) + o_p(1). \end{aligned}$$

From Proposition 7 and the asymptotic independence between estimators $\hat{\beta}_i$ and $\hat{\Lambda}$, we deduce that, conditionally on Z_{t-1}, $\sqrt{T}\left(\widehat{CE}_{i,t} - CE_{i,t}\right) \Rightarrow N\left(0, \Sigma_{CE_{i,t}}\right)$, where

$$\begin{aligned} \Sigma_{CE_{i,t}} &= \tau_i^2\left(\lambda'_t \otimes Z'_{t-1}\right) E'_2 Q_x^{-1} S_{ii} Q_x^{-1} E_2\left(\lambda_t \otimes Z_{t-1}\right) \\ &\quad + \left(Z'_{t-1} \otimes b'_{i,t}\right) W_{d,K} \Sigma_\Lambda W_{K,d}\left(Z_{t-1} \otimes b_{i,t}\right). \end{aligned}$$

We can use Proposition 7 to build confidence intervals. It suffices to replace the unknown quantities $Q_x, Q_z, \mu_{\beta_3}, \mu_{\beta_3}^{(2)}, \Sigma_{\beta_3}, \Sigma_u, S_1$ and ν by their empirical counterparts and S_2 by the thresholded estimator $\tilde{S}_2$ as in Section 2.3. Then we can extend Proposition 3 to the conditional case under Assumptions B.1-B.3, A.4 and C.1-C.8.

Since Equation (10) corresponds to the asset pricing restriction (4), the null hypothesis of correct specification of the conditional model is

$$\mathcal{H}_0 : \text{ there exists } \nu \in \mathbb{R}^{dK} \text{ such that } \beta_1 = \beta_3\nu, \text{ with } vec\left[\beta_3\right] = P\beta_2,$$
$$\text{for almost all } \beta = \left(\beta'_1, \beta'_2\right)' \in \mathbb{B}.$$

Under $\mathcal{H}_0$, we have $E_G\left[\left(\beta_1 - \beta_3\nu\right)'\left(\beta_1 - \beta_3\nu\right)\right] = 0$. The alternative hypothesis is

$$\mathcal{H}_1 : \quad \inf_{\nu \in \mathbb{R}^{dK}} E_G\left[\left(\beta_1 - \beta_3\nu\right)'\left(\beta_1 - \beta_3\nu\right)\right] > 0.$$

As in Section 2.4, we build the SSR: $\hat{Q}_e = \frac{1}{n}\sum_i \hat{e}_i'\hat{e}_i$, with $\hat{e}_i = \hat{\beta}_{1,i} - \hat{\beta}_{3,i}\hat{\nu} = \hat{C}_\nu\hat{\beta}_i$ and the statistic $\hat{\xi}_{nT} = T\sqrt{n}\left(\hat{Q}_e - \frac{1}{T}\hat{B}_\xi\right)$, where $\hat{B}_\xi = \frac{1}{n}\sum_i \tau_i^2 tr\left[\hat{C}_\nu'\hat{Q}_{x,i}\hat{S}_{ii}\hat{Q}_{x,i}\hat{C}_\nu\right]$ and $\hat{C}_\nu = \left(E_1' - (\hat{\nu}\otimes I_p)PE_2'\right)'$. Thus, under $\mathcal{H}_0$ and Assumptions B.1-B.3, A.5 and C.1-C.10, we have $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT} \Rightarrow N(0,1)$, where $\tilde{\Sigma}_\xi = 2\frac{1}{n}\sum_{i,j}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^4} tr\left[\hat{C}_\nu'\hat{Q}_x^{-1}\tilde{S}_{ij}\hat{Q}_x^{-1}\hat{C}_\nu\right]^2$ as in Proposition 4. Under $\mathcal{H}_1$, we have $\hat{\xi} \xrightarrow{p} +\infty$, as in Proposition 5.

As in Section 2.4, the null hypothesis when the factors are tradable assets becomes:

$$\mathcal{H}_0: \quad \beta_1 = 0 \text{ for almost all } \beta = (\beta_1', \beta_2')' \in \mathbb{B},$$

against the alternative hypothesis

$$\mathcal{H}_1: \quad E_G\left[\beta_1'\beta_1\right] > 0.$$

We only have to substitute $\hat{Q}_a = \frac{1}{n}\sum_i \hat{\beta}_{1,i}'\hat{\beta}_{1,i}$ for $\hat{Q}_e$, and $E_1 = (I_p : 0)'$ for $\hat{C}_\nu$. This gives an extension of Gibbons, Ross and Shanken (1989) to the conditional case.

4 Empirical results

4.1 Asset pricing model and data description

Our baseline asset pricing model is a four-factor model with $f_t = (r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t})'$ where $r_{m,t}$ is the month t excess return on CRSP NYSE/AMEX/Nasdaq value-weighted market portfolio over the risk free rate (proxied by the monthly 30-day T-bill beginning-of-month yield), and $r_{smb,t}$, $r_{hml,t}$ and $r_{mom,t}$ are the month t returns on zero-investment factor-mimicking portfolios for size, book-to-market, and momentum (see Fama and French (1993), Jegadeesh and Titman (1993), Carhart (1997)). To account for time-varying alpha, betas and risk premia we use the instruments $Z_t = (1, Z_t^{*\prime})'$ where Z_t^* is a 4×1 vector of predictive variables measured at the end of month t (minus their averages over the sample). The four predictive variables are the one-month T-bill yield, the dividend yield of the CRSP value-weighted NYSE/AMEX stock index, the term spread, proxied by the difference between yields on 10-year Treasurys and three-month T-bills, and the default spread, proxied by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds (Ferson and Schadt (1996)). We use monthly stock returns data provided by the CRSP between January 1960 and December 2009. Our original sample is free of survivorship bias, but we further select only stocks having at least 60 monthly return observations in order to obtain precise alphas and betas estimates. These returns need not be contiguous. However, when we observe a missing return, we delete the following-month return since CRSP fills it with the cumulated return since the last non-missing return. Our final universe has 12,952 stocks. Stocks are classified into 25 industry sectors (Ferson and Harvey (1999)). For comparing results, we consider the 25 Fama-French (FF) portfolios and 48 industry portfolios as base assets with asymptotics for fixed n and $T \to \infty$.

4.2 Estimation results

We first present unconditional estimates before looking at the path of the time-varying estimates. Table 1 gathers the estimated annual risk premia for the following unconditional models: four-factor model, the Fama-French model and CAPM. Estimation results for the individual stocks ($n = 12,952$) and for the portfolios ($n = 25$ and $n = 48$) are strikingly different for some factors. The estimated risk premia for the market factor are of the same magnitude and all positive across the three universe of assets and the three models. The 95% confidence interval is larger by construction for fixed n, and the one built from the 25 portfolios contains the interval for large n. Let us now focus on the estimated results of the four-factor model. For the individual stocks the size factor is remunerated (4.05%) and the value factor is not remunerated. Indeed, the estimated $\hat{\lambda}_{hml}$ is negative (-2.82%), but not significantly different from zero. For the 25 portfolios we observe the reverse, i.e., the size factor is not remunerated and the value factor is remunerated (5.72%). For the 48 portfolios, the size and value factors are not remunerated. The momentum factor is only remunerated for the 25 portfolios. The large, but imprecise, estimate for the momentum premium when $n = 25$ comes from the estimate for $\hat{\nu}_{mom}$ (34.32%) that is larger and less accurate than the estimates for $\hat{\nu}_m$,$\hat{\nu}_{smb}$ and $\hat{\nu}_{hml}$ (0.60%, 0.36% and 0.72% respectively). Moreover, the 25 momentum loadings are not statistically significant for most of the portfolios. For $\hat{\lambda}_m$, $\hat{\lambda}_{smb}$, and $\hat{\lambda}_{hml}$ we obtain similar inferential results when we consider the Fama-French model. Our point estimates for $\hat{\lambda}_m$, $\hat{\lambda}_{smb}$, and $\hat{\lambda}_{hml}$ for large n agree with Ang, Liu and Schwarz (2008). Inferential conclusions based on the confidence intervals however differ since they find a remunerated hml factor under fixed n asymptotics, while our large n treatment does not. Our point estimates and confidence intervals for $\hat{\lambda}_m$, $\hat{\lambda}_{smb}$, and $\hat{\lambda}_{hml}$ for $n = 25$ agree with Shanken and Zhou (2007).

Figure 1 shows that a potential explanation of the discrepancies revealed in Table 1 is the much larger heterogeneity of the factor loadings for the individual stocks. The portfolio betas are all concentrated in the middle of the cross-sectional distribution obtained from the individual stocks. Creating portfolios distorts information by shrinking the dispersion of betas.

Figure 2 plots the estimated time-varying path of the four risk premia from the individual stocks. We also plot the unconditional estimates and the average lambda over time. The discrepancy between the unconditional estimate and the average over time is explained by a well-known bias coming from market-timing and volatility-timing (Jagannathan and Wang (1996), Lewellen and Nagel (2006), Boguth, Carlson, Fisher and Simutin (2010)). The risk premia for the market and size factors feature a counter-cyclical pattern. Indeed, the risk premia increase during economic contractions and decreases near business cycle troughs. On the contrary, the risk premia for value and momentum factors are pro-cyclical. Furthermore, conditional estimates of the value premium take negative values albeit most of the time not significantly different from zero. This contrasts with previous empirical results based on portfolios (Petkova and Zhang (2005)). Phalippou (2007) points to similar discrepancies in the unconditional value premium by obtaining a growth premium when portfolios are built on stocks with a high institutional ownership.

Figure 3 plots the estimated time-varying path of the four risk premia from the 25 portfolios. We also plot the unconditional estimates and the average lambda over time. The discrepancy between the unconditional estimate and the averages over time is also observed for $n = 25$. The conditional point estimates for $\hat{\lambda}_{mom}$ are larger and more imprecise than the unconditional estimate in Table 1. Indeed, the pointwise confidence intervals contain the confidence interval of the unconditional estimate for $\hat{\lambda}_{mom}$. Finally, we compare Figures 2 and 3 for each factor and we observe that the patterns of risk premia look similar but their levels are different. By definition of $vec\left[\hat{\Lambda}'\right]$, the risk premia are affected by the estimates $\hat{\nu}$ and $\hat{F}$. The estimates $\hat{F}$ only rely on the time-series information and not on the cross-sectional information. They impact the patterns of risk premia by channeling the behavior of the instruments irrespective of using portfolios or individual stocks. The estimates $\hat{F}$ impact the levels of that channelling since $vec\left[\Lambda'\right] = \hat{\nu} + vec\left[\hat{F}'\right]$ but to a less extent.

4.3 Specification test results

As already mentioned Figure 1 shows that the 25 FF portfolios all have four-factor market and momentum betas close to one and zero, respectively, so the model can be thought as a two-factor model consisting of smb and hml for the purposes of explaining cross-sectional variation in expected returns. For the 48 industry portfolios the dispersion around one and zero is slightly larger. As depicted in Figure 1 by Lewellen, Nagel and Shanken (2010), this empirical concentration implies that it is easy to get artificially large estimates $\hat{\rho}^2$ of the cross-sectional R^2 for three- and four-factor models. On the contrary, the observed heterogeneity in the betas coming from the individual stocks impedes this. This suggests that it is much less easy to find factors that explain expected returns on individual stocks than on portfolios. Reporting large $\hat{\rho}^2$, or small SSR $\hat{Q}_e$, when n is large, is much more impressive than when n is small.

Table 2 gathers specification test results for unconditional factor models. As already mentioned, we prefer working with test statistics based on the SSR $\hat{Q}_e$ instead of $\hat{\rho}^2$ since the population R^2 is not well-defined with tradable factors under the null hypothesis of well-specification (its denominator is zero). For the individual stocks, we compute the test statistic $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT}$ as well as its associated p-value. For the 25 FF and 48 industry portfolios, we compute the test statistic $T\hat{Q}_e$ as well as its associated p-value. We do similarly for the test statistics relying on $\hat{Q}_a$. As expected the rejection of the well specification is strong on the individual stocks. This suggests that the unconditional models do not describe the behavior of individual stocks. For the 25 portfolios, the test statistics $T\hat{Q}_e$ rejects the well specification for the three-factor model and the CAPM. For the 48 industry portfolios, the test statistics $T\hat{Q}_e$ rejects the well specification for the CAPM. This exemplifies the need of proper inference when gauging model explanation. We also observe a stronger rejection for the specification test based on $\hat{Q}_a$, which looks at the stronger restriction for tradable factors.

Table 3 gathers specification test results for conditional factor models. For the four-factor models, the test statistics $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT}$ and $T\hat{Q}_e$ fail to reject the well specification. This suggest that the conditional models do a good, almost perfect, job in modelling the behavior of individual stocks. For the 25 portfolios, the test statistics $T\hat{Q}_e$ and $T\hat{Q}_a$ also fail to reject the conditional specification for the three-factor model.

References

D.W.K. Andrews and J.C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix extimator. *Econometrica*, 60(4):953–966, 1992.

A. Ang, J. Liu, and K. Schwarz. Using individual stocks or portfolios in tests of factor models. *Working Paper*, 2008.

D. Avramov and T. Chordia. Asset pricing models and financial market anomalies. *Review of Financial Studies*, 19(3):1000–1040, 2006.

J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

J. Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1): 191–221, 2002.

J. Bai and S. Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150, 2006.

J.B. Berk. Sorting out sorts. *Journal of Finance*, 55(1):407–427, 2000.

P.J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008.

F. Black, M. Jensen, and M. Scholes. *The Capital Asset Pricing Model: Some Empirical Findings In Jensen, M.C. (Ed.), Studies in the Theory of Capital Markets. Praeger, New York.* Praeger, New York, 1972.

O. Boguth, Carlson M., Fisher A., and Simutin M. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *forthcoming Journal of Financial Economics*, 2010.

D. Bosq. *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag New York, 1998.

S.J. Brown, W.N. Goetzmann, and S.A. Ross. Survival. *The Journal of Finance*, 50(3):853–873, 1995.

M. Carhart. On persistence of mutual fund performance. *Journal of Finance*, 52(1):57–82, 1997.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

J. H. Cochrane. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, 104(3):572–621, 1996.

G. Connor. A unified beta pricing theory. *Journal of Economic Theory*, 34(1):13–31, 1984.

G. Connor and R.A. Korajczyk. Estimating pervasive economic factors with missing observations. *Working Paper No. 34, Department of Finance, Northwestern University*, 1987.

J. Conrad, M. Cooper, and G. Kaul. Value versus glamour. *The Journal of Finance*, 58(5):1969–1995, 2003.

N. El Karoui. Operator norm consistent estimation of large dimensional sparce covariance matrices. *Annals of Statistics*, 36(6):2717–2756, 2008.

E.F. Fama and K.R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

E.F. Fama and K.R. French. Industry costs of equity. *Journal of Financial Economics*, 43:153–193, 1997.

E.F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–36, 1973.

W.E. Ferson and C.R. Harvey. The variation of economic risk premiums. *Journal of Political Economy*, 99 (2):385–415, 1991.

W.E. Ferson and C.R. Harvey. Conditioning variables and the cross section of stock returns. *Journal of Finance*, 54(4):1325–1360, 1999.

W.E. Ferson and R.W. Schadt. Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, 51(2):425–61, 1996.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–54, 2000.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics*, 119(2):231–55, 2004.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(4):830–40, 2005.

E. Ghysels. On stable factor structures in the pricing of risk: do rime-varying betas help or hurt? *Jounal of Finance*, 53(2):549–573, 1998.

R. Gibbons, S.A. Ross, and J. Shanken. A test of the efficiency of a given portfolio. *Econometrica*, 57(5): 1121–1152, 1989.

C. Gourieroux, A. Monfort, and Trognon A. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700, 1984.

J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979.

R. Jagannathan and Z. Wang. The conditional capm and the cross-section of expected returns. *Journal of Finance*, 51(1):3–53, 1996.

R. Jagannathan and Z. Wang. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, 53(4):1285–1309, 1998.

R Jagannathan, G. Skoulakis, and Z. Wang. *Handbook of Financial Econometrics*, volume 2, chapter The analysis of the cross section of security returns, pages 73–134. North-Holland, 2009.

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48:65–91, 1993.

R. Kan, C. Robotti, and J. Shanken. Pricing model performance and the two-pass cross-sectional regression methodology. *Working Paper*, 2009.

