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

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

First draft: December 2010. This version: September 2012.

Abstract

We develop an econometric methodology to infer the path of risk premia from a 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 both 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 analysis 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 unconditional estimates, follow the macroeconomic cycles, and do not match risk premia estimates on standard sets of portfolios. The asset pricing restrictions are rejected for a conditional four-factor model capturing market, size, value and momentum effects.

JEL Classification: C12, C13, C23, C51, C52 , G12.

Keywords: large panel, factor model, risk premium, asset pricing.

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

*Acknowledgements: We gratefully acknowledge the financial support of the Swiss National Science Foundation (Prodoc project PDFM11-114533 and NCCR FINRISK). We thank Y. Amihud, L. Barras, S. Bonhomme, A. Buraschi, V. Chernozhukov, G. Dhaene, R. Engle, J. Fan, E. Ghysels, C. Gouriéroux, S. Heston, E. Renault, E. Sentana, F. Trojani, participants at the Cass conference 2010, CIRPEE conference 2011, ECARES conference 2011, CORE conference 2011, Montreal panel data conference 2011, ESEM 2011, SoFiE annual conference 2012, and participants at seminars at Columbia, NYU, Georgetown, Maryland, McGill, GWU, ULB, UCL, Humboldt, Orléans, Imperial College, CREST, Athens, CORE, Bernheim center, Brown, Princeton, Duke, Northwestern, HEC Montreal, CEMFI, Leuven, Rotterdam, Bern, Zurich, Cergy-Pontoise, Luxembourg for helpful comments.

1 Introduction

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

The workhorse to estimate equity risk premia in a linear multi-factor setting is the two-pass cross-sectional regression method developed by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973). A series of papers address its large and finite sample properties for unconditional linear factor models, see e.g. Shanken (1985, 1992), Jagannathan and Wang (1998), Shanken and Zhou (2007), Kan, Robotti and Shanken (2012), and the review paper of Jagannathan, Skoulakis and Wang (2009). The literature has not yet formally addressed statistical inference for equity risk premia in conditional linear factor models despite its empirical relevance.

In this paper, we study how we can infer the time-varying behaviour of equity risk premia from large stock returns databases under 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 (2007, 2009) exemplify this promising route when studying bond risk premia. Connor, Hagmann, and Linton (2012) show that large cross-sections exploit data more efficiently in a semiparametric characteristic-based factor model of stock returns. Our approach relying on individual stocks returns is also inspired by the theoretical framework underlying the Arbitrage Pricing Theory (APT). In this setting, approximate factor structures with nondiagonal error covariance matrices (Chamberlain and Rothschild (1983, CR)) answer 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 idiosyncratic error terms, such approximate factor models 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. Indeed, we let the number of assets grow to infinity mimicking the large economies of financial theory.

The potential loss of information and bias induced by grouping stocks to build portfolios in asset pricing tests further motivate our approach (e.g. 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 we loose a lot of efficiency 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 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)) under an approximate factor structure (Chamberlain and Rothschild (1983)) with a continuum of assets. We explicitly show the relationship between the ruling out of asymptotic arbitrage opportunities and an empirically testable restriction for large economies in a conditional setting. We also formalize the sampling scheme so that observed assets are random draws from an underlying population (Andrews (2005)). Such a construction is close to the setting advocated by Al-Najjar (1995, 1998, 1999) in a static framework with exact factor structure. He discusses several key advantages of using a continuum economy in arbitrage pricing and risk decomposition. 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 estimators 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 in 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 test of the asset pricing restrictions 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 (2012) for a related approach in unconditional models and asymptotics with fixed n). 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)). We know from standard textbooks in corporate finance 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, we first present the three theoretical contributions in an unconditional setting before the extension 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), Ferson and Harvey (1999)). For the conditional versions, we use both macrovariables and firm characteristics as instruments. The estimated paths show 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 recent financial crisis. Furthermore, the conditional risk premia estimates exhibit large positive and negative strays from unconditional estimates, follow the macroeconomic cycles, and do not match risk premia estimates on standard sets of portfolios. The asset pricing restrictions are rejected for a conditional 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 either common to all stocks, or stock specific. Section 4 contains the empirical results. In the Appendices, we gather the technical assumptions and some proofs. We use high-level assumptions to get our results and show in Appendix 4 that we meet all of them under a block cross-sectional dependence structure on the error terms in a serially i.i.d. framework. We place all omitted proofs and the Monte Carlo simulation results in the online supplementary materials. There, we also include some empirical results on the cost of equity and robustness checks.

