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

Patrick Gagliardini
Elisa Ossola
Olivier Scaillet

First version: December 2010
Current version: August 2011

This research has been carried out within the NCCR FINRISK projects on
“Financial Econometrics For Risk Management”
“New Methods in Theoretical and Empirical Asset Pricing”
TIME-VARYING RISK PREMIUM IN LARGE CROSS-SECTIONAL EQUITY DATASETS

Patrick Gagliardini\textsuperscript{a}, Elisa Ossola\textsuperscript{b} and Olivier Scaillet\textsuperscript{c*}

First draft: December 2010
This version: August 2011

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 where the conditioning includes instruments common to all assets and asset specific instruments. The estimator uses simple weighted two-pass cross-sectional regressions, and we show its 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 induced by the no-arbitrage assumption in large economies. The empirical illustration on returns for about ten thousands US stocks from July 1964 to December 2009 shows that conditional risk premia 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.

\textit{JEL Classification:} C12, C13, C23, C51, C52 , G12.

\textit{Keywords:} large panel, factor model, risk premium, asset pricing.

\textsuperscript{a}University of Lugano and Swiss Finance Institute, \textsuperscript{b}University of Lugano, \textsuperscript{c}University of Genève and Swiss Finance Institute.

\textsuperscript{*}Acknowledgements: We gratefully acknowledge the financial support of the Swiss National Science Foundation (Prodoc project PDFM11-114533). We thank Y. Amihud, V. Chernozhukov, B. Engle, J. Fan, E. Ghysels, C. Gouriéroux, S. Heston, participants at the Cass conference 2010, CIRPEE conference 2011, ECARES conference 2011, Core conference 2011, Montreal panel data conference 2011, ESEM 2011, and participants at seminars at Columbia, NYU, Georgetown, Maryland, McGill, GWU, ULB, UCL, Humboldt, Orléans, Imperial College, CREST, Athens for helpful comments.
1 Introduction


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 we 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 (2009) show that it is a promising route to follow to study bond risk premia. Connor, Hagmann, and Linton (2011) show that large cross-section helps to exploit data more efficiently in a semiparametric characteristic-based factor model of stock returns. It is also inspired by the framework underlying the Arbitrage Pricing Theory (APT). Approximate factor structures with nondiagonal error covariance matrices (Chamberlain and Rothschild (1983, CR)) address the potential empirical mismatch of exact factor structures with diagonal error covariance matrices underlying the original APT of Ross (1976). Under weak cross-sectional dependence among error terms, they generate no-arbitrage restrictions in large economies where the number of assets grows to infinity. Our paper develops an econometric methodology tailored to the APT framework.
We let the number of assets grow to infinity mimicking the large economies of financial theory.

Our approach is further motivated 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 threefold. First we derive no-arbitrage restrictions in a multi-period economy (Hansen and Richard (1987)) with a continuum of assets and an approximate factor structure (Chamberlain and Rothschild (1983)). We explicitly show the relationship between the ruling out of asymptotic arbitrage opportunities and a testable restriction for large economies in a conditional setting. We also formalize the sampling scheme when observed assets are random draws from an underlying population (Andrews (2005)). Second we derive a new weighted two-pass cross-sectional estimator of the path over time of the risk premia from large unbalanced panels of excess returns. We study its large sample properties in conditional linear factor models where the conditioning includes instruments common to all assets and asset specific instruments. 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 relate the results to bias-corrected estimation (Hahn and Kuersteiner (2002), Hahn and Newey (2004)) accounting for the well-known incidental parameter problem of the panel literature (Neyman and Scott (1948)). 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, unless imposing strong ad hoc structural restrictions. Third we provide a goodness-of-fit test for the conditional factor model underlying the estimation. The test exploits the asymptotic distribution of a weighted sum of squared residuals of the second-pass cross-sectional regression (see Lewellen, Nagel and Shanken (2010), Kan, Robotti and Shanken (2009) for a related approach in unconditional models and asymptotics with fixed \( n \)). The construction of the test statistic relies on consistent estimation of large-dimensional sparse covariance matrices by thresholding (Bickel and Levina (2008), El Karoui (2008), Fan, Liao, and Mincheva (2011)). 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 pedagogical purposes the three theoretical contributions are first presented in an unconditional setting before being extended to a conditional setting.

For our empirical contributions, we consider the Center for Research in Security Prices (CRSP) database and take the Compustat database to match firm characteristics. The merged dataset comprises about ten thousands stocks with monthly returns from July 1964 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)). For the conditional versions we use both macrovariables and firm characteristics as instruments. 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.

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 extend all results to cover a conditional linear factor model where the instruments inducing time varying coefficients can be common to all stocks or stock specific. Section 4 contains the empirical results. Section 5 contains the simulation results. Finally, Section 6 concludes. In the Appendix, we gather the technical assumptions and some proofs. We place all omitted proofs in the online supplementary materials. We use high-level assumptions to get our results and show in Appendix 4 that they are all met under a block cross-sectional dependence structure on the error terms in a serially i.i.d. framework.

## 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 Excess return generation and asset pricing restrictions

We start by describing how excess returns are generated before examining the implications of absence of arbitrage opportunities in terms of asset pricing restrictions in large economies. We combine the construction of Hansen and Richard (1987) for multi periods and the construction of Andrews (2005) for cross-section dependence. We need such a formal construction to guarantee that all information sets are well-defined when we condition on them, and that, under a sampling mechanism (see the next section), cross-sectional limits exist and are invariant to reordering of the assets.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. The random vector $f$ admitting values in $\mathbb{R}^K$, and the collection of random variables $\varepsilon(\gamma)$, $\gamma \in [0, 1]$, are defined on this probability space. Moreover, let $\beta = (a, b)'$ be a vector function defined on $[0, 1]$, with values in $\mathbb{R} \times \mathbb{R}^K$. The dynamics is described by the measurable and
measure-preserving transformation $S$ mapping $\Omega$ into itself. If $\omega \in \Omega$ is the state of the world at time $0$, then $S^t(\omega)$ is the state at time $t$, where $S^t$ denotes the transformation $S$ applied $t$ times successively.

**Assumption APR.1** The excess returns $R_t(\gamma)$ of asset $\gamma \in [0, 1]$ at date $t = 1, 2, \ldots$ satisfy the unconditional linear factor model:

$$R_t(\gamma) = a(\gamma) + b(\gamma)'f_t + \varepsilon_t(\gamma),$$

(1)

where the random variables $\varepsilon_t(\gamma)$ and $f_t$ are defined by $\varepsilon_t(\gamma, \omega) = \varepsilon(\gamma, S^t(\omega))$ and $f_t(\omega) = f(S^t(\omega))$.

Assumption APR.1 defines the excess return processes for an economy with a continuum of assets. The index set is the interval $[0, 1]$ without loss of generality. Vector $f_t$ gathers the values of the $K$ observable factors at date $t$, while the intercept $a(\gamma)$ and factor sensitivities $b(\gamma)$ of asset $\gamma \in [0, 1]$ are time invariant. Since $S$ is measure-preserving, all processes are stationary. Let further $x_t = (1, f_t)'$ which yields the compact formulation:

$$R_t(\gamma) = \beta(\gamma)'x_t + \varepsilon_t(\gamma).$$

(2)

In order to define the information sets, let $\mathcal{F}_0 \subset \mathcal{F}$ be a sub sigma-field. Random vector $f$ is assumed measurable w.r.t. $\mathcal{F}_0$. Define $\mathcal{F}_t = \{S^{-t}(A), A \in \mathcal{F}_0\}, t = 1, 2, \ldots$, and assume that $\mathcal{F}_1$ contains $\mathcal{F}_0$. Then, the filtration $\mathcal{F}_t, t = 1, 2, \ldots$ characterizes the information available to investors.

Let us now introduce supplementary assumptions on factors, factor loadings and error terms.

**Assumption APR.2** The matrix $\int b(\gamma)b(\gamma)'d\gamma$ is positive definite.

Assumption APR.2 implies non-degeneracy in the factor loadings across assets.

**Assumption APR.3** For any $\gamma \in [0, 1]$: $E[\varepsilon_t(\gamma)|\mathcal{F}_{t-1}] = 0$ and $Cov[\varepsilon_t(\gamma), f_t|\mathcal{F}_{t-1}] = 0$.

Hence, the error terms have mean zero and are uncorrelated with the factors conditionally on information $\mathcal{F}_{t-1}$. In Assumption APR.4 (i) below, we build on CR and impose an approximate factor structure for the conditional distribution of the error terms given $\mathcal{F}_{t-1}$ in any countable collection of assets.

**Assumption APR.4** The conditional variance-covariance matrix $\Sigma_{\varepsilon,t,n}$ of the errors $\varepsilon_t(\gamma_1), \ldots, \varepsilon_t(\gamma_n)$ given $\mathcal{F}_{t-1}$ is such that: (i) $\text{eig}_{\max}(\Sigma_{\varepsilon,t,n}) \leq c_1, \text{ P-a.s.}$, (ii) $\text{eig}_{\min}(\Sigma_{\varepsilon,t,n}) \geq c_2, \text{ P-a.s., for all } \gamma_1, \ldots, \gamma_n$
\[ n \in \mathbb{N}, \text{ and some constants } 0 < c_2 \leq c_1 < \infty, \text{ where } \text{eig}_{\text{min}}(\Sigma_{\varepsilon,t,n}) \text{ and } \text{eig}_{\text{max}}(\Sigma_{\varepsilon,t,n}) \text{ denote the smallest and the largest eigenvalues of matrix } \Sigma_{\varepsilon,t,n}, (iii) \text{eig}_{\text{max}}(V[f_t|\mathcal{F}_{t-1}]) \leq c_3 \text{ and } \text{eig}_{\text{min}}(V[f_t|\mathcal{F}_{t-1}]) \geq c_4, \text{ P-a.s., for some constants } 0 < c_4 \leq c_3 < \infty. \]

Assumptions APR.4 (ii)-(iii) are mild regularity conditions used in the proof of Proposition 1.

Absence of asymptotic arbitrage opportunities generates asset pricing restrictions in large economies (Ross (1976), CR). We define asymptotic arbitrage opportunities in terms of sequences of portfolios \( p_n, n \in \mathbb{N} \). Portfolio \( p_n \) is defined by the share \( \alpha_0 \) invested in the riskfree asset and the shares \( \alpha_i \) invested in the risky assets \( \gamma_i \) for \( i = 1, \ldots, n \). The shares are measurable w.r.t. \( \mathcal{F}_0 \). Then \( C(p_n) = \sum_{i=0}^{n} \alpha_i \) is the portfolio cost at \( t = 0 \), and \( p_n = C(p_n)R_0 + \sum_{i=1}^{n} \alpha_i R_1(\gamma_i) \) is the portfolio payoff at \( t = 1 \), where \( R_0 \) denotes the riskfree gross return measurable w.r.t. \( \mathcal{F}_0 \). We can work with \( t = 1 \) because of stationarity.

**Assumption APR.5** There are no asymptotic arbitrage opportunities in the economy, that is, there exists no portfolio sequence \( (p_n) \) such that \( \lim_{n \to \infty} P[p_n \geq 0] = 1 \) and \( \lim_{n \to \infty} P[C(p_n) \leq 0, p_n > 0] > 0. \)

Assumption APR.5 excludes portfolios that approximate arbitrage opportunities when the number of included assets increases. Arbitrage opportunities are investments with non-positive cost and non-negative payoff in each state of the world, and positive payoff in some states of the world (Hansen and Richard (1987), Definition 2.4). Then, the asset pricing restriction is given in the next Proposition 1.

**Proposition 1** Under Assumptions APR.1-APR.5, there exists a unique vector \( \nu \in \mathbb{R}^K \) such that:

\[ a(\gamma) = b(\gamma)'\nu, \tag{3} \]

for almost all \( \gamma \in [0, 1] \).

The asset pricing restriction in Proposition 1 can be rewritten as

\[ E[R_t(\gamma)] = b(\gamma)'\lambda, \tag{4} \]

for almost all \( \gamma \in [0, 1] \), where \( \lambda = \nu + E[f_t] \) is the vector of the risk premia. In the CAPM, we have \( K = 1 \) and \( \nu = 0 \). When a factor \( f_{k,t} \) is a portfolio excess return, we also have \( \nu_k = 0, k = 1, \ldots, K \).
Proposition 1 differs from CR Theorem 3 in terms of both the definition of asymptotic arbitrage opportunities and the derived asset pricing restriction. We prefer the definition underlying Assumption APR.5 since it extends more easily to the conditional setting (see Section 3). In Appendix 2 we show that Assumption APR.5 implies the no-arbitrage conditions in Assumptions A.1 i) and ii) of CR, for a.e. conditional economy given \( \mathcal{F}_0 \) with a countable collection of assets. Then, by applying CR Theorem 3 to each such conditional economy, we deduce that \( \sum_{i=1}^{\infty} \left( a(\gamma_i) - b(\gamma_i) \nu \right)^2 < \infty \) for some \( \nu \in \mathbb{R}^K \), for a.e. sequence \( (\gamma_i) \). Finally, by assuming that the asset characteristics are defined by the underlying function \( \beta = (a, b) \nu \) defined on \([0, 1]\), we transform this summability condition into the equivalent restriction (3). Proposition 1 characterizes the functions \( \beta \) that are compatible with absence of asymptotic arbitrage opportunities. The restriction in Proposition 1 is testable with large equity datasets and large sample sizes (Section 2.5), and therefore is not affected by the Shanken (1982) critique. The next section describes how we get the data from sampling the continuum of assets.

2.2 The sampling scheme

We estimate the risk premia from a sample of observations on returns and factors for \( n \) assets and \( T \) dates. In available databases, asset returns are not observed for all firms at all dates. We account for the unbalanced nature of the panel through a collection of indicator variables \( I(\gamma), \gamma \in [0, 1] \), and define \( I_t(\gamma, \omega) = I[\gamma, S^t(\omega)] \). Then \( I_t(\gamma) = 1 \) if the return of asset \( \gamma \) is observable by the econometrician at date \( t \), and 0 otherwise (Connor and Korajczyk (1987)). To ease exposition and to keep the factor structure linear, we assume a missing-at-random design (Rubin (1976), Heckman (1979)), that is, independence between unobservability and returns generation.

**Assumption SC.1** The random variables \( I_t(\gamma), \gamma \in [0, 1] \), are independent of \( \epsilon_t(\gamma), \gamma \in [0, 1] \), and \( f_t \).

Another design would require an explicit modeling of the link between the unobservability mechanism and the continuum of assets; this would yield a nonlinear factor structure.

Assets are randomly drawn from the population according to a probability distribution \( G \) on \([0, 1]\). We use a single distribution \( G \) in order to avoid the notational burden when working with different distributions on different subintervals of \([0, 1]\).
Assumption SC.2 The random variables \( \gamma_i, i = 1, \ldots, n \), are i.i.d. indices, independent of \( \varepsilon_t(\gamma), I_t(\gamma) \), \( \gamma \in [0, 1] \) and \( f_t \), each with continuous distribution \( G \) with support \([0, 1] \).

For any \( n, T \in \mathbb{N} \), the excess returns are \( R_{i,t} = R_t(\gamma_i) \) and the observability indicators are \( I_{i,t} = I_t(\gamma_i) \), for \( i = 1, \ldots, n \), and \( t = 1, \ldots, T \). The excess return \( R_{i,t} \) is observed if and only if \( I_{i,t} = 1 \). Similarly, let \( \beta_i = \beta(\gamma_i) = (a_i, b_i)' \) be the characteristics, \( \varepsilon_{i,t} = \varepsilon_t(\gamma_i) \) the error terms and \( \sigma_{ij,t} = E[\varepsilon_{i,t}^2|x_L, \gamma_i, \gamma_j] \) the conditional variances and covariances of the assets in the sample, where \( x_L = \{x_t, x_{t-1}, \ldots\} \). By random sampling, we get a random coefficient panel model (e.g. Wooldridge (2002)). The characteristic \( \beta_i \) of asset \( i \) is random, and potentially correlated with the error terms \( \varepsilon_{i,t} \) and the observability indicators \( I_{i,t} \), as well as the conditional variances \( \sigma_{ii,t} \), through the index \( \gamma_i \). Random elements \( (\beta_i, \sigma_{ii,t}, \varepsilon_{i,t}, I_{i,t})' \), \( i = 1, \ldots, n \), are exchangeable (Andrews (2005)). Hence, assets randomly drawn from the population have ex-ante the same features. However, given a specific realization of the indices in the sample, assets have ex-post heterogeneous features.

2.3 Asymptotic properties of risk premium estimation

We consider a two-pass approach (Fama and MacBeth (1973), Black, Jensen and Scholes (1972)) building on Equations (1) and (3).

First Pass: The first pass consists in computing time-series OLS estimators \( \hat{\beta}_i = (\hat{a}_i, \hat{b}_i)' = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_{t} I_{i,t} x_t R_{i,t}, \) for \( i = 1, \ldots, n \), where \( \hat{Q}_{x,i} = \frac{1}{T_i} \sum_{t} I_{i,t} x_t x_t' \) and \( T_i = \sum_{t} I_{i,t} \). In available panels the random sample size \( T_i \) for asset \( i \) can be small, and the inversion of matrix \( \hat{Q}_{x,i} \) can be numerically unstable. This can yield unreliable estimates of \( \beta_i \). To address this, we introduce a trimming device: \( 1^X = 1 \{ CN (\hat{Q}_{x,i}) \leq \chi_{1,T}, \tau_{i,T} \leq \chi_{2,T} \} \), where \( CN (\hat{Q}_{x,i}) = \sqrt{eig_{\max} (\hat{Q}_{x,i}) / eig_{\min} (\hat{Q}_{x,i})} \) denotes the condition number of matrix \( \hat{Q}_{x,i} \), \( \tau_{i,T} = T/T_i \), and the two sequences \( \chi_{1,T} > 0 \) and \( \chi_{2,T} > 0 \) diverge asymptotically. The first trimming condition \( \{ CN (\hat{Q}_{x,i}) \leq \chi_{1,T} \} \) keeps in the cross-section only assets for which the time series regression is not too badly conditioned. A too large value of \( CN (\hat{Q}_{x,i}) = 1/CN (\hat{Q}_{x,i}^{-1}) \) indicates multicollinearity problems and ill-conditioning (Belsley, Kuh, and Welsch (2004), Greene (2008)). The second trimming condition \( \{ \tau_{i,T} \leq \chi_{2,T} \} \) keeps in the cross-section only assets for which the time series is not too short.
Second Pass: The second pass consists in computing a cross-sectional estimator of \( \nu \) by regressing the \( \hat{a}_i \)'s on the \( \hat{b}_i \)'s keeping the non-trimmed assets only. We use a WLS approach. The weights are estimates of \( w_i = v_i^{-1} \), where the \( v_i \) are the asymptotic variances of the standardized errors \( \sqrt{T} \left( \hat{a}_i - \hat{b}_i \nu \right) \) in the cross-sectional regression for large \( T \). We have \( v_i = \tau_i c'_i Q^{-1}_x S_{ii} Q^{-1}_x c_i \), where \( Q_x = E \left[ x'_i x'_i \right] \), \( S_{ii} = \text{plim} \frac{1}{T} \sum_{t} \sigma_{ii,t} x'_i x'_t = E \left[ \varepsilon^2_{i,t} x'_i x'_t | \gamma_i \right] \), \( \tau_i = \text{plim} \tau_{i,T} = E \left[ I_{i,t} | \gamma_i \right]^{-1} \), and \( c_i = (1, -\nu')' \). We use the estimates \( \hat{v}_i = \tau_i c'_i Q^{-1}_x \hat{S}_{ii} Q^{-1}_x c_i \), where \( \hat{S}_{ii} = \frac{1}{T_i} \sum_{t} I_{i,t} \hat{\varepsilon}^2_{i,t} x'_i x'_t \), \( \hat{\varepsilon}_{i,t} = R_{i,t} - \hat{b}'_i x_t \) and \( \hat{c}_{i} = (1, -\hat{\nu}')' \). To estimate \( \nu \), we use the OLS estimator \( \hat{\nu}_1 = \left( \sum_i 1^T \hat{b}_i \hat{b}_i' \right)^{-1} \sum_i 1^T \hat{b}_i \hat{a}_i \), i.e., a first-step estimator with unit weights. The WLS estimator is:

\[
\hat{\nu} = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i \hat{a}_i,
\]

where \( \hat{Q}_b = \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i \hat{b}_i' \) and \( \hat{w}_i = \hat{1}_i \hat{\nu}_i^{-1} \). Weighting accounts for the statistical precision of the first-pass estimates. Under conditional homoskedasticity \( \sigma_{ii,t} = \sigma_{ii} \) and a balanced panel \( \tau_{i,T} = 1 \), we have \( v_i = c'_i Q^{-1}_x c_i \sigma_{ii} \). There, \( v_i \) is directly proportional to \( \sigma_{ii} \), and we can simply pick the weights as \( \hat{w}_i = \hat{\sigma}_{ii}^{-1} \), where \( \hat{\sigma}_{ii} = \frac{1}{T} \sum_t \hat{\varepsilon}^2_{i,t} \) (Shanken (1992)). The final estimator of the risk premia is

\[
\hat{\lambda} = \hat{\nu} + \frac{1}{T} \sum_t f_t.
\]

Starting from the asset pricing restriction (4), another estimator of \( \lambda \) is \( \hat{\lambda} = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i \hat{R}_i \), where \( \hat{R}_i = \frac{1}{T_i} \sum_{t} I_{i,t} R_{i,t} \). This estimator is numerically equivalent to \( \hat{\lambda} \) in the balanced case, where \( I_{i,t} = 1 \) for all \( i \) and \( t \). In the general unbalanced case, it is equal to \( \hat{\lambda} = \hat{\nu} + \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i \hat{f}_i \), where \( \hat{f}_i = \frac{1}{T_i} \sum_{t} I_{i,t} f_t \). Estimator \( \hat{\lambda} \) is often studied by the literature (see, e.g., Shanken (1992), Kandel and Stambaugh (1995), Jagannathan and Wang (1998)), and is also consistent. Estimating \( E \left[ f_i \right] \) 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 2.4). This explains our preference for \( \hat{\lambda} \) over \( \bar{\lambda} \).