S. Kandel and S. Stambaugh. Portfolio inefficiency and the cross-section of expected returns. *The Journal of Finance*, 50(1):157–184, 1995.

M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56:815–849, 2001.

J. Lewellen and S. Nagel. The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82:289–314, 2006.

J. Lewellen, S. Nagel, and J. Shanken. A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics*, 96:175–194, 2010.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.

R. H. Litzenberger and K. Ramaswamy. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, 7(2):163–195, 1979.

A. W. Lo and A. C. MacKinlay. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3(3):431–67, 1990.

S. Ludvigson and S. Ng. Macro factors in bond risk premia. *Review of Financial Studies, forthcoming*, 2010.

J.R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2007.

W.K. Newey and K.D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4):631–653, 1994.

M. H. Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

M. Petersen. Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies*, 22(1):451–480, 2008.

R. Petkova and L. Zhang. Is value riskier than growth? *Journal of Financial Economics*, 78(1):187–202, 2005.

L. Phalippou. Can risk-based theories explain the value premium? *Review of Finance*, 11(2):143–166, 2007.

S.A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

D.B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

J. Shanken. Multivariate tests of the zero-beta capm. *Journal of Financial Economics*, 14(3):327–348, 1985.

J. Shanken. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45:99–120, 1990.

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

J. Shanken and G. Zhou. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics*, 84(1):40–86, 2007.

W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

J.H. Stock and M.W. Watson. Forecasting using principal components from a large number of predictors. *Journal of American Statistical Association*, 97(460):1167–1179, 2002 a.

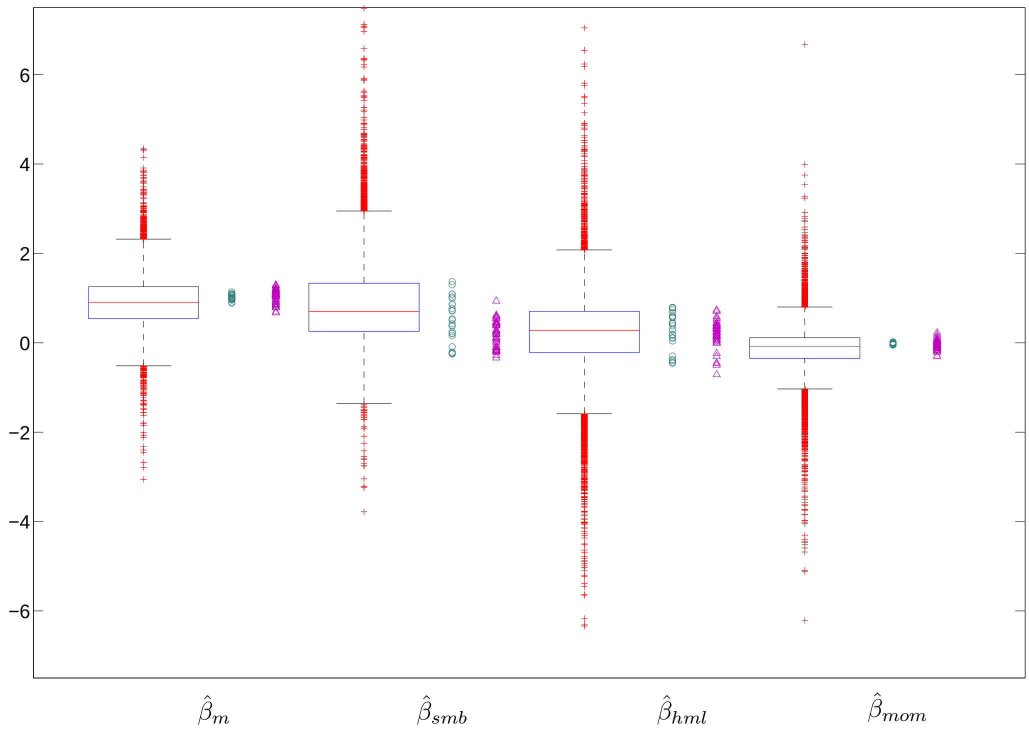
J.H. Stock and M.W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–62, 2002 b.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

H. White. *Asymptotic Theory for Econometricians*. Academic Press, 2001.

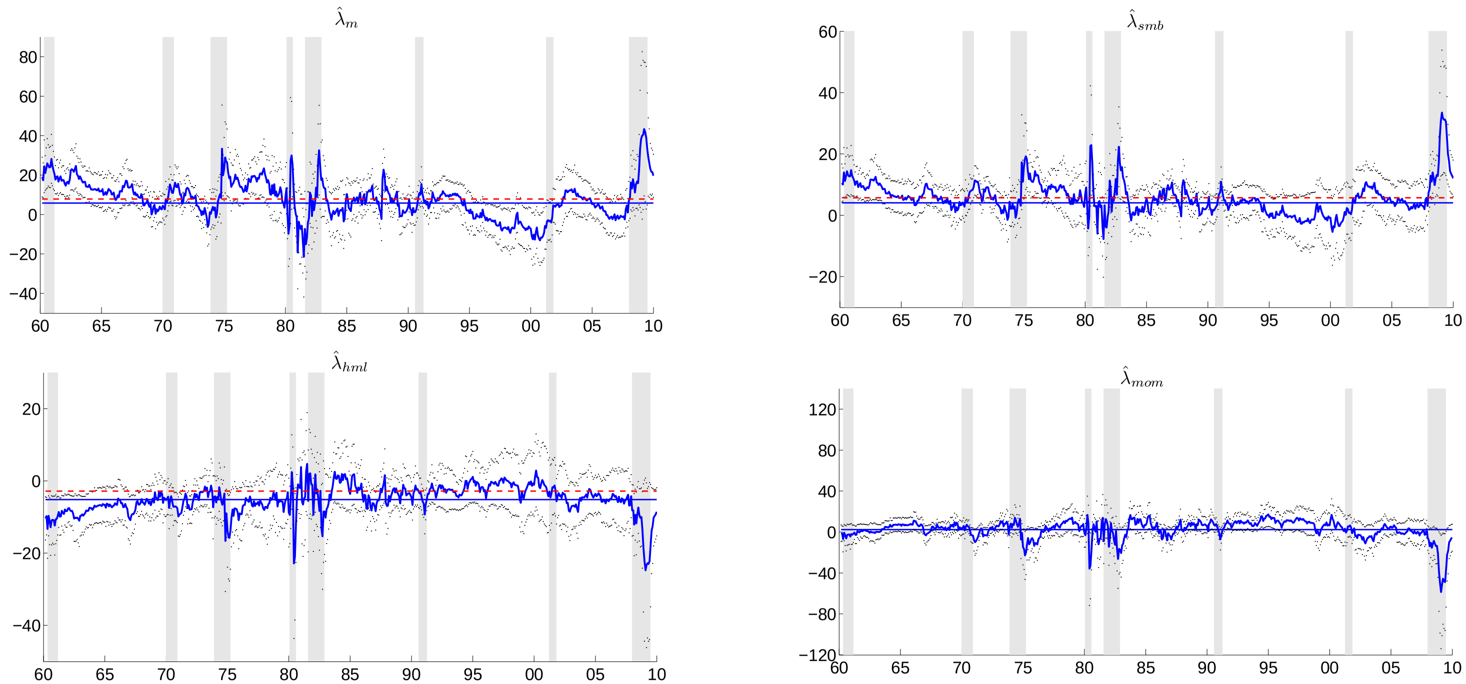
J.M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2002.

Figure 1: **Distribution of the factor loadings**



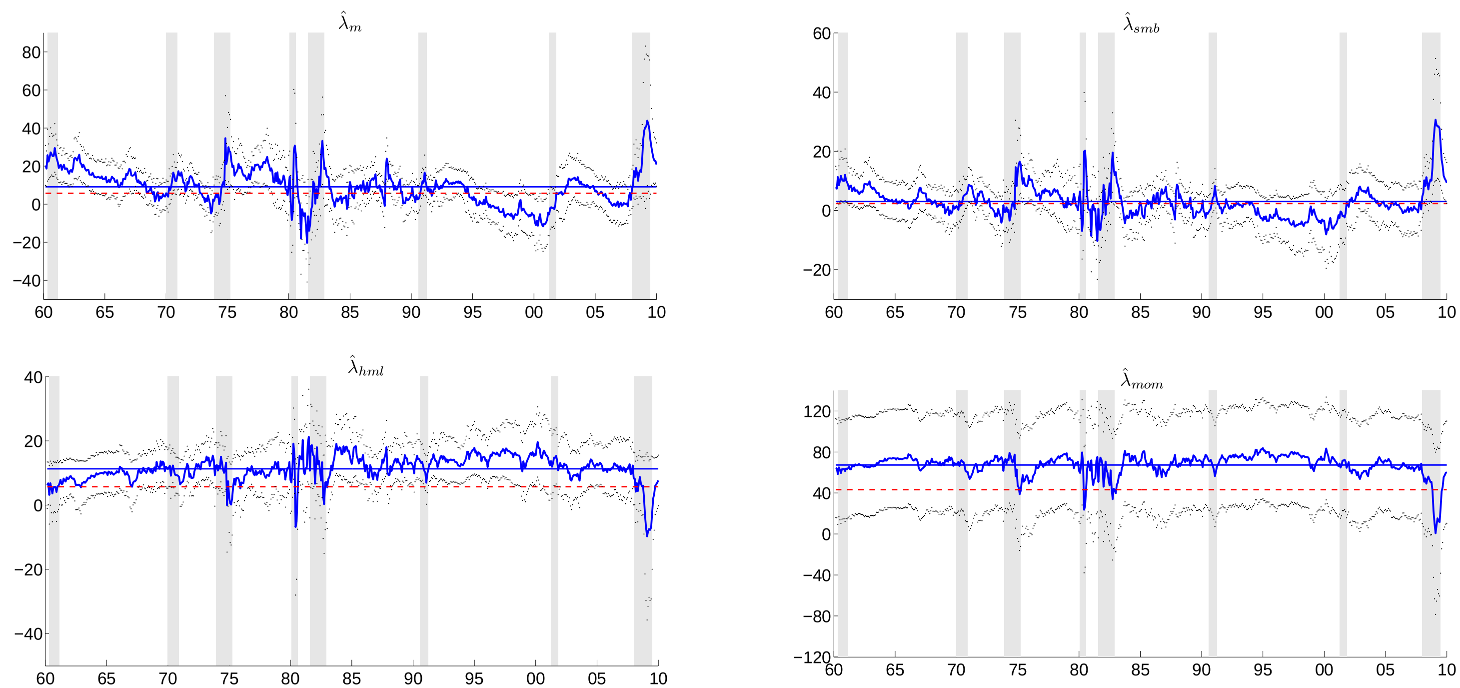
The figure displays box-plots for the distribution of factor loadings $\hat{\beta}_m, \hat{\beta}_{smb}, \hat{\beta}_{hml}$ and $\hat{\beta}_{mom}$. The factor loadings are estimated by running the time-series OLS regression in equation (2) for $n = 12,952$ from 1960/01 to 2009/12. Moreover, next to each box-plot we report the estimated factor loadings for the 25 Fama-French portfolios (circles) and for the 48 industry portfolios (triangles).

Figure 2: **Path of estimated annualized risk premia with $n = 12,952$**



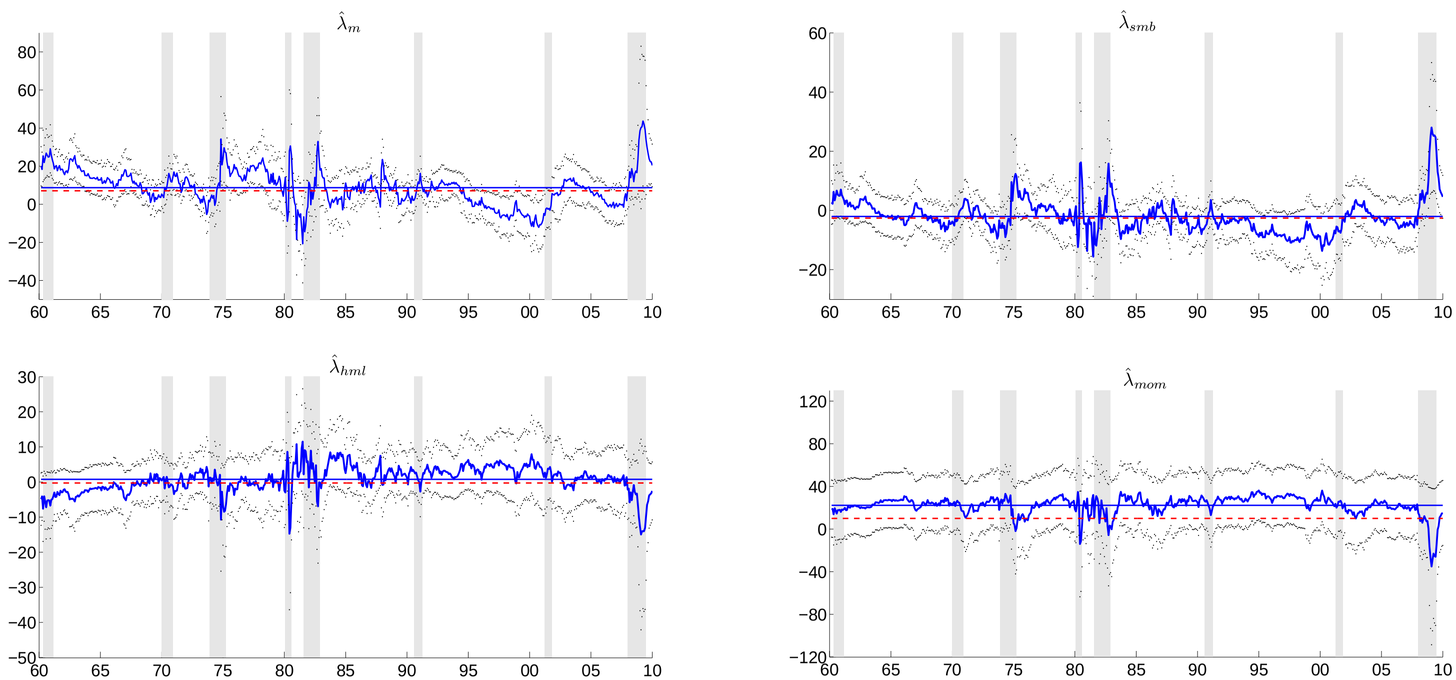
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_m$, $\hat{\lambda}_{smb}$, $\hat{\lambda}_{hml}$ and $\hat{\lambda}_{mom}$ and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dotted horizontal line) and the average conditional estimate (solid horizontal line). We consider all stocks ($n = 12,952$) as base assets. The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER). The recessions start at the peak of a business cycle and end at the trough.