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 the generating process for the excess returns before examining the implications of absence of arbitrage opportunities in terms of model restrictions. We combine the constructions of Hansen and Richard (1987) and Andrews (2005) to define a multi-period economy with a continuum of assets having strictly stationary and ergodic return processes. We use such a formal construction to guarantee that (i) the economy is invariant to time shifts, so that we can establish all properties by working at $t = 1$, (ii) time series averages converge almost surely to population expectations, (iii) under a suitable sampling mechanism (see the next section), cross-sectional limits exist and are invariant to reordering of the assets, (iv) the derived no-arbitrage restriction is empirically testable. This construction allows reconciling finance and econometric analysis in a coherent framework.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ 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 time-shift transformation S mapping Ω 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. Transformation S is assumed to be measure-preserving and ergodic (i.e., any set in $\mathcal{F}$ invariant under S has measure either 1, or 0).

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

$$R_t(\gamma) = a(\gamma) + b(\gamma)' f_t + \varepsilon_t(\gamma), \tag{1}$$

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 transformation S is measure-preserving and ergodic, all processes are strictly stationary and ergodic (Doob (1953)). Let further define $x_t = (1, f_t')'$ which yields the compact formulation:

$$R_t(\gamma) = \beta(\gamma)' x_t + \varepsilon_t(\gamma). \tag{2}$$

In order to define the information sets, let $\mathcal{F}_0 \subset \mathcal{F}$ be a sub sigma-field. We assume that random vector f is measurable w.r.t. $\mathcal{F}_0$. Define $\mathcal{F}_t = \{S^{-t}(A), A \in \mathcal{F}_0\}$, $t = 1, 2, ...$, through the inverse mapping S^{-t} and assume that $\mathcal{F}_1$ contains $\mathcal{F}_0$. Then, the filtration $\mathcal{F}_t$, $t = 1, 2, ...$, characterizes the flow of 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 impose an approximate factor structure for the conditional distribution of the error terms given $\mathcal{F}_{t-1}$ in almost any countable collection of assets. More precisely, for any sequence (γ_i) in $[0,1]$, let $\Sigma_{\varepsilon,t,n}$ denote the $n \times n$ conditional variance-covariance matrix of the error vector $[\varepsilon_t(\gamma_1), ..., \varepsilon_t(\gamma_n)]'$ given $\mathcal{F}_{t-1}$, for $n \in \mathbb{N}$. Let μ_Γ be the probability measure on the set $\Gamma = [0,1]^{\mathbb{N}}$ of sequences (γ_i) in $[0,1]$ induced by i.i.d. random sampling from a continuous distribution G with support $[0,1]$.

Assumption APR.4 *For any sequence* (γ_i) *in set* $\mathcal{J}$*: (i)* $eig_{\max}(\Sigma_{\varepsilon,t,n}) = o(n)$*, as* $n \to \infty$*,* P*-a.s., (ii)* $\inf_{n\geq 1} eig_{\min}(\Sigma_{\varepsilon,t,n}) > 0$*,* P*-a.s., where* $\mathcal{J} \subset \Gamma$ *is such that* $\mu_\Gamma(\mathcal{J}) = 1$*, and* $eig_{\min}(\Sigma_{\varepsilon,t,n})$ *and* $eig_{\max}(\Sigma_{\varepsilon,t,n})$ *denote the smallest and the largest eigenvalues of matrix* $\Sigma_{\varepsilon,t,n}$*, (iii)* $eig_{\min}(V[f_t|\mathcal{F}_{t-1}]) > 0$*,* P*-a.s.*