We derive the asymptotic properties under assumptions on the conditional distribution of the error terms.
Assumption A.1 There exists a positive constant $M$ such that for all $n$:

a) $E \left[ \varepsilon_{i,t} | \{ \varepsilon_{j,t-1}, \gamma_j, j = 1, \ldots, n \} \right] = 0$, with $\varepsilon_{i,t-1} = \{ \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots \}$ and $x_t = \{ x_t, x_{t-1}, \ldots \}$;

b) $\sigma_{i,t} \leq M$, $i = 1, \ldots, n$;

c) $E \left[ \frac{1}{n} \sum_{i,j} E \left[ \sigma_{i,j,t}^2 | \gamma_i, \gamma_j \right] \right]^{1/2} \leq M$, where $\sigma_{i,j,t} = E \left[ \varepsilon_{i,t} \varepsilon_{j,t} | x_L, \gamma_i, \gamma_j \right]$.

Assumption A.1 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)). More general error structures are possible but complicate consistent estimation of the asymptotic variances of the estimators (see Section 2.4).

Proposition 2 summarizes consistency of estimators $\hat{\nu}$ 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 2 Under Assumptions APR.1-APR.5, SC.1-SC.2, A.1 and C.1a), C.2-C.5, we get $a) \| \hat{\nu} - \nu \| = o_p (1)$ and $b) \| \hat{\lambda} - \lambda \| = o_p (1)$, when $n, T \to \infty$ such that $n \asymp T^{\gamma}$ for $\gamma > 0$.

The conditions in Proposition 2 allow for large w.r.t. $T$ (short panel asymptotics) when $\gamma > 1$. Shanken (1992) shows consistency of $\hat{\nu}$ and $\hat{\lambda}$ for a fixed $n$ and $T \to \infty$. This consistency does not imply Proposition 2. Shanken (1992) (see also Litzenberger and Ramaswamy (1979)) further shows that we can estimate $\nu$ consistently in the second pass with a modified cross-sectional estimator for a fixed $T$ and $n \to \infty$. Since $\lambda = \nu + E \left[ f_t \right]$, consistent estimation of the risk premia themselves is impossible for a fixed $T$ (see Shanken (1992) for the same point).

Proposition 3 below gives the large-sample distributions under the double asymptotics $n, T \to \infty$. Let us define $\tau_{i,j,T} = T / T_{i,j}$, where $T_{i,j} = \sum_t I_{i,j,t}$ and $I_{i,j,t} = I_{i,t} I_{j,t}$ for $i, j = 1, \ldots, n$. Let us further define $\tau_{ij} = \text{plim}_{T \to \infty} \tau_{i,j,T} = E \left[ I_{i,j,t} | \gamma_i, \gamma_j \right]^{-1}$, $S_{ij} = \text{plim}_{T \to \infty} \frac{1}{T} \sum_t \sigma_{i,j,t} x_t x_t' = E \left[ \varepsilon_{i,t} \varepsilon_{j,t} | x_t, x_t' \right]$ and $Q_b = \text{plim}_{n \to \infty} \frac{1}{n} \sum_{i,j} w_i b_i b_i' = E \left[ w_i b_i b_i' \right]$. 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 (see Appendix 4).
Assumption A.2 As \( n, T \to \infty \) such that \( n \asymp T^\gamma \) for \( \gamma \in \Gamma_1 \subset \mathbb{R}^+ \), a) \( \frac{1}{\sqrt{n}} \sum_i w_i \tau_i (Y_{i,T} \otimes b_i) \Rightarrow N (0, S_b) \), where \( Y_{i,T} = \frac{1}{\sqrt{T}} \sum_t I_{i,t} x_{i,t} \) and \( S_b = \lim_{n \to \infty} E \left[ \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} S_{ij} \otimes b_i b_j' \right] \)

\[
= \text{plim}_{n \to \infty} \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} S_{ij} \otimes b_i b_j' \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} \text{Cov} (f_t, f_s) .
\]

Proposition 3 Under Assumptions APR.1-APR.5, SC.1-SC.2, A.1-A.2, and C.1(a), C.2-C.5, we get:

a) \( \sqrt{nT} \left( \hat{\nu} - \nu - \frac{1}{T} \hat{B}_\nu \right) \Rightarrow N (0, \Sigma_\nu) \) , where \( \Sigma_\nu = Q_b^{-1} \lim_{n \to \infty} E \left[ \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} (c_\nu Q_x^{-1} S_{ij} Q_x^{-1} c_\nu) b_i b_j' \right] Q_b^{-1} \) and the bias term is \( \hat{B}_\nu = \hat{Q}_b^{-1} \left( \frac{1}{n} \sum_i \hat{w_i} \tau_i T \hat{E}_2 \hat{Q}_b^{-1} \hat{S}_{i,i} \hat{Q}_b^{-1} \hat{c}_\nu \right) \), with \( E_2 = (0 : Id_K)' \) and \( c_\nu = (1, -\nu') ' \);

b) \( \sqrt{T} (\hat{\lambda} - \lambda) \Rightarrow N (0, \Sigma_f) \) when \( n, T \to \infty \) such that \( n \asymp T^\gamma \) for \( \gamma \in \Gamma_1 \cap (0, 3) .
\]

The asymptotic variance matrix in Proposition 3 can be rewritten as:

\[
\Sigma_\nu = \text{plim}_{n \to \infty} \left( \frac{1}{n} B_n' W_n B_n \right)^{-1} \frac{1}{n} B_n' W_n V_n W_n B_n \left( \frac{1}{n} B_n' W_n B_n \right)^{-1}
\]

where \( B_n = (b_1, \ldots, b_n)' \), \( W_n = \text{diag}(w_1, \ldots, w_n) \) and \( V_n = [v_{ij}]_{i,j=1,\ldots,n} \) with \( v_{ij} = \frac{\tau_i \tau_j}{\tau_{ij}} c_\nu Q_x^{-1} S_{ij} Q_x^{-1} c_\nu \), which gives \( v_{ii} = v_i \). In the homoskedastic and balanced case, we have \( c_\nu Q_x^{-1} c_\nu = 1 + \lambda' V[f_t]^{-1} \lambda \) and \( V_n = (1 + \lambda' V[f_t]^{-1} \lambda) \Sigma_{\epsilon,n} \), with \( \Sigma_{\epsilon,n} = [\sigma_{ij}]_{i,j=1,\ldots,n} \). Then, the asymptotic variance of \( \hat{\nu} \) reduces to \( \text{plim}_{n \to \infty} \left( 1 + \lambda' V[f_t]^{-1} \lambda \right) \left( \frac{1}{n} B_n' W_n B_n \right)^{-1} \frac{1}{n} B_n' W_n \Sigma_{\epsilon,n} W_n B_n \left( \frac{1}{n} B_n' W_n B_n \right)^{-1} \). In particular, in the CAPM we have \( K = 1 \) and \( \nu = 0 \), which implies that \( \sqrt{\frac{\lambda^2}{V[f_t]}} \) is equal to the slope of the Capital Market Line \( \sqrt{\frac{E[f_t]^2}{V[f_t]}} \), i.e., the Sharpe Ratio of the market portfolio.

Proposition 3 shows that the estimator \( \hat{\nu} \) has a fast convergence rate \( \sqrt{nT} \) and features an asymptotic bias term. Both \( \hat{a}_i \) and \( \hat{b}_i \) in the definition of \( \hat{\nu} \) contain an estimation error; for \( \hat{b}_i \), this is the well-known Error-In-Variable (EIV) problem. The EIV problem does not impede consistency since we let \( T \) grow to infinity. However, it induces the bias term \( \hat{B}_\nu/T \) which centers the asymptotic distribution of \( \hat{\nu} \). We have \( \Gamma_1 = \mathbb{R}^+ \) in Assumption A.2, when \( (\epsilon_{i,t}) \) and \( (x_t) \) are i.i.d. across time and errors \( (\epsilon_{i,t}) \) feature a cross-sectional block dependence structure (see Appendix 4). Then, the upper bound on the relative expansion
rates of $n$ and $T$ is $n = o(T^3)$. The control of first-pass estimation errors uniformly across assets requires that the cross-section dimension $n$ should not be too large w.r.t. the time series dimension $T$.

If we knew the true factor mean, for example $E[f_t] = 0$, and did not need to estimate it, the estimator $\hat{\nu} + E[f_t]$ of the risk premia would have the same fast rate $\sqrt{nT}$ as the estimator of $\nu$, and would inherit its asymptotic distribution. Since we do not know the true factor mean, the asymptotic distribution of $\hat{\lambda}$ is driven only by the variability of the factor since the convergence rate $\sqrt{T}$ of the sample average $\frac{1}{T} \sum_t f_t$ dominates the convergence rate $\sqrt{nT}$ 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 $\nu$. 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 $\Sigma_{\lambda,n} = \Sigma_f + (1 + \lambda'V[f_t]^{-1}\lambda)\left(\frac{1}{n}B_n'W_nB_n\right)^{-1} \left(\frac{1}{n^2}B_n'W_n\Sigma_{\epsilon,n}W_nB_n\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}\right)$, for fixed $n$ and $T \to \infty$. The two components in $\Sigma_{\lambda,n}$ come from estimation of $E[f_t]$ and $\nu$, respectively. In the heteroskedastic setting with fixed $n$, a slight extension of Theorem 1 in Jagannathan and Wang (1998), or Theorem 3.2 in Jagannathan, Skoulikis, and Wang (2009), to the unbalanced case yields $\Sigma_{\lambda,n} = \Sigma_f + \left(\frac{1}{n}B_n'W_nB_n\right)^{-1} \left(\frac{1}{n^2}B_n'W_nW_nB_n\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}\right)$. Letting $n \to \infty$ gives $\Sigma_f$ 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. Using a large number of assets instead of a small number of portfolios does help to eliminate the EIV contribution.

Proposition 3 suggests exploiting the analytical bias correction $\hat{B}_\nu / T$ and using $\hat{\nu}_B = \hat{\nu} - \frac{1}{T} \hat{B}_\nu$ instead of $\hat{\nu}$. Furthermore, $\hat{\lambda}_B = \hat{\nu}_B + \frac{1}{T} \sum_t f_t$ delivers a bias-free estimator of $\lambda$ at order $1/T$, which shares the same root-$T$ asymptotic distribution as $\hat{\lambda}$.

Finally, we can relate the results of Proposition 3 to bias-corrected estimation accounting for the well-known incidental parameter problem of the panel literature (Neyman and Scott (1948), see Lancaster (2000) for a review). Model (1) under restriction (3) can be written as $R_{i,t} = b'_i (f_t + \nu) + \varepsilon_{i,t}$. In the likelihood setting of Hahn and Newey (2004) (see also Hahn and Kuersteiner (2002)), the $b_i$ correspond to the individual effects and $\nu$ to the common parameter of interest. Available results tell us: (i) the estimator of $\nu$ is inconsistent if $n$ goes to infinity while $T$ is held fixed; (ii) the estimator of $\nu$ is asymptotically biased even if $T$ grows at the same rate as $n$; (iii) an analytical bias correction may yield an estimator of $\nu$ that is root-$(nT)$
asymptotically normal and centered at the truth if $T$ grows faster than $n^{1/3}$. The two-pass estimators $\nu$ and $\nu_B$ exhibits the properties (i)-(iii) as expected by analogy with unbiased estimation in large panels.

### 2.4 Confidence intervals

We can use Proposition 3 to build confidence intervals by means of consistent estimation of the asymptotic variances. 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 traded. We estimate $\Sigma_f$ 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 $\hat{\lambda}$ is straightforward. On the contrary, getting a HAC estimator for $\hat{\Sigma}_f$ appearing in the asymptotic distribution of $\hat{\lambda}$ is not obvious in the unbalanced case.

The construction of confidence intervals for the components of $\hat{\nu}$ is more difficult. Indeed, $\Sigma_{\nu}$ 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 $\hat{S}_{ij} = \frac{1}{T_{ij}} \sum_t I_{ij,t} \hat{\epsilon}_{i,t} \hat{\epsilon}_{j,t} x_t x_t' W_t$, 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)). Fan, Liao, and Mincheva (2011) have recently focused on the estimation of $E[\epsilon_t' \epsilon_t]$ in large balanced panel.

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 $S_{ij}$. In the following assumption we use the notion of sparsity suggested by Bickel and Levina (2008) adapted to our framework with random coefficients.

**Assumption A.3** There exist constants $q, \delta \in [0, 1)$ such that $\max_i \sum_j \|S_{ij}\|^q = O_p\left(n^\delta\right)$.

Assumption A.3 tells us that most cross-asset contributions $\|S_{ij}\|$ can be neglected. As sparsity increases, we can choose coefficients $q$ and $\delta$ closer to zero. Assumption A.3 does not impose sparsity of the covariance matrix of the returns themselves. Assumption A.1 c) is also a sparsity condition, which ensures that the limit
matrix $\Sigma_\nu$ is well-defined when combined with Assumption C.3. Both sparsity assumptions are satisfied under weak cross-sectional dependence between the error terms, for instance, under a block dependence structure (see Appendix 4).

As in Bickel and Levina (2008), let us introduce the thresholded estimator $\tilde{S}_{ij} = \hat{S}_{ij} 1 \{ \| \hat{S}_{ij} \| \geq \kappa \}$ of $S_{ij}$, which we refer to as $\hat{S}_{ij}$ thresholded at $\kappa = \kappa_{n,T}$. 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 4** Under Assumptions APR.1-APR.5, SC.1-SC.2, A.1-A.3, C.1-C.5, we have

$$\tilde{\Sigma}_\nu^{-1/2} \sqrt{nT} \left( \hat{\nu} - \frac{1}{T} \hat{B}_\nu - \nu \right) \Rightarrow N \left( 0, \text{Id}_K \right),$$

where $\tilde{\Sigma}_\nu = \hat{Q}_b^{-1} \sum_{i,j} \hat{w}_i \hat{w}_j (\tau_{ij}, T) (c^*_i \hat{Q}_x^{-1} \hat{S}_{ij} \hat{Q}_x^{-1} \tau_{ij}, T) \hat{b}_i \hat{b}_j^{\prime} \hat{Q}_b^{-1}$, when $n, T \to \infty$ such that $n \propto T^\gamma$ for $\gamma \in \Gamma_1 \cap \left( 0, \min \left\{ 1 + \eta, \frac{1 - q}{2\delta} \right\} \right)$, and $\kappa = M \sqrt{\frac{\log n}{T^\eta}}$ for a constant $M$ and $\eta \in (0, 1]$ as in Assumption C.1.

Constant $\eta \in (0, 1]$ is defined in Assumption C.1 and is related to the time series dependence of processes $(\varepsilon_{i,t})$ and $(x_t)$. We have $\eta = 1$, when $(\varepsilon_{i,t})$ and $(x_t)$ are serially i.i.d. as in Appendix 4 and Bickel and Levina (2008). The matrix made of thresholded blocks $\tilde{S}_{ij}$ is not guaranteed to be semi definite positive (sdp). However, we expect that the double summation on $i$ and $j$ makes $\tilde{\Sigma}_{\nu}$ sdp in empirical applications. In case it is not, El Karoui (2008) discusses a few solutions based on shrinkage.

### 2.5 Tests of asset pricing restrictions

The null hypothesis underlying the asset pricing restriction (3) is

$$H_0 : \text{there exists } \nu \in \mathbb{R}^K \text{ such that } a(\gamma) = b(\gamma)'\nu, \text{ for almost all } \gamma \in [0, 1].$$

Under $H_0$, we have $E_G \left[ (a_i - b_i'\nu)^2 \right] = 0$. Since $\nu$ is estimated via the WLS cross-sectional regression of the estimates $\hat{a}_i$ on the estimates $\hat{b}_i$, we suggest a test based on the weighted sum of squared residuals SSR of the cross-sectional regression. The weighted SSR is $\hat{Q}_e = \frac{1}{n} \sum_i \hat{w}_i \hat{e}_i^2$, with $\hat{e}_i = c_i^* \hat{\beta}_i$, which is an empirical counterpart of $E_G \left[ w_i (a_i - b_i'\nu)^2 \right]$. Let us define $S_{ii,T} = \frac{1}{T} \sum_t I_{i,t} \sigma_{ii,t} x_{i,t} x_{i,t}'$, and introduce the commutation matrix $W_{m,n}$ of order $mn \times mn$ such that $W_{m,n} \mathrm{vec}[A] = \mathrm{vec}[A']$ for any matrix $A \in \mathbb{R}^{m \times n}$, where the vector operator $\mathrm{vec}[\cdot]$ stacks the
elements of an \( m \times n \) matrix as a \( mn \times 1 \) vector. If \( m = n \), we write \( W_n \) instead of \( W_{n,n} \). For two \((K+1)\times(K+1)\) matrices \( A \) and \( B \), equality \( W_{(K+1)}(A \otimes B) = (B \otimes A)W_{(K+1)} \) also holds (see Chapter 3 of Magnus and Neudecker (2007) for other properties).

**Assumption A.4** For \( n, T \to \infty \) such that \( n \asymp T^{\bar{\gamma}} \) for \( \bar{\gamma} \in \Gamma_2 \subset \Gamma_1 \), we have

\[
\frac{1}{\sqrt{n}} \sum_i w_i \tau_i^2 (Y_{i,T} \otimes Y_{i,T} - \text{vec}[S_{ii,T}]) = N(0, \Omega),
\]

where the asymptotic variance matrix is:

\[
\Omega = \lim_{n \to \infty} E \left[ \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i^2 \tau_j^2}{\tau_i^2} \left[ S_{ij} \otimes S_{ij} + (S_{ij} \otimes S_{ij})W_{(K+1)} \right] \right] = \text{plim} \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i^2 \tau_j^2}{\tau_i^2} \left[ S_{ij} \otimes S_{ij} + (S_{ij} \otimes S_{ij})W_{(K+1)} \right].
\]

Assumption A.4 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 Appendix A.x that Assumption A.4 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 Appendix A.x that Assumption A.4 holds under a cross-sectional block dependence structure for the errors. Intuitively, the expression of the variance-covariance matrix \( \Omega \) is related to the result that, for random \((K+1) \times 1\) vectors \( Y_1 \) and \( Y_2 \) which are jointly normal with covariance matrix \( S \), we have \( \text{Cov}(Y_1 \otimes Y_1, Y_2 \otimes Y_2) = S \otimes S + (S \otimes S)W_{(K+1)} \).