Figure 3: **Path of estimated annualized risk premia with** $n = 25$



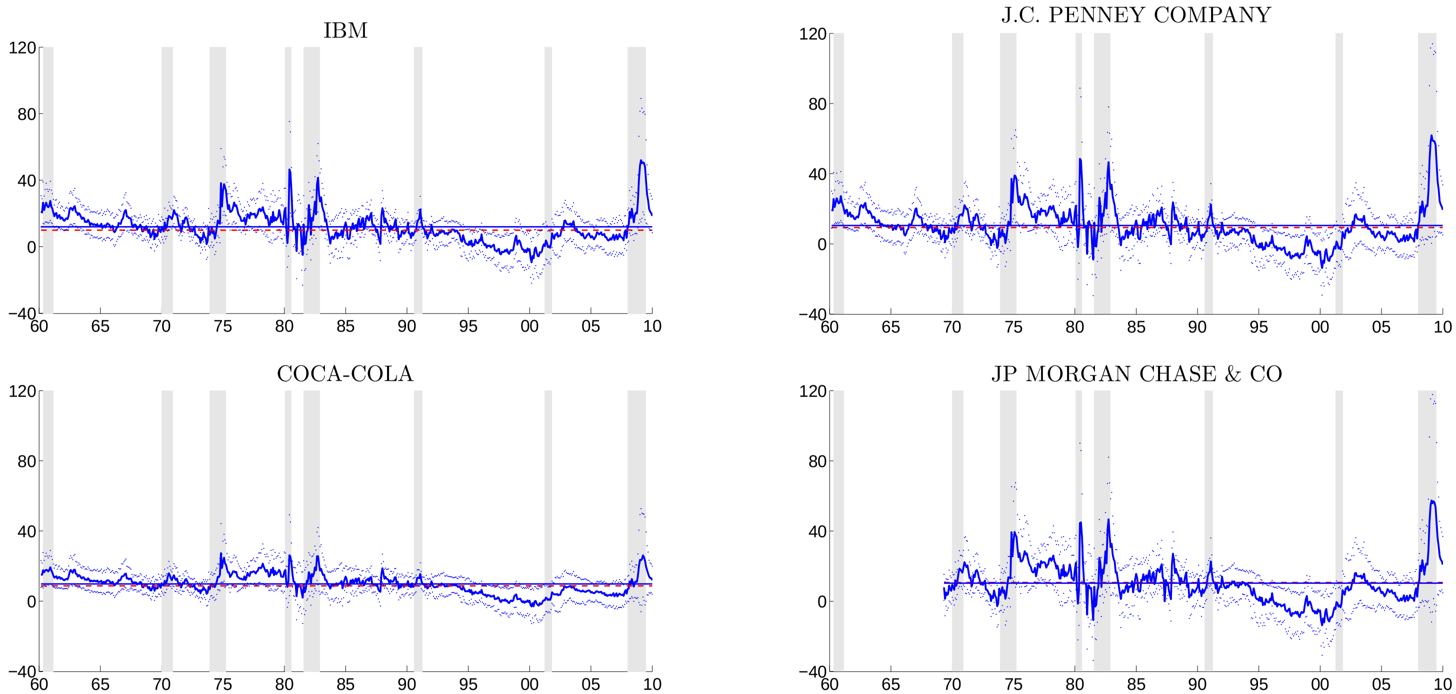
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_m$, $\hat{\lambda}_{smb}$, $\hat{\lambda}_{hml}$ and $\hat{\lambda}_{mom}$ and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dotted horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER).

Figure 4: **Path of estimated annualized risk premia with $n = 48$**



The figure plots the path of estimated annualized risk premia $\hat{\lambda}_m$, $\hat{\lambda}_{smb}$, $\hat{\lambda}_{hml}$ and $\hat{\lambda}_{mom}$ and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dotted horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER).

Figure 5: Path of estimated annualized cost of equities



The figure plots the path of estimated annualized cost of equities for IBM, J.C. Penney Company, Coca-Cola and JPMorgan Chase & Co and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dotted horizontal line) and the average conditional estimate (solid horizontal line). The vertical shaded areas denote recessions determinated by the National Bureau of Economic Research (NBER).

Table 1: **Estimated annualized risk premia for the unconditional models**

	Stocks ($n = 12,952$)		Portfolios ($n = 25$)		Portfolios ($n = 48$)	
	point estimate (%)	95% conf. interval	point estimate (%)	95% conf. interval	point estimate (%)	95% conf. interval
			Four-factor model			
$\hat{\lambda}_m$	5.82	(1.21, 10.43)	5.72	(0.94, 10.50)	7.10	(2.36, 11.85)
$\hat{\lambda}_{smb}$	4.05	(0.94, 7.15)	2.31	(-1.14, 5.76)	-2.59	(-6.80, 1.63)
$\hat{\lambda}_{hml}$	-2.82	(-6.03, 0.38)	5.72	(2.30, 9.14)	-0.29	(-4.65, 4.06)
$\hat{\lambda}_{mom}$	2.25	(-1.96, 6.46)	43.22	(16.51, 69.93)	10.00	(-2.95, 22.94)
			Fama-French model			
$\hat{\lambda}_m$	5.88	(1.27, 10.49)	4.83	(0.17, 9.50)	6.78	(2.07, 11.49)
$\hat{\lambda}_{smb}$	3.82	(0.71, 6.93)	2.36	(-0.84, 5.55)	-3.13	(-7.12, 0.8)
$\hat{\lambda}_{hml}$	-2.66	(-5.87, 0.54)	5.72	(2.43, 9.01)	-1.06	(-5.11, 2.98)
			CAPM			
$\hat{\lambda}_m$	7.22	(2.60, 11.83)	7.05	(2.17, 11.93)	5.88	(1.10, 10.67)

The point estimates of the risk premia for the unconditional models are reported for all stocks ($n = 12,952$), as well as for the 25 Fama-French portfolios and for the 48 industry portfolios. The cross-sectional regressions are based on estimates of the factor loadings. The table contains the estimated annualized risk premia referred to the market $\left(\hat{\lambda}_m\right)$, book-to-market $\left(\hat{\lambda}_{smb}\right)$, size $\left(\hat{\lambda}_{hml}\right)$ and momentum $\left(\hat{\lambda}_{mom}\right)$ factors. In order to build the confidence intervals, we compute $\hat{\Sigma}_\nu$ in Proposition 3 for $n = 12,952$. When we consider 25 and 48 portfolios as base assets, we compute an estimate of the covariance matrix $\Sigma_{\lambda,n}$ defined in Section 2.2.

Table 2: **Specification test results for the unconditional models**

	Test statistic based on $\hat{Q}_e, \mathcal{H}_0: \ a = b'\nu$			Test statistic based on $\hat{Q}_a, \mathcal{H}_0: \ a = 0$		
	Stocks ($n = 12,952$)	Portfolios ($n = 25$)	Portfolios ($n = 48$)	Stocks ($n = 12,952$)	Portfolios ($n = 25$)	Portfolios ($n = 48$)
			Four-factor model			
Test statistic	-4.7132	0.0006	0.0015	2.0142	0.0009	0.0035
p-value	0.0000	0.2707	0.6507	0.0466	0.0036	0.0783
			Fama-French model			
Test statistic	-5.8384	0.0010	0.0020	-1.3756	0.0010	0.0048
p-value	0.0000	0.0688	0.2818	0.1689	0.0489	0.0110
			CAPM			
Test statistic	-6.4676	0.0044	0.0026	-6.0488	0.0062	0.0029
p-value	0.0000	0.0000	0.0886	0.0000	0.0000	0.1972

The test statistic $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ defined in Proposition 4 is computed for $n = 12,952$. For $n = 25$ and $n = 48$, the test statistic $T\hat{Q}_e$ is reported. The test statistic based on $\hat{Q}_a$ is also computed as described in Section 2.4. The table reports the p-values, respectively.

Table 3: **Specification test results for the conditional models**

	Test statistic based on $\hat{Q}_e, \mathcal{H}_0: \beta_1 = \beta_3\nu$			Test statistic based on $\hat{Q}_a$ $\mathcal{H}_0: \beta_1 = 0$		
	Stocks ($n = 12,952$)	Portfolios ($n = 25$)	Portfolios ($n = 48$)	Stocks ($n = 12,952$)	Portfolios ($n = 25$)	Portfolios ($n = 48$)
Four-factor model						
Test statistic	0.0341	0.2993	4.5344	0.0352	0.3502	5.5282
p-value	0.9728	0.3351	0.2684	0.9719	0.6186	0.3158
Fama-French model						
Test statistic	0.0354	0.5924	4.8583	0.0380	0.7623	6.3139
p-value	0.9717	0.0060	0.1969	0.9697	0.0306	0.1835
CAPM						
Test statistic	0.0331	1.1877	5.6696	0.0367	2.5684	7.1874
p-value	0.9736	0.0000	0.0848	0.9707	0.0000	0.0954

The test statistic $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ defined in Section 3.2 is computed for $n = 12,952$. For $n = 25$ and $n = 48$, the test statistic is $T\hat{Q}_e$. The test statistic based on $\hat{Q}_a$ is also computed. The table reports the p-values, respectively.

Appendix 1

The assumptions used to derive the large sample properties of the estimators are given below:

Assumption C.1 $\sup_i \left\| \hat{Q}_{x,i} - Q_x \right\| = o_p(1)$, *with* $\hat{Q}_{x,i} = \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t x_t'$.

Assumption C.2 *There exists constants* $\eta, \bar{\eta} \in (0,1)$ *such that for all* $\delta > 0$, $i = 1, ..., n$, *and* $n, T \in \mathbb{N}$, *we have* $\mathbb{P}\left[\left\| \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \right\| \geq \delta\right] \leq C_1 T \exp\left\{-C_2 \delta^2 T^\eta\right\} + C_3 \delta^{-1} \exp\left\{-C_4 T^{\bar{\eta}}\right\}$, *where* $C_1, C_2, C_3, C_4 > 0$ *are constants.*

Assumption C.3 *a)* $E\left[\sup_i \left\| \hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1} \right\|^4\right] = o(1)$; *b)* $E\left[\|x_t\|^4\right] < M$, *for all* t *and a constant* M.

Assumption C.4 $\frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{Q}_{x,i}^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_{x,i}^{-1} = O_p(1)$.

Assumption C.5 $\sup_i \|S_{ii,T} - S_{ii}\| = O_p\left(\frac{1}{\sqrt{T}}\right)$, *where* $S_{ii,T} = \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} \sigma_{ii,t} x_t x_t'$.

Assumption C.6 $\frac{1}{n} \sum_{i,j} \|S_{ij}\| = O(1)$.

Assumption C.7 *There exist constants* $\eta, \bar{\eta} \in (0,1]$ *such that for all* $\delta > 0$, $i = 1, ..., n$, *and* $n, T \in \mathbb{N}$, *we have*

a) $\max_{i,j} \mathbb{P}\left[\left\| \hat{S}_{ij}^0 - S_{ij} \right\| \geq \delta\right] \leq C_1 T \exp\left\{-C_2 \delta^2 T^\eta\right\} + C_3 \delta^{-1} \exp\left\{-C_4 T^{\bar{\eta}}\right\}$,

where $C_1, C_2, C_3, C_4 > 0$ *are constants and* $\hat{S}_{ij}^0 = \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} \varepsilon_{j,t} x_t x_t'$.

Further, the same upper bound holds also for:

b) $\max_{i,j} \max_{k,l,m} \mathbb{P}\left[\left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_{k,t} x_{l,t} x_{m,t} \right| \geq \delta\right]$, *with* $1 \leq k, l, m \leq K$;

c) $\mathbb{P}\left[\left\| \frac{1}{T_i} \sum_t \left(\mathbf{1}_{i,t} x_t x_t' - E\left[x_t x_t'\right]\right) \right\| \geq \delta\right]$;

d) $\max_{i,j} \max_{k,l,m,p} \mathbb{P}\left[\left| \frac{1}{T_{ij}} \sum_t \left(\mathbf{1}_{ij,t} x_{k,t} x_{l,t} x_{m,t} x_{p,t} - E\left[x_{k,t} x_{l,t} x_{m,t} x_{p,t}\right]\right) \right| \geq \delta\right]$, *with* $1 \leq k, l, m, p \leq K$;

Assumption C.8 $\sup_i \left\| \hat{Q}_{x,i}^{-1} - Q_x^{-1} \right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$, *and* $\frac{1}{n} \sum_i \tau_i^2 S_{ii} - S_1 = O\left(\frac{1}{\sqrt{n}}\right)$.

Assumption C.9 $\frac{1}{\sqrt{nT}} \sum_i \sum_t \tau_i^4 \mathbf{1}_{i,t} \left(\varepsilon_{i,t}^2 - \sigma_{ii,t} \right) \hat{Q}_{x,i}^{-1} x_t x_t' \hat{Q}_{x,i}^{-1} = O_p(1)$.

Assumption C.1 concerns the existence of the moments of x_t and defines the uniform consistency for $\hat{Q}_{x,i}$. Assumption C.2 requires that the average $\frac{1}{T} \sum_t x_t \varepsilon_{i,t}$ satisfies a large deviation bound for large T uniformly in i. This assumption holds under standard conditions on the higher order moments and the serial dependence of process $x_t \varepsilon_{i,t}$ such as strong mixing (see, for instance, Theorem 1.4 in Bosq (1998)).

Appendix 2

A.2.1 Proof of Proposition 1

a) Uniform consistency of $\hat{\beta}_i$. By using $\hat{\beta}_i - \beta_i = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t}$, we have

$$\sup_i \left\| \hat{\beta}_i - \beta_i \right\| \leq \sup_i \left\| \hat{Q}_{x,i}^{-1} \right\| \sup_i \left\| \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \right\|.$$

Lemma 1 *Let* $\hat{A}_{i,n}$, $i = 1, ..., n$ *be a square matrix-valued stochastic array such that* $\sup_i \left\| \hat{A}_{i,n} - A \right\| = o_p(1)$ *as* $n \to \infty$, *where matrix* A *is non-singular. Then* $\sup_i \left\| \hat{A}_{i,n}^{-1} - A^{-1} \right\| = o_p(1)$.

By using Lemma 1 and Assumption C.1, we get $\sup_i \left\| \hat{Q}_{x,i}^{-1} \right\| = O_p(1)$. Then, we have to prove

$$\mathbb{P}\left[\sup_i \left\| \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \right\| \geq \delta \right] = o_p(1), \quad \forall \delta > 0.$$

We have:

$$\mathbb{P}\left[\sup_i \left\| \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \right\| \geq \delta \right] \leq n \sup_i \mathbb{P}\left[\left\| \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \right\| \geq \delta \right],$$

and by Assumption C.2, we deduce:

$$\mathbb{P}\left[\left\| \frac{1}{T_i} \sum_t \mathbf{1}_{i,t} x_t \varepsilon_{i,t} \right\| \geq \delta \right] \leq C_1 T \exp\left\{ -C_2 \delta^2 T^{\eta} \right\} + C_3 \delta^{-1} \exp\left\{ -C_4 T^{\bar{\eta}} \right\}.$$

Then, $\sup_i \left\| \hat{\beta}_i - \beta_i \right\| = o_p(1)$ follows.

b) Consistency of $\hat{\nu}$. By definition of $\hat{\nu}$, we have

$$\begin{aligned} \hat{\nu} - \nu &= \left(\frac{1}{n} \sum_i \hat{b}_i \hat{b}_i' \right)^{-1} \frac{1}{n} \sum_i \hat{b}_i \left(\hat{a}_i - \hat{b}_i' \nu \right) \\ &= \left(\frac{1}{n} \sum_i \hat{b}_i \hat{b}_i' \right)^{-1} \frac{1}{n} \sum_i \hat{b}_i c_\nu' \left(\hat{\beta}_i - \beta_i \right), \end{aligned} \tag{12}$$

where $c_\nu = (1, -\nu')'$. Now, $\frac{1}{n} \sum_i \hat{b}_i c' \left(\hat{\beta}_i - \beta_i \right) = o_p(1)$, and $\frac{1}{n} \sum_i \hat{b}_i \hat{b}_i' = E_G \left[b_i b_i' \right] + o_p(1)$, since $\sup_i \left\| \hat{\beta}_i - \beta_i \right\| = o_p(1)$, and $\sup_i \left\| \hat{b}_i \right\| = O_p(1)$ by Assumption A.2 and part a). Thus, $\|\hat{\nu} - \nu\| = o_p(1)$.

c) Consistency of $\hat{\lambda}$. By $\frac{1}{T}\sum_t (f_t) - E[f_t] = o_p(1)$, we get

$$\left\|\hat{\lambda}-\lambda\right\| \leq \|\hat{\nu}-\nu\| + \left\|\frac{1}{T}\sum_t f_t - E[f_t]\right\| = o_p(1).$$

■

A.2.2 Proof of Proposition 2

a) Asymptotic normality of $\hat{\beta}_i$. By definition of $\hat{\beta}_i$, we have

$$\sqrt{T_i}\left(\hat{\beta}_i-\beta_i\right) = \hat{Q}_{x,i}^{-1}Y_{i,T}.$$

By Assumption C.1, $\hat{Q}_{x,i}$ converges in probability to Q_x, for any $i = 1, ..., n$. By Assumption A.3a), and Slutsky theorem, the conclusion follows.

b) Asymptotic normality of $\hat{\nu}$. From Equation (12), and by using $\sqrt{T_i}\left(\hat{b}_i - b_i\right) = E_2'\hat{Q}_{x,i}^{-1}Y_{i,T}$, we get:

$$\begin{aligned}
\sqrt{nT}\left(\hat{\nu}-\nu\right) &= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_i\hat{b}_i c_\nu'\hat{Q}_{x,i}^{-1}Y_{i,T} \\
&= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_i b_i c_\nu'\hat{Q}_{x}^{-1}Y_{i,T} \\
&\quad + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_i b_i c_\nu'\left(\hat{Q}_{x,i}^{-1}-\hat{Q}_{x}^{-1}\right)Y_{i,T} \\
&\quad + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{nT}}\sum_i \tau_i^2\left(E_2'\hat{Q}_{x,i}^{-1}Y_{i,T}\right)\left(c_\nu'\hat{Q}_{x,i}^{-1}Y_{i,T}\right) \\
&=: I_{21}+I_{22}+I_{23}.
\end{aligned}$$

By the properties of the vec operator and Kronecker product, we have

$$\begin{aligned} I_{21} &= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i \tau_i b_i Y_{i,T}'\right)\hat{Q}_x^{-1}c_\nu \\ &= \left(c_\nu'\hat{Q}_x^{-1}\otimes\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\right)\frac{1}{\sqrt{n}}\sum_i \tau_i vec\left(b_i Y_{i,T}'\right) \\ &= \left(c_\nu'\hat{Q}_x^{-1}\otimes\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\right)\frac{1}{\sqrt{n}}\sum_i \tau_i \left(Y_{i,T}\otimes b_i\right). \end{aligned}$$

Then $I_{21}\Rightarrow N\left(0,\Sigma_\nu\right)$, by Proposition 1 and Assumptions A.3b) and C.7c).