Assumption APR.4 (i) is weaker than boundedness of the largest eigenvalue, i.e., $\sup_{n\geq 1} eig_{\max}(\Sigma_{\varepsilon,t,n}) < \infty$, P-a.s., as in CR. This is useful for the checks of Appendix 4 under a block cross-sectional dependence structure. Assumptions APR.4 (ii)-(iii) are mild regularity conditions for 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 $\alpha_{0,n}$ invested in the riskfree asset and the shares $\alpha_{i,n}$ invested in the selected risky assets γ_i, for $i = 1, ..., n$. The shares are measurable w.r.t. $\mathcal{F}_0$. Then, $C(p_n) = \sum_{i=0}^{n} \alpha_{i,n}$ is the portfolio cost at $t = 0$, and $p_n = C(p_n)R_0 + \sum_{i=1}^{n} \alpha_{i,n} R_1(\gamma_i)$ 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* $\lim_{n\to\infty} \mathbb{P}[p_n \geq 0] = 1$ *and* $\lim_{n\to\infty} \mathbb{P}[C(p_n) \leq 0, p_n > 0] > 0$.

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

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 $\gamma \in [0, 1]$.

We can rewrite the asset pricing restriction as

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

for almost all $\gamma \in [0, 1]$, where $\lambda = \nu + E\left[f_t\right]$ 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 $\nu_k = 0$, $k = 1, ..., K$.

Proposition 1 is already stated by Al-Najjar (1998) Proposition 2 for a strict factor structure in an unconditional economy (static case) with the definition of arbitrage as in CR. We extend his result to an approximate factor structure in a conditional economy (dynamic case) with the definition of arbitrage as in Hansen and Richard (1987). Proposition 1 differs from CR Theorem 3 in terms of the returns generating framework, the definition of asymptotic arbitrage opportunities, and the derived asset pricing restriction. Specifically, we consider a multi-period economy with conditional information as opposed to a single period unconditional economy as in CR. We extend such a setting to time varying risk premia in Section 3. We prefer the definition underlying Assumption APR.5 since it corresponds to the definition of arbitrage that is standard in dynamic asset pricing theory (e.g., Duffie (2001)). As pointed out by Hansen and Richard (1987), Ross (1978) has already chosen that type of definition. It also eases the proof based on new arguments. However, in Appendix 2, we derive the link between the no-arbitrage conditions in Assumptions A.1 i) and ii) of CR, written P-a.s. w.r.t. the conditional information $\mathcal{F}_0$ and for almost every countable collection of assets, and the asset pricing restriction (3) valid for the continuum of assets. Hence, we are able to characterize the functions $\beta = (a, b')'$ defined on $[0, 1]$ that are compatible with absence of asymptotic arbitrage opportunities under both definitions of arbitrage in the continuum economy. CR derive the pricing restriction $\sum_{i=1}^{\infty}\left(a(\gamma_i)-b(\gamma_i)'\nu\right)^2<\infty$, for some $\nu\in\mathbb{R}^K$ and for a given sequence (γ_i), while we derive the restriction (3), for almost all $\gamma\in[0,1]$. In Appendix 2, we show that the set of sequences (γ_i) such that $\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}\left(a(\gamma_i)-b(\gamma_i)'\nu\right)^2<\infty$ has measure 1 under μ_Γ, when the asset pricing restriction (3) holds, and measure 0, otherwise. This result is a consequence of the Kolmogorov zero-one law (see e.g. Billingsley (1995)). In other words, validity of the summability condition in CR for a countable collection of assets without validity of the asset pricing restriction (3) is an impossible event. From the proofs in Appendix 2, we also get a reverse implication compared to Proposition 1: when the asset pricing restriction (3) does not hold, asymptotic arbitrage in the sense of Assumption APR.5, or of Assumptions A.1 i) and ii) of CR, exists for μ_Γ-almost any countable collection of assets. The restriction in Proposition 1 is testable with large equity datasets and large sample sizes (Section 2.5). Therefore, we are not affected by the Shanken (1982) critique, namely the problem that finiteness of the sum $\sum_{i=1}^{\infty}\left(a(\gamma_i)-b(\gamma_i)'\nu\right)^2$ for a given countable economy cannot be tested empirically. 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, we do not observe asset returns 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 γ is observable by the econometrician at date t, and 0 otherwise (Connor and Korajczyk (1987)). To keep the factor structure linear, we assume a missing-at-random design (Rubin (1976)), that is, independence between unobservability and returns generation.