Let us now introduce the following statistic \( \hat{\xi}_{n,T} = T^{1/2} \left( \hat{Q}_e - \frac{1}{T} \hat{B}_\xi \right) \), where the centering term simplifies to \( \hat{B}_\xi = 1 \) thanks to the weighting scheme. Under the null hypothesis \( \mathcal{H}_0 \), we prove that

\[
\hat{\xi}_{n,T} \Rightarrow N(0, \Sigma_\xi), \quad \text{where} \quad \Sigma_\xi = 2 \lim_{n \to \infty} E \left[ \frac{1}{n} \sum_{i,j} w_i w_j v_i^2 \right] = 2 \text{plim} \frac{1}{n} \sum_{i,j} w_i w_j v_i^2 \text{ as } n, T \to \infty \text{ (see Appendix A.2.5).}
\]

Then a feasible testing procedure exploits the consistent estimator \( \hat{\Sigma}_\xi = 2 \sum_{i,j} \hat{w}_i \hat{w}_j \hat{v}_{ij}^2 \)

of the asymptotic variance \( \Sigma_\xi \), where \( \hat{v}_{ij} = \frac{\tau_i \tau_j}{\tau_i \tau_j} \left( \hat{Q}_e - \hat{S}_{ij} \right) \hat{Q}_e^{-1} \).

**Proposition 5** Under \( \mathcal{H}_0 \) and Assumptions APR.1-APR.5, SC.1-SC.2, A.1-A.4 and C.1-C.5, we have

\[
\hat{\Sigma}_\xi^{-1/2} \hat{\xi}_{n,T} \Rightarrow N(0, 1), \text{ as } n, T \to \infty \text{ such that } n \asymp T^{\bar{\gamma}} \text{ for } \bar{\gamma} \in \Gamma_2 \cap \left( 0, \min \left\{ 2\eta, \eta - \frac{q}{2\delta} \right\} \right).
\]

In the homoskedastic case, the asymptotic variance of \( \hat{\xi}_{n,T} \) reduces to \( \Sigma_\xi = 2 \lim_{n \to \infty} \frac{1}{n} \sum_{i,j} \tau_i \tau_j \sigma_{ij}^2 \).

For fixed \( n \), we can rely on the test statistic \( T \hat{Q}_e \), which is asymptotically distributed as

\[
\frac{1}{n} \sum_{j} e_{ij} g_{ij} \chi_j^2
\]
for \( j = 1, \ldots, (n - K) \), where the \( \chi_j^2 \) are i.i.d. chi-square variables with 1 degree of freedom, and the coefficients \( e_i g_j \) are the non-zero eigenvalues of matrix \( V_n^{1/2} (W_n - W_n B_n (B_n' W_n B_n)^{-1} B_n' W_n) V_n^{1/2} \) (see Kan et al. (2009)). By letting \( n \) grow, the sum of chi-square variables converges to a Gaussian variable after recentering and rescaling, which yields heuristically the result of Proposition 5.

The alternative hypothesis is

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

Let us define the pseudo-true value \( \nu_\infty = \arg \inf_{\nu \in \mathbb{R}^K} Q_w^w(\nu) \), where \( Q_w^w(\nu) = E_G \left[ w_i (a_i - b_i' \nu)^2 \right] \) (White (1982), Gourieroux, Monfort and Trognon (1984)) and population errors \( e_i = a_i - b_i' \nu_\infty = d_i' \beta_i \), \( i = 1, \ldots, n \), 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 6** Under \( \mathcal{H}_1 \) and Assumptions APR.1-APR.5, SC.1-SC.2, A.1-A.4 and C.1-C.5, we have \( \hat{\xi}_{nT} \overset{p}{\to} +\infty \), and \( \sqrt{n} (\hat{\nu} - \nu_\infty) \Rightarrow N(0, \Sigma_{\nu_\infty}) \), where \( \Sigma_{\nu_\infty} = Q_b^{-1} E_G [w_i^2 e_i^2 b_i' b_i] Q_b^{-1} \) and \( \sqrt{T} (\hat{\lambda} - \lambda_\infty) \Rightarrow N(0, \Sigma_f) \), and \( \lambda_\infty = \nu_\infty + E [f_i] \), as \( n, T \to \infty \) such that \( n \asymp T^{\bar{\gamma}} \) for \( \bar{\gamma} \in \Gamma_2 \cap \left( 1, \min \left\{ 2\eta, \eta \frac{1 - q}{2\delta} \right\} \right) \).

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 \( b_i \) has no impact on the asymptotic distribution of \( \hat{\nu} \) since the bias induced by the EIV problem is of the order \( O(1/T) \), and \( \sqrt{n}/T = o(1) \). The lower bound 1 on rate \( \bar{\gamma} \) in Proposition 6 ensures that cross-sectional estimation of \( \nu \) has asymptotically no impact on the estimation of \( \lambda \).

To study the local asymptotic power, we can adopt the following local alternative:

\[
\mathcal{H}_{1,nT} : \inf_{\nu \in \mathbb{R}^K} Q_w^w(\nu) = \frac{\psi}{\sqrt{nT}} > 0, \text{ for a constant } \psi > 0. \]

Then we can show (see the supplementary materials) that \( \xi_{nT} \Rightarrow N(\psi, \Sigma_{\xi}) \), and the test is locally asymptotically powerful. Pesaran and Yamagata (2008) consider a similar local analysis for a test of slope homogeneity in large panels.
Finally, we can derive a test for the null hypothesis when the factors come from tradable assets, i.e., are portfolio excess returns:

\[ \mathcal{H}_0 : a(\gamma) = 0 \text{ for almost all } \gamma \in [0, 1] \iff E_G[a_t^2] = 0, \]

against the alternative hypothesis

\[ \mathcal{H}_1 : E_G[a_t^2] > 0. \]

We only have to substitute \( \hat{a}_i \) for \( \hat{e}_i \), and \( E_1 = (1, 0)' \) for \( c_0 \) in Proposition 5.

### 3 Conditional factor model

In this section we extend the setting of Section 2 to conditional specifications 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 Excess return generation and asset pricing restrictions

The following assumptions are the analogues of Assumptions APR.1 and APR.2, and Proposition 7 is the analogue of Proposition 1. Its proof derives easily from the steps in A.2.1 for the unconditional case.

**Assumption APR.6** The excess returns \( R_t(\gamma) \) of asset \( \gamma \in [0, 1] \) at date \( t = 1, 2, \ldots \) satisfy the conditional linear factor model:

\[ R_t(\gamma) = a_t(\gamma) + b_t(\gamma)' f_t + \varepsilon_t(\gamma), \quad (7) \]

where \( a_t(\gamma) \) and \( b_t(\gamma) \) are \( \mathcal{F}_{t-1} \)-measurable scalar and vector functions defined on \( [0, 1] \).

**Assumption APR.7** The matrix \( \int b_t(\gamma)b_t(\gamma)'d\gamma \) is positive definite, \( P \)-a.s., for any date \( t = 1, 2, \ldots \).
Proposition 7  Under Assumptions APR.3-APR.7, there exists a unique vector $\nu_t \in \mathbb{R}^K$ at date $t = 1, 2, \ldots$ such that $\nu_t$ is $\mathcal{F}_{t-1}$-measurable and:
\[
a_t(\gamma) = b_t(\gamma)'\nu_t,
\]
for almost all $\gamma \in [0, 1]$.

The asset pricing restriction in Proposition 7 can be rewritten as
\[
E[R_t(\gamma)|\mathcal{F}_{t-1}] = b_t(\gamma)'\lambda_t,
\]
for almost all $\gamma \in [0, 1]$, where $\lambda_t = \nu_t + E[f_t|\mathcal{F}_{t-1}]$ is the vector of the conditional risk premia.

The sampling scheme is the same as in Section 2.2, and we use the same type of notations, for example $b_{i,t} = b_t(\gamma_i)$. To have a workable version of (7), we further specify the conditioning information and how coefficients depend on it. The conditioning information is generated by lagged instruments $Z_{t-1} \in \mathbb{R}^p$ and $Z_{i,t-1} \in \mathbb{R}^q$. To end up with a linear regression model we specify that the vector of factor sensitivities $b_{i,t}$ is a linear function of lagged instruments $Z_{t-1}$ (Shanken (1990), Ferson and Harvey (1991)) and $Z_{i,t-1}$ (Avramov and Chordia (2006)): $b_{i,t} = B_t Z_{t-1} + C_i Z_{i,t-1}$, where $B_t \in \mathbb{R}^{K \times p}$ and $C_i \in \mathbb{R}^{K \times q}$, for any $i, t$. The lagged instruments $Z_{t-1}$ are common to all stocks. They may include the constant and past observations of the factors and some additional variables such as macroeconomic variables. The lagged instruments $Z_{i,t-1}$ are specific to stock $i$. They may include past observations of firm characteristics and stock returns. The lagged instruments may include powers to account for possible nonlinearities. 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 $\Lambda \in \mathbb{R}^{K \times p}$, for any $t$. Furthermore, we assume that the conditional expectation of $Z_t$ given the past $Z_{t-1}$ depends on $Z_{t-1}$ only and is linear, as, for instance, in a Vector Autoregressive (VAR) model of order 1. Then $E[f_t|Z_{t-1}] = F Z_{t-1}$, where $F \in \mathbb{R}^{K \times p}$, for any $t$. Under these functional specifications the asset pricing restriction linking the expected excess returns and the factor sensitivities
\[
E[R_{i,t} | \{Z_{i,t-1} : i = 1, \ldots, n\} , Z_{t-1}] = b'_{i,t} \lambda_t,
\]
implies that the intercept $a_{i,t}$ is a quadratic form in lagged instruments $Z_{t-1}$ and $Z_{i,t-1}$, that is, $a_{i,t} = Z_{t-1}' B_{i,t} (\Lambda - F) Z_{t-1} + Z_{i,t-1}' C_{i,t} (\Lambda - F) Z_{i,t-1}$. This shows that assuming a priori linearity of $a_{i,t}$
in the lagged instruments $Z_{t-1}$ and $Z_{i,t-1}$ is in general not compatible with linearity of $b_{i,t}$ and $E [f_t | Z_{t-1}]$.

The conditional factor model (7) then becomes

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

(11)

which is nonlinear in the parameters $\Lambda, F, B_t,$ and $C_t$. In order to implement the two-pass methodology in a conditional context it is useful to rewrite model (11) as a model that is linear in transformed parameters and new regressors. The regressors include $x_{2,i,t} = \left( f_t' \otimes Z_{t-1}', f_t' \otimes Z_{i,t-1}' \right)' \in \mathbb{R}^{d_2}$ with $d_2 = K(p + q)$. The first components with common instruments take the interpretation of scaled factors, while the second components do not since they depend on $i$. The regressors also include the predetermined variables $x_{1,i,t} = \left( vech \left[ X_t \right]', vec \left[ X_t \right] \right)' \in \mathbb{R}^{d_1}$ with $d_1 = p(p + 1)/2 + pq$, where the symmetric matrix $X_t = \left[ X_{t,k,l} \right] \in \mathbb{R}^{p \times p}$ is such that $X_{t,k,l} = Z_{t-1,k}^2$, if $k = l$, and $X_{t,k,l} = 2 Z_{t-1,k} Z_{t-1,l}$, otherwise, $k, l = 1, \ldots, p$, and the matrix $X_{i,t} = Z_{t-1} Z_{i,t-1} \in \mathbb{R}^{p \times q}$. The vector-half operator $vech \left[ \cdot \right]$ stacks the lower elements of a $p \times p$ matrix as a $p(p + 1)/2 \times 1$ 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 (11) as in (2) through appropriate redefinitions of the regressors and loadings (see Appendix 3):

$$R_{i,t} = \beta_i' x_{i,t} + \varepsilon_{i,t},$$

(12)

where $x_{i,t} = \left( x_{1,i,t}', x_{2,i,t}' \right)'$ has dimension $d = d_1 + d_2$, and $\beta_i = \left( \beta_{1,i}', \beta_{2,i}' \right)'$ is such that

$$\beta_{1,i} = \Psi \beta_{2,i}, \quad \beta_{2,i} = \left( vec \left[ B_t \right]', vec \left[ C_t \right] \right)',$$

(13)

$$\Psi = \begin{pmatrix} \frac{1}{2} D_p^+ \left( (\Lambda - F)' \otimes I_p + I_p \otimes (\Lambda - F)' W_{p,K} \right) & 0 \\ 0 & (\Lambda - F)' \otimes I_q \end{pmatrix}.$$}

The matrix $D_p^+$ is the $p(p + 1)/2 \times p^2$ Moore-Penrose inverse of the duplication matrix $D_p$, such that $vech \left[ A \right] = D_p^+ vec \left[ A \right]$ for any $A \in \mathbb{R}^{p \times p}$ (see Chapter 3 in Magnus and Neudecker (2007)). When $Z_t = 1$ and $Z_{i,t} = 0$, we have $p = p(p + 1)/2 = 1$ and $q = 0$, and model (12) reduces to model (2).

In (13) the $d_1 \times 1$ vector $\beta_{1,i}$ is a linear transformation of the $d_2 \times 1$ vector $\beta_{2,i}$. This clarifies that the asset pricing restriction (10) implies a constraint on the coefficient distributions via their supports. The coefficients of this linear transformation depend on matrix $\Lambda - F$. For the purpose of estimating the loading
coefficients of the risk premia in matrix $\Lambda$, the parameter restrictions can be written as (see Appendix 3):

$$
\beta_{1,i} = \beta_{3,i} \nu, \quad \nu = \text{vec} \left[ \Lambda' - F' \right], \quad \beta_{3,i} = \left( \left[ D_p^+ \left( B'_i \otimes I_p \right) \right]' \left[ W_{p,q} \left( C'_i \otimes I_p \right) \right]' \right)'.$$

Furthermore, we can relate the $d_1 \times Kp$ matrix $\beta_{3,i}$ to the vector $\beta_{2,i}$ (see Appendix 3):

$$
\text{vec} \left[ \beta'_{3,i} \right] = J_a \beta_{2,i},
$$

where the $d_1pK \times d_2$ block-diagonal matrix of constants $J_a$ is given by

$$
J_a = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix}
$$

with diagonal blocks $J_{11} = W_{p(p+1)/2,pK} \left( I_K \otimes \left[ \left( I_p \otimes D_p^+ \right) \left( W_p \otimes I_p \right) \right] \right)$ and $J_{22} = W_{pq,pK} \left( I_K \otimes \left[ \left( I_p \otimes W_{p,q} \right) \left( W_{p,q} \otimes I_p \right) \right] \right)$. The link (15) 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 $J_a$ is equal to $I_K$. Equations (14) and (15) in the conditional setting are the counterparts of restriction (3) in the static setting.

### 3.2 Asymptotic properties of time-varying risk premium estimation

We consider a two-pass approach building on Equations (12) and (14).

**First Pass:** The first pass consists in computing time-series OLS estimators

$$
\hat{\beta}_i = \left( \hat{\beta}'_{1,i}, \hat{\beta}'_{2,i} \right)' = \hat{Q}_x^{-1} \frac{1}{T} \sum_t I_{i,t} x_{i,t} R_{i,t}, \quad \text{for } i = 1, \ldots, n, \quad \text{where } \hat{Q}_x = \frac{1}{T} \sum_t I_{i,t} x_{i,t} x_{i,t}'
$$

We use the same trimming device as in Section 2.

**Second Pass:** The second pass consists in computing a cross-sectional estimator of $\nu$ by regressing the $\hat{\beta}_{1,i}$ on the $\hat{\beta}_{3,i}$ keeping non-trimmed assets only. We use a WLS approach. The weights are estimates of $w_i = (\text{diag} \left[ \nu_i \right])^{-1}$, where $\nu_i$ are the asymptotic variances of the standardized errors $\sqrt{T} \left( \hat{\beta}_{1,i} - \beta_{3,i} \nu \right)$ in the cross-sectional regression for large $T$. We have

$$
\nu_i = \tau_i C'_i Q_{x,i}^{-1} S_{i,t} Q_{x,i}^{-1} C'_i, \quad \text{where } Q_{x,i} = E \left[ x_{i,t} x_{i,t}' | \gamma_i \right],
$$

$S_{i,t} = \text{plim} \frac{1}{T} \sum_t \sigma_{i,t} x_{i,t} x_{i,t}' = E \left[ \varepsilon_{i,t}^2 x_{i,t} x_{i,t}' | \gamma_i \right], \quad \sigma_{i,t} = E \left[ \varepsilon_{i,t}^2 | \gamma_i \right], \quad \text{and } C'_i = \left( E'_i - \left( I_{d_1} \otimes \nu' \right) J_a E'_2 \right)'$

with $E_1 = (I_{d_1}, 0_{d_1 \times d_2})'$, $E_2 = (0_{d_2 \times d_1}, I_{d_2})'$. We use the estimates

$$
\hat{\nu}_i = \tau_i J_{i,T} C_{\hat{\nu}_1} \hat{Q}_{x,i}^{-1} \hat{S}_{i,t} \hat{Q}_{x,i}^{-1} C_{\hat{\nu}_1}, \quad \text{where } \hat{S}_{i,t} = \frac{1}{T} \sum_t I_{i,t} \varepsilon_{i,t}^2 x_{i,t} x_{i,t}' = R_{i,t} - \hat{\beta}'_{2,i} x_{i,t} \quad \text{and } C_{\hat{\nu}_1} = \left( E'_1 - \left( I_{d_1} \otimes \hat{\nu}'_1 \right) J_a E'_2 \right)'.$$

To estimate $C_{\nu}$, we use the OLS estimator

$$
\hat{\nu}_1 = \left( \sum_i 1^X \beta_{3,i} \beta_{3,i} \right)^{-1} \sum_i 1^X \beta_{3,i} \beta_{3,i} \hat{\beta}_{1,i}, \quad \text{i.e., a first-step estimator with unit weights.}
$$
The WLS estimator is:
\[
\hat{\nu} = Q_{\beta_3}^{-1} \frac{1}{n} \sum_i \hat{\beta}_{3,i} \hat{\nu}_i \hat{\beta}_{1,i},
\]
where \( Q_{\beta_3} = \frac{1}{n} \sum_i \beta_{3,i} \hat{\nu}_i \beta_{1,i} \) and \( \hat{\nu}_i = 1^X_i \left( \text{diag} \left[ \hat{\nu}_i \right] \right)^{-1} \). The final estimator of the risk premia is \( \hat{\lambda}_t = \hat{\Lambda} Z_{t-1} \) where we deduce \( \hat{\Lambda} \) from the relationship \( \text{vec} \left[ \hat{\Lambda} \right] = \hat{\nu} + \text{vec} \left[ \hat{\beta}^T \right] \) with the estimator \( \hat{\beta} \) obtained by a SUR regression of factors \( f_t \) on lagged instruments \( Z_{t-1} \):
\[
\hat{F} = \sum_t f_t Z_{t-1}' \left( \sum_t Z_{t-1} Z_{t-1}' \right)^{-1}.
\]

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) \( E \left[ \varepsilon_{i,j,t}, Z_{j,t-1}, j = 1, \ldots, n \right] = 0 \), with \( Z_L = \{ Z_t, Z_{t-1}, \ldots \} \) and \( Z_L = \{ Z_j, Z_{j,t-1}, \ldots \} \)

b) \( \sigma_{i,t} \leq M, i = 1, \ldots, n; \) c) \( E \left[ \frac{1}{n} \sum_{i,j} E \left[ \left| \varepsilon_{i,j,t} \right|^2 \right] \right]^{1/2} \leq M, \) where \( \sigma_{i,t} = E \left[ \varepsilon_{i,t} \varepsilon_{j,t} \right] \).