Let us now consider I_{22}, and define $\zeta_{nT}=\frac{1}{\sqrt{n}}\sum_i \tau_i b_i c_\nu'\left(\hat{Q}_{x,i}^{-1}-\hat{Q}_x^{-1}\right)Y_{i,T}$. We have by Assumption A.1a)

$$E\left[\zeta_{nT}^2|\left(x_t\right)\right] = \frac{1}{n}\sum_{i,j}\frac{1}{\sqrt{T_iT_j}}\sum_t \tau_i\tau_j b_i c_\nu'\left(\hat{Q}_{x,i}^{-1}-\hat{Q}_x^{-1}\right)\sigma_{ij,t}\mathbf{1}_{ij,t}x_tx_t'\left(\hat{Q}_{x,j}^{-1}-\hat{Q}_x^{-1}\right)c_\nu b_j'.$$

By using Assumptions A.1c) and $T_i\geq cT$, we get

$$\left\|E\left[\zeta_{nT}^2|\left(x_t\right)\right]\right\|\leq C\sup_i\left\|\hat{Q}_{x,i}^{-1}-\hat{Q}_x^{-1}\right\|^2\frac{1}{T}\sum_t\left\|x_t\right\|^2,$$

for a constant C. By using Assumption C.3 and Cauchy-Schwarz inequality, we deduce $I_{22}=o_p\left(1\right)$.

Let us consider I_{23}. We have

$$\begin{aligned} I_{23} &= \sqrt{\frac{n}{T}}B_v \\ &\quad+\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{nT}}\sum_i \tau_i^2E_2'\hat{Q}_{x,i}^{-1}\left(Y_{i,T}Y_{i,T}'-S_{ii,T}\right)\hat{Q}_{x,i}^{-1}c_\nu \\ &\quad+\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{nT}}\sum_i \tau_i^2E_2'\hat{Q}_{x,i}^{-1}\left(S_{ii,T}-S_{ii}\right)\hat{Q}_{x,i}^{-1}c_\nu \\ &\quad+\left\{\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}-\left(\mu_b^{(2)}\right)^{-1}\right\}\frac{1}{\sqrt{nT}}\sum_i \tau_i^2E_2'\hat{Q}_{x,i}^{-1}S_{ii}\hat{Q}_{x,i}^{-1}c_\nu \\ &\quad+\left(\mu_b^{(2)}\right)^{-1}\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_i \tau_i^2E_2'\hat{Q}_{x,i}^{-1}S_{ii}\hat{Q}_{x,i}^{-1}c_\nu-E_2'Q_x^{-1}S_1Q_x^{-1}c_\nu\right) \\ &= \sqrt{\frac{n}{T}}B_v+I_{24}+I_{25}+I_{26}+I_{27}. \end{aligned}$$

The asymptotic distribution of $\hat{\nu}$ follows if terms I_{24}, I_{25}, I_{26} and I_{27} are $o_p(1)$. We have $I_{24} = O_p\left(\frac{1}{\sqrt{T}}\right)$ from Assumption C.4. Moreover,

$$\left|\frac{1}{n}\sum_i \tau_i^2 E_2' \hat{Q}_{x,i}^{-1}(S_{ii,T} - S_{ii})\hat{Q}_{x,i}^{-1}c_\nu\right| \leq \|E_2\|\|c_\nu\| \sup_i \left\|\hat{Q}_{x,i}^{-1}\right\|^2 \sup_i \|S_{ii,T} - S_{ii}\|.$$

Thus, from Assumption C.5 and $\frac{\sqrt{n}}{T} = o(1)$, we get $I_{25} = O_p\left(\frac{\sqrt{n}}{T}\right) = o_p(1)$. By using $\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1} - \left(\mu_b^{(2)}\right)^{-1} = -\left(\mu_b^{(2)}\right)^{-1}\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i' - \mu_b^{(2)}\right)\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1} = O_p\left(\frac{1}{\sqrt{T}}\right)$, and $\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}S_{ii}\hat{Q}_{x,i}^{-1}c_\nu = O_p(1)$, we get $I_{26} = O_p\left(\frac{\sqrt{n}}{T}\right) = o_p(1)$. Finally, by using $\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}S_{ii}\hat{Q}_{x,i}^{-1}c_\nu - E_2'Q_x^{-1}S_1Q_x^{-1}c_\nu = O_p\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{n}}\right)$ from Assumption C.8, we get $I_{27} = o_p(1)$.

c) Asymptotic normality of $\hat{\lambda}$. We have

$$\sqrt{T}\left(\hat{\lambda} - \lambda\right) = \frac{1}{\sqrt{T}}\sum_t (f_t - E[f_t]) + \sqrt{T}(\hat{\nu} - \nu).$$

By using $\sqrt{T}(\hat{\nu} - \nu) = O_p\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}}\right) = o_p(1)$, the conclusion follows from Assumption A.3c). ■

A.2.3 Proof of Proposition 3

We have to show: (i) $\sqrt{\frac{n}{T}}\left(\hat{B}_\nu - B_\nu\right) = o_p(1)$, (ii) $\hat{\Sigma}_\nu - \Sigma_\nu = o_p(1)$.

Let us first prove (i). We have:

$$\begin{aligned}
\sqrt{\frac{n}{T}}\left(\hat{B}_\nu - B_\nu\right) &= \sqrt{\frac{n}{T}}\left(\hat{\mu}_b^{(2)}\right)^{-1}\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}\hat{S}_{ii}\hat{Q}_{x,i}^{-1}(\hat{c}_\nu - c_\nu) \\
&\quad + \sqrt{\frac{n}{T}}\left(\hat{\mu}_b^{(2)}\right)^{-1}\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}\left(\hat{S}_{ii} - S_{ii}\right)\hat{Q}_{x,i}^{-1}c_\nu \\
&\quad + \sqrt{\frac{n}{T}}\left(\left(\hat{\mu}_b^{(2)}\right)^{-1} - \left(\mu_b^{(2)}\right)^{-1}\right)\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}S_{ii}\hat{Q}_{x,i}^{-1}c_\nu \\
&\quad + \sqrt{\frac{n}{T}}\left(\mu_b^{(2)}\right)^{-1}\left(\frac{1}{n}\sum_i \tau_i^2 E_2'\hat{Q}_{x,i}^{-1}S_{ii}\hat{Q}_{x,i}^{-1}c_\nu - E_2'Q_x^{-1}SQ_x^{-1}c_\nu\right) \\
&=: I_{31} + I_{32} + I_{33} + I_{34}
\end{aligned}$$

where I_{33} and I_{34} are $o_p(1)$ from the proof of Proposition 2. Moreover, $I_{31} = O_p\left(\frac{\sqrt{n}}{T}\right) = o_p(1)$ and $I_{32} = O_p\left(\frac{\sqrt{n}}{T}\right) = o_p(1)$ by using $\hat{c}_\nu - c_\nu = O_p\left(\frac{1}{T}\right)$ and by the following lemma.

Lemma 2 *Under Assumptions...* $\sup_i \left\| \hat{S}_{ii} - S_{ii} \right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$, *when* $n, T \to \infty$.

Let us now prove (ii). This follows if $\tilde{S}_{bb} = S_{bb} + o_p(1)$. From Proposition 1 and Assumption C.6, we have to prove

$$\frac{1}{n}\sum_{i,j} \left\| \tilde{S}_{ij} - S_{ij} \right\| = o_p(1).$$

For this purpose, we introduce the following Lemmas 2 and 3 that extend results in Bickel and Levina (2008) from the i.i.d. case to the time series case.

Lemma 3 *Let* $\psi_{nT} = \max_{i,j} \left\| \hat{S}_{ij} - S_{ij} \right\|$, *and* $\Psi_{nT}(\delta) = \max_{i,j} \mathbb{P}\left[\left\| \hat{S}_{ij} - S_{ij} \right\| \geq \delta \right]$. *Under Assumption A.4,* $\frac{1}{n}\sum_{i,j} \left\| \tilde{S}_{ij} - S_{ij} \right\| = O_p\left(\psi_{nT} c_0(n) \kappa^{-q} + c_0(n) \kappa^{1-q} + \psi_{nT} n^2 \Psi_{nT}((1-v)\kappa) \right)$, *for any* $v \in (0,1)$.

Lemma 4 *Under Assumption C.7, if* $\kappa = M\sqrt{\frac{\log n}{T^\eta}}$ *with* M *large,* $n^2 \Psi_{nT}((1-v)\kappa) = O(1)$, *for any* $v \in (0,1)$, *and* $\psi_{nT} = O_p\left(\sqrt{\frac{\log n}{T^\eta}}\right)$.

From Lemma 3 and 4 it follows $\frac{1}{n}\sum_{i,j} \left\| \tilde{S}_{ij} - S_{ij} \right\| = O_p\left(\left(\frac{\log n}{T^\eta} \right)^{(1-q)/2} c_0(n) \right) = o_p(1)$. ■

A.2.4 Proof of Proposition 4

By definition of $\hat{Q}_e$, we get the following result:

Lemma 5 *Under Assumptions A.3b), and* $\mathcal{H}_0$, *we have* $\hat{Q}_e = \frac{1}{n}\sum_i \left[\hat{c}_\nu' \left(\hat{\beta}_i - \beta_i \right) \right]^2 + O_p\left(\frac{1}{nT} + \frac{1}{T^2} \right)$.

From Lemma 5 it follows:

$$\hat{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \left\{ \left[\hat{c}_\nu' \sqrt{T_i} \left(\hat{\beta}_i - \beta_i \right) \right]^2 - \hat{c}_\nu' \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} \hat{c}_\nu \right\} + o_p(1).$$

By using $\sqrt{T_i}\left(\hat{\beta}_i - \beta_i\right) = \hat{Q}_{x,i}^{-1} Y_{i,T}$, we get

$$\begin{aligned}
\hat{\xi}_{nT} &= \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_{x,i}^{-1} \left(Y_{i,T} Y_{i,T}' - \hat{S}_{ii}\right) \hat{Q}_{x,i}^{-1} \hat{c}_\nu + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_{x,i}^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_{x,i}^{-1} \hat{c}_\nu \\
&\quad - \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_{x,i}^{-1} \left(\hat{S}_{ii} - S_{ii,T}\right) \hat{Q}_{x,i}^{-1} \hat{c}_\nu + o_p(1).
\end{aligned}$$

Lemma 6 *Under Assumptions C.1 and C.9,* $\frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_{x,i}^{-1} \left(\hat{S}_{ii} - S_{ii,T}\right) \hat{Q}_{x,i}^{-1} \hat{c}_\nu = o_p(1)$.

From Lemma 6, we get

$$\begin{aligned}
\hat{\xi}_{nT} &= \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_{x,i}^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_{x,i}^{-1} \hat{c}_\nu + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_x^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_x^{-1} \hat{c}_\nu \\
&\quad + \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \left(\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1}\right) \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_x^{-1} \hat{c}_\nu \\
&\quad + \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_x^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \left(\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1}\right) \hat{c}_\nu + o_p(1).
\end{aligned}$$

Lemma 7 *Under Assumptions ...,* $\frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \left(\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1}\right) \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_x^{-1} \hat{c}_\nu = o_p(1)$.

Then, $\hat{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_x^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) \hat{Q}_x^{-1} \hat{c}_\nu + o_p(1)$. By using that $tr\left[A'B\right] = vec\left[A\right]' vec\left[B\right]$, and $vec\left[YY'\right] = (Y \otimes Y)$ for a vector Y, we get

$$\begin{aligned}
\hat{\xi}_{nT} &= \frac{1}{\sqrt{n}} \sum_i \tau_i^2 tr\left[\hat{Q}_x^{-1} \hat{c}_\nu \hat{c}_\nu' \hat{Q}_x^{-1} \left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right)\right] + o_p(1) \\
&= \left(vec\left[\hat{Q}_x^{-1} \hat{c}_\nu \hat{c}_\nu' \hat{Q}_x^{-1}\right]\right)' \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \left(Y_{i,T} \otimes Y_{i,T} - vec\left[S_{ii,T}\right]\right) + o_p(1).
\end{aligned}$$

By using Assumptions A.4 and C.1, and by consistency of $\hat{\nu}$, statistic $\hat{\xi}_{nT} \Rightarrow N(0, \Sigma_\xi)$. By using MN Theorem 3 Chapter 2, we have

$$\begin{aligned}
vec\left[Q_x^{-1} c_\nu c_\nu' Q_x^{-1}\right]' \left(S_{ij} \otimes S_{ij}\right) vec\left[Q_x^{-1} c_\nu c_\nu' Q_x^{-1}\right] &= tr\left[S_{ij} Q_x^{-1} c_\nu c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu c_\nu' Q_x^{-1}\right] \\
&= \left(c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)^2, \qquad (13)
\end{aligned}$$

and

$$vec\left[Q_x^{-1}c_\nu c_\nu' Q_x^{-1}\right]'\left(S_{ij}\otimes S_{ij}\right)W_{(Kd+p)}vec\left[Q_x^{-1}c_\nu c_\nu' Q_x^{-1}\right] \quad = \quad \left(c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)^2. \qquad (14)$$

Then, from the definition of Ω and Equations (13) and (14), we deduce $\Sigma_\xi = 2\lim_{n,T\to\infty}\frac{1}{n}\sum_{i,j}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^4}\left(c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)^2$. ■

A.2.5 Proof of Proposition 5

a) Asymptotic normality of $\hat{\nu}$. By definition of $\hat{\nu}$ and under $\mathcal{H}_1$, we have

$$\begin{aligned}
\hat{\nu}-\nu_\infty &= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{n}\sum_i \hat{b}_i c_\infty'\hat{\beta}_i \\
&= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{n}\sum_i \hat{b}_i c_\infty'\left(\hat{\beta}_i-\beta_i\right) + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{n}\sum_i \hat{b}_i e_i \\
&= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{n}\sum_i \hat{b}_i c_\infty'\left(\hat{\beta}_i-\beta_i\right) \\
&\quad + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{n}\sum_i b_i e_i + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{n}\sum_i \left(\hat{b}_i-b_i\right)e_i
\end{aligned}$$

Thus we get:

$$\begin{aligned}
\sqrt{n}\left(\hat{\nu}-\nu_\infty\right) &= \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \frac{1}{\sqrt{T_i}}\hat{b}_i c_\infty'\hat{Q}_{x,i}^{-1}Y_{i,T} \\
&\quad + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i b_i e_i \\
&\quad + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \frac{1}{\sqrt{T_i}} e_i E_2'\hat{Q}_{x,i}^{-1}Y_{i,T} \\
&=: I_{51}+I_{52}+I_{53}.
\end{aligned}$$

From Assumption A.2 and $E_G\left[b_i e_i\right]=0$, we get $\frac{1}{\sqrt{n}}\sum_i b_i e_i \Rightarrow N\left(0, E_G\left[b_i b_i' e_i^2\right]\right)$ by the CLT. Then the asymptotic distribution of $\hat{\nu}$ follows if terms I_{51} and I_{53} are $o_p(1)$. From the proof of Proposition 2, we have:

$$
\begin{aligned}
I_{51} &= \frac{1}{\sqrt{T}}\left[\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_i b_i c_\infty' \hat{Q}_x^{-1}Y_{i,T} + \sqrt{\frac{n}{T}}B_\nu + o_p(1)\right] \\
&= O_p\left(\frac{1}{\sqrt{T}} + \frac{\sqrt{n}}{T}\right) = o_p(1),
\end{aligned}
$$

since $\frac{\sqrt{n}}{T} = o_p(1)$. Let us now consider I_{53}:

$$
\begin{aligned}
I_{53} &= \frac{1}{\sqrt{T}}\left[\left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_i e_i E_2' \hat{Q}_x^{-1}Y_{i,T}\right. \\
&\quad \left. + \left(\frac{1}{n}\sum_i \hat{b}_i\hat{b}_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_i e_i E_2'\left(\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1}\right)Y_{i,T}\right] \\
&= \frac{1}{\sqrt{T}}\left(I_{54} + I_{55}\right).
\end{aligned}
$$

By Proposition 1 and Assumptions 2 and B.9c), $I_{54} = O_p(1)$. Moreover, I_{55} is $o_p(1)$ by similar arguments applied on I_{22} in the proof of Proposition 2. Thus, $I_{53} = o_p(1)$.

b) Asymptotic normality of $\hat{\lambda}$. We have

$$
\sqrt{T}\left(\hat{\lambda} - \lambda_\infty\right) = \sqrt{T}\left(\hat{\nu} - \nu_\infty\right) + \frac{1}{\sqrt{T}}\sum_t \left(f_t - E\left[f_t\right]\right).
$$

By using $\sqrt{T}\left(\hat{\nu} - \nu_\infty\right) = O_p\left(\sqrt{\frac{T}{n}}\right) = o_p(1)$, the conclusion follows.

c) Consistency of the test. By definition of $\hat{Q}_e$, we get the following result:

Lemma 8 *Under Assumptions A.3b), and $\mathcal{H}_1$, we have*
$\hat{Q}_e = \frac{1}{n}\sum_i \left[\hat{c}_\nu'\left(\hat{\beta}_i - \beta_i\right)\right]^2 + \frac{1}{n}\sum_i e_i^2 + O_p\left(\frac{1}{\sqrt{nT}}\right)$.