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

Another design would require an explicit modeling of the link between the unobservability mechanism and the returns process of the continuum of assets (Heckman (1979)); 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 γ_i, $i = 1, ..., n$, 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 $i = 1, ..., n$, and $t = 1, ..., T$. 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 $\sigma_{ij,t} = E[\varepsilon_{i,t}\varepsilon_{j,t}|x_{\underline{t}}, \gamma_i, \gamma_j]$ the conditional variances and covariances of the assets in the sample, where $x_{\underline{t}} = \{x_t, x_{t-1}, ...\}$. By random sampling, we get a random coefficient panel model (e.g. Hsiao (2003), Chapter 6). The characteristic β_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 γ_i. If the a_is and b_is were treated as given parameters, and not as realizations of random variables, invoking cross-sectional LLNs and CLTs as in some assumptions and parts of the proofs would have no sense. Moreover, cross-sectional limits would be dependent on the selected ordering of the assets. Instead, our assumptions and results do not rely on a specific ordering of assets. Random elements $(\beta_i', \sigma_{ii,t}, \varepsilon_{i,t}, I_{i,t})'$, $i = 1, ..., n$, 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 $i = 1, ..., n$, 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 β_i. To address this, we introduce a trimming device: $\mathbf{1}_i^{\chi} = \mathbf{1}\left\{CN\left(\hat{Q}_{x,i}\right) \leq \chi_{1,T}, \tau_{i,T} \leq \chi_{2,T}\right\}$, where $CN\left(\hat{Q}_{x,i}\right) = \sqrt{eig_{\max}\left(\hat{Q}_{x,i}\right)/eig_{\min}\left(\hat{Q}_{x,i}\right)}$

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 $\{CN\left(\hat{Q}_{x,i}\right) \leq \chi_{1,T}\}$ keeps in the cross-section only assets for which the time series regression is not too badly conditioned. A too large value of $CN\left(\hat{Q}_{x,i}\right)$ 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. We use both trimming conditions in the proofs of the asymptotic results.

Second Pass: The second pass consists in computing a cross-sectional estimator of ν 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 $v_i = \tau_i c_\nu' Q_x^{-1} S_{ii} Q_x^{-1} c_\nu$, where $Q_x = E\left[x_t x_t'\right]$, $S_{ii} = \underset{T\to\infty}{\text{plim}} \frac{1}{T}\sum_t \sigma_{ii,t} x_t x_t' = E\left[\varepsilon_{i,t}^2 x_t x_t' | \gamma_i\right]$, $\tau_i = \underset{T\to\infty}{\text{plim}}\, \tau_{i,T} = E[I_{i,t}|\gamma_i]^{-1}$, and $c_\nu = (1, -\nu')'$. We use the estimates $\hat{v}_i = \tau_{i,T} c_{\hat{\nu}_1}' \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} c_{\hat{\nu}_1}$, where $\hat{S}_{ii} = \frac{1}{T_i}\sum_t I_{i,t}\hat{\varepsilon}_{i,t}^2 x_t x_t'$, $\hat{\varepsilon}_{i,t} = R_{i,t} - \hat{\beta}_i' x_t$ and $c_{\hat{\nu}_1} = (1, -\hat{\nu}_1')'$. To estimate c_ν, we use the OLS estimator $\hat{\nu}_1 = \left(\sum_i \mathbf{1}_i^\chi \hat{b}_i \hat{b}_i'\right)^{-1} \sum_i \mathbf{1}_i^\chi \hat{b}_i \hat{a}_i$, i.e., a first-step estimator with unit weights. The WLS estimator is:

$$\hat{\nu} = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i \hat{a}_i, \tag{5}$$

where $\hat{Q}_b = \frac{1}{n}\sum_i \hat{w}_i \hat{b}_i \hat{b}_i'$ and $\hat{w}_i = \mathbf{1}_i^\chi \hat{v}_i^{-1}$. 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 $v_i = c_\nu' Q_x^{-1} c_\nu \sigma_{ii}$. There, v_i is directly proportional to σ_{ii}, and we can simply pick the weights as $\hat{w}_i = \hat{\sigma}_{ii}^{-1}$, where $\hat{\sigma}_{ii} = \frac{1}{T}\sum_t \hat{\varepsilon}_{i,t}^2$ (Shanken (1992)). The final estimator of the risk premia vector is