Proposition 8 summarizes consistency of estimators \( \hat{\nu} \) and \( \hat{\Lambda} \) under the double asymptotics \( n, T \rightarrow \infty \). It extends Proposition 2 to the conditional case.

**Proposition 8** Under Assumptions APR.3-APR.7, SC.1-SC.2, B.1 and C.1a), C.2-C.6, we get

a) \( \| \hat{\nu} - \nu \| = o_p \left( 1 \right) \), b) \( \| \hat{\Lambda} - \Lambda \| = o_p \left( 1 \right) \), when \( n, T \rightarrow \infty \) such that \( n \propto T^{\bar{\gamma}} \) for \( \bar{\gamma} > 0 \).

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

Proposition 9 below gives the large-sample distributions under the double asymptotics \( n, T \rightarrow \infty \). It extends Proposition 3 to the conditional case through adequate use of selection matrices. The following assumption is similar to Assumption A.2. We make use of \( Q_{\beta_3} = E_G \left[ \beta_{3,i} \beta_{3,i} \right] \), \( Q_z = E \left[ Z_t Z_t' \right] \), \( S_{ij} = \frac{1}{T} \sum_t \sigma_{i,j,t} x_{i,t} x_{j,t}' = E \left[ \varepsilon_{i,t} \varepsilon_{j,t} x_{i,t} x_{j,t}' \right] \) and \( S_{Q,ij} = Q_{x,ij}^{-1} S_{ij} Q_{x,ij}^{-1} \), otherwise, we keep the same notations as in Section 2.

**Assumption B.2** As \( n, T \rightarrow \infty \) such that \( n \propto T^{\bar{\gamma}} \) for \( \bar{\gamma} \in \Gamma_1 \subset \mathbb{R}^+ \), a) \( \frac{1}{\sqrt{n}} \sum_i \tau_i \left[ \left( Q_{x,i}^{-1} Y_{i,T} \right) \otimes v_{3,i} \right] \Rightarrow \)
\[
N \left( 0, S_{v3} \right), \text{ with } Y_{i,T} = \frac{1}{\sqrt{T}} \sum_t I_{i,t} x_{i,t} \varepsilon_{i,t}, v_{3,i} = \text{vec} \left[ \beta_{3,i} w_i \right] \text{ and } S_{v3} = \lim_{n \rightarrow \infty} E \left[ \frac{1}{n} \sum_{i,j} \tau_{ij} S_{Q,ij} \otimes v_{3,i} v_{3,j}' \right]
\]
\[
= \lim_{n \rightarrow \infty} \sum_{i,j} \frac{\tau_{ij}}{\tau_{ij}} \left[ S_{Q,ij} \otimes v_{3,i} v_{3,j}' \right]; \) b) \( \frac{1}{\sqrt{T}} \sum_t u_t \otimes Z_{t-1} \Rightarrow N \left( 0, \Sigma_u \right), \text{ 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 9 Under Assumptions APR.3-APR.7, SC.1-SC.2, B.1-B.2 and C.1a), C.2-C.6, we have

a) \( \sqrt{nT} \left( \hat{v} - v - \frac{1}{T} \tilde{B}_v \right) \Rightarrow N(0, \Sigma_v) \) where \( \tilde{B}_v = \tilde{Q}_{\beta_3} J_b \frac{1}{n} \sum_{i} \tau_{i,T} \text{vec} \left[ E_{z_i}^2 T \tilde{S}_i \tilde{Q}_{\beta_3} C_{\nu} \tilde{w}_i \right] \) and

\[
\Sigma_v = \left( \text{vec} \left[ C_{\nu}' \bigotimes Q_{\beta_3}^{-1} \right] \right)' S_{v3} \left( \text{vec} \left[ C_{\nu}' \bigotimes Q_{\beta_3}^{-1} \right] \right), \quad \text{with } J_b = \left( \text{vec} [I_{d_1}]' \bigotimes I_{K'p} \right) \left( I_{d_1} \bigotimes J_a \right) \text{ and } C_{\nu} = \left( E_1' - (I_{d_1} \bigotimes \nu) J_a E_2' \right)'.
\]

b) \( \sqrt{T} \text{vec} \left[ \tilde{N} - \tilde{N}' \right] \Rightarrow N(0, \Sigma_\lambda) \) where \( \Sigma_\lambda = (I_K \bigotimes Q_{\nu}^{-1}) \Sigma_u (I_K \bigotimes Q_{\nu}^{-1}) \),

when \( n, T \to \infty \) such that \( n \times T^\gamma \) for \( \gamma \in \Gamma_1 \cap (0, 3) \).

Since \( \lambda_t = \Lambda Z_{t-1} = (Z_{t-1}' \bigotimes I_K) W_{p,K} \text{vec} [A'] \), part b) implies conditionally on \( Z_{t-1} \) that

\( \sqrt{T} \left( \tilde{\lambda}_t - \lambda_t \right) \Rightarrow N(0, (Z_{t-1}' \bigotimes I_K) W_{p,K} \Sigma_\lambda W_{K,p} (Z_{t-1}' \bigotimes I_K)) \).

We can use Proposition 9 to build confidence intervals. It suffices to replace the unknown quantities \( Q_x \), \( Q_z \), \( \nu \) by their empirical counterparts. For matrix \( S_{v3} \), we use the thresholded estimator \( \tilde{S}_{ij} \) as in Section 2.4. Then we can extend Proposition 4 to the conditional case under Assumptions B.1-B.2, A.3, A.4 and C.1-C.6.

Since Equation (14) corresponds to the asset pricing restriction (3), the null hypothesis of correct specification of the conditional model is

\[
H_0 : \text{there exists } \nu \in \mathbb{R}^{pK} \text{ such that } \beta_1(\gamma) = \beta_3(\gamma) \nu, \quad \text{with } \text{vec} [\beta_3(\gamma)'] = J_a \beta_2(\gamma),
\]

for almost all \( \gamma \in [0, 1] \).

Under \( H_0 \), we have \( E_G \left[ (\beta_1, - \beta_3, \nu)' (\beta_1, - \beta_3, \nu) \right] = 0 \). The alternative hypothesis is

\[
H_1 : \inf_{\nu \in \mathbb{R}^{pK}} E_G \left[ (\beta_1, - \beta_3, \nu)' (\beta_1, - \beta_3, \nu) \right] > 0.
\]

As in Section 2.5, we build the SSR \( \hat{Q}_e = \frac{1}{n} \sum_{i} \hat{e}_i \tilde{w}_i \hat{e}_i \), with \( \hat{e}_i = \hat{\beta}_{1,i} - \hat{\beta}_{3,i} \hat{v} = C_{\nu}' \hat{\beta}_i \) and

the statistic \( \hat{\xi}_{nT} = T \sqrt{n} \left( \hat{Q}_e - \frac{1}{T} \tilde{B}_\xi \right) \), where \( \tilde{B}_\xi = d_1 \).

Assumption B.3 For \( n, T \to \infty \) such that \( n \times T^{-\gamma} \) for \( \gamma \in \Gamma_2 \subset \Gamma_1 \), we have

\[
\frac{1}{\sqrt{n}} \sum_{i} \tau_i^2 \left[ (Q_{x,i}' \bigotimes Q_{x,i}^{-1}) (Y_{i,T} \bigotimes Y_{T,1} - \text{vec} [S_{x,i,T}]) \right] \bigotimes \text{vec} [w_i] \Rightarrow N(0, \Omega), \quad \text{where the asymptotic variance matrix is:}
\]

\[
\Omega = \lim_{n \to \infty} \frac{1}{n} \sum_{i,j} \frac{\tau_i^2 \tau_j^2}{\tau_{ij}^2} \left[ S_{Q_{i,j}} \bigotimes S_{Q_{i,j}} + (S_{Q_{i,j}} \bigotimes S_{Q_{i,j}}) W_d \bigotimes (\text{vec} [w_i] \text{vec} [w_j]') \right]
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i,j} \frac{\tau_i^2 \tau_j^2}{\tau_{ij}^2} \left[ S_{Q_{i,j}} \bigotimes S_{Q_{i,j}} + (S_{Q_{i,j}} \bigotimes S_{Q_{i,j}}) W_d \bigotimes (\text{vec} [w_i] \text{vec} [w_j]') \right].
\]
Proposition 10  Under \( \mathcal{H}_0 \) and Assumptions APR.3-APR.7, SC.1-SC.2, B.1-B.2, A.3, A.4 and C.1-C.6, we have \( \sum_{\xi}^{-1/2} \xi_n \to N(0,1) \), where \( \sum_{\xi} = 2 \frac{1}{n} \sum_{i,j} \hat{\tau}_{ij}^2 T \),

\[
\left[ \hat{w}_i \left( C_{ij} \hat{Q}_{x,i}^{-1} \hat{S}_{ij} \hat{Q}_{x,j}^{-1} \hat{C}_v \right) \hat{w}_j \left( C_{ij} \hat{Q}_{x,j}^{-1} \hat{S}_{ji} \hat{Q}_{x,i}^{-1} \hat{C}_v \right) \right]
\]

as \( n, T \to \infty \) such that \( n \asymp T^\gamma \) for \( \gamma \in \Gamma_2 \cap \left( 0, \min \left\{ 2, \eta \frac{1-\bar{q}}{2\delta} \right\} \right) \).

Under \( \mathcal{H}_1 \), we have \( \xi \to +\infty \), as in Proposition 6.

As in Section 2.5, the null hypothesis when the factors are tradable assets becomes:

\[
\mathcal{H}_0 : \quad \beta_1(\gamma) = 0 \quad \text{for almost all } \gamma \in [0,1],
\]

against the alternative hypothesis

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

We only have to substitute \( \hat{Q}_a = \frac{1}{n} \sum \hat{\beta}_1,i \hat{w}_i \hat{\beta}_1,i \) for \( \hat{Q}_e \), and \( E_1 = (I_{d_1}, 0)' \) for \( C_v \). This gives an extension of Gibbons, Ross and Shanken (1989) to the conditional case and with double asymptotics. Implementing the original Gibbons, Ross and Shanken (1989) test, which uses a weighting matrix corresponding to an inverted estimated covariance matrix, becomes quickly problematic; each \( \beta_1,i \) is of dimension \( d_1 \times 1 \), and the inverted matrix is of dimension \( nd_1 \times nd_1 \). We expect to compensate the potential loss of power induced by a diagonal weighting thanks to the large number \( nd_1 \) of restrictions. Our preliminary unreported Monte Carlo simulations show that the test exhibits good power properties for a couple of hundreds of assets.

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 alphas, betas and risk premia, we use a specification based on two common variables and two firm-level variables. We take the instruments \( Z_t = (1, Z_t^*)' \), where bivariate vector \( Z_t^* \) includes the term spread,
proxied by the difference between yields on 10-year Treasury and three-month T-bill, and the default spread, proxied by the yield difference between Moody’s Baa-rated and Aaa-rated corporate bonds. We take $Z_{i,t}$ as a bivariate vector made of the market capitalization and the book-to-market equity of firm $i$. We refer to Avramov and Chordia (2006) for convincing theoretical and empirical arguments in favor of the chosen conditional specification. The vector $x_{i,t}$ is of dimension $d = 32$. The firm characteristics are computed as in the appendix of Fama and French (2008) from Compustat. We use monthly stock returns data provided by CRSP and we exclude financial firms (Standard Industrial Classification Codes between 6000 and 6999) as in Fama and French (2008). The dataset after matching CRSP and Compustat contents comprises $n = 9,936$ stocks and covers the period from July 1964 to December 2009 with $T = 546$. For comparison purposes with a standard methodology for small $n$, we consider the 25 and 100 Fama-French (FF) portfolios as base assets. We have downloaded the time series of factors, portfolios and portfolio characteristics from the website of Kenneth French.

4.2 Estimation results

We first present unconditional estimates before looking at the path of the time-varying estimates. We use $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/12$ for the unconditional estimation and $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/36$ for the conditional estimation. In the reported results for the four-factors model, we denote by $n^\chi$ the dimension of the cross-section after trimming. We use a data-driven threshold selected by cross-validation as in Bickel and Levina (2008). Table 1 gathers the estimated annual risk premia for the following unconditional models: the four-factor model, the Fama-French model, and the CAPM. In Table 2, we display the estimates of the components of $\nu$. When $n$ is large, we use bias-corrected estimates for $\lambda$ and $\nu$. When $n$ is small, we use asymptotics for fixed $n$ and $T \to \infty$. The estimated risk premia for the market factor are of the same magnitude and all positive across the three universes of assets and the three models. The 95% confidence intervals are larger by construction for fixed $n$, and they often contain the interval for large $n$. For the four-factor model and the individual stocks the size factor is positively remunerated (2.91%) and it is not significantly different from zero. The value factor commands a significant negative reward (-4.55%). Phalippou (2007) obtained a similar result, indeed he got a growth premium when portfolios are built on stocks with a high institutional ownership. The momentum factor is largely remunerated (7.34%) and
significantly different from zero. For the 25 and 100 FF portfolios we observe that the size factor is not significantly positively remunerated while the value factor is significantly positively remunerated (4.81% and 5.11%). The momentum factor bears a significant positive reward (34.03% and 17.29%). The large, but imprecise, estimate for the momentum premium when $n = 25$ and $n = 100$ comes from the estimate for $\nu_{mom}$ (25.40% and 8.66%) that is much larger and less accurate than the estimates for $\nu_m, \nu_{smb}$ and $\nu_{hml}$ (0.85%, -0.26%, 0.03%, and 0.55%, 0.01%, 0.33%). Moreover, while for portfolios the estimates of $\nu_m, \nu_{smb}$ and $\nu_{hml}$ are statistically not significant, for individual stocks these estimates are statistically different from zero. In particular, the estimate of $\nu_{hml}$ is large and negative, which explains the negative estimate on the value premium displayed in Table 1.

A potential explanation of the discrepancies revealed in Tables 1 and 2 between individual stocks and portfolios is the much larger heterogeneity of the factor loadings for the former. 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. The estimation results for the momentum factor exemplify the problems related to a small number of portfolios exhibiting a tight factor structure (Lewellen, Nagel and Shanken (2010)). For $\lambda_m, \lambda_{smb}$, and $\lambda_{hml}$, we obtain similar inferential results when we consider the Fama-French model. Our point estimates for $\lambda_m, \lambda_{smb}$ and $\lambda_{hml}$, for large $n$ agree with Ang, Liu and Schwarz (2008). Our point estimates and confidence intervals for $\lambda_m, \lambda_{smb}$ and $\lambda_{hml}$, agree with the results reported by Shanken and Zhou (2007) for the 25 portfolios.

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, size and value factors feature a counter-cyclical pattern. Indeed, these risk premia increase during economic contractions and decrease during economic booms. Gomes, Kogan and Zhang (2003) and Zhang (2005) construct equilibrium models exhibiting a countercyclical behavior in size and book-to-market effects. On the contrary, the risk premium for momentum factor is pro-cyclical. Furthermore, conditional estimates of the value premium take stable and positive values. They are not significantly different from zero during economic booms. The conditional estimates of the size
premium are most of the time slightly positive, and not significantly different from zero.

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 \( \lambda_{\text{mom},t} \) 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 \( \lambda_{\text{mom}} \). Finally, by comparing Figures 2 and 3, we observe that the patterns of risk premia look similar except for the book-to-market factor. Indeed, the risk premium for the value effect estimated from the 25 portfolios is pro-cyclical, contradicting the counter-cyclical behavior predicted by finance theory. By comparing Figures 3 and 4, we observe that increasing the number of portfolios to 100 does not help in reconciling the discrepancy.

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 100 FF 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 the cross-sectional variation of 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, when \( n \) is large, 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 \( \Sigma_{\xi}^{-1/2} \hat{\xi} nT \) as well as its associated \( p \)-value. For the 25 and 100 FF portfolios, we compute weighted test statistics (Gibbons, Ross and Shanken (1989)) as well as their associated \( p \)-value. We do similarly for the test statistics relying on the alphas \( a \). 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 Gibbons-Ross-Shanken test statistic rejects the well specification for the CAPM and the three-factor model. The four-factor model is not rejected at 1% level, but it is rejected at 5% level.

4.4 Cost of equity

The results in Section 3 can be used for estimation and inference on the cost of equity in conditional factor models. We can estimate the time varying cost of equity $CE_{i,t} = r_{f,t} + b_{i,t}' \lambda_t$ of firm $i$ with $CE_{i,t} = r_{f,t} + \hat{b}_{i,t}' \hat{\lambda}_t$, where $r_{f,t}$ is the risk-free rate. We have (see Appendix 3)

$$\sqrt{T} \left( CE_{i,t} - CE_{i,t} \right) = \psi_{i,t}' E^t \sqrt{T} \left( \hat{\beta}_i - \beta_i \right) + \left( Z_{t-1}' \otimes b_{i,t}' \right) W_p, K \sqrt{T \text{vec} \left[ \hat{\Lambda}' - \Lambda' \right]} + o_p(1), \quad (17)$$

where $\psi_{i,t} = \left( \lambda_t' \otimes Z_{t-1}' \otimes Z_{i,t}' \right)'$. Standard results on OLS imply that estimator $\hat{\beta}_i$ is asymptotically normal, $\sqrt{T} \left( \hat{\beta}_i - \beta_i \right) \Rightarrow N \left( 0, \tau^2_i Q^{-1}_x S_i Q^{-1}_x \right)$, and independent of estimator $\hat{\Lambda}$. Then, from Proposition 7 we deduce that $\sqrt{T} \left( CE_{i,t} - CE_{i,t} \right) \Rightarrow N \left( 0, \Sigma_{CE_{i,t}} \right)$, conditionally on $Z_{t-1}$, where

$$\Sigma_{CE_{i,t}} = \tau^2_i \psi_{i,t}' E^t Q^{-1}_x \left( S_i, Q^{-1}_x, E^t \right) \psi_{i,t} + \left( Z_{t-1}' \otimes b_{i,t}' \right) W_p, K \Sigma_{\Lambda} W_K, p \left( Z_{t-1} \otimes b_{i,t} \right).$$

Figure 5 plots the path of the estimated annualized costs of equity for Ford Motor, Disney, Motorola and Sony. The cost of equity has risen tremendously during the recent subprime crisis.
References


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 $n = 9,936$ from 1964/07 to 2009/12. Moreover, next to each box-plot we report the estimated factor loadings for the 25 and 100 Fama-French portfolios (circles and triangles, respectively).
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 (dashed horizontal line) and the average conditional estimate (solid horizontal line). We consider all stocks ($n = 9,926$ and $n^x = 2,612$) 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.
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_{m,t}$, $\hat{\lambda}_{smb,t}$, $\hat{\lambda}_{hml,t}$ and $\hat{\lambda}_{mom,t}$ and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dashed 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).
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 (dashed 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).
The figure plots the path of estimated annualized cost of equities for Ford Motor, Disney Walt, Motorola and Sony and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dashed 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).
Table 1: Estimated annualized risk premia for the unconditional models

<table>
<thead>
<tr>
<th></th>
<th>Stocks ($n = 9,936$, $n^x = 9,902$)</th>
<th>Portfolios ($n = n^x = 25$)</th>
<th>Portfolios ($n = n^x = 100$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias corrected estimate (%)</td>
<td>95% conf. interval</td>
<td>point estimate (%)</td>
</tr>
<tr>
<td>Four-factor model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_m$</td>
<td>8.08</td>
<td>(3.20, 12.99)</td>
<td>5.70</td>
</tr>
<tr>
<td>$\lambda_{smb}$</td>
<td>2.91</td>
<td>(-0.45, 6.26)</td>
<td>3.02</td>
</tr>
<tr>
<td>$\lambda_{hml}$</td>
<td>-4.55</td>
<td>(-8.01, -1.08)</td>
<td>4.81</td>
</tr>
<tr>
<td>$\lambda_{mom}$</td>
<td>7.34</td>
<td>(2.74, 11.94)</td>
<td>34.03</td>
</tr>
<tr>
<td>Fama-French model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_m$</td>
<td>7.60</td>
<td>(2.72, 12.49)</td>
<td>5.04</td>
</tr>
<tr>
<td>$\lambda_{smb}$</td>
<td>2.73</td>
<td>(-0.62, 6.09)</td>
<td>3.00</td>
</tr>
<tr>
<td>$\lambda_{hml}$</td>
<td>-4.95</td>
<td>(-8.42, -1.49)</td>
<td>5.20</td>
</tr>
<tr>
<td>CAPM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_m$</td>
<td>7.39</td>
<td>(2.50, 12.27)</td>
<td>6.98</td>
</tr>
</tbody>
</table>