By similar arguments as in the proof of Proposition 4, we get:

$$
\begin{aligned}
\hat{\xi}_{nT} &= \frac{1}{\sqrt{n}}\sum_i \tau_i^2 \hat{c}_\nu' \hat{Q}_x^{-1}\left(Y_{i,T}Y_{i,T}' - S_{ii,T}\right)\hat{Q}_x^{-1}\hat{c}_\nu + T\frac{1}{\sqrt{n}}\sum_i e_i^2 + O_p\left(\sqrt{T}\right) \\
&= O_p(1) + O\left(T\sqrt{n}E_G\left[\left(a - b\nu_\infty\right)^2\right]\right) + O_p(T).
\end{aligned}
$$

■

Appendix 3

The assumptions used to derive the large sample properties of the estimators of the conditional factor model are given below:

Assumption C.10 *a) There exists a positive definite matrix* Q_{zz} *such that* $\left\| \frac{1}{T} \sum_{t} Z_t Z_t' - Q_{zz} \right\| = o_p(1)$*; b)* $\left\| E\left[u_t u_t' | Z_{t-1}\right] \right\| \leq M$*, where* $u_t = f_t - F Z_{t-1}$.

Appendix 4

A.4.1 Derivation of Equations (8) and (9)

From Equation (7) and by using $vec\,[ABC] = \left[C' \otimes A\right] vec\,[B]$ (MN Theorem 2, p.35), we get

$$Z'_{t-1}B'_i f_t = vec\left[Z'_{t-1}B'_i f_t\right] = \left[f'_t \otimes Z'_{t-1}\right] vec\left[B'_i\right] = x'_{2,t}\beta_{2,i}.$$

By definition of matrix X_t in Section 3.1, we have

$$\begin{aligned} Z'_{t-1}B'_i\left(\Lambda - F\right)Z_{t-1} &= \frac{1}{2}Z'_{t-1}\left[B'_i\left(\Lambda - F\right) + \left(\Lambda - F\right)' B_i\right] Z_{t-1} \\ &= \frac{1}{2}vech\,[X_t]'\, vech\left[B'_i\left(\Lambda - F\right) + \left(\Lambda - F\right)' B_i\right] \\ &= \frac{1}{2}x'_{1,t} vech\left[B'_i\left(\Lambda - F\right) + \left(\Lambda - F\right)' B_i\right]. \end{aligned}$$

By using the Moore-Penrose inverse of the duplication matrix D_d, we get

$$vech\left[B'_i\left(\Lambda - F\right) + \left(\Lambda - F\right)' B_i\right] = D_d^+\left[vec\left[B'_i\left(\Lambda - F\right)\right] + vec\left[\left(\Lambda - F\right)' B_i\right]\right].$$

Finally, by the properties of the vec operator and the commutation matrix $W_{d,K}$, we obtain

$$\begin{aligned} \frac{1}{2}D_d^+\left[vec\left[B'_i\left(\Lambda - F\right)\right] + vec\left[\left(\Lambda - F\right)' B_i\right]\right] &= \frac{1}{2}D_d^+\left[\left(\Lambda - F\right)' \otimes I_d + I_d \otimes \left(\Lambda - F\right)' W_{d,K}\right]\beta_{2,i} \\ &= \beta_{1,i}. \end{aligned} \tag{15}$$

The conclusion follows. ■

A.4.2 Derivation of Equation (10)

From (15), and the properties of the vec operator, we get

$$\beta_{1,i} = \frac{1}{2}D_d^+\left[\left(I_d \otimes B'_i\right) vec\left[\Lambda - F\right] + \left(B'_i \otimes I_d\right) vec\left[\Lambda' - F'\right]\right].$$

Since $vec\,[\Lambda - F] = W_{K,d} vec\,[\Lambda' - F']$, we can factorize $\nu = vec\left[\Lambda' - F'\right]$ to obtain

$$\beta_{1,i} = \frac{1}{2}D_d^+\left[\left(I_d \otimes B'_i\right) W_{d,K} + B'_i \otimes I_d\right] vec\left[\Lambda' - F'\right].$$

By properties of commutation and duplication matrices (MN p.54-58), we have $\left(I_d \otimes B_i'\right) W_{d,K} = W_d \left(B_i' \otimes I_d\right)$ and $D_d^+ W_d = D_d^+$, then

$$\beta_{1,i} \quad = \quad D_d^+ \left(B_i' \otimes I_d\right) vec \left[\Lambda' - F'\right] = \beta_{3,i}\nu.$$

■

A.4.2 Derivation of Equation (11)

From the definition of $\beta_{3,i}$, we have $vec\left[\beta_{3,i}\right] = vec\left[D_d^+ \left(B_i' \otimes I_d\right)\right]$. By MN Theorem 2 p. 35 and Example 1 p. 56, we obtain

$$\begin{aligned} vec\left[\beta_{3,i}\right] &= \left(I_{dK} \otimes D_d^+\right) vec\left[B_i' \otimes I_d\right] \\ &= \left(I_{dK} \otimes D_d^+\right) \left\{I_K \otimes \left[\left(W_d \otimes I_d\right)\left(I_d \otimes vec\left[I_d\right]\right)\right]\right\} vec\left[B_i'\right]. \end{aligned}$$

By writing $I_{dK} = I_K \otimes I_d$, we get

$$vec\left[\beta_{3,i}\right] = \left\{I_K \otimes \left[\left(I_d \otimes D_d^+\right)\left(W_d \otimes I_d\right)\left(I_d \otimes vec\left[I_d\right]\right)\right]\right\} vec\left[B_i'\right].$$

■

A.4.3 Proof of Proposition 6

a) Uniform consistency of $\hat{\beta}_i$. See proof of Proposition 1a).

b) Uniform consistency of $\hat{\nu}$. By definition of $\hat{\nu}$ we have: $\hat{\nu} - \nu = \left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1} \frac{1}{n}\sum_i \hat{\beta}_{3,i}' \left(\hat{\beta}_{1,i} - \hat{\beta}_{3,i}\nu\right)$. From Equation (11) and MN Theorem 2 p.35, we get $\hat{\beta}_{3,i}\nu = \left(\nu' \otimes I_p\right) P\hat{\beta}_{2,i}$. Moreover, by using matrices E_1 and E_2, we obtain $\left(\hat{\beta}_{1,i} - \hat{\beta}_{3,i}\nu\right) = \left[E_1' - \left(\nu' \otimes I_p\right) PE_2'\right]\hat{\beta}_i = C_\nu'\hat{\beta}_i = C_\nu'\left(\hat{\beta}_i - \beta_i\right)$, from Equation (10). It follows that

$$\hat{\nu} - \nu \quad = \quad \left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1} \frac{1}{n}\sum_i \hat{\beta}_{3,i}' C_\nu' \left(\hat{\beta}_i - \beta_i\right). \tag{16}$$

By using part a) the result follows.

c) Uniform consistency of Λ'. By definition of ν we deduce

$$\left\| vec\left[\hat{\Lambda}' - \Lambda'\right]\right\| \leq \|\hat{\nu} - \nu\| + \left\| vec\left[\hat{F}' - F'\right]\right\|.$$

By part b), $\|\hat{\nu} - \nu\| = o_p(1)$. By LLN and Assumptions C.10, we have $\frac{1}{T}\sum_t Z_{t-1}Z'_{t-1} = O_p(1)$ and $\frac{1}{T}\sum_t u_t Z'_{t-1} = o_p(1)$. Then, by Slustky theorem, we conclude that $\left\| vec\left[\hat{F}' - F'\right]\right\| = o_p(1)$. Then, the result follows. ■

A.4.4 Proof of Proposition 7

a) Asymptotic normality of $\hat{\beta}_i$. See proof of Proposition 2a).

b) Asymptotic normality of $\hat{\nu}$. From Equation (16) and by using $\sqrt{T_i}\left(\hat{\beta}_i - \beta_i\right) = \hat{Q}_{x,i}^{-1}Y_{i,T}$, we get

$$\begin{aligned}
\sqrt{nT}\left(\hat{\nu} - \nu\right) &= \left(\frac{1}{n}\sum_i \hat{\beta}'_{3,i}\hat{\beta}_{3,i}\right)^{-1} \frac{1}{\sqrt{n}}\sum_i \tau_i \hat{\beta}'_{3,i} C'_\nu \hat{Q}_{x,i}^{-1} Y_{i,T} \\
&= \left(\frac{1}{n}\sum_i \hat{\beta}'_{3,i}\hat{\beta}_{3,i}\right)^{-1} \frac{1}{\sqrt{n}}\sum_i \tau_i \beta'_{3,i} C'_\nu \hat{Q}_{x}^{-1} Y_{i,T} \\
&\quad + \left(\frac{1}{n}\sum_i \hat{\beta}'_{3,i}\hat{\beta}_{3,i}\right)^{-1} \frac{1}{\sqrt{n}}\sum_i \tau_i \beta'_{3,i} C'_\nu \left(\hat{Q}_{x,i}^{-1} - \hat{Q}_{x}^{-1}\right) Y_{i,T} \\
&\quad + \left(\frac{1}{n}\sum_i \hat{\beta}'_{3,i}\hat{\beta}_{3,i}\right)^{-1} \frac{1}{\sqrt{n}}\sum_i \tau_i \left(\hat{\beta}_{3,i} - \beta_{3,i}\right)' C'_\nu \hat{Q}_{x,i}^{-1} Y_{i,T} \\
&= I_{71} + I_{72} + I_{73}.
\end{aligned}$$

By MN Theorem 2 p. 35, we have

$$\begin{aligned}
I_{71} &= \left(\frac{1}{n}\sum_i \hat{\beta}'_{3,i}\hat{\beta}_{3,i}\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_i \tau_i \left(Y'_{i,T} \otimes \beta'_{3,i}\right)\right) vec\left[C'_\nu \hat{Q}_x^{-1}\right] \\
&= \left(vec\left[C'_\nu \hat{Q}_x^{-1}\right]' \otimes \left(\frac{1}{n}\sum_i \hat{\beta}'_{3,i}\hat{\beta}_{3,i}\right)^{-1}\right) \frac{1}{\sqrt{n}}\sum_i \tau_i vec\left[Y'_{i,T} \otimes \beta'_{3,i}\right].
\end{aligned}$$

By applying MN Theorem 10 p. 55 and $W_{p,1} = I_p$, we get

$$\begin{aligned}
vec\left[Y'_{i,T} \otimes \beta'_{3,i}\right] &= \left(I_{(Kd+p)} \otimes W_{p,1} \otimes I_{dK}\right)\left(vec\left[Y'_{i,T}\right] \otimes vec\left[\beta'_{3,i}\right]\right) \\
&= Y_{i,T} \otimes vec\left[\beta'_{3,i}\right].
\end{aligned}$$

Then:

$$I_{71} = \left(vec\left[C_\nu' \hat{Q}_x^{-1}\right]' \otimes \left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1}\right) \frac{1}{\sqrt{n}}\sum_i \tau_i Y_{i,T} \otimes vec\left[\beta_{3,i}'\right].$$

$I_{71} \Rightarrow N(0, \Sigma_\nu)$ follows by Proposition 6 and Assumption B.3b) and C.7c).

Let us consider I_{72}. By similar arguments applied on I_{12} in the proof of Proposition 2, $I_{72} = o_p(1)$.

Let us consider I_{73}. We introduce the following lemma:

Lemma 9 *Let A be a $m \times n$ matrix and b be a $n \times 1$ vector. Then,*

$$Ab = \left(vec\left[I_n\right]' \otimes I_m\right) vec\left[vec\left[A\right] b'\right].$$

By Lemma 9, Equation (11) and $vec\left[\left(\hat{\beta}_{3,i} - \beta_{3,i}\right)'\right] = W_{p,dK} P E_2' \hat{Q}_{x,i}^{-1} Y_{i,T}$, we have

$$\begin{aligned} I_{73} &= \left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1} \frac{1}{\sqrt{nT}}\sum_i \tau_i^2 \left(vec\left[I_p\right]' \otimes I_{dK}\right) vec\left[W_{p,dK} P E_2' \hat{Q}_{x,i}^{-1} Y_{i,T} Y_{i,T}' \hat{Q}_{x,i}^{-1} C_\nu\right] \\ &=: \sqrt{\frac{n}{T}} B_\nu + I_{74} + I_{75} + I_{76} + I_{77}, \end{aligned}$$

where $I_{74} = \left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1} J_1 \frac{1}{\sqrt{nT}}\sum_i \tau_i^2 vec\left[W_{p,dK} P E_2' \hat{Q}_{x,i}^{-1}\left(Y_{i,T}Y_{i,T}' - S_{ii,T}\right)\hat{Q}_{x,i}^{-1} C_\nu\right] = o_p(1)$,
$I_{75} = \left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1} J_1 \frac{1}{\sqrt{nT}}\sum_i \tau_i^2 vec\left(W_{p,dK} P E_2' \hat{Q}_{x,i}^{-1}\left(S_{ii,T} - S_{ii}\right)\hat{Q}_{x,i}^{-1} C_\nu\right) = o_p(1)$,
$I_{76} = \left\{\left(\frac{1}{n}\sum_i \hat{\beta}_{3,i}'\hat{\beta}_{3,i}\right)^{-1} - \left(\mu_{\beta_3}^{(2)}\right)^{-1}\right\} J_1 \frac{1}{\sqrt{nT}}\sum_i \tau_i^2 vec\left(W_{p,dK} P E_2' \hat{Q}_{x,i}^{-1} S_{ii} \hat{Q}_{x,i}^{-1} C_\nu\right) = o_p(1)$ and
$I_{77=} \left(\mu_{\beta_3}^{(2)}\right)^{-1} \sqrt{\frac{n}{T}} J_1 \left(\frac{1}{n}\sum_i \tau_i^2 vec\left(W_{p,dK} P E_2' \hat{Q}_{x,i}^{-1} S_{ii} \hat{Q}_{x,i}^{-1} C_\nu - W_{p,dK} P E_2' Q_x^{-1} S_1 Q_x^{-1} C_\nu\right)\right) = o_p(1)$
by similar arguments applied on I_{24}, I_{25}, I_{26} and I_{27}.By using similar arguments on I_{13} as in the proof of Proposition 2, we can show that I_{74}, I_{75}, I_{76} and I_{77} are $o_p(1)$.