$$\hat{\lambda} = \hat{\nu} + \frac{1}{T}\sum_t f_t. \tag{6}$$

We can avoid the trimming on the condition number if we substitute $\hat{Q}_x^{-1}$ for $\hat{Q}_{x,i}^{-1}$ in the first-pass estimator definition. However, this increases the asymptotic variance of the bias corrected estimator of ν, and does not extend to the conditional case. Starting from the asset pricing restriction (4), another estimator of λ is $\bar{\lambda} = \hat{Q}_b^{-1} \frac{1}{n}\sum_i \hat{w}_i \hat{b}_i \bar{R}_i$, where $\bar{R}_i = \frac{1}{T_i}\sum_t I_{i,t} R_{i,t}$. This estimator is numerically equivalent to $\hat{\lambda}$ in the balanced case, where $I_{i,t} = 1$ for all i and t. In the unbalanced case, it is equal to $\bar{\lambda} = \hat{\nu} + \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i \hat{b}_i' \bar{f}_i$, where $\bar{f}_i = \frac{1}{T_i} \sum_t I_{i,t} f_t$. Estimator $\bar{\lambda}$ is often studied by the literature (see, e.g., Shanken (1992), Kandel and Stambaugh (1995), Jagannathan and Wang (1998)), and is also consistent. Estimating $E[f_t]$ with a simple average of the observed factor instead of a weighted average based on estimated betas simplifies the form of the asymptotic distribution in the unbalanced case (see below and Section 2.4). 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) $E\left[\varepsilon_{i,t} | \{\varepsilon_{j,\underline{t-1}}, \gamma_j, j = 1, ..., n\}, x_{\underline{t}}\right] = 0$, *with* $\varepsilon_{j,\underline{t-1}} = \{\varepsilon_{j,t-1}, \varepsilon_{j,t-2}, \cdots\}$ *and* $x_{\underline{t}} = \{x_t, x_{t-1}, \cdots\}$;
b) $\frac{1}{M} \leq \sigma_{ii,t} \leq M, \quad i = 1, ..., n;$ *c)* $E\left[\frac{1}{n} \sum_{i,j} E\left[\sigma_{ij,t}^2 | \gamma_i, \gamma_j\right]^{1/2}\right] \leq M$, *where* $\sigma_{ij,t} = E\left[\varepsilon_{i,t}\varepsilon_{j,t} | x_{\underline{t}}, \gamma_i, \gamma_j\right]$.

Assumption A.1 allows for a martingale difference sequence for the error terms (part a)) including potential conditional heteroskedasticity (part b)) as well as weak cross-sectional dependence (part c)). In particular, Assumption A.1 c) is the same as Assumption C.3 in Bai and Ng (2002), except that we have an expectation w.r.t. the random draws of assets. 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 $n, T \to \infty$.

Proposition 2 *Under Assumptions APR.1-APR.5, SC.1-SC.2, A.1 b) and C.1, C.4, C.5, we get a)* $\|\hat{\nu} - \nu\| = o_p(1)$ *and b)* $\left\|\hat{\lambda} - \lambda\right\| = o_p(1)$, *when* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $\bar{\gamma} > 0$.

Consistency of the estimators holds under double asymptotics such that the cross-sectional size n grows not faster than a power of the time series size T. For instance, the conditions in Proposition 2 allow for n large w.r.t. T (short panel asymptotics) when $\bar{\gamma} > 1$. 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 ν consistently in the second pass with a modified cross-sectional estimator for a fixed T and $n \to \infty$. Since $\lambda = \nu + E[f_t]$, consistent estimation of the risk premia themselves is impossible for a fixed T (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 $\tau_{ij,T} = T/T_{ij}$, where $T_{ij} = \sum_t I_{ij,t}$ and $I_{ij,t} = I_{i,t}I_{j,t}$ for $i, j = 1, ..., n$. Let us further define $\tau_{ij} = \operatorname*{plim}_{T\to\infty} \tau_{ij,T} = E[I_{ij,t}|\gamma_i, \gamma_j]^{-1}$, $S_{ij} = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_t \sigma_{ij,t}x_t x_t' = E[\varepsilon_{i,t}\varepsilon_{j,t}x_t x_t'|\gamma_i, \gamma_j]$ and $Q_b = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_i w_i b_i b_i' = E[w_i b_i b_i']$. The following assumption describes the CLTs underlying the proof of the distributional properties.