The table contains the estimated annualized risk premia for the market ($\lambda_m$), size ($\lambda_{smb}$), book-to-market ($\lambda_{hml}$) and momentum ($\lambda_{mom}$) factors. The bias corrected estimates $\lambda_B$ of $\lambda$ are reported for individual stocks ($n = 9,936$). In order to build the confidence intervals for $n = 9,936$, we use $\Sigma_f$. When we consider 25 and 100 portfolios as base assets, we compute an estimate of the covariance matrix $\Sigma_{\lambda,n}$ defined in Section 2.3.
Table 2: Estimated annualized $\nu$ for the unconditional models

<table>
<thead>
<tr>
<th></th>
<th>Stocks ($n = 9,936, n^x = 9,902$)</th>
<th>Portfolios ($n = n^x = 25$)</th>
<th>Portfolios ($n = n^x = 100$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias corrected estimate (%)</td>
<td>95% conf. interval</td>
<td>point estimate (%)</td>
</tr>
<tr>
<td>Four-factor model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_m$</td>
<td>3.22</td>
<td>(2.95, 3.50)</td>
<td>0.85</td>
</tr>
<tr>
<td>$\nu_{smb}$</td>
<td>-0.37</td>
<td>(-0.67, -0.06)</td>
<td>-0.26</td>
</tr>
<tr>
<td>$\nu_{hml}$</td>
<td>-9.33</td>
<td>(-9.67, -8.90)</td>
<td>0.03</td>
</tr>
<tr>
<td>$\nu_{mom}$</td>
<td>-1.29</td>
<td>(-1.88, -0.70)</td>
<td>25.40</td>
</tr>
<tr>
<td>Fama-French model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_m$</td>
<td>2.75</td>
<td>(2.48, 3.02)</td>
<td>0.18</td>
</tr>
<tr>
<td>$\nu_{smb}$</td>
<td>-0.54</td>
<td>(-0.85, -0.22)</td>
<td>-0.27</td>
</tr>
<tr>
<td>$\nu_{hml}$</td>
<td>-9.74</td>
<td>(-10.08, -9.39)</td>
<td>0.41</td>
</tr>
<tr>
<td>CAPM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_m$</td>
<td>2.53</td>
<td>(2.32, 2.74)</td>
<td>2.12</td>
</tr>
</tbody>
</table>

The table contains the annualized estimates of the components of vector $\nu$ for the market ($\nu_m$), size ($\nu_{smb}$), book-to-market ($\nu_{hml}$) and momentum ($\nu_{mom}$) factors. The bias corrected estimates $\hat{\nu}_B$ of $\nu$ are reported for individual stocks ($n = 9,936$). In order to build the confidence intervals, we compute $\hat{\Sigma}_\nu$ in Proposition 4 for $n = 9,936$. When we consider 25 and 100 portfolios as base assets, we compute an estimate of the covariance matrix $\Sigma_{\nu,n}$ defined in Section 2.3.
### Table 3: Specification test results for the unconditional models

<table>
<thead>
<tr>
<th></th>
<th>Test statistic based on $\tilde{Q}_e$, $H_0 : a = b'\nu$</th>
<th>Test statistic based on $\tilde{Q}_a$, $H_0 : a = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stocks ($n = 9,936$)</td>
<td>Portfolios ($n = 25$)</td>
</tr>
<tr>
<td><strong>Four-factor model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>22.9551</td>
<td>35.2231</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0267</td>
</tr>
<tr>
<td><strong>Fama-French model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>20.8816</td>
<td>83.6846</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td><strong>CAPM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test statistic</td>
<td>22.3152</td>
<td>110.8368</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The test statistic $\tilde{\Sigma}_\xi^{-1/2} \tilde{\xi}_n T$ defined in Proposition 5 is computed for $n = 9,936$. For $n = 25$ and $n = 100$, the test statistic $T\tilde{e}' \Omega^{-1} \tilde{e}$ is reported. The test statistic $T\tilde{a}' \Omega^{-1} \tilde{a}$ is also computed. The table reports the p-values, respectively.
Appendix 1: Regularity conditions

In this Appendix, we list and comment the additional assumptions used to derive the large sample properties of the estimators and test statistics. For unconditional models, we use Assumptions C.1-C.5 below with \( x_t = (1, f_t')' \).

**Assumption C.1** There exists constants \( \eta, \bar{\eta} \in (0, 1] \) and \( C_1, C_2, C_3, C_4 > 0 \) such that for all \( \delta > 0 \) and \( T \in \mathbb{N} \) we have:

\[
\begin{align*}
  & a) \ P \left[ \left\| \frac{1}{T} \sum_t (x_t x_t' - E [x_t x_t']) \right\| \geq \delta \right] \leq C_1 T \exp \{-C_2 \delta^2 T^m\} + C_3 \delta^{-1} \exp \{-C_4 T^m\}. \\
  & b) \sup_{\gamma \in [0, 1]} \mathbb{P} \left[ \left\| \frac{1}{T} \sum_t I_t(\gamma) \left( x_t x_t' - E [x_t x_t'] \right) \right\| \geq \delta \right] ; \\
  & c) \sup_{\gamma \in [0, 1]} \mathbb{P} \left[ \left\| \frac{1}{T} \sum_t I_t(\gamma) x_t \varepsilon_t(\gamma) \right\| \geq \delta \right] ; \\
  & d) \sup_{\gamma, \gamma' \in [0, 1]} \mathbb{P} \left[ \left\| \frac{1}{T} \sum_t I_t(\gamma) I_t(\gamma') \left( \varepsilon_t(\gamma) \varepsilon_t(\gamma') x_t x_t' - E [\varepsilon_t(\gamma) \varepsilon_t(\gamma') x_t x_t'] \right) \right\| \geq \delta \right] ; \\
  & e) \sup_{\gamma, \gamma' \in [0, 1]} \mathbb{P} \left[ \left\| \frac{1}{T} \sum_t I_t(\gamma) I_t(\gamma') x_k t x_l, t x_m, t \varepsilon_t(\gamma) \right\| \geq \delta \right].
\end{align*}
\]

**Assumption C.2** There exists \( c > 0 \) such that \( \sup_{\gamma \in [0, 1]} \mathbb{E} \left[ \left\| \frac{1}{T} \sum_t I_t(\gamma) \left( x_t x_t' - E [x_t x_t'] \right) \right\|^4 \right] = O(T^{-c}) \).

**Assumption C.3** a) There exists a constant \( M \) such that \( \| x_t \| \leq M, \ P \text{-a.s.} \). Moreover, b) \( \sup_{\gamma \in [0, 1]} \| \beta(\gamma) \| < \infty \) and c) \( \inf_{\gamma \in [0, 1]} \mathbb{E} [I_t(\gamma)] > 0 \).

**Assumption C.4** There exists a constant \( M \) such that for all \( n, T \):

\[
\begin{align*}
  & a) \frac{1}{n T^2} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} |E [\varepsilon_{i,t_1} \varepsilon_{i,t_2} \varepsilon_{j,t_3} \varepsilon_{j,t_4} | \gamma_i, \gamma_j]| \leq M; \\
  & b) \frac{1}{n T^2} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} \|E [\eta_{i,t_1} \varepsilon_{i,t_2} \varepsilon_{j,t_3} \varepsilon_{j,t_4} | \gamma_i, \gamma_j]\| \leq M, \ \text{where} \ \eta_{i,t} = \varepsilon_{i,t} x_t x_t' - E [\varepsilon_{i,t}^2 x_t x_t'| \gamma_i]; \\
  & c) \frac{1}{n T^2} \sum_{i,j} \sum_{t_1, \ldots, t_6} |E [\varepsilon_{i,t_1} \varepsilon_{i,t_2} \varepsilon_{j,t_3} \varepsilon_{j,t_4} \varepsilon_{j,t_5} \varepsilon_{j,t_6} | \gamma_i, \gamma_j]| \leq M;
\end{align*}
\]

**Assumption C.5** The trimming constants satisfy \( \chi_{1,T} = O ((\log T)^{\kappa_1}), \chi_{2,T} = O ((\log T)^{\kappa_2}), \text{with} \ \kappa_1, \kappa_2 > 0 \).
For conditional models, Assumptions C.1-C.5 are used with $x_t$ replaced by $x_{i,t}$ as defined in Section 3.1. More precisely, for Assumptions C.1a) and C.3a) we replace $x_t$ by $x_t(\gamma)$ and require the bound to be valid uniformly w.r.t. $\gamma \in [0, 1]$; for Assumptions C.1b)-c) and C.2 we replace $x_t$ by $x_t(\gamma)$; for Assumption C.4b) we replace $x_t$ by $x_{i,t}$. Furthermore, we use:

**Assumption C.6** There exists a constant $M$ such that $\| E [u_t u_t'|Z_{t-1}] \| \leq M$ for all $t$, where $u_t = f_t - F Z_{t-1}$.
Appendix 2: Unconditional factor model

A.2.1 Proof of Proposition 1

For given assets \( \gamma_1, ..., \gamma_n \in [0, 1] \), let \( \mu_n = A_n + B_n E[f_1|\mathcal{F}_0] \) and \( V_n = B_n V[f_1|\mathcal{F}_0] B_n' + \Sigma_{\epsilon,1,n} \) be the mean vector and the variance-covariance matrix of asset excess returns \( (R_1(\gamma_1), ..., R_1(\gamma_n))' \) conditional on \( \mathcal{F}_0 \), where \( A_n = [a(\gamma_1), ..., a(\gamma_n)]' \), \( B_n = [b(\gamma_1), ..., b(\gamma_n)]' \). Let \( e_n = \mu_n - B_n \left( B_n' B_n \right)^{-1} B_n' A_n \) be the residual of the orthogonal projection of \( \mu_n \) (and \( A_n \)) onto the columns of \( B_n \). Then, for portfolio \( p_n \) we have:

\[
C(p_n) = \alpha_{0,n} + \alpha_n' \epsilon_n, \quad E[p_n|\mathcal{F}_0] = R_0 C(p_n) + \alpha_n' \mu_n, \quad V[p_n|\mathcal{F}_0] = \alpha_n' V_n \alpha_n,
\]

where \( \epsilon_n = (1, ..., 1)' \) and \( \alpha_n = (\alpha_{1,n}, ..., \alpha_{n,n})' \). We denote by \( \mathcal{J} \) the set of asset sequences \( (\gamma_i) \) in \([0, 1]\) such that \( B_n' B_n/n \) converges to a positive definite matrix as \( n \to \infty \), and by \( \mathcal{P} \) the set of portfolio sequences based on asset sequences in \( \mathcal{J} \).

The proof involves three steps.

**Step 1:** The absence of asymptotic arbitrage opportunities (Assumption APR.5) implies that the two conditions:

(i) If \( V[p_n|\mathcal{F}_0] \to 0 \) and \( C(p_n) \to 0 \), then \( E[p_n|\mathcal{F}_0] \to 0 \);

(ii) If \( V[p_n|\mathcal{F}_0] \to 0 \), \( C(p_n) \to 1 \) and \( E[p_n|\mathcal{F}_0] \to \delta \), then \( \delta \geq 0 \),

hold for any portfolio sequence \( (p_n) \) in \( \mathcal{P} \), \( P \)-a.s.

**Step 2:** Conditions (i) and (ii) imply that \( \inf_{\nu \in \mathbb{R}^K} \sum_{i=1}^{\infty} \left( a(\gamma_i) - b(\gamma_i)\nu \right)^2 < \infty \) for any sequence \( (\gamma_i) \) in \( \mathcal{J} \).

**Step 3:** There exists a unique vector \( \nu \in \mathbb{R}^K \) such that \( a(\gamma) = b(\gamma)\nu \) for almost all \( \gamma \in [0, 1] \).

**Step 1:** The proof is by contradiction. Suppose that Condition (i) does not hold. Then, with probability strictly larger than zero, there exists a portfolio sequence \( (p_n) \) in \( \mathcal{P} \) such that \( V[p_n|\mathcal{F}_0] \to 0 \), \( C(p_n) \to 0 \) and \( E[p_n|\mathcal{F}_0] \to 0 \). Let us denote by \( S \in \mathcal{F} \) the set of events where this fact occurs. The condition \( E[p_n|\mathcal{F}_0] \to 0 \) implies that either \( \liminf E[p_n|\mathcal{F}_0] \) or \( \limsup E[p_n|\mathcal{F}_0] \) is different from zero. Then, when \( S \) occurs, there exists a subsequence such that \( \inf_{n \geq 0} E[p_n|\mathcal{F}_0] > 0 \) possibly after changing the sign of the portfolio weights.

45
First, we prove that \( \limsup_{n \to \infty} \| e_n \|^2 = +\infty \) on \( S \). Since \( \| e_n \| \) is non-decreasing w.r.t. \( n \), we have \( \sup_{n \geq 0} \| e_n \| = \lim_{n \to \infty} \sup_{n \geq 0} \| e_n \|. \) Suppose that \( \sup_{n \geq 0} \| e_n \| < \infty \). By using that:

\[
V(p_n|F_0) = (B_n^\prime \alpha_n)^\prime V[f_1|F_0](B_n^\prime \alpha_n) + \alpha_n^\prime \Sigma_{\varepsilon,1,n} \alpha_n \geq c \left( \| B_n^\prime \alpha_n \|^2 + \| \alpha_n \|^2 \right),
\]

for \( c = \min\{\text{eig}_{\text{min}}(V[f_1|F_0]), \text{eig}_{\text{min}}(\Sigma_{\varepsilon,1,n})\} \geq 0 \) (Assumptions APR.4 (ii)-(iii)), we deduce \( \| \alpha_n \| \to 0 \), \( B_n^\prime \alpha_n \to 0 \), and \( \inf_{n \geq 0} \alpha_n^\prime \mu_n > 0 \), P-a.s. on \( S \). Write \( \alpha_n^\prime \mu_n = \alpha_n^\prime B_n \left( \frac{B_n^\prime B_n}{n} \right)^{-1} \frac{B_n^\prime \mu_n}{n} + \alpha_n^\prime e_n \). Now, by Cauchy-Schwarz \( \alpha_n^\prime e_n \leq \| \alpha_n \| \| e_n \| = o(1) \). Then, since \( B_n^\prime B_n/n \) converges to a positive definite matrix and \( B_n^\prime \mu_n/n \) is bounded, it follows \( \alpha_n^\prime \mu_n = o(1) \), which is in contradiction with \( \inf_{n \geq 0} \alpha_n^\prime \mu_n > 0 \). Thus, \( \limsup_{n \to \infty} \| e_n \|^2 = +\infty \) must hold on \( S \).

Let us now show that an asymptotic arbitrage portfolio exists. Define the portfolio sequence \( (q_n) \) with investments \( \alpha_n = \frac{1}{\| e_n \|^2} e_n \) and \( \alpha_{0,n} = -\iota_n^\prime \alpha_n \) if \( S \) occurs, and zero investment otherwise. This portfolio has zero cost, i.e., \( C(q_n) = 0 \), and \( E[q_n|F_0] = 1 \) and \( V[q_n|F_0] \leq c_1 \| e_n \|^{-2} \) on \( S \) (Assumption APR.4 (i)). Moreover, we have:

\[
\begin{align*}
V[q_n|F_0] &= E \left[ (q_n - E[q_n|F_0])^2 | F_0 \right] \geq E \left[ (q_n - E[q_n|F_0])^2 | F_0, q_n \leq 0 \right] \cdot P[q_n \leq 0|F_0] \\
&\geq P[q_n \leq 0|F_0] \geq 1 - V[q_n|F_0] \geq 1 - c_1 \| e_n \|^{-2}
\end{align*}
\]

if \( S \) occurs, and \( q_n = 0 \) otherwise. Thus, \( P[q_n > 0|F_0] \to 1 \) as \( n \to \infty \), P-a.s. By using the Law of Iterated Expectation and the Lebesgue dominated convergence theorem, \( P[q_n > 0] \to P[S] \) and \( P[q_n \geq 0] \to 1 \).

Since by Assumption APR.5 no asymptotic arbitrage portfolio exists, we deduce that Condition (i) holds. By a similar argument, we can show that Condition (ii) also holds.

**Step 2:** Conditions (i) and (ii) correspond to conditions A.1 (i) and (ii) in CR written conditionally on \( F_0 \). Step 2 follows by using Assumption APR.4 and applying CR Theorem 3 conditionally on \( F_0 \) for any countable collection of assets \( (\gamma_i) \) in \( \mathcal{F} \).

**Step 3:** The proof is by contradiction. Suppose that \( \int [a(\gamma) - b(\gamma)^\prime \nu]^2 d\gamma > 0 \) for any \( \nu \in \mathbb{R}^K \). Let sequence \( (\gamma_i) \) in \([0,1]\) be a realization of i.i.d. uniform random draws. Then this sequence is in \( \mathcal{F} \) with probability 1 by the LLN. Moreover, we have

\[
\sum_{i=1}^{\infty} \left( a(\gamma_i) - b(\gamma_i)^\prime \nu \right)^2 \geq \sum_{i=1}^{n} \left( a(\gamma_i) - b(\gamma_i)^\prime \nu \right)^2 = nQ_n(\nu),
\]
where $Q_n(\nu) = \frac{1}{n} \sum_{i=1}^{n} \left( a(\gamma_i) - b(\gamma_i)^{\nu} \right)^2$, for any $n \in \mathbb{N}$ and $\nu \in \mathbb{R}^K$. Thus, for any $n \in \mathbb{N}$:

$$\inf_{\nu} \sum_{i=1}^{\infty} \left( a(\gamma_i) - b(\gamma_i)^{\nu} \right)^2 \geq n \inf_{\nu} Q_n(\nu) = nQ_n(\nu_n),$$

(18)

where $\nu_n = \left( \sum_{i=1}^{n} b(\gamma_i)b(\gamma_i)^\nu \right)^{-1} \sum_{i=1}^{n} b(\gamma_i)a(\gamma_i)$. Consider the LHS of the inequality (18). From Step 2, $\inf_{\nu} \sum_{i=1}^{\infty} \left( a(\gamma_i) - b(\gamma_i)^{\nu} \right)^2 < \infty$ with probability 1. Consider now the RHS of (18). By the strong LLN, we have $Q_n(\nu_n) \to Q_\infty(\nu_\infty) = \int [a(\gamma) - b(\gamma)^{\nu_\infty}]^2 d\gamma$, with probability 1, where $\nu_\infty = \left( \int b(\gamma)b(\gamma)^{\nu} d\gamma \right)^{-1} \cdot \int b(\gamma)a(\gamma) d\gamma$. Since $Q_\infty(\nu_\infty) > 0$, we deduce that $nQ_n(\nu_n) \to \infty$, P-a.s.. Then, Inequality (18) cannot hold for large $n$ with probability 1. Thus, it must be that $\int [a(\gamma) - b(\gamma)^{\nu}]^2 d\gamma = 0$ for some $\nu \in \mathbb{R}^K$. Since matrix $\int b(\gamma)b(\gamma)^{\nu} d\gamma$ is non-singular (Assumption APR.2), such $\nu$ is unique.