c) Asymptotic normality of $vec\left(\hat{\Lambda}'\right)$. We have

$$\sqrt{T} vec\left[\hat{\Lambda}' - \Lambda'\right] = \sqrt{T} vec\left[\hat{F}' - F'\right] + \sqrt{T}\left(\hat{\nu} - \nu\right).$$

By using, $\sqrt{T}\left(\hat{\nu} - \nu\right) = O_p\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}}\right) = o_p(1)$, the conclusion follows from Assumption B.3c). ■

Supplementary Materials

Time-Varying Risk Premium in Large Cross-Sectional Equity Datasets

P. Gagliardini, E. Ossola and O. Scaillet

A Proof of Lemma 1

Let $\epsilon > 0$ be given. We have to prove that

$$\mathbb{P}\left(\sup_i \left\|\hat{A}_{i,n}^{-1} - A^{-1}\right\| > \epsilon\right) = o(1).$$

Let $\eta = \frac{1}{2}\left\|A^{-1}\right\|^{-1}$ and consider the event $\sup_i \left\|\hat{A}_{i,n} - A\right\| \leq \eta$. Write:

$$\begin{aligned} \hat{A}_{i,n}^{-1} - A^{-1} &= \left[A\left(I - A^{-1}\left(A - \hat{A}_{i,n}\right)\right)\right]^{-1} - A^{-1} \\ &= \left\{\left[I - A^{-1}\left(A - \hat{A}_{i,n}\right)\right]^{-1} - I\right\} A^{-1} \end{aligned}$$

and use that, for a square matrix B such that $\|B\| < 1$ we have

$$(I - B)^{-1} = I + B + B^2 + B^3 + ...$$

and

$$\left\|(I - B)^{-1} - I\right\| \leq \|B\| + \|B\|^2 + ... \leq \frac{\|B\|}{1 - \|B\|}.$$

Thus, we get for any i:

$$\begin{aligned} \left\|\hat{A}_{i,n}^{-1} - A^{-1}\right\| &\leq \frac{\left\|A^{-1}\left(A - \hat{A}_{i,n}\right)\right\|}{1 - \left\|A^{-1}\left(A - \hat{A}_{i,n}\right)\right\|} \left\|A^{-1}\right\| \\ &\leq \frac{\left\|A^{-1}\right\|^2 \left\|A - \hat{A}_{i,n}\right\|}{1 - \left\|A^{-1}\right\| \left\|A - \hat{A}_{i,n}\right\|} \\ &\leq 2\left\|A^{-1}\right\|^2 \left\|\hat{A}_{i,n} - A\right\|. \end{aligned}$$

Thus, if $\sup_i \left\|\hat{A}_{i,n} - A\right\| \leq \eta$ for $\eta = \left\|A^{-1}\right\|^{-1}/2$, we have

$$\sup_i \left\|\hat{A}_{i,n}^{-1} - A^{-1}\right\| \leq 2\left\|A^{-1}\right\|^2 \sup_i \left\|\hat{A}_{i,n} - A\right\|.$$

We get

$$\begin{aligned} \mathbb{P}\left[\sup_i \left\|\hat{A}_{i,n}^{-1} - A^{-1}\right\| > \epsilon\right] &\leq \mathbb{P}\left[\sup_i \left\|\hat{A}_{i,n} - A\right\| > \eta\right] \\ &\quad + \mathbb{P}\left[\sup_i \left\|\hat{A}_{i,n} - A\right\| > \frac{\epsilon}{2\left\|A^{-1}\right\|^2}\right] \\ &= o(1). \end{aligned}$$

■

B Proof of Lemma 2

C Proof of Lemma 3

By definition of $\tilde{S}_{ij}$, we have

$$\begin{aligned}\frac{1}{n}\sum_{i,j}\left\|\tilde{S}_{ij}-S_{ij}\right\| &= \frac{1}{n}\sum_{i,j}\left\|\hat{S}_{ij}\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa\right\}}-S_{ij}\right\| \\ &\leq \frac{1}{n}\sum_{i,j}\left\|S_{ij}\mathbf{1}_{\{\|S_{ij}\|\geq\kappa\}}-S_{ij}\right\|+\frac{1}{n}\sum_{i,j}\left\|\hat{S}_{ij}\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa\right\}}-S_{ij}\mathbf{1}_{\{\|S_{ij}\|\geq\kappa\}}\right\| \\ &=: I_{31}+I_{32}.\end{aligned}$$

By Assumption A.4,

$$I_{31}=\frac{1}{n}\sum_{i,j}\|S_{ij}\|\,\mathbf{1}_{\{\|S_{ij}\|<\kappa\}}\leq\max_i\sum_j\|S_{ij}\|^q\,\kappa^{1-q}\leq\kappa^{1-q}c_0\left(n\right). \tag{17}$$

Let us now consider II:

$$\begin{aligned}I_{32} &= \frac{1}{n}\sum_{i,j}\left\|\hat{S}_{ij}\right\|\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa,\|S_{ij}\|<\kappa\right\}}+\frac{1}{n}\sum_{i,j}\|S_{ij}\|\,\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|<\kappa,\|S_{ij}\|\geq\kappa\right\}} \\ &\quad+\frac{1}{n}\sum_{i,j}\left\|\hat{S}_{ij}-S_{ij}\right\|\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa,\|S_{ij}\|\geq\kappa\right\}} \\ &\leq \max_i\sum_j\left\|\hat{S}_{ij}\right\|\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa,\|S_{ij}\|<\kappa\right\}}+\max_i\sum_j\|S_{ij}\|\,\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|<\kappa,\|S_{ij}\|\geq\kappa\right\}} \\ &\quad+\max_i\sum_j\left\|\hat{S}_{ij}-S_{ij}\right\|\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa,\|S_{ij}\|\geq\kappa\right\}} \\ &=: I_{33}+I_{34}+I_{35}.\end{aligned}$$

From Assumption A.4, we have:

$$I_{35}\leq\max_{i,j}\left\|\hat{S}_{ij}-S_{ij}\right\|\max_i\sum_j\|S_{ij}\|^q\,\kappa^{-q}\leq O_p\left(\psi_{nT}c_0\left(n\right)\kappa^{-q}\right). \tag{18}$$

Let us study I_{33}:

$$\begin{aligned}I_{33} &\leq \max_i\sum_j\left\|\hat{S}_{ij}-S_{ij}\right\|\mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\|\geq\kappa,\|S_{ij}\|<\kappa\right\}}+\max_i\sum_j\|S_{ij}\|\,\mathbf{1}_{\{\|S_{ij}\|<\kappa\}} \\ &=: I_{36}+I_{37}.\end{aligned}$$

By Assumption A.4,

$$I_{37} \leq \kappa^{1-q} c_0(n). \tag{19}$$

Now take $v \in (0,1)$. Let $N_i(\delta) = \sum_j \mathbf{1}_{\left\{\left\|\hat{S}_{ij} - S_{ij}\right\| > \delta\right\}}$, then

$$\begin{aligned} I_{36} &= \max_i \sum_j \left\|\hat{S}_{ij} - S_{ij}\right\| \mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\| \geq \kappa, \left\|S_{ij}\right\| \leq v\kappa\right\}} + \max_i \sum_j \left\|\hat{S}_{ij} - S_{ij}\right\| \mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\| \geq \kappa, v\kappa < \left\|S_{ij}\right\| < \kappa\right\}} \\ &\leq \max_{i,j} \left\|\hat{S}_{ij} - S_{ij}\right\| \max_i N_i\left((1-v)\kappa\right) + \max_{i,j} \left\|\hat{S}_{ij} - S_{ij}\right\| c_0(n) (v\kappa)^{-q} \end{aligned}$$

and $\max_i N_i\left((1-v)\kappa\right) = O_p\left(n^2 \max_{i,j} \mathbb{P}\left[\left\|\hat{S}_{ij} - S_{ij}\right\| \geq (1-v)\kappa\right]\right)$. Thus,

$$I_{36} = O_p\left(\psi_{nT} n^2 \Psi_{nT}\left((1-v)\kappa\right) + \psi_{nT} c_0(n) (v\kappa)^{-q}\right). \tag{20}$$

Finally, we consider I_{34}. We have

$$\begin{aligned} I_{34} &\leq \max_i \sum_j \left(\left\|\hat{S}_{ij} - S_{ij}\right\| + \left\|\hat{S}_{ij}\right\|\right) \mathbf{1}_{\left\{\left\|\hat{S}_{ij}\right\| < \kappa, \left\|S_{ij}\right\| \geq \kappa\right\}} \\ &\leq \max_{i,j} \left\|\hat{S}_{ij} - S_{ij}\right\| \max_i \sum_j \mathbf{1}_{\left\{\left\|S_{ij}\right\| \geq \kappa\right\}} + \kappa \max_i \sum_j \mathbf{1}_{\left\{\left\|S_{ij}\right\| \geq \kappa\right\}} \\ &= O_p\left(\psi_{nT} c_0(n) \kappa^{-q} + c_0(n) \kappa^{1-q}\right). \end{aligned} \tag{21}$$

Combining (17)-(21) the result follows. ■

D Proof of Lemma 4

By using $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} - x_t'\left(\hat{\beta}_i - \beta_i\right)$ and $\hat{S}_{ij}^0 = \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} \varepsilon_{j,t} x_t x_t'$, we have:

$$\begin{aligned} \hat{S}_{ij} &= \hat{S}_{ij}^0 - \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_t'\left(\hat{\beta}_j - \beta_j\right) x_t x_t' - \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{j,t} x_t'\left(\hat{\beta}_i - \beta_i\right) x_t x_t' \\ &\quad + \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \left(\hat{\beta}_i - \beta_i\right)' x_t x_t' \left(\hat{\beta}_j - \beta_j\right) x_t x_t' \\ &= \hat{S}_{ij}^0 - A_{ij} - B_{ij} + C_{ij} \end{aligned}$$

where $A_{ij} = B_{ji}$. Then, for any i, j, we have $\left\| \hat{S}_{ij} - S_{ij} \right\| \leq \left\| \hat{S}^0_{ij} - S_{ij} \right\| + \|A_{ij}\| + \|B_{ij}\| + \|C_{ij}\|$. We get for any $\delta \geq 0$:

$$\begin{aligned}
\Psi_{nT}(\delta) &\leq \max_{i,j} \mathbb{P}\left[\left\|\hat{S}^0_{ij} - S_{ij}\right\| \geq \frac{\delta}{4}\right] + \max_{i,j} \mathbb{P}\left[\|A_{ij}\| \geq \frac{\delta}{4}\right] \\
&\quad + \max_{i,j} \mathbb{P}\left[\|B_{ij}\| \geq \frac{\delta}{4}\right] + \max_{i,j} \mathbb{P}\left[\|C_{ij}\| \geq \frac{\delta}{4}\right] \\
&= \Psi^0_{nT}(\delta/4) + 2P_{1,nT}(\delta/4) + P_{2,nT}(\delta/4),
\end{aligned}$$

where $\Psi^0_{nT}(\delta/4) = \max_{i,j} \mathbb{P}\left[\left\|\tilde{S}^0_{ij} - S_{ij}\right\| \geq \frac{\delta}{4}\right]$, $P_{1,nT}(\delta/4) = \max_{i,j} \mathbb{P}\left[\|A_{ij}\| \geq \frac{\delta}{4}\right]$, and $P_{2,nT}(\delta/4) = \max_{i,j} \mathbb{P}\left[\|C_{ij}\| \geq \frac{\delta}{4}\right]$. From Assumption **??**a), we have:

$$\Psi^0_{nT}(\delta/4) \leq C_1 T \exp\left\{-\frac{C_2\delta^2}{16} T^\eta\right\} + \frac{4C_3}{\delta} \exp\left\{-C_4 T^\gamma\right\}.$$

Let us consider $P_{1,nT}(\delta/4)$. For some constant C, we have

$$\|A_{ij}\| \leq C \max_{k,l,m} \left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_{k,t} x_{l,t} x_{m,t} \right| \left\| \hat{\beta}_j - \beta_j \right\|.$$

Thus,

$$\begin{aligned}
P_{1,nT}(\delta/4) &\leq \max_{i,j} \mathbb{P}\left[\max_{k,l,m} \left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_{k,t} x_{l,t} x_{m,t} \right| \left\| \hat{\beta}_j - \beta_j \right\| \geq \frac{\delta}{4C}\right] \\
&\leq \max_{i,j} \mathbb{P}\left[\text{either } \max_{k,l,m} \left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_{k,t} x_{l,t} x_{m,t} \right| \geq \sqrt{\frac{\delta}{4C}} \right. \\
&\qquad \left. \text{or } \left\| \hat{\beta}_j - \beta_j \right\| \geq \sqrt{\frac{\delta}{4C}}\right] \\
&\leq \max_{i,j} \mathbb{P}\left[\max_{k,l,m} \left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_{k,t} x_{l,t} x_{m,t} \right| \geq \sqrt{\frac{\delta}{4C}}\right] \\
&\quad + \max_j \mathbb{P}\left[\left\| \hat{\beta}_j - \beta_j \right\| \geq \sqrt{\frac{\delta}{4C}}\right] \\
&\leq (K+1)^3 \max_{i,j} \max_{k,l,m} \mathbb{P}\left[\left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} \varepsilon_{i,t} x_{k,t} x_{l,t} x_{m,t} \right| \geq \sqrt{\frac{\delta}{4C}}\right] \\
&\quad + \max_j \mathbb{P}\left[\left\| \hat{\beta}_j - \beta_j \right\| \geq \sqrt{\frac{\delta}{4C}}\right],
\end{aligned}$$

where $K = 1$. By Assumption **??**b),

$$\max_{i,j} \max_{k,l,m} \mathbb{P}\left[\left|\frac{1}{T_{ij}}\sum_t \mathbf{1}_{ij,t}\varepsilon_{i,t}x_{k,t}x_{l,t}x_{m,t}\right| \geq \sqrt{\frac{\delta}{4C}}\right] \leq C_1 T \exp\left\{-\frac{C_2\delta}{4C}T^{\eta}\right\} + C_3\sqrt{\frac{4C}{\delta}}\exp\left\{-C_4 T^{\gamma}\right\}.$$

Let us focus on $\max_j \mathbb{P}\left[\left\|\hat{\beta}_j - \beta_j\right\| \geq \sqrt{\frac{\delta}{4C}}\right]$. By using

$$\left\|\hat{\beta}_j - \beta_j\right\| \leq \|Q_{xx}^{-1}\| \left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| + \left\|\hat{Q}_{xx,j}^{-1} - Q_{xx}^{-1}\right\| \left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\|,$$

we get

$$\begin{aligned}
\mathbb{P}\left[\left\|\hat{\beta}_j - \beta_j\right\| \geq \sqrt{\frac{\delta}{4C}}\right] &\leq \mathbb{P}\left[\left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| \geq \sqrt{\frac{\delta}{16C}}\left\|Q_{xx}^{-1}\right\|^{-1}\right] \\
&\quad + \mathbb{P}\left[\left\|\hat{Q}_{xx,j}^{-1} - Q_{xx}^{-1}\right\| \left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| \geq \sqrt{\frac{\delta}{16C}}\right] \\
&\leq \mathbb{P}\left[\left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| \geq \sqrt{\frac{\delta}{16C}}\left\|Q_{xx}^{-1}\right\|^{-1}\right] \\
&\quad + \mathbb{P}\left[\left\|\hat{Q}_{xx,j}^{-1} - Q_{xx}^{-1}\right\| \geq \left(\frac{\delta}{16C}\right)^{1/4}\right] \\
&\quad + \mathbb{P}\left[\left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| \geq \left(\frac{\delta}{16C}\right)^{1/4}\right] \\
&\leq 2\mathbb{P}\left[\left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| \geq \sqrt{\frac{\delta}{16C}}\left\|Q_{xx}^{-1}\right\|^{-1}\right] \\
&\quad + \mathbb{P}\left[\left\|\hat{Q}_{xx,j}^{-1} - Q_{xx}^{-1}\right\| \geq \left(\frac{\delta}{16C}\right)^{1/4}\right],
\end{aligned}$$

for small δ. From Assumption **??**b),

$$\max_j \mathbb{P}\left[\left\|\frac{1}{T_j}\sum_t \mathbf{1}_{j,t}x_t\varepsilon_{j,t}\right\| \geq \sqrt{\frac{\delta}{16C}}\left\|Q_{xx}^{-1}\right\|^{-1}\right] \leq C_1 T \exp\left\{-\frac{C_2\delta}{16C}\left\|Q_{xx}^{-1}\right\|^{-2}T^{\eta}\right\} + C_3\sqrt{\frac{16C}{\delta}}\left\|Q_{xx}^{-1}\right\|\exp\left\{-C_4 T^{\gamma}\right\}.$$