Assumption A.2 *As $n, T \to \infty$, a)* $\frac{1}{\sqrt{n}}\sum_i w_i\tau_i\left(Y_{i,T}\otimes b_i\right) \Rightarrow N\left(0, S_b\right)$, *where* $Y_{i,T} = \frac{1}{\sqrt{T}}\sum_t I_{i,t}x_t\varepsilon_{i,t}$ *and* $S_b = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i\tau_j}{\tau_{ij}} S_{ij}\otimes b_i b_j'\right] =$ *a.s.-* $\lim_{n\to\infty}\frac{1}{n}\sum_{i,j} w_i w_j\frac{\tau_i\tau_j}{\tau_{ij}}S_{ij}\otimes b_i b_j'$;
b) $\frac{1}{\sqrt{T}}\sum_t\left(f_t - E\left[f_t\right]\right) \Rightarrow N\left(0, \Sigma_f\right)$, *where* $\Sigma_f = \lim_{T\to\infty}\frac{1}{T}\sum_{t,s} Cov\left(f_t, f_s\right)$.

Assumptions A.2a) and b) require the asymptotic normality of cross-sectional and time series averages of scaled error terms, and of time-series averages of factor values, respectively. These CLTs hold under weak serial and cross-sectional dependencies such as temporal mixing and block dependence (see Appendix 4).

Assumption A.3 *For any* $1 \leq t, s \leq T$, $T \in \mathbb{N}$ *and* $\gamma \in [0, 1]$, *we have* $E\left[\varepsilon_t(\gamma)^2\varepsilon_s(\gamma)|x_{\underline{T}}\right] = 0$.

Assumption A.3 is a symmetry condition on the error distribution given the factors. It is used to prove that the sampling variability of the estimated weights $\hat{w}_i$ does not impact the asymptotic distribution of estimator $\hat{\nu}$. Our setting differs from the standard feasible WLS framework since we have to estimate each incidental parameter S_{ii}. We can dispense with Assumption A.3 if we use OLS to estimate parameter ν, i.e., estimator $\hat{\nu}_1$, or if we put a more restrictive condition on the relative rate of n w.r.t. T.

Proposition 3 *Under Assumptions APR.1-APR.5, SC.1-SC.2, A.1-A.3, and C.1-C.5, we get:*
a) $\sqrt{nT}\left(\hat{\nu} - \nu - \frac{1}{T}\hat{B}_\nu\right) \Rightarrow N\left(0, \Sigma_\nu\right)$, *where* $\Sigma_\nu =$ *a.s.-* $\lim_{n\to\infty} Q_b^{-1}\left(\frac{1}{n}\sum_{i,j} w_i w_j\frac{\tau_i\tau_j}{\tau_{ij}}(c_\nu' Q_x^{-1}S_{ij}Q_x^{-1}c_\nu)b_i b_j'\right)Q_b^{-1}$ *and the bias term is* $\hat{B}_\nu = \hat{Q}_b^{-1}\left(\frac{1}{n}\sum_i \hat{w}_i\tau_{i,T}E_2'\hat{Q}_{x,i}^{-1}\hat{S}_{ii}\hat{Q}_{x,i}^{-1}c_{\hat{\nu}}\right)$, *with* $E_2 = (0 : I_K)'$, $c_{\hat{\nu}} = (1, -\hat{\nu}')'$, *and*
b) $\sqrt{T}\left(\hat{\lambda} - \lambda\right) \Rightarrow N\left(0, \Sigma_f\right)$, *when* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $0 < \bar{\gamma} < 3$.