A.2.2 Proof of Proposition 2

a) Consistency of $\hat{\nu}$. From equation (5) and the asset pricing restriction (3), we have:

$$\hat{\nu} - \nu = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{\beta}_i \hat{c}_i \left( \beta_i - \beta_i \right)$$

$$= Q_b^{-1} \frac{1}{n} \sum_i \hat{w}_i b_i c_i \left( \beta_i - \beta_i \right) + \left( \hat{Q}_b^{-1} - Q_b^{-1} \right) \frac{1}{n} \sum_i \hat{w}_i b_i c_i \left( \beta_i - \beta_i \right)$$

$$+ \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \left( \hat{\beta}_i - \beta_i \right) c_i \left( \hat{\beta}_i - \beta_i \right).$$

By using $\hat{\beta}_i - \beta_i = \frac{\tau_{i,T}}{\sqrt{T}} Q_{x,i}^{-1} Y_{i,T}$ and $\hat{Q}_b^{-1} - Q_b^{-1} = -\hat{Q}_b^{-1} \left( \hat{Q}_b - Q_b \right) Q_b^{-1}$, we get:

$$\hat{\nu} - \nu = \frac{1}{\sqrt{nT}} Q_b^{-1} \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i c_i \hat{Q}_{x,i}^{-1} Y_{i,T} - \frac{1}{\sqrt{nT}} \hat{Q}_b^{-1} \left( \hat{Q}_b - Q_b \right) Q_b^{-1} \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i c_i \hat{Q}_{x,i}^{-1} Y_{i,T}$$

$$+ \frac{1}{T} \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \tau_{i,T}^2 E_2 \hat{Q}_{x,i}^{-1} Y_{i,T} Y_{i,T}^\prime \hat{Q}_{x,i}^{-1} c_i$$

$$= \frac{1}{\sqrt{nT}} Q_b^{-1} I - \frac{1}{\sqrt{nT}} \hat{Q}_b^{-1} \left( \hat{Q}_b - Q_b \right) Q_b^{-1} I + \frac{1}{T} \hat{Q}_b^{-1} I_2.$$

(20)
To control \( I_1 \), we use the decomposition:

\[
I_1 = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i c'_v Q_x^{-1} Y_{i,T} + \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i c'_v \left( \hat{Q}_x^{-1} - Q_x^{-1} \right) Y_{i,T} =: I_{111} + I_{112}.
\]

Write \( I_{11} = I_{111} \hat{Q}_x^{-1} c_v \) and decompose \( I_{111} := \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i Y'_{i,T} \) as:

\[
I_{111} = \frac{1}{\sqrt{n}} \sum_i w_i \tau_{i,T} Y_{i,T} + \frac{1}{\sqrt{n}} \sum_i (1 - 1_i^{\chi}) w_i \tau_{i,T} Y'_{i,T} + \frac{1}{\sqrt{n}} \sum_i 1_i^{\chi} w_i (\tau_{i,T} - \tau_i) b_i Y'_{i,T}
\]
\[
+ \frac{1}{\sqrt{n}} \sum_i 1_i^{\chi} (v_i^{-1} - v_i^{-2}) \tau_{i,T} b_i Y'_{i,T} =: I_{1111} + I_{1112} + I_{1113} + I_{1114}.
\]

We have \( E \left[ ||I_{1111}||^2 | x_{T}, I_T, \{ \gamma_i \} \right] = \frac{1}{nT} \sum_{i,j} \sum_t w_i w_j \tau_{i,j} \sigma_{i,j} ||x_t||^2 b'_j b_i \) by Assumption A.1 a).

Then, by using \( ||x_t|| \leq M, ||b|| \leq M, \tau_i \leq M, w_i \leq M \) from Assumption C.3, and Assumption A.1 c), we get \( E \left[ ||I_{1111}||^2 | \{ \gamma_i \} \right] \leq C \). Then \( I_{1111} = O_p(1) \). To control \( I_{1112} \), we use the next Lemma.

**Lemma 1** Under Assumption C.2: \( \sup_i \mathbb{P}[ I_i^{\chi} = 0 ] = O(T^{-\bar{b}}) \), for any \( \bar{b} > 0 \).

By using \( ||I_{1112}|| \leq \frac{C}{\sqrt{n}} \sum_i (1 - 1_i^{\chi}) ||Y_{i,T}|| \), \( \sup_i E[||Y_{i,T}|| | x_T, I_T, \{ \gamma_i \} ] \leq C \) from Assumption A.1, and Lemma 1, it follows \( I_{1112} = O_p(\sqrt{n}T^{-\bar{b}}) \), for any \( \bar{b} > 0 \). Since \( n \propto T^\gamma \), we get \( I_{1112} = o_p(1) \). We have \( E \left[ ||I_{1113}||^2 | x_T, I_T, \{ \gamma_i \} \right] \leq \frac{C}{nT} \sum_{i,j} \sum_t 1_i^{\chi} 1_j^{\chi} ||\sigma_{i,j}|| ||x_t||^2 \) by Assumption A.1 c), we get \( E \left[ ||I_{1113}||^2 | {\gamma_i} \right] \leq C M \sup_{\gamma \in [0,1]} E \left[ 1_i^{\chi} ||\tau_{i,T} - \tau_i|| \gamma_i = \gamma \right] \). By using \( \tau_{i,T} - \tau_i = -\tau_{i,T} \hat{\tau}_i T \sum_t (I_t(\gamma) - E[I_t(\gamma)]) \) we get \( \sup_{\gamma \in [0,1]} E \left[ 1_i^{\chi} ||\tau_{i,T} - \tau_i|| ^4 | \gamma_i = \gamma \right] \)

\[
\leq C \chi_{2,T} \sup_{\gamma \in [0,1]} E \left[ \left| \frac{1}{T} \sum_t (I_t(\gamma) - E[I_t(\gamma)]) \right|^4 \right] = o(1) \) from Assumptions C.2 and C.5. Then \( I_{1113} = o_p(1) \). From \( \hat{v}_i^{-1} - v_i^{-1} = -v_i^{-2} (\hat{v}_i - v_i) + \hat{v}_i^{-1} v_i^{-2} (\hat{v}_i - v_i)^2 \), we get:

\[
I_{1114} = -\frac{1}{\sqrt{n}} \sum_i 1_i^{\chi} v_i^{-2} (\hat{v}_i - v_i) \tau_{i,T} b_i Y'_{i,T} + \frac{1}{\sqrt{n}} \sum_i 1_i^{\chi} \hat{v}_i^{-1} v_i^{-2} (\hat{v}_i - v_i)^2 \tau_{i,T} b_i Y'_{i,T}
\]
\[
=: I_{11141} + I_{11142}.
\]
Let us first consider $I_{11141}$. We have:

$$
\hat{v}_i - v_i = \tau_i T c_{i,t} \hat{Q}_{x,i}^{-1} (\hat{S}_{i,t} - S_{i,t}) \hat{Q}_{x,i}^{-1} c_{i,t} + 2\tau_i T (c_{i,t} - c_v) (\hat{Q}_{x,i}^{-1} S_{i,t} \hat{Q}_{x,i}^{-1} c_{i,t}) \\
+ \tau_i T (c_{i,t} - c_v) (\hat{Q}_{x,i}^{-1} S_{i,t} \hat{Q}_{x,i}^{-1} (c_{i,t} - c_v) + 2\tau_i T c_{i,t} (\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}) S_{i,t} \hat{Q}_{x,i}^{-1} c_{i,t}) \\
+ \tau_i T c_{i,t} (\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}) S_{i,t} ((\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}) c_{i,t} + (\tau_i T - \tau_i) c_{i,t} Q_{x,i}^{-1} S_{i,t} \hat{Q}_{x,i}^{-1} c_{i,t})
$$

and we get for the first two terms:

$$
I_{111411} = -\frac{1}{\sqrt{n}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T c_{i,t} \hat{Q}_{x,i}^{-1} (\hat{S}_{i,t} - S_{i,t}) \hat{Q}_{x,i}^{-1} c_{i,t} b_i Y_i', T \\
I_{111412} = -\frac{2}{\sqrt{n}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T (c_{i,t} - c_v) (\hat{Q}_{x,i}^{-1} S_{i,t} \hat{Q}_{x,i}^{-1} c_{i,t}) b_i Y_i', T.
$$

We first show $I_{111412} = o_p(1)$. For this purpose, it is enough to show that $c_{i,t} - c_v = O_p(T^{-c})$, for some $c > 0$, and $\frac{1}{\sqrt{n}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T (\hat{Q}_{x,i}^{-1} S_{i,t} \hat{Q}_{x,i}^{-1}) b_i Y_i', T = O_p(\chi_1^2 \chi_2^2 T)$, for any $k, l$. The first statement is implied by arguments showing consistency without estimated weights. The second statement follows from $1_i \| \hat{Q}_{x,i}^{-1} \| \le C \chi_1 T$, $1_i \tau_i T \le \chi_2 T$ (see control of $I_{12}$ below), and an argument as for $I_{1111}$. Let us now prove that $I_{111411} = o_p(1)$. For this purpose, it is enough to show that

$$
J_1 := \frac{1}{\sqrt{n}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T (\hat{Q}_{x,i}^{-1} (S_{i,t} - S_{i,t}) \hat{Q}_{x,i}^{-1}) b_i Y_i', T = o_p(1),
$$

for any $k, l$. By using $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} - x_i' (\hat{\beta}_i - \beta_i) = \varepsilon_{i,t} - \frac{\tau_i T}{\sqrt{T}} x_i' \hat{Q}_{x,i}^{-1} Y_i', T$, we get:

$$
\hat{S}_{i,t} - S_{i,t} = \frac{1}{T_i} \sum_t I_{i,t} (\varepsilon_{i,t}^2 - \varepsilon_{i,t}^2) x_i x_i' + \frac{1}{T_i} \sum_t I_{i,t} (\varepsilon_{i,t}^2 x_i x_i' - S_{i,t}) \\
= \frac{\tau_i T}{\sqrt{T}} W_{i,t} - \frac{2\tilde{\tau}_i^2 T}{T} W_{i,t} Q_{x,i}^{-1} Y_i', T + \frac{\tau_i^2 T}{T} \hat{Q}_{x,i}^{(4)} Q_{x,i}^{-1} Y_i', T \hat{Q}_{x,i} Q_{x,i}^{-1},
$$

where $W_{i,t} := \frac{1}{\sqrt{T}} \sum_t I_{i,t} \eta_{i,t}$, $\eta_{i,t} = \varepsilon_{i,t}^2 x_i x_i' - E[\varepsilon_{i,t}^2 x_i x_i' \mid \gamma_i]$, $W_{i,t} := \frac{1}{\sqrt{T}} \sum_t I_{i,t} \varepsilon_{i,t} x_i^2$,

$$
\hat{Q}_{x,i}^{(4)} := \frac{1}{\sqrt{T}} \sum_t I_{i,t} x_i x_i' and x_i has been treated as a scalar to ease notation. Then:

$$
J_1 = \frac{1}{\sqrt{nT}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T \hat{Q}_{x,i}^{-1} W_{i,t} \hat{Q}_{x,i}^{-1} b_i Y_i', T - \frac{2}{\sqrt{nT}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T \hat{Q}_{x,i}^{-1} W_{i,t} \hat{Q}_{x,i}^{-1} b_i Y_i', T \\
+ \frac{1}{\sqrt{nT}} \sum_i 1_i v_i^{-2} \tilde{\tau}_i^2 T \hat{Q}_{x,i}^{-1} \hat{Q}_{x,i}^{(4)} Y_i', T \hat{Q}_{x,i}^{-1} b_i Y_i', T =: J_{11} + J_{12} + J_{13}.
$$
Let us consider $J_{11}$. We have:

\[
E \left[ \| J_{11} \|^2 | x_T, I_T, \{ \gamma_i \} \right] \leq \frac{C}{nT^3} \sum_{i,j} \sum_{t,i,t_2,t_3,t_4} 1_i^T \gamma_i^2 \tau_{i,T}^2 \tau_{i,T}^2 \| \hat{Q}_{x,i}^{-1} \|^2 \| \hat{Q}_{x,j}^{-1} \|^2 \| E \left[ \eta_{i,t_1} \xi_{i,t_2} \xi_{j,t_3} \xi_{j,t_4} | x_T, \gamma_i, \gamma_j \right] \|.
\]

By using $1_i^T \| \hat{Q}_{x,i}^{-1} \| \leq C_{1} \chi_1, 1_i^T \tau_{i,T} \leq \chi_2$, the Law of Iterated Expectations and Assumptions C.4 b) and C.5, we get $E \left[ \| J_{11} \|^2 \right] = o(1)$. Thus $J_{11} = o_p(1)$. By similar argument and using Assumptions C.4 a), c), we get $J_{12} = o_p(1)$ and $J_{13} = o_p(1)$. Hence $J_1 = o_p(1)$. Paralleling the detailed arguments provided above, we can show that all other remaining terms making $I_{1114}$ are also $o_p(1)$. Hence $I_{11} = O_p(1)$.

To control $I_{12}$, we have:

\[
I_{12} = \frac{1}{\sqrt{n}} \sum_i 1_i^T v_i^{-1} \tau_{i,T} b_i c_i \left( \hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1} \right) Y_{i,T}
\]

\[
+ \frac{1}{\sqrt{n}} \sum_i 1_i^T (\hat{v}_i - v_i^{-1}) \tau_{i,T} b_i c_i \left( \hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1} \right) Y_{i,T} =: I_{121} + I_{122}.
\]

From $\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1} = -\hat{Q}_x^{-1} \left( \frac{1}{T} \sum_t I_{i,t} x_t x_t' - Q_x \right) \hat{Q}_x^{-1} = -\tau_{i,T} \hat{Q}_x^{-1} W_{i,T} \hat{Q}_{x,i}^{-1} + \hat{Q}_x^{-1} W_T \hat{Q}_{x,i}^{-1},$ where

\[
W_{i,T} = \frac{1}{T} \sum_t I_{i,t} (x_t x_t' - Q_x) \text{ and } W_T = \frac{1}{T} \sum_t (x_t x_t' - Q_x), \text{ we can write:}
\]

\[
I_{121} = -\frac{1}{\sqrt{n}} \sum_i 1_i^T v_i^{-1} \tau_{i,T} \tau_{i,T} b_i c_i \hat{Q}_x^{-1} W_{i,T} \hat{Q}_{x,i}^{-1} Y_{i,T} + \frac{1}{\sqrt{n}} \sum_i 1_i^T v_i^{-1} \tau_{i,T} b_i c_i \hat{Q}_x^{-1} W_T \hat{Q}_{x,i}^{-1} Y_{i,T}
\]

\[
= \left( -\frac{1}{\sqrt{n}} \sum_i 1_i^T v_i^{-1} \tau_{i,T} \tau_{i,T} b_i c_i \hat{Q}_x^{-1} W_{i,T} \hat{Q}_{x,i}^{-1} Y_{i,T} + \frac{1}{\sqrt{n}} \sum_i 1_i^T v_i^{-1} \tau_{i,T} b_i c_i \hat{Q}_x^{-1} W_T \hat{Q}_{x,i}^{-1} Y_{i,T} \right) \hat{Q}_x^{-1} c_i
\]

\[
= (I_{1211} + I_{1212}) \hat{Q}_x^{-1} c_i.
\]

Let us consider term $I_{1211}$. From Assumption C.3 we have:

\[
E \left[ \| I_{1211} \|^2 | x_T, I_T, \{ \gamma_i \} \right] \leq \frac{C_4 \chi_2 \chi_2 T}{nT} \sum_{i,j} \sum_t 1_i^T 1_j^T | \sigma_{i,t} | \| \hat{Q}_{x,i}^{-1} \| \| \hat{Q}_{x,j}^{-1} \| \| W_{i,T} \| \| W_{j,T} \|.
\]

Now, by using that $\| \hat{Q}_{x,i}^{-1} \|^2 = Tr \left( \hat{Q}_{x,i}^{-2} \right) = \sum_{k=1}^{K+1} \lambda_k^2 \leq \frac{K+1}{\min \left( \lambda_{x,i} \right)} = \frac{K+1}{\max \left( \lambda_{x,i} \right)} C \left( \hat{Q}_{x,i} \right)^2$, where the $\lambda_k$ are the eigenvalues of matrix $\hat{Q}_{x,i}$, and $\min \left( \hat{Q}_{x,i} \right) \geq 1$, we get $1_i^T \| \hat{Q}_{x,i}^{-1} \| \leq C \chi_1$ and:

\[
E \left[ \| I_{1211} \|^2 | x_T, I_T, \{ \gamma_i \} \right] \leq \frac{C_4 \chi_2 \chi_2 T}{nT} \sum_{i,j} \sum_t | \sigma_{i,t} | \| W_{i,T} \| \| W_{j,T} \|.
\]

50
Then, from Cauchy-Schwartz inequality and Assumption A.1 c), we get
\[ E \left[ \left| I_{1211} \right|^2 \right| \gamma_i \right] \leq C M_x^2 \gamma_i^4 \sup_i \left[ \left| W_{i,T} \right|^4 |\gamma_i| \right]^{1/2}. \]
From Assumption C.2 we have \( \sup_{\gamma \in [0,1]} E \left[ \left| W_{i,T} \right|^4 |\gamma_i| \right] \leq \sup_{\gamma \in [0,1]} E \left[ \left| W_{i,T} \right|^4 |\gamma_i| \right] = O(T^{-c}) \). It follows \( I_{1211} = o_p(1) \). Similarly \( I_{1212} = o_p(1) \), and then \( I_{121} = o_p(1) \). We can also show that \( I_{122} = o_p(1) \), which yields \( I_{12} = o_p(1) \). Hence, \( I_1 = O_p(1) \).

Consider now \( I_2 \). We have:
\[
\frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} = \frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} + \frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \left( \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \right) Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1}
\]
\[
+ \frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \left( \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \right) Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \left( \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \right) Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1}
\]
\[=: I_{21} + I_{22} + I_{23} + I_{24}. \]

Let us control the four terms. We get \( I_{21} = O_p(1) \) by using a decomposition similar to \( I_{111} \) and for the leading term
\[
\left\| \frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \right\| \leq C \left\| \tilde{Q}_i^{-1} \right\|^2 \frac{1}{n} \sum_i \left\| Y_{i,T} \right\|^2 \text{ and } E \left[ \left\| Y_{i,T} \right\|^2 |x_T, I_T, \{\gamma_i\} \right] \leq C.
\]
Moreover, we get \( I_{22} = o_p(1) \) by using a decomposition similar to \( I_{111} \) and for the leading term
\[
\left\| \frac{1}{n} \sum_i w_i^2 T_{x,i}^2 \left( \tilde{Q}_i^{-1} Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \right) Y_{i,T} Y_{x,i}^{-1} \tilde{Q}_i^{-1} \right\| \leq C \left\| \tilde{Q}_i^{-1} \right\|^2 \chi_1 \frac{1}{n} \sum_i \left\| Y_{i,T} \right\|^2 \text{ (see control of term } I_{211} \text{).}
\]
Similarly, we get that \( I_{23} = o_p(1) \) and \( I_{24} = o_p(1) \). Hence, \( I_2 = O_p(1) \).