The argument in the proof of Lemma 1 implies that either $\left\| Q_{xx}^{-1} \right\| \left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \frac{1}{2}$ or $\left\| \hat{Q}_{xx,j}^{-1} - Q_{xx}^{-1} \right\| \leq 2 \left\| Q_{xx}^{-1} \right\|^2 \left\| \hat{Q}_{xx,j} - Q_{xx} \right\|$. Then,

$$\begin{aligned} \mathbb{P}\left[\left\| \hat{Q}_{xx,j}^{-1} - Q_{xx}^{-1} \right\| \geq \left(\frac{\delta}{16C} \right)^{1/4} \right] &\leq \mathbb{P}\left[\left\| Q_{xx}^{-1} \right\| \left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \frac{1}{2} \right] \\ &\quad + \mathbb{P}\left[2 \left\| Q_{xx}^{-1} \right\|^2 \left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \left(\frac{\delta}{16C} \right)^{1/4} \right] \\ &= \mathbb{P}\left[\left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \frac{1}{2} \left\| Q_{xx}^{-1} \right\|^{-1} \right] \\ &\quad + \mathbb{P}\left[\left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \frac{1}{2} \left(\frac{\delta}{16C} \right)^{1/4} \left\| Q_{xx}^{-1} \right\|^{-2} \right] \\ &\leq 2\mathbb{P}\left[\left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \frac{1}{2} \left(\frac{\delta}{16C} \right)^{1/4} \left\| Q_{xx}^{-1} \right\|^{-2} \right] \end{aligned}$$

for small $\delta > 0$. From Assumption B.9c),

$$\begin{aligned} \mathbb{P}\left[\left\| \hat{Q}_{xx,j} - Q_{xx} \right\| \geq \frac{1}{2} \left(\frac{\delta}{16C} \right)^{1/4} \left\| Q_{xx}^{-1} \right\|^{-2} \right] &\leq C_1 T \exp\left\{ -\frac{C_2}{4} \sqrt{\frac{\delta}{16C}} \left\| Q_{xx}^{-1} \right\|^{-4} T^{\eta} \right\} \\ &\quad + 2C_3 \left(\frac{16C}{\delta} \right)^{1/4} \left\| Q_{xx}^{-1} \right\|^2 \exp\left\{ -C_4 T^{\gamma} \right\}. \end{aligned}$$

Then, $P_{1,nT}(\delta/4) \leq \tilde{C}_1 T \exp\left\{ -\tilde{C}_2 \delta T^{\eta} \right\} + \frac{\tilde{C}_3}{\sqrt{\delta}} \exp\left\{ -C_4 T^{\gamma} \right\}$, for small $\delta > 0$ and constants $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3, C_4$.

Let us now study $P_{2,nT}(\delta/4)$. We have

$$\begin{aligned} \|C_{ij}\| &\leq \left\| \hat{\beta}_i - \beta_i \right\| \left\| \hat{\beta}_j - \beta_j \right\| \sup_{k,l,m,p} \left| \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} x_{k,t} x_{l,t} x_{m,t} x_{p,t} \right| \\ &\leq \left\| \hat{\beta}_i - \beta_i \right\| \left\| \hat{\beta}_j - \beta_j \right\| \max_{k,l,m,p} \left| E\left[x_{k,t} x_{l,t} x_{m,t} x_{p,t} \right] \right| \\ &\quad + \left\| \hat{\beta}_i - \beta_i \right\| \left\| \hat{\beta}_j - \beta_j \right\| \max_{k,l,m,p} \left| \frac{1}{T_{ij}} \sum_t \left(\mathbf{1}_{ij,t} x_{k,t} x_{l,t} x_{m,t} x_{p,t} - E\left[x_{k,t} x_{l,t} x_{m,t} x_{p,t} \right] \right) \right| \left\| \hat{\beta}_j - \beta_j \right\|. \end{aligned}$$

Thus, for $C = \max_{k,l,m,p} |E[x_{k,t}x_{l,t}x_{m,t}x_{p,t}]| \leq E\left[\|x_t\|^4\right]$, we have:

$$\begin{aligned}
P_{2,nT}(\delta/4) &\leq \max_{i,j} \mathbb{P}\left[C\left\|\hat{\beta}_i - \beta_i\right\|\left\|\hat{\beta}_j - \beta_j\right\| \geq \frac{\delta}{8}\right] \\
&\quad + \max_{i,j} \mathbb{P}\left[\max_{k,l,m,p}\left|\frac{1}{T_{ij}}\sum_t \left(\mathbf{1}_{ij,t}x_{k,t}x_{l,t}x_{m,t}x_{p,t} - E[x_{k,t}x_{l,t}x_{m,t}x_{p,t}]\right)\right|\right. \\
&\quad \left.\left\|\hat{\beta}_i - \beta_i\right\|\left\|\hat{\beta}_j - \beta_j\right\| \geq \frac{\delta}{8}\right] \\
&\leq 4\max_i \mathbb{P}\left[\left\|\hat{\beta}_i - \beta_i\right\| \geq \left(\frac{\delta}{8C}\right)^{1/3}\right] + \\
&\quad + (K+1)^4 \max_{i,j}\max_{k,l,m,p} \mathbb{P}\left[\left|\frac{1}{T_{ij}}\sum_t \left(\mathbf{1}_{ij,t}x_{k,t}x_{l,t}x_{m,t}x_{p,t} - E[x_{k,t}x_{l,t}x_{m,t}x_{p,t}]\right)\right| \geq \left(\frac{\delta}{8}\right)^{1/3}\right].
\end{aligned}$$

From Assumption **??**b)-d), then $P_{2,nT}(\delta/4) \leq \tilde{C}_1 T \exp\left\{-\tilde{C}_2\delta^{2/3}T^\eta\right\} + \tilde{C}_3\delta^{-1/3}\exp\{-C_4T^\gamma\}$, for small $\delta > 0$ and some constants $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3, C_4$.

We deduce $\Psi_{nT}(\delta) \leq C_1^* T \exp\left\{-C_2^*\delta^2 T^\eta\right\} + C_3^*\delta^{-1}\exp\{-C_4^*T^\gamma\}$, for some constants $C_1^*, C_2^*, C_3^*, C_4^*$. For $\delta = (1-v)\kappa$ and $\kappa = M\sqrt{\frac{\log n}{T^\eta}}$, we get

$$\begin{aligned}
n^2\Psi_{nT}((1-v)\kappa) &\leq C_1^* n^2 T \exp\left\{-C_2^* M^2 (1-v)^2 \log n\right\} \\
&\quad + \frac{n^2 C_3^*}{(1-v)M}\sqrt{\frac{T^\eta}{\log n}}\exp\{-C_4^* T^\gamma\} \\
&= O(1),
\end{aligned}$$

for M sufficiently large.

Finally, let us prove that $\psi_{nT} = O_p\left(\sqrt{\frac{\log n}{T^\eta}}\right)$. Let $\epsilon > 0$. Then,

$$\begin{aligned}
\mathbb{P}\left[\psi_{nT} \geq \sqrt{\frac{\log n}{T^\eta}}\epsilon\right] &\leq n^2 \max_{ij} \mathbb{P}\left[\left\|\hat{S}_{ij} - S_{ij}\right\| \geq \sqrt{\frac{\log n}{T^\eta}}\epsilon\right] \\
&= n^2 \Psi_{nT}\left(\sqrt{\frac{\log n}{T^\eta}}\epsilon\right) \\
&\leq n^2 \Psi_{nT}(\kappa) = O(1)
\end{aligned}$$

for large ϵ. The conclusion follows. ■

E Check of Assumption A.5 under block-dependence

In the next result we show that Assumption A.5 holds under a cross-sectional dependence structure of the errors.

LemmaSM. 1 *Suppose that: (1) the factors (f_t) are iid, such that $E\left[|f_t|^6\right] < \infty$; (2) the errors $(\varepsilon_{i,t})$ are independent of the factor process (f_t), iid over time, with zero mean and such that $\sup_i E\left[|\varepsilon_{i,t}|^6\right] = O(1)$, and satisfy the following cross-sectional dependence structure: for any $n \in \mathbb{N}$, there exists a partition of the n assets in $M_n \leq n$ blocks $C_1, ..., C_{M_n}$, such that $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$ are independent if $i \in C_l$ and $j \in C_r$ with $l \neq r$, such that*

$$\frac{1}{n}\sum_{m=1}^{M_n} |C_m|^2 = O(1), \ \frac{1}{n^{3/2}}\sum_{m=1}^{M_n} |C_m|^3 = o(1), \tag{22}$$

where $|C_m|$ denotes the size of block C_m. Then, Assumption **??** *is satisfied.*

The conditions (22) on block sizes and block number require that there are not too many large blocks, that is, the partition in independent blocks is sufficiently fine grained asymptotically. The proof of Lemma SM.1 is given next.

Proof of Lemma SM.1 Under the conditions of Lemma SM.1, we have $S_{ii,T} = \sigma_{ii}^2 \hat{Q}_{xx,i}$ and $S_{ij} = \sigma_{ij} Q_{xx}$. The proof is in two steps.

STEP 1: We first show that conditional on the factor path (f_t), we have

$$\Upsilon_{nT} := \frac{1}{\sqrt{n}}\sum_i \tau_i^2 \left[Z_{i,T} \otimes Z_{i,T} - \widetilde{S}_{ii,T}\right] \Rightarrow N\left(0, \Omega\right), \tag{23}$$

P-a.s., where $\widetilde{S}_{ii,T} = \sigma_{ii}^2 vec(\hat{Q}_{xx,i})$ and $\Omega = \lim_{n\to\infty} \frac{1}{n}\sum_{i,j} \frac{\tau_i^4 \tau_j^4}{\tau_{ij}^4} \sigma_{ij}^2 \left[Q_{xx} \otimes Q_{xx} + \left(Q_{xx} \otimes Q_{xx}\right) W\right]$. For this purpose, we apply the Lyapunov CLT for heterogenous independent arrays [see Davidson (1994), Theorem 23.11]. Write

$$\Upsilon_{nT} = \frac{1}{\sqrt{n}}\sum_m \sum_{i \in C_m} \tau_i^2 \left[Z_{i,T} \otimes Z_{i,T} - \widetilde{S}_{ii,T}\right] = \frac{1}{\sqrt{M_n}}\sum_m W_{m,nT}, \quad m = 1, ..., M_n,$$

where:

$$W_{m,nT} = \sqrt{\frac{M_n}{n}}\sum_{i \in C_m} \tau_i^2 \left[Z_{i,T} \otimes Z_{i,T} - \widetilde{S}_{ii,T}\right].$$

Conditional on (f_t), the variables $W_{m,nT}$, for $m = 1, ..., M_n$ are independent, with zero mean. The conclusion follows if we prove:

i) $\lim_{n,T} \frac{1}{M_n} \sum_m V\left[W_{m,nT} | (f_t)\right] = \Omega$, P-a.s, and

ii) $\lim_{n,T} \frac{1}{M_n^{3/2}} \sum_m E\left[\|W_{m,nT}\|^3 | (f_t)\right] = 0$, P-a.s..

To show i), we use:

$$\begin{aligned} V\left[W_{m,nT} | (f_t)\right] &= \frac{M_n}{n} \sum_{i,j \in C_m} \tau_i^2 \tau_j^2 Cov\left[Z_{i,T} \otimes Z_{i,T}, Z_{j,T} \otimes Z_{j,T} | (f_t)\right] \\ &= \frac{M_n}{n} \sum_{i,j \in C_m} \tau_i^2 \tau_j^2 \left\{ E\left[\left(Z_{i,T} \otimes Z_{i,T}\right)\left(Z_{j,T} \otimes Z_{j,T}\right)' | (f_t)\right] - \widetilde{S}_{ii,T} \widetilde{S}_{jj,T}' \right\}. \end{aligned}$$

Now, we have by the independence property over time:

$$\begin{aligned} E\left[\left(Z_{i,T} \otimes Z_{i,T}\right)\left(Z_{j,T} \otimes Z_{j,T}\right)' | (f_t)\right] &= \frac{1}{T_i T_j} \sum_t \sum_s \sum_p \sum_q E\left[\varepsilon_{i,t}\varepsilon_{i,p}\varepsilon_{j,s}\varepsilon_{j,q} | (f_t)\right] \mathbf{1}_{i,t}\mathbf{1}_{i,p}\mathbf{1}_{j,s}\mathbf{1}_{j,q} \\ &\quad \cdot \left(x_t x_s' \otimes x_p x_q'\right) \\ &= E\left[\varepsilon_{it}^2 \varepsilon_{jt}^2\right] \frac{1}{T_i T_j} \sum_t \mathbf{1}_{i,t}\mathbf{1}_{j,t} \left(x_t x_t' \otimes x_t x_t'\right) \\ &\quad + \sigma_{ij}^2 \frac{1}{T_i T_j} \sum_t \sum_{p \neq t} \mathbf{1}_{ij,t}\mathbf{1}_{ij,p} \left(x_t x_t' \otimes x_p x_p'\right) \\ &\quad + \sigma_i^2 \sigma_j^2 \frac{1}{T_i T_j} \sum_t \sum_{s \neq t} \mathbf{1}_{i,t}\mathbf{1}_{j,s} \left(x_t x_s' \otimes x_t x_s'\right) \\ &\quad + \sigma_{ij}^2 \frac{1}{T_i T_j} \sum_t \sum_{s \neq t} \mathbf{1}_{ij,t}\mathbf{1}_{ij,s} \left(x_t x_s' \otimes x_s x_t'\right) \\ &=: E\left[\varepsilon_{it}^2 \varepsilon_{jt}^2\right] A_{1,T} + \sigma_{ij}^2 A_{2,T} + \sigma_i^2 \sigma_j^2 A_{3,T} + \sigma_{ij}^2 A_{4,T}. \end{aligned}$$

Moreover, $A_{1,T} = \frac{T_{ij}}{T_i T_j} \sum_t \frac{\mathbf{1}_{ij,t}}{T_{ij}} \left(x_t x_t' \otimes x_t x_t'\right) = O\left(T_{ij}/T_i T_j\right)$. Let us define $\hat{Q}_{xx,ij} = \frac{1}{T_{ij}} \sum_t \mathbf{1}_{ij,t} x_t x_t'$, then

$$\begin{aligned} A_{2,T} &= \frac{1}{T_i T_j} \sum_t \sum_p \mathbf{1}_{ij,t}\mathbf{1}_{ij,p} \left(x_t x_t' \otimes x_p x_p'\right) - A_{1,T} = \frac{T_{ij}^2}{T_i T_j} \left(\hat{Q}_{xx,ij} \otimes \hat{Q}_{xx,ij}\right) + O\left(T_{ij}/T_i T_j\right), \\ A_{3,T} &= \frac{1}{T_i T_j} \sum_t \sum_s \mathbf{1}_{i,t}\mathbf{1}_{j,s} \left(x_t x_s' \otimes x_t x_s'\right) - A_{1,T} = vec\left(\hat{Q}_{xx,i}\right) vec\left(\hat{Q}_{xx,j}\right)' + O\left(T_{ij}/T_i T_j\right), \end{aligned}$$

and

$$
\begin{aligned}
A_{4,T} &= \frac{1}{T_iT_j}\sum_t\sum_s \mathbf{1}_{ij,t}\mathbf{1}_{ij,s}\left(x_tx_s^{'}\otimes x_sx_t^{'}\right)-A_{1,T}\\
&= \frac{1}{T_iT_j}\sum_t\sum_s \mathbf{1}_{ij,t}\mathbf{1}_{ij,s}\left(x_t\otimes x_s\right)\left(x_s\otimes x_t\right)^{'}-A_{1,T}\\
&= \frac{1}{T_iT_j}\sum_t\sum_s \mathbf{1}_{ij,t}\mathbf{1}_{ij,s}\left(x_t\otimes x_s\right)\left(x_t\otimes x_s\right)^{'}W-A_{1,T}\\
&= \frac{T_{ij}^2}{T_iT_j}\left(\hat{Q}_{xx,ij}\otimes\hat{Q}_{xx,ij}\right)W+O\left(T_{ij}/T_iT_j\right).
\end{aligned}
$$

Then, it follows that:

$$
V\left[W_{m,nT}|\left(f_t\right)\right]=\frac{M_n}{n}\left[\sum_{i,j\in C_m}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^4}\sigma_{ij}^2\left(\hat{Q}_{xx,ij}\otimes\hat{Q}_{xx,ij}+\hat{Q}_{xx,ij}\otimes\hat{Q}_{xx,ij}W\right)+O\left(\frac{1}{T}\sum_{i,j\in C_m}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^2}\right)\right].
$$

Since $T_i\geq cT$, it follows

$$
\begin{aligned}
\frac{1}{M_n}\sum_m V\left[W_{m,nT}|\left(f_t\right)\right] &= \frac{1}{n}\sum_m\left(\sum_{i,j\in C_m}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^4}\tilde{\sigma}_{ij}^2\right)+O\left(\frac{1}{nT}\sum_m|C_m|^2\right)\\
&= \frac{1}{n}\sum_{i,j}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^4}\tilde{\sigma}_{ij}^2+o_p\left(1\right),
\end{aligned}
$$

with $\tilde{\sigma}_{ij}^2:=\sigma_{ij}^2\left(\hat{Q}_{xx,ij}\otimes\hat{Q}_{xx,ij}+\hat{Q}_{xx,ij}\otimes\hat{Q}_{xx,ij}W\right)$. The sequence $\frac{1}{n}\sum_{i,j}\frac{\tau_i^4\tau_j^4}{\tau_{ij}^4}\tilde{\sigma}_{ij}^2$ is upper bounded by a constant times $\frac{1}{n}\sum_m|C_m|^2=O(1)$, so it converges to Ω by the monotone convergence theorem. Point i) follows.