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

$$\Sigma_\nu = a.s.\text{-}\lim_{n\to\infty} \Sigma_{\nu,n}, \qquad \Sigma_{\nu,n} := \left(\frac{1}{n}B_n'W_nB_n\right)^{-1}\frac{1}{n}B_n'W_nV_nW_nB_n\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}, \tag{7}$$

where $B_n = (b_1, ..., b_n)'$, $W_n = diag(w_1, ..., w_n)$ and $V_n = [v_{ij}]_{i,j=1,...,n}$ 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 $V_n = (1 + \lambda'V[f_t]^{-1}\lambda)\Sigma_{\varepsilon,n}$, where $\Sigma_{\varepsilon,n} = [\sigma_{ij}]_{i,j=1,...,n}$. Then, the asymptotic variance of $\hat{\nu}$ reduces to a.s.-$\lim_{n\to\infty}(1 + \lambda'V[f_t]^{-1}\lambda)\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}\frac{1}{n}B_n'W_n\Sigma_{\varepsilon,n}W_nB_n\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}$. In particular, in the CAPM, we have $K = 1$ and $\nu = 0$, which implies that $\sqrt{\lambda^2/V[f_t]}$ is equal to the slope of the Capital Market Line $\sqrt{E[f_t]^2/V[f_t]}$, 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 $\hat{\nu}$. The upper bound on the relative expansion rates of n and T in Proposition 3 is $n = O(T^{\bar{\gamma}})$ for $\bar{\gamma} < 3$. The control of first-pass estimation errors uniformly across assets requires that the cross-section dimension n is not 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 ν, and would inherit its asymptotic distribution. Since we do not know the true factor mean, only the variability of the factor drives the asymptotic distribution of $\hat{\lambda}$, since the estimation error $O_p\left(1/\sqrt{T}\right)$ of the sample average $\frac{1}{T}\sum_t f_t$ dominates the estimation error $O_p\left(1/\sqrt{nT} + 1/T\right)$ of $\hat{\nu}$. This result is an oracle property for $\hat{\lambda}$, namely that its asymptotic distribution is the same irrespective of the knowledge of ν. This property is in sharp difference with the single asymptotics with a fixed n and $T \to \infty$. In the balanced case and with homoskedastic errors, Theorem 1 of Shanken (1992) shows that the rate of convergence of $\hat{\lambda}$ is $\sqrt{T}$ and that its asymptotic variance is $\Sigma_{\lambda,n} = \Sigma_f + \frac{1}{n}(1 + \lambda'V[f_t]^{-1}\lambda)\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}\frac{1}{n}B_n'W_n\Sigma_{\varepsilon,n}W_nB_n\left(\frac{1}{n}B_n'W_nB_n\right)^{-1}$, for fixed n and $T \to \infty$. The two components in $\Sigma_{\lambda,n}$ come from estimation of $E[f_t]$ and ν, 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, Skoulakis, and Wang (2009), to the unbalanced case yields

$\Sigma_{\lambda,n} = \Sigma_f + \frac{1}{n}\Sigma_{\nu,n}$, where $\Sigma_{\nu,n}$ is defined in (7). Letting $n \to \infty$ gives Σ_f under weak cross-sectional dependence. Thus, exploiting the full cross-section of assets improves efficiency asymptotically, and the positive definite matrix $\Sigma_{\lambda,n} - \Sigma_f$ corresponds to the efficiency gain. Using a large number of assets instead of a small number of portfolios does help to eliminate the contribution coming from estimation of ν.