Finally, we have:
\[
\tilde{Q}_b - Q_b = \left( \frac{1}{n} \sum_i w_i^2 b_i b_i' - Q_b \right) + \frac{1}{n} \sum_i \left( \tilde{w}_i - w_i \right) b_i b_i'
\]
\[
+ \frac{1}{n} \sum_i \tilde{w}_i \left( \tilde{b}_i - b_i \right) b_i' + \frac{1}{n} \sum_i \tilde{w}_i \left( \tilde{b}_i - b_i \right)' + \frac{1}{n} \sum_i \tilde{w}_i \left( \tilde{b}_i - b_i \right) \left( \tilde{b}_i - b_i \right)'
\]
\[
= \left( \frac{1}{n} \sum_i w_i^2 b_i b_i' - Q_b \right) + \frac{1}{n} \sum_i \left( \tilde{w}_i - w_i \right) b_i b_i' + \frac{1}{n \sqrt{T}} \sum_i \tilde{w}_i \tau_{x,i} E_2 \tilde{Q}_x^{-1} Y_{i,T} \tilde{b}_i
\]
\[
+ \frac{1}{n \sqrt{T}} \sum_i \tilde{w}_i \tau_{x,i} b_i Y_{i,T} \tilde{Q}_x^{-1} E_2 + \frac{1}{n \sqrt{T}} \sum_i \tilde{w}_i \tau_{x,i} E_2 \tilde{Q}_x^{-1} Y_{i,T} \tilde{b}_i \tilde{Y}_{i,T} \tilde{Q}_x^{-1} E_2
\]
\[=: I_3 + I_4 + I_5 + I_5' + I_6. \tag{21}
\]

From Assumption SC.2, we have \( I_3 = o_p(1) \), and \( I_4 = o_p(1) \) follows from Lemma 1. Moreover, by similar arguments as for terms \( I_1 \) and \( I_2 \), we can show that \( I_5 \) and \( I_6 \) are \( o_p(1) \). Then, from Equation (21), we get
$\hat{Q}_b - Q_b = o_p(1)$. Thus, from (20) we deduce that $\|\hat{\nu} - \nu\| = O_p\left(\frac{1}{\sqrt{nT}} + \frac{1}{T}\right) = o_p(1)$.

b) Consistency of $\hat{\lambda}$. By Assumption C.1a), we have $\frac{1}{T} \sum_t f_t - E[f_t] = o_p(1)$, and thus

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

### A.2.3 Proof of Proposition 3

a) Asymptotic normality of $\hat{\nu}$. From Equation (20), we have:

$$
\sqrt{nT} \left(\hat{\nu} - \frac{1}{T} \hat{B}_\nu - \nu\right) = Q_b^{-1} I_1 + Q_b^{-1} \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_i^2 \left( \hat{E}^2 Q_{x,i}^{-1} Y_i, T Y_i, T \hat{Q}_{x,i}^{-1} c_0 - \tau_i^{-1} \hat{E}^2 \hat{Q}_{x,i}^{-1} \hat{S}_i \hat{Q}_{x,i}^{-1} c_0 \right)
+ o_p(1) =: Q_b^{-1} I_1 + Q_b^{-1} I_7 + o_p(1).
$$

Let us first show that $Q_b^{-1} I_1$ is asymptotically normal. From the proof of Proposition 2 and the properties of the vec operator and Kronecker product, we have:

$$
Q_b^{-1} I_1 = Q_b^{-1} \left( \frac{1}{\sqrt{n}} \sum_i w_i \tau_i b_i Y_i, T \right) \hat{Q}_{x,i}^{-1} c_0 + o_p(1) = \left( c_i \hat{Q}_{x,i}^{-1} \otimes Q_b^{-1} \right) \frac{1}{\sqrt{n}} \sum_i w_i \tau_i vec[b_i Y_i, T] + o_p(1)
$$

Then we deduce $Q_b^{-1} I_1 \Rightarrow N(0, \Sigma_\nu)$, by Assumptions A.2a) and C.1a).

Let us now show that $I_7 = o_p(1)$. We have:

$$
I_7 = \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_i^2 E^2 \hat{Q}_{x,i}^{-1} \left( Y_i, T Y_i, T - S_{ii, T} \right) \hat{Q}_{x,i}^{-1} c_0 - \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_i^2 E^2 \hat{Q}_{x,i}^{-1} \left( \tau_i^{-1} \hat{S}_{ii} - S_{ii, T} \right) \hat{Q}_{x,i}^{-1} c_0
$$

$$
- \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_i T E^2 \hat{Q}_{x,i}^{-1} \left( \hat{S}_{ii} - \hat{S}_{ii} \right) \hat{Q}_{x,i}^{-1} c_0 - \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_i T E^2 \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} (c_0 - c_\nu)
$$

$$
=: I_{71} - I_{72} - I_{73} - I_{74},
$$

where $\hat{S}_{ii}^0 = \frac{1}{T_i} \sum_t I_{i,t} e_{i,t} x_t x'_t$ and $S_{ii, T} = \frac{1}{T} \sum_t I_{i,t} \sigma_{ii, T} x_t x'_t$. The four terms are bounded in the next Lemma.

**Lemma 2** Under Assumptions C.1a), b), C.3-C.5, $I_{71} = O_p\left(\frac{1}{\sqrt{T}}\right)$, $I_{72} = O_p\left(\frac{1}{T}\right)$, $I_{73} = O_p\left(\frac{\sqrt{n}}{T^{\sqrt{T}}}\right)$ and $I_{74} = O_p\left(\frac{1}{T} + \frac{\sqrt{n}}{T^{\sqrt{T}}}\right)$.
Then, from \( n = O(T^3) \), we get \( I_T = o_p(1) \) and the conclusion follows.

b) Asymptotic normality of \( \lambda \). We have \( \sqrt{T} (\hat{\lambda} - \lambda) = \frac{1}{\sqrt{T}} \sum_t (f_t - E[f_t]) + \sqrt{T} (\hat{\nu} - \nu) \). 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.2b).

**A.2.4 Proof of Proposition 4**

From Proposition 3, we have to show that \( \tilde{\Sigma}_\nu - \Sigma_\nu = o_p(1) \). By \( \Sigma_\nu = (c' Q_x^{-1} \otimes Q_b^{-1}) S_b (Q_x^{-1} c \otimes Q_b^{-1}) \) and \( \tilde{\Sigma}_\nu = (c' \tilde{Q}_x^{-1} \otimes \tilde{Q}_b^{-1}) \tilde{S}_b (\tilde{Q}_x^{-1} c \otimes \tilde{Q}_b^{-1}) \), where \( \tilde{S}_b = \frac{1}{n} \sum_{i,j} \tilde{w}_i \tilde{w}_j \frac{\tau_i \tau_j}{\tau_{ij}} (\tilde{S}_{ij} - S_{ij}) \otimes \tilde{b}_i \tilde{b}_j \), the statement follows if \( \tilde{S}_b - S_b = o_p(1) \). The leading term in \( \tilde{S}_b - S_b \) is given by \( I_8 = \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} (\tilde{S}_{ij} - S_{ij}) \otimes b_i b_j \), while the other ones can be shown to be \( o_p(1) \) by arguments similar to the proofs of Propositions 2 and 3. By using that \( \tau_i \leq M, \tau_{ij} \geq 1, w_i \leq M \) and \( \|b_i\| \leq M, I_8 = o_p(1) \) follows if we show: \( \frac{1}{n} \sum_{i,j} \|\tilde{S}_{ij} - S_{ij}\| = o_p(1) \).

For this purpose, we introduce the following Lemmas 3 and 4 that extend results in Bickel and Levina (2008) from the i.i.d. case to the time series case.

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

**Lemma 4** Under Assumptions C.1 and C.3, if \( \kappa = M \sqrt{\frac{\log n}{T \eta}} \) with \( M \) large, then \( n^2 \Psi_nT ((1 - v) \kappa) = O(1) \), for any \( v \in (0, 1) \), and \( \psi_nT = O_p \left( \sqrt{\frac{\log n}{T \eta}} \right) \).

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

**A.2.5 Proof of Proposition 5**

By definition of \( \tilde{Q}_e \), we get the following result:

**Lemma 5** Under \( \mathcal{H}_0 \) and Assumption A.2a), we have \( \tilde{Q}_e = \frac{1}{n} \sum_i \tilde{w}_i \left[ c' \left( \tilde{\beta}_i - \beta_i \right) \right]^2 + O_p \left( \frac{1}{nT} + \frac{1}{T^2} \right) \).
From Lemmas 1 and 5, it follows: \( \dot{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \left[ c'_p \sqrt{T} \left( \hat{\beta}_i - \beta_i \right) \right] - \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} c_p \right \} + o_p (1). \)

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

\[
\dot{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( Y_{i,T} \hat{Y}_{i,T} - \tau_{i,T} \hat{S}_{ii} \right) \hat{Q}_{x,i}^{-1} c_p + o_p (1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( Y_{i,T} \hat{Y}_{i,T} - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_p - \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( \tau_{i,T} \hat{S}_{ii} - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_p
\]

\[
+ o_p (1) =: I_{g1} + I_{g2} + o_p (1).
\]

We have \( I_{g1} = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( Y_{i,T} \hat{Y}_{i,T} - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_p + o_p (1) \) by arguments similar to the proof of Proposition 2 (see control of \( I_{111} \)). By using \( \tau_{i,T} \hat{S}_{ii} - S_{ii,T} = \frac{1}{T} \sum_t I_{i,t} \left( \hat{\beta}_{i,t} - \beta_{i,t} \right) x_{i,t} x_{i,t}' + \frac{1}{T} \sum_t I_{i,t} \left( \hat{\sigma}_{ii,t} - \sigma_{ii,t} \right) x_{i,t} x_{i,t}' \) and an argument similar to the proof of Proposition 2 (see control of \( J_1 \)), we can show that \( I_{g2} = O_p (\sqrt{n}/T + 1/\sqrt{T}) \). By using \( n = o (T^2) \), it follows \( I_{g2} = o_p (1) \). Then,

\[
\dot{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( Y_{i,T} \hat{Y}_{i,T} - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_p + o_p (1). \]

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

\[
\dot{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( Y_{i,T} \hat{Y}_{i,T} - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_p + o_p (1)
\]

\[
= \left( vec \left[ \hat{Q}_{x,i}^{-1} c_p c'_p \hat{Q}_{x,i}^{-1} \right] \right)' \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} c'_p \hat{Q}_{x,i}^{-1} \left( Y_{i,T} \otimes Y_{i,T} - vec \left[ S_{ii,T} \right] \right) + o_p (1).
\]

By using Assumption A.4, and by consistency of \( \hat{\nu} \) and \( \hat{Q}_x \), we get \( \dot{\xi}_{nT} \Rightarrow N (0, \Sigma_{\xi}) \), where \( \Sigma_{\xi} = \left( vec \left[ Q_x^{-1} c_p c'_p Q_x^{-1} \right] \right)' \Omega \left( vec \left[ Q_x^{-1} c_p c'_p Q_x^{-1} \right] \right) \). By using MN Theorem 3 Chapter 2, we have

\[
vec \left[ Q_x^{-1} c_p c'_p Q_x^{-1} \right]' (S_{ij} \otimes S_{ij}) vec \left[ Q_x^{-1} c_p c'_p Q_x^{-1} \right] = tr \left[ S_{ij} Q_x^{-1} c_p c'_p Q_x^{-1} S_{ij} Q_x^{-1} c_p c'_p Q_x^{-1} \right] = \left( c'_p Q_x^{-1} S_{ij} Q_x^{-1} c_p \right)^2,
\]

(23)

and

\[
vec \left[ Q_x^{-1} c_p c'_p Q_x^{-1} \right]' (S_{ij} \otimes S_{ij}) W_{(K+1)} vec \left[ Q_x^{-1} c_p c'_p Q_x^{-1} \right] = \left( c'_p Q_x^{-1} S_{ij} Q_x^{-1} c_p \right)^2.
\]

(24)

Then, from the definition of \( \Omega \) and Equations (23) and (24), we deduce \( \Sigma_{\xi} = 2 \lim_{n \to \infty} \frac{1}{n} \sum_{i,j} \hat{w}_i \hat{w}_j \frac{\tau_{i,j}^2 + \tau_{j,i}^2}{\tau_{i,j}^2} \left( c'_p Q_x^{-1} S_{ij} Q_x^{-1} c_p \right)^2. \) Finally, \( \Sigma_{\xi} = \Sigma_{\xi} + o_p (1) \) follows from
\[
\frac{1}{n} \sum_{i,j} \| \hat{S}_{ij} - S_{ij} \| = o_p(1) \text{ and } \frac{1}{n} \sum_{i,j} \| \hat{S}_{ij} - S_{ij} \|^2 = o_p(1).
\]

### A.2.6 Proof of Proposition 6

#### a) Asymptotic normality of \( \hat{\nu} \)

By definition of \( \hat{\nu} \) and under \( \mathcal{H}_1 \), we have

\[
\hat{\nu} - \nu_{\infty} = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \tilde{w}_i \tilde{b}_i \epsilon_{\nu_{\infty}} \beta_i = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \tilde{w}_i \tilde{b}_i \epsilon_{\nu_{\infty}} \left( \hat{\beta}_i - \beta_i \right) + \hat{Q}_b^{-1} \frac{1}{n} \sum_i \tilde{w}_i \tilde{b}_i e_i.
\]

Thus we get:

\[
\sqrt{n} (\hat{\nu} - \nu_{\infty}) = \hat{Q}_b^{-1} \frac{1}{\sqrt{nT}} \sum_i \tilde{w}_i \tau_i, T \hat{b}_i \epsilon_{\nu_{\infty}} \hat{Q}_{x,i}^{-1} Y_{i,T} + \hat{Q}_b^{-1} \frac{1}{\sqrt{n}} \sum_i \tilde{w}_i \hat{b}_i e_i + \hat{Q}_b^{-1} \frac{1}{\sqrt{nT}} \sum_i \tilde{w}_i \tau_i, T e_i E_2^i \hat{Q}_{x,i}^{-1} Y_{i,T}.
\]

From Assumption SC.2 and \( E_G \left[ w_i b_i e_i \right] = 0 \), we get \( \frac{1}{\sqrt{n}} \sum_i \tilde{w}_i \hat{b}_i e_i \Rightarrow N \left( 0, E_G \left[ b_i^2 e_i^2 \right] \hat{Q}_b^{-1} \right) \) by the CLT. Thus \( I_{102} \Rightarrow N \left( 0, Q_b^{-1} E_G \left[ w_i^2 e_i^2 b_i^2 \right] Q_b^{-1} \right) \). Then the asymptotic distribution of \( \hat{\nu} \) follows if terms \( I_{101}, I_{103} \) and \( I_{104} \) are \( o_p(1) \). From similar arguments as in the proof of Proposition 2 (control of term \( I_1 \)), we have \( \frac{1}{\sqrt{n}} \sum_i \tilde{w}_i \tilde{b}_i \tilde{e}_i \hat{Q}_{x,i}^{-1} Y_{i,T} = O_p(1) \) and \( \frac{1}{\sqrt{n}} \sum_i \tilde{w}_i \tilde{e}_i \hat{Q}_{x,i}^{-1} Y_{i,T} = O_p(1) \). Thus \( I_{101} = o_p(1) \) and \( I_{104} = o_p(1) \). Moreover, term \( I_{103} \) is \( o_p(1) \) from Lemma 1.

#### b) Asymptotic normality of \( \hat{\lambda} \)

We have \( \sqrt{T} \left( \hat{\lambda} - \lambda_{\infty} \right) = \sqrt{T} (\hat{\nu} - \nu_{\infty}) + \frac{1}{\sqrt{T}} \sum_t (f_t - E[f_t]) \). By using \( \sqrt{T} (\hat{\nu} - \nu_{\infty}) = o_p \left( \sqrt{T} \frac{T}{n} \right) = o_p(1) \), the conclusion follows.

#### c) Consistency of the test

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

**Lemma 6** Under \( \mathcal{H}_1 \) and Assumption A.2a), we have \( \hat{Q}_e = \frac{1}{n} \sum_i \tilde{w}_i \left[ \epsilon_{\nu_{\infty}} \left( \hat{\beta}_i - \beta_i \right) \right]^2 + \frac{1}{n} \sum i \tilde{w}_i e_i^2 + O_p \left( \frac{1}{\sqrt{nT}} \right) \).
By similar arguments as in the proof of Proposition 4, we get:

\[
\hat{\xi}_{nT} = \frac{1}{\sqrt{n}} \sum_i w_i \tau_i^2 c_0' \hat{Q}_x^{-1} (Y_i, T Y_i' - S_{ii, T}) \hat{Q}_x^{-1} c_0 + T \frac{1}{\sqrt{n}} \sum_i w_i e_i^2 + O_p \left( \sqrt{T} \right)
\]

= \quad O_p (1) + O \left( T \sqrt{n} \mathbb{E}_G \left[ w_i (a_i - b_i \nu_{\infty})^2 \right] \right) + O_p (T).

Under $\mathcal{H}_1$ we have $\mathbb{E}_G \left[ w_i (a_i - b_i \nu_{\infty})^2 \right] > 0$, since $w_i > 0$ and $(a_i - b_i \nu_{\infty})^2 > 0$, P-a.s. ■
Appendix 3: Conditional factor model

A.3.1 Derivation of Equations (12) and (13)

From Equation (11) and by using $vec[ABC] = [C' \otimes A] vec[B]$ (MN Theorem 2, p. 35), we get

$$Z'_{t-1}B'_if_t = vec[Z'_{t-1}B'_if_t] = [f'_i \otimes Z'_{t-1}] vec[B'_i],$$

and

$$Z'_{t-1}C'_if_t = [f'_i \otimes Z'_{t-1}] vec[C'_i],$$

which gives

$$Z'_{t-1}B'_if_t + Z'_{t-1}C'_if_t = x'_{2,i,t} \beta_{2,i}.$$

a) By definition of matrix $X_t$ in Section 3.1, we have

$$Z'_{t-1}B'_i(\Lambda - F)Z_{t-1} = \frac{1}{2}Z'_{t-1} [B'_i(\Lambda - F) + (\Lambda - F)' B_i] Z_{t-1}$$

$$= \frac{1}{2} \text{vech} [X_t]' \text{vech} [B'_i(\Lambda - F) + (\Lambda - F)' B_i].$$

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

$$\text{vech} [B'_i(\Lambda - F) + (\Lambda - F)' B_i] = D^+_p \text{[vec} B'_i(\Lambda - F)] + \text{vec} [(\Lambda - F)' B_i].$$

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

$$\frac{1}{2}D^+_p \text{[vec} B'_i(\Lambda - F)] + \text{vec} [(\Lambda - F)' B_i] = \frac{1}{2}D^+_p \text{[(\Lambda - F)' \otimes I_p + I_p \otimes (\Lambda - F)' W_{p,K} vec} B'_i].$$

b) By definition of matrix $X_{i,t}$ in Section 3.1, we have

$$Z'_{i-1,t}C'_i(\Lambda - F)Z_{t-1} = vec[Z_{t-1}Z'_{i-1,t-1}] vec[C'_i(\Lambda - F)] = vec [X_{i,t}] [(\Lambda - F)' \otimes I_q] vec [C'_i].$$

By combining a) and b), we deduce $Z'_{t-1}B'_i(\Lambda - F)Z_{t-1} + Z'_{i-1,t}C'_i(\Lambda - F)Z_{t-1} = x'_{1,i,t} \beta_{1,i}$ and

$$\beta_{1,i} = \Psi \beta_{2,i}.$$

A.3.2 Derivation of Equation (14)

a) From the properties of the vec operator, we get

$$\text{vec} [B'_i(\Lambda - F)] + \text{vec} [(\Lambda - F)' B_i] = (I_p \otimes B'_i) vec [\Lambda - F] + (B'_i \otimes I_p) vec [\Lambda' - F'].$$

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

$$\frac{1}{2}D^+_p \text{[vec} B'_i(\Lambda - F)] + \text{vec} [(\Lambda - F)' B_i] = \frac{1}{2}D^+_p \text{[(I_p \otimes B'_i) W_{p,K} + B'_i \otimes I_p]} \nu.$$
By properties of commutation and duplication matrices (MN p. 54-58), we have \((I_p \otimes B_i') W_{p,K} = W_p (B_i' \otimes I_p)\) and \(D_p^+ W_p = D_p^+\), then \(\frac{1}{2} D_p^+ \left[ (I_p \otimes B_i') W_{p,K} + B_i' \otimes I_p \right] = D_p^+ (B_i' \otimes I_p)\).

b) From the properties of the vec operator, we get

\[
\text{vec} \left[ C'_i (\Lambda - F) \right] = (I_p \otimes C'_i) \text{vec} [\Lambda - F] = (I_p \otimes C'_i) W_{p,K} \text{vec} [\Lambda' - F'] = W_{p,q} (C'_i \otimes I_p) \nu.
\]

A.3.3 Derivation of Equation (15)

a) By MN Theorem 2 p. 35 and Exercise 1 p. 56, and by writing \(I_{pK} = I_K \otimes I_p\), we obtain

\[
\text{vec} \left[ D_p^+ (B_i' \otimes I_p) \right] = (I_{pK} \otimes D_p^+) \text{vec} [B_i' \otimes I_p]
\]

\[
= (I_{pK} \otimes D_p^+) \{I_K \otimes [(W_p \otimes I_p) (I_p \otimes \text{vec} [I_p])] \} \text{vec} [B_i']
\]

\[
= \{ I_K \otimes [(I_p \otimes D_p^+) (W_p \otimes I_p) (I_p \otimes \text{vec} [I_p])] \} \text{vec} [B_i'].
\]

Moreover, \(\text{vec} \left[ \{ D_p^+ (B_i' \otimes I_p) \} \right] \) = \(W_{p(p+1)/2,pK} \text{vec} \left[ D_p^+ (B_i' \otimes I_p) \right] \).

b) Similarly, \(\text{vec} \left[ W_{p,q} (C'_i \otimes I_p) \right] = \{ I_K \otimes [(I_p \otimes W_{p,q}) (W_{p,q} \otimes I_p) (I_q \otimes \text{vec} [I_p])] \} \text{vec} [C'_i]\) and \(\text{vec} \left[ \{ W_{p,q} (C'_i \otimes I_p) \} \right] \) = \(W_{pq,pK} \text{vec} \left[ W_{p,q} (C'_i \otimes I_p) \right] \).