Let us now prove point ii). We have:

$$
\begin{aligned}
\frac{1}{M_n^{3/2}}\sum_m E\left[\|W_{m,nT}\|^3|\left(f_t\right)\right] &\leq \frac{1}{n^{3/2}}\sum_m\left[\sum_{i\in C_m}\tau_i^2\left(E\left[\|\left(Z_{i,T}\otimes Z_{i,T}\right)\|^3|\left(f_t\right)\right]^{1/3}+\left\|\widetilde{S}_{ii,T}\right\|\right)\right]^3\\
&\leq \frac{1}{n^{3/2}}\left(\sum_m\sum_{i\in C_m}\tau_i^2\right)\left(\sup_i E\left[\|Z_{i,T}\otimes Z_{i,T}\|^3|\left(f_t\right)\right]^{1/3}+\sup_i\left\|\widetilde{S}_{ii,T}\right\|\right)^3.
\end{aligned}
$$

Now,

$$
\begin{aligned}
E\left[\|Z_{i,T}\otimes Z_{i,T}\|^3|\left(f_t\right)\right] &\leq E\left[\|Z_{i,T}\|^6|\left(f_t\right)\right]=E\left[\left(Z_{i,T}^{'}Z_{i,T}\right)^3|\left(f_t\right)\right]\\
&= \frac{1}{T_i^3}\sum_{t_1,\ldots,t_6}\mathbf{1}_{i,t_1}\ldots\mathbf{1}_{i,t_6}E\left[\varepsilon_{i,t_1}\ldots\varepsilon_{i,t_6}\right]\left(x_{t_1}'x_{t_2}\right)\left(x_{t_3}'x_{t_4}\right)\left(x_{t_5}'x_{t_6}\right).
\end{aligned}
$$

By the independence property, the non-zero terms $E\left[\varepsilon_{i,t_1}...\varepsilon_{i,t_6}\right]$ involve at most 3 different time indices, which implies $\sup_i E\left[\|Z_{i,T}\otimes Z_{i,T}\|^3 \,|\, (f_t)\right] = O(1)$, P-a.s. Similarly $\sup_i \left\|\widetilde{S}_{iT}\right\| = O\left(1\right)$, P-a.s. Thus, we get:

$$\frac{1}{M_n^{3/2}}\sum_{m=1}^{M_n} E\left[\|W_{m,nT}\|^3 \,|\, (f_t)\right] = O\left(\frac{1}{n^{3/2}}\sum_{m=1}^{M_n}\left(\sum_{i\in C_m}\tau_i^2\right)^3\right) = O\left(\frac{1}{n^{3/2}}\sum_{m=1}^{M_n}|C_m|^3\right) = o(1).$$

STEP 2: We show that (23) implies the statement of Lemma SM.**1**. Indeed, from (23) we have:

$$\lim_{n,T\to\infty} P\left[\alpha'\Upsilon_{nT}\le z|\,(f_t)\right] = \Phi\left(\frac{z}{\sqrt{\alpha'\Omega\alpha}}\right),$$

for any $\alpha\in\mathbb{R}^4$ and for any $z\in\mathbb{R}$ and any (f_t), P-a.s. We now apply the Lebesgue dominated convergence theorem, by using that the sequence of random variables $P\left[\alpha'\Upsilon_{nT}\le z|\,(f_t)\right]$ are such that $P\left[\alpha'\Upsilon_{nT}\le z|\,(f_t)\right]\le 1$, uniformly in n and T. We conclude that, for any $z\in\mathbb{R}$:

$$\lim_{n,T\to\infty} P\left[\alpha'\Upsilon_{nT}\le z\right] = \lim_{n,T\to\infty} E\left(P\left[\alpha'\Upsilon_{nT}\le z|\,(f_t)\right]\right) = \Phi\left(\frac{z}{\sqrt{\alpha'\Omega\alpha}}\right),$$

since $\Phi\left(\frac{z}{\sqrt{\alpha'\Omega\alpha}}\right)$ is independent of the factor path. The conclusion follows. ■

F Proof of Lemma 5

Under the null hypothesis $\mathcal{H}_0$, and by definition of the fitted residual $\hat{e}_i$, we have

$$\begin{aligned}
\hat{e}_i &= a_i - b_i'\hat{\nu} + \hat{c}'\left(\hat{\beta}_i-\beta_i\right)\\
&= a_i - b_i'\nu + \hat{c}'\left(\hat{\beta}_i-\beta_i\right) - b_i'\left(\hat{\nu}-\nu\right)\\
&= \hat{c}'\left(\hat{\beta}_i-\beta_i\right) - b_i'\left(\hat{\nu}-\nu\right).
\end{aligned}$$

By definition of $\hat{Q}_{ee}$, it follows

$$\begin{aligned}
\hat{Q}_{ee} &= \frac{1}{n}\sum_i\left[\hat{c}'\left(\hat{\beta}_i-\beta_i\right)\right]^2 - 2\left(\hat{\nu}-\nu\right)'\frac{1}{n}\sum_i b_i\left(\hat{\beta}_i-\beta_i\right)'\hat{c} + \left(\hat{\nu}-\nu\right)'\frac{1}{n}\sum_i b_ib_i'\left(\hat{\nu}-\nu\right)\\
&= \frac{1}{n}\sum_i\left[\hat{c}'\left(\hat{\beta}_i-\beta_i\right)\right]^2 + I_{51} + I_{52}.
\end{aligned}$$

Let us study the second term in the RHS:

$$\begin{aligned} I_{51} &= \frac{1}{\sqrt{n}}\left(\hat{\nu}-\nu\right)' \frac{1}{\sqrt{nT}}\sum_i \tau_i b_i Z'_{i,T}\hat{Q}_{xx,i}^{-1}\hat{c} \\ &= \frac{1}{\sqrt{n}}\left(\hat{\nu}-\nu\right)' \frac{1}{\sqrt{T}}\left\{\frac{1}{\sqrt{n}}\sum_i \tau_i b_i Z'_{i,T}\hat{Q}_{xx}^{-1}\right. \\ &\quad \left.+\frac{1}{\sqrt{n}}\sum_i \tau_i b_i Z'_{i,T}\left(\hat{Q}_{xx,i}^{-1}-\hat{Q}_{xx}^{-1}\right)\right\}\hat{c} \\ &=: \frac{1}{\sqrt{n}}\left(\hat{\nu}-\nu\right)' \frac{1}{\sqrt{T}}\left(I_{53}+I_{54}\right)\hat{c}. \end{aligned}$$

Now, we have $(\hat{\nu}-\nu) = O_p\left(\frac{1}{\sqrt{nT}}+\frac{1}{T}\right)$ and $\hat{c} = O_p(1)$ by Proposition 2. Moreover, $I_{53} = O_p(1)$ by Assumption A.3b), and $I_{54} = o_p(1)$ from the Proof of Proposition 2. Thus, $I_{51} = O_p\left(\frac{1}{nT}+\frac{1}{T\sqrt{nT}}\right)$.

Let us now consider I_{52}. From Proposition 2, we have $I_{52} = O_p\left(\frac{1}{nT}+\frac{1}{T^2}\right)$. The conclusion follows.

■

G Proof of Lemma 6

By using the properties of the trace operator,

$$\frac{1}{\sqrt{n}}\sum_i \tau_i^2 \hat{c}'\hat{Q}_{xx,i}^{-1}\left(\hat{S}_{ii}-S_{ii,T}\right)\hat{Q}_{xx,i}\hat{c} = tr\left[\hat{c}\hat{c}'\frac{1}{\sqrt{n}}\sum_i \tau_i^2 \hat{Q}_{xx,i}^{-1}\left(\hat{S}_{ii}-S_{ii,T}\right)\hat{Q}_{xx,i}\right].$$

Thus, it is enough to show

$$\frac{1}{\sqrt{n}}\sum_i \tau_i^2 \hat{Q}_{xx,i}^{-1}\left(\hat{S}_{ii}-S_{ii,T}\right)\hat{Q}_{xx,i} = o_p(1).$$

We have $\hat{\varepsilon}_{i,t} = R_{i,t}-\hat{\beta}_i' x_t = \varepsilon_{i,t}-\left(\hat{\beta}_i-\beta_i\right)' x_t$. Then,

$$\begin{aligned} \hat{S}_{ii}-S_{ii,T} &= \frac{1}{T}\sum_t \tau_i^2 \mathbf{1}_{i,t}\left(\varepsilon_{i,t}^2-\sigma_{ii,t}\right)x_t x_t' - 2\frac{1}{T}\sum_t \tau_i^2 \mathbf{1}_{i,t}\left(\hat{\beta}_i-\beta_i\right)' x_t'\varepsilon_{i,t}x_t x_t' \\ &\quad +\left(\hat{\beta}_i-\beta_i\right)'\hat{Q}_{xx,i}\left(\hat{\beta}_i-\beta_i\right)'. \end{aligned}$$

Lemma 10 *Under Assumptions...* $\sup_i \left\|\hat{\beta}_i-\beta_i\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$.

From Assumption **??** and Lemma SM 2, we have

$$\hat{S}_{ii} - S_{ii,T} = \frac{1}{T}\sum_t \tau_i^2 \mathbf{1}_{i,t}\left(\varepsilon_{i,t}^2 - \sigma_{ii,t}\right) x_t x_t' + O_p\left(\frac{1}{T}\right),$$

Thus, by using Assumption B.11(?), we get

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_i \tau_i^2 \hat{Q}_{xx,i}^{-1}\left(\hat{S}_{ii} - S_{ii,T}\right)\hat{Q}_{xx,i}^{-1} &= \frac{1}{\sqrt{n}}\sum_i \tau_i^2 \hat{Q}_{xx,i}^{-1}\left[\frac{1}{T}\sum_t \tau_i^2 \mathbf{1}_{i,t}\left(\varepsilon_{i,t}^2 - \sigma_{ii,t}\right) x_t x_t'\right]\hat{Q}_{xx,i}^{-1} \\
&\quad + O_p\left(\frac{1}{T}\frac{1}{\sqrt{n}}\sum_i \tau_i^2\right) \\
&= \frac{1}{\sqrt{T}}\frac{1}{\sqrt{nT}}\sum_i\sum_t \tau_i^4 \mathbf{1}_{i,t}\left(\varepsilon_{i,t}^2 - \sigma_{ii,t}\right)\hat{Q}_{xx,i}^{-1} x_t x_t' \hat{Q}_{xx,i}^{-1} \\
&\quad + O_p\left(\frac{\sqrt{n}}{T}\right) \\
&= O_p\left(\frac{1}{\sqrt{T}} + \frac{\sqrt{n}}{T}\right).
\end{aligned}$$

The conclusion follows by the condition $\frac{\sqrt{n}}{T} = o_p(1)$. ■

H Proof of Lemma 7

I Proof of Lemma 8

Under $\mathcal{H}_1$, we have $\hat{e}_i = e_i + \hat{c}'\left(\hat{\beta}_i - \beta_i\right) - b_i'(\hat{\nu} - \nu)$. By definition of $\hat{Q}_{ee}$, it follows:

$$\begin{aligned}
\hat{Q}_{ee} &= \frac{1}{n}\sum_i e_i^2 + 2\frac{1}{n}\sum_i \hat{c}'\left(\hat{\beta}_i - \beta_i\right) e_i - 2(\hat{\nu} - \nu)'\frac{1}{n}\sum_i b_i e_i \\
&\quad + \frac{1}{n}\sum_i \left[\hat{c}'\left(\hat{\beta}_i - \beta_i\right)\right]^2 - 2(\hat{\nu} - \nu)'\frac{1}{n}\sum_i b_i\left(\hat{\beta}_i - \beta_i\right)'\hat{c} + (\hat{\nu} - \nu)'\frac{1}{n}\sum_i b_i b_i'(\hat{\nu} - \nu) \\
&=: I_{81} + I_{82} + I_{83} + I_{84} + I_{85} + I_{86}. \qquad (24)
\end{aligned}$$

From proof of Proposition 5a), we have

$$
\begin{aligned}
I_{82} &= \frac{1}{\sqrt{nT}}\left\{\frac{1}{\sqrt{n}}\sum_i \tau_i e_i E_2' \hat{Q}_{xx}^{-1} Z_{i,T}\right.\\
&\left. +\frac{1}{\sqrt{n}}\sum_i \tau_i e_i E_2'\left(\hat{Q}_{xx,i}^{-1}-\hat{Q}_{xx}^{-1}\right) Z_{i,T}\right\}\\
&= \frac{1}{\sqrt{nT}}\left(I_{87}+I_{88}\right).
\end{aligned}
$$

The term I_{87} is $O_p(1)$. Moreover, I_{88} is $o_p(1)$. Thus, $I_{82} = O_p\left(\frac{1}{\sqrt{nT}}\right)$. By using Proposition 5, we have $I_{83} = O_p\left(\frac{1}{n}\right)$. Let us now consider I_{85}. From proof of Lemma 5, we get $I_{85} = O_p\left(\frac{1}{n\sqrt{T}}\right)$. Let us cosider I_{86}. From Proposition 5, we have $I_{86} = O_p\left(\frac{1}{n}\right)$. The conclusion follows. ∎

J Proof of Lemma 9

By applying MN Theorem 2 p.35, Theorem 10 p. 55 and using $W_{n,1} = I_n$, we have

$$
\begin{aligned}
Ab = vec(Ab) &= \left(b' \otimes A\right) vec\left(I_n\right)\\
&= vec\left[\left(b' \otimes A\right) vec\left(I_n\right)\right]\\
&= \left(vec\left(I_n\right)' \otimes I_m\right) vec\left(b' \otimes A\right)\\
&= \left(vec\left(I_n\right)' \otimes I_m\right)\left(I_n \otimes W_{n,1} \otimes I_m\right)\left(vec\left(b'\right) \otimes vec\left(A\right)\right)\\
&= \left(vec\left(I_n\right)' \otimes I_m\right)\left(I_n \otimes I_m\right) vec\left(vec\left(A\right) b'\right)\\
&= \left(vec\left(I_n\right)' \otimes I_m\right) vec\left(vec\left(A\right) b'\right)
\end{aligned}
$$

∎