By combining a) and b) and using \(\text{vec} \left[ \beta_{3,i} \right] = \left( \text{vec} \left[ \{ D_p^+ (B_i' \otimes I_p) \} \right] , \text{vec} \left[ \{ W_{p,q} (C'_i \otimes I_p) \} \right] \right)\) the conclusion follows.

A.3.4 Proof of Proposition 8

a) Consistency of \(\hat{\nu}\). By definition of \(\hat{\nu}\) we have: \(\hat{\nu} - \nu = \hat{Q}^{-1}_{\beta_3} \frac{1}{n} \sum_i \hat{\beta}_{3,i} \hat{w}_i \left( \hat{\beta}_{1,i} - \hat{\beta}_{3,i} \nu \right)\). From Equation (15) and MN Theorem 2 p. 35, we get \(\hat{\beta}_{3,i} \nu = \text{vec}[\nu' \hat{\beta}_{3,i}'] = (I_{d_1} \otimes \nu') \text{vec}[\hat{\beta}_{3,i}'] = (I_{d_1} \otimes \nu') J_{a} \hat{\beta}_{3,i}\).

Moreover, by using matrices \(E_1\) and \(E_2\), we obtain \(\left( \hat{\beta}_{1,i} - \hat{\beta}_{3,i} \nu \right) = [E'_1 - (I_{d_1} \otimes \nu') J_{a} E'_2] \hat{\beta}_i = C'_{\nu} \hat{\beta}_i = C'_{\nu} \left( \hat{\beta}_i - \hat{\beta}_i \right),\) from Equation (14). It follows that

\[
\hat{\nu} - \nu = \hat{Q}^{-1}_{\beta_3} \frac{1}{n} \sum_i \hat{\beta}_{3,i} \hat{w}_i C'_{\nu} \left( \hat{\beta}_i - \hat{\beta}_i \right).
\]

By comparing with Equation (19) and using the same arguments as in the proof of Proposition 1 applied to \(\beta'_3\) instead of \(b\), the result follows.
b) Consistency of \( \Lambda' \). By definition of \( \nu \), we deduce \( \| \text{vec} \left[ \hat{\Lambda}' - \Lambda' \right] \| \leq \| \hat{\nu} - \nu \| + \| \text{vec} \left[ \hat{F}' - F' \right] \| \). By part a), \( \| \hat{\nu} - \nu \| = o_p(1) \). By LLN and Assumptions C.1a),b) and C.6, 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 \( \| \text{vec} \left[ \hat{F}' - F' \right] \| = o_p(1) \). The result follows.

A.3.5 Proof of Proposition 9

a) Asymptotic normality of \( \hat{\nu} \). From Equation (25) and by using \( \sqrt{T} \left( \hat{\beta}_i - \beta_i \right) = \tau_i T \hat{Q}_{x,i}^{-1} Y_{i,T} \), we get

\[
\sqrt{nT} (\hat{\nu} - \nu) = \hat{Q}_{\beta_3}^{-1} \frac{1}{\sqrt{n}} \sum_i \tau_i T \hat{\beta}_{3,i} \hat{w}_i (\nu C'_{\nu} \hat{Q}_{x,i}^{-1} Y_{i,T})
\]

\[
= \hat{Q}_{\beta_3}^{-1} \frac{1}{\sqrt{n}} \sum_i \tau_i T \hat{\beta}_{3,i} \hat{w}_i (\nu C'_{\nu} \hat{Q}_{x,i}^{-1} Y_{i,T}) + \hat{Q}_{\beta_3}^{-1} \frac{1}{\sqrt{n}} \sum_i \tau_i T \hat{\beta}_{3,i} \hat{w}_i (\nu C'_{\nu} (\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}) Y_{i,T})
\]

\[
+ \hat{Q}_{\beta_3}^{-1} \frac{1}{\sqrt{n}} \sum_i \tau_i T (\hat{\beta}_{3,i} - \beta_{3,i})' \hat{w}_i (\nu C'_{\nu} \hat{Q}_{x,i}^{-1} Y_{i,T}) =: I_{T1} + I_{T2} + I_{T3}.
\]

By MN Theorem 2 p. 35, we have \( I_{T1} = \hat{Q}_{\beta_3}^{-1} \left( \frac{1}{\sqrt{n}} \sum_i \tau_i T \left[ (Y'_{i,T} Q_{x,i}^{-1}) \otimes (\beta_{3,i} \hat{w}_i) \right] \right) \text{vec} [C'_{\nu}] \).

As in the proof of Propositions 2 and 3, we have \( I_{T1} = \hat{Q}_{\beta_3}^{-1} \left( \frac{1}{\sqrt{n}} \sum_i \tau_i T \left[ (Y'_{i,T} Q_{x,i}^{-1}) \otimes (\beta_{3,i} \hat{w}_i) \right] \right) \text{vec} [C'_{\nu}] \)

\[+ o_p(1) =: I_{T11} + o_p(1) \]. We can rewrite \( I_{T11} = \left( \text{vec} [C'_{\nu}] \otimes \hat{Q}_{\beta_3}^{-1} \right) \frac{1}{\sqrt{n}} \sum_i \tau_i \text{vec} \left[ (Y'_{i,T} Q_{x,i}^{-1}) \otimes (\beta_{3,i} \hat{w}_i) \right] \).

Moreover, by using \( \text{vec} \left[ (Y'_{i,T} Q_{x,i}^{-1}) \otimes (\beta_{3,i} \hat{w}_i) \right] = (Q_{x,i}^{-1} Y_{i,T}) \otimes \text{vec} [\beta_{3,i} \hat{w}_i] \) (see MN Theorem 10 p. 55), we get \( I_{T11} = \left( \text{vec} [C'_{\nu}] \otimes \hat{Q}_{\beta_3}^{-1} \right) \frac{1}{\sqrt{n}} \sum_i \tau_i \left[ (Q_{x,i}^{-1} Y_{i,T}) \otimes v_3 \right] \). Then \( I_{T11} \Rightarrow N (0, \Sigma_{\nu}) \) follows from Assumption B.2 a).

Let us consider \( I_{T2} \). By similar arguments as in the proof of Proposition 3, \( I_{T2} = o_p(1) \).

Let us consider \( I_{T3} \). We introduce the following lemma:

**Lemma 7** Let \( A \) be a \( m \times n \) matrix and \( b \) be a \( n \times 1 \) vector. Then, \( Ab = (\text{vec} [I_n]' \otimes I_m) \text{vec} [\text{vec} [A] b'] \).

59
By Lemma 7, Equation (15) and $\sqrt{T} \text{vec} \left[ \left( \hat{\beta}_{3,i} - \beta_{3,i} \right) \right] = \tau_{i,T} J_a E_2 Q_{x,i}^{-1} Y_{i,T}$, we have

$$I_{73} = Q_{\beta_3}^{-1} \frac{1}{\sqrt{nT}} \sum_i \tau_{i,T}^2 \left( \text{vec} [I_{d_1}]' \otimes I_{K_p} \right) \text{vec} \left[ J_a E_2 Q_{x,i}^{-1} Y_{i,T}' Y_{i,T}' Q_{x,i}^{-1} C_{\nu} \hat{w}_i \right]$$

$$= Q_{\beta_3}^{-1} \frac{1}{\sqrt{nT}} \sum_i \tau_{i,T}^2 J_b \text{vec} \left[ E_2 Q_{x,i}^{-1} Y_{i,T}' Y_{i,T}' Q_{x,i}^{-1} C_{\nu} \hat{w}_i \right] =: \sqrt{\frac{n}{T}} \hat{B}_\nu + I_{74},$$

where $I_{74} = o_p(1)$ by similar arguments as in the proof of Proposition 3.

b) Asymptotic normality of $\text{vec} \left( \hat{\Lambda}' \right)$. We have $\sqrt{T} \text{vec} \left[ \hat{\Lambda}' - \Lambda' \right] = \sqrt{T} \text{vec} \left[ \hat{F}' - F' \right] + \sqrt{T} (\hat{\nu} - \nu)$. By using $\sqrt{T} \text{vec} [\hat{F}' - F'] = \left[ I_K \otimes \left( \frac{1}{T} \sum_t Z_{t-1} Z_{t-1}' \right)^{-1} \right] \frac{1}{\sqrt{T}} \sum_t u_t \otimes Z_{t-1}$ and $\sqrt{T} (\hat{\nu} - \nu) = O_p \left( \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} \right) = o_p(1)$, the conclusion follows from Assumption B.2b).

**A.3.6 Proof of Proposition 10**

By similar arguments as in the proof of Proposition 5, we have:

$$\hat{Q}_e = \frac{1}{n} \sum_i \left( \hat{\beta}_i - \beta_i \right)' C_{\nu} \hat{w}_i \hat{C}_{\nu}' \left( \hat{\beta}_i - \beta_i \right) + O_p \left( \frac{1}{nT} + \frac{1}{T^2} \right)$$

$$= \frac{1}{nT} \sum_i \tau_{i,T}^2 \text{tr} \left[ C_{\nu} \hat{Q}_{x,i}^{-1} Y_{i,T}' Y_{i,T}' \hat{Q}_{x,i}^{-1} C_{\nu} \hat{w}_i \right] + O_p \left( \frac{1}{nT} + \frac{1}{T^2} \right).$$

By using that $\tau_{i,T} \left[ C_{\nu} \hat{Q}_{x,i}^{-1} \hat{S}_{i,T} \hat{Q}_{x,i}^{-1} C_{\nu} \hat{w}_i \right] = 1_i d_1$ and Lemma 1 in the conditional case, we get:

$$\hat{Q}_e = \frac{1}{\sqrt{n}} \sum_i \tau_{i,T}^2 \text{tr} \left[ C_{\nu} \hat{Q}_{x,i}^{-1} (Y_{i,T}' Y_{i,T}' - \tau_{i,T}^{-1} \hat{S}_{i,T}) \hat{Q}_{x,i}^{-1} C_{\nu} \hat{w}_i \right] + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_i \tau_{i,T}^2 \text{tr} \left[ C_{\nu} \hat{Q}_{x,i}^{-1} (Y_{i,T}' Y_{i,T}' - S_{i,T}) \hat{Q}_{x,i}^{-1} C_{\nu} \hat{w}_i \right] + o_p(1).$$
Now, by using $tr(ABCD) = vec(D')'(C' \otimes A)vec(B)$ (MN Theorem 3, p. 31) and $vec(ABC) = (C' \otimes A)vec(B)$ for conformable matrices, we have:

$$
tr \left[ C'_v Q^{-1}_{x,i} (Y_{i,T} Y'_{i,T} - S_{i,i}) Q^{-1}_{x,i} C_v w_i \right] = vec[w_i]' (C'_v \otimes C'_v) vec \left[ Q^{-1}_{x,i} (Y_{i,T} Y'_{i,T} - S_{i,i}) Q^{-1}_{x,i} \right] = vec[w_i]' \left( C'_v \otimes C'_v \right) \left( Q^{-1}_{x,i} \otimes Q^{-1}_{x,i} \right) vec \left[ Y_{i,T} Y'_{i,T} - S_{i,i} \right] = vec \left[ C'_v \otimes C'_v \right] \left\{ \left[ \left( Q^{-1}_{x,i} \otimes Q^{-1}_{x,i} \right) (Y_{i,T} \otimes Y_{i,T} - vec[S_{i,i}]) \right] \otimes vec[w_i] \right\}.
$$

Thus, we get $\xi_{nT} = vec \left[ C'_v \otimes C'_v \right]' \left[ \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \left( \left( Q^{-1}_{x,i} \otimes Q^{-1}_{x,i} \right) (Y_{i,T} \otimes Y_{i,T} - vec[S_{i,i}]) \right) \otimes vec[w_i] \right]$. From Assumption B.3, we get $\xi_{nT} \rightarrow N(0, \Sigma_\xi)$, where $\Sigma_\xi = vec \left[ C'_v \otimes C'_v \right]' \Omega vec \left[ C'_v \otimes C'_v \right]$. Now, by using that $tr(ABCD) = vec(D')'(A \otimes C')vec(B')$ (see Theorem 3, p. 31, in MN) we have:

$$
vec \left[ C'_v \otimes C'_v \right] \left[ (S_{Q,ij,i} \otimes S_{Q,ij,j}) \otimes vec[w_i] vec[w_j] \right] vec \left[ C'_v \otimes C'_v \right] = tr \left[ (S_{Q,ij,i} \otimes S_{Q,ij,j}) (C_v \otimes C_v) vec[w_j] vec[w_i]' (C'_v \otimes C'_v) \right] = vec[w_i]' \left( C'_v S_{Q,ij,j} C_v \otimes (C'_v S_{Q,ij,i} C_v) \right) vec[w_j] = tr \left[ (C'_v S_{Q,ij,j} C_v) w_j \right] \left( C'_v S_{Q,ij,i} C_v \right) w_i = tr \left[ (C'_v Q^{-1}_{x,i,j} S_{ij} Q^{-1}_{x,j,i} C_v) w_j \right] \left( C'_v Q^{-1}_{x,i,j} S_{ij} Q^{-1}_{x,j,i} C_v \right) w_i \right],
$$

and similarly $vec \left[ C'_v \otimes C'_v \right] \left[ \left( S_{Q,ij,i} \otimes S_{Q,ij,j} \right) W_q \otimes vec[w_i] \right] vec[w_j]' vec \left[ C'_v \otimes C'_v \right] = tr \left[ \left( C'_v Q^{-1}_{x,i,j} S_{ij} Q^{-1}_{x,j,i} C_v \right) w_j \right] \left( C'_v Q^{-1}_{x,i,j} S_{ij} Q^{-1}_{x,j,i} C_v \right) w_i$. Thus, we get the asymptotic variance matrix

$$
\Sigma_\xi = 2 \lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{1}{n} \sum_{i,j} \tau_i^2 \tau_j^2 tr \left[ \left( C'_v Q^{-1}_{x,i,j} S_{ij} Q^{-1}_{x,j,i} C_v \right) w_j \left( C'_v Q^{-1}_{x,i,j} S_{ij} Q^{-1}_{x,j,i} C_v \right) w_i \right] \right].
$$

From $\hat{\Sigma}_\xi = \Sigma_\xi + o_p(1)$, the conclusion follows. ■

A.3.7 Proof of Equation (17)

We have:

$$
\hat{b}_{i,t} \tilde{\lambda}_i = tr \left[ Z_{i-1} Z'_{i-1} \hat{B}_i \hat{A} \right] + tr \left[ Z_{i-1} Z'_{i-1} \hat{C}_i \hat{A} \right] = (Z'_{i-1} \otimes Z'_{i-1}) vec \left[ \hat{B}_i \hat{A} \right] + (Z'_{i-1} \otimes Z'_{i-1}) vec \left[ \hat{C}_i \hat{A} \right].
$$
Thus, we get:

\[
\sqrt{T} \left( CE_{i,t} - CE_{i,t-1} \right) \\
= (Z'_{t-1} \otimes Z'_{t-1}) \sqrt{T} \left( \text{vec} \left[ \hat{B}' \hat{\Lambda} \right] - \text{vec} \left[ B' \Lambda \right] \right) + (Z'_{i,t-1} \otimes Z'_{i,t-1}) \sqrt{T} \left( \text{vec} \left[ \hat{C}' \hat{\Lambda} \right] - \text{vec} \left[ C' \Lambda \right] \right) \\
= (Z'_{t-1} \otimes Z'_{t-1}) \left[ (\hat{\Lambda}' \otimes I_p) \sqrt{T} \text{vec} \left[ \hat{B}' - B' \right] + (I_p \otimes B'_i) \sqrt{T} \text{vec} \left[ \hat{\Lambda} - \Lambda \right] \right] \\
+ (Z'_{i,t-1} \otimes Z'_{i,t-1}) \left[ (\hat{\Lambda}' \otimes I_q) \sqrt{T} \text{vec} \left[ \hat{C}' - C'_i \right] + (I_p \otimes C'_i) \sqrt{T} \text{vec} \left[ \hat{\Lambda} - \Lambda \right] \right].
\]

By using that \( \hat{\Lambda} = \Lambda + o_p(1) \) and \( \text{vec} \left[ \hat{\Lambda} - \Lambda \right] = W_{p,K'} \text{vec} \left[ \hat{\Lambda}' - \Lambda' \right] \), equation (17) follows. \( \blacksquare \)
Appendix 4: Check of Assumptions A under block dependence

In this appendix we verify that the cross-sectional dependence and asymptotic normality conditions in Assumptions A.1-A.4 are satisfied under a block-dependence structure in a serially i.i.d. framework. Let us assume that:

**BD.1** The errors $\varepsilon_t(\gamma)$ are i.i.d. over time with $E[\varepsilon_t(\gamma)] = 0$, for all $\gamma \in [0,1]$. For any $n$, there exists a partition of the interval $[0,1]$ into $J_n \leq n$ subintervals $I_1, ..., I_{J_n}$, such that $\varepsilon_t(\gamma)$ and $\varepsilon_t(\gamma')$ are independent if $\gamma$ and $\gamma'$ belong to different subintervals, and $J_n \to \infty$ as $n \to \infty$.

**BD.2** The blocks are such that $n \sum_{m=1}^{J_n} |B_m|^2 = O(1)$, $n^{3/2} \sum_{m=1}^{J_n} |B_m|^3 = o(1)$, where $B_m = \int_{I_m} dG(\gamma)$.

**BD.3** The factors $(f_t)$ are i.i.d. over time and independent of the errors $(\varepsilon_t(\gamma))$, $\gamma \in [0,1]$.

**BD.4** There exist a constant $M$ such that $\|f_t\| \leq M$, P-a.s.. Moreover, $\sup_{\gamma \in [0,1]} E[|\varepsilon_t(\gamma)|^6] < \infty$, $\sup_{\gamma \in [0,1]} \|\beta(\gamma)\| < \infty$ and $\inf_{\gamma \in [0,1]} E[I_t(\gamma)] > 0$.

The block-dependence structure as in Assumption BD.1 is satisfied for instance when there are unobserved industry-specific factors independent among industries and over time, as Ang, Liu, Schwartz (2010). In empirical applications, blocks can match industrial sectors. Then, the number $J_n$ of blocks amounts to a couple of dozens, and the number of assets $n$ amounts to a couple of thousands. There are approximately $nB_m$ assets in block $m$, when $n$ is large. In the asymptotic analysis, Assumption BD.2 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.

**Lemma 8** Let Assumptions BD.1-4 on block dependence and Assumptions SC.1-SC.2 on random sampling hold. Then, Assumptions A.1, A.2 (with $\Gamma_1 = \mathbb{R}^+$), A.3 (with any $q \in (0,1)$ and $\delta = 1/2$) and A.4 (with $\Gamma_2 = \mathbb{R}^+$) are satisfied.

In Lemma 8, we have $\Gamma_1 = \Gamma_2 = \mathbb{R}^+$, which means that there is no condition on the relative expansion rates of $n$ and $T$. 

63