Proposition 3 suggests exploiting the analytical bias correction $\hat{B}_\nu/T$ and using estimator $\hat{\nu}_B = \hat{\nu} - \frac{1}{T}\hat{B}_\nu$ instead of $\hat{\nu}$. Furthermore, $\hat{\lambda}_B = \hat{\nu}_B + \frac{1}{T}\sum_t f_t$ delivers a bias-free estimator of λ 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 (Neyman and Scott (1948)) in the panel literature (see Lancaster (2000) for a review). We can write model (1) under restriction (3) 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_is correspond to the individual fixed effects and ν to the common parameter of interest. Available results on the fixed-effects approach tell us: (i) the Maximum Likelihood (ML) estimator of ν is inconsistent if n goes to infinity while T is held fixed, (ii) the ML estimator of ν is asymptotically biased even if T grows at the same rate as n, (iii) an analytical bias correction may yield an estimator of ν that is root-(nT) asymptotically normal and centered at the truth if T grows faster than $n^{1/3}$. The two-pass estimators $\hat{\nu}$ and $\hat{\nu}_B$ exhibit the properties (i)-(iii) as expected by analogy with unbiased estimation in large panels. This clear link with the incidental parameter literature highlights another advantage of working with ν in the second pass regression. Chamberlain (1992) considers a general random coefficient model nesting Model (1) under restriction (3). He establishes asymptotic normality of an estimator of ν for fixed T and balanced panel data. His estimator does not admit a closed-form and requires a numerical optimization. This leads to computational difficulties in the conditional extension of Section 3. This also makes the study of his estimator under double asymptotics and cross-sectional dependence challenging. Recent advances on the incidental parameter problem in random coefficient models for fixed T are Arellano and Bonhomme (2012) and Bonhomme (2012).

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 Σ_f by a standard HAC estimator $\hat{\Sigma}_f$ such as in Newey and West (1994) or Andrews and Monahan (1992). Hence, the construction of confidence intervals with valid asymptotic coverage for components of $\hat{\lambda}$ is straightforward. On the contrary, getting a HAC estimator for $\bar{\Sigma}_f$ appearing in the asymptotic distribution of $\bar{\lambda}$ is not obvious in the unbalanced case.

The construction of confidence intervals for the components of $\hat{\nu}$ is more difficult. Indeed, Σ_ν 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 $\hat{S}_{ij} = \frac{1}{T_{ij}} \sum_t I_{ij,t} \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{j,t} x_t x_t'$, 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) focus on the estimation of the variance-covariance matrix of the errors in large balanced panel with nonrandom coefficients.

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; the absence of any natural cross-sectional ordering among the matrices S_{ij} motivates this choice of estimator. 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.4 *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.4 tells us that we can neglect most cross-asset contributions $\|S_{ij}\|$. As sparsity increases, we can choose coefficients q and δ closer to zero. Assumption A.4 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 Σ_ν is well-defined when combined with Assumption C.4. We meet both sparsity assumptions, as well as the approximate factor structure Assumption APR.4 (i), 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 $\tilde{S}_{ij} = \hat{S}_{ij} \mathbf{1}\left\{\left\|\hat{S}_{ij}\right\| \geq \kappa\right\}$ 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.4, C.1-C.5, we have* $\tilde{\Sigma}_{\nu}^{-1/2}\sqrt{nT}\left(\hat{\nu}-\frac{1}{T}\hat{B}_{\nu}-\nu\right)\Rightarrow N\left(0,I_K\right)$ *where* $\tilde{\Sigma}_{\nu}=\hat{Q}_b^{-1}\left[\frac{1}{n}\sum_{i,j}\hat{w}_i\hat{w}_j\frac{\tau_{i,T}\tau_{j,T}}{\tau_{ij,T}}(c_{\hat{\nu}}'\hat{Q}_x^{-1}\tilde{S}_{ij}\hat{Q}_x^{-1}c_{\hat{\nu}})\hat{b}_i\hat{b}_j'\right]\hat{Q}_b^{-1}$, *when* $n,T\to\infty$ *such that* $n=O\left(T^{\bar{\gamma}}\right)$ *for* $0<\bar{\gamma}<\min\left\{3,\eta\frac{1-q}{2\delta}\right\}$, *and* $\kappa=M\sqrt{\frac{\log n}{T^{\eta}}}$ *for a constant* $M>0$ *and* $\eta\in(0,1]$ *as in Assumption C.1.*

In Assumption C.1, we define constant $\eta\in(0,1]$ which 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 stronger the time series dependence (smaller η) and the lower the sparsity (q and δ closer to 1), the more restrictive the condition on the relative rate $\bar{\gamma}$. We cannot guarantee the matrix made of thresholded blocks $\tilde{S}_{ij}$ 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

$$\mathcal{H}_0:\ \text{there exists } \nu\in\mathbb{R}^K \text{ such that } a(\gamma)=b(\gamma)'\nu,\quad \text{for almost all } \gamma\in[0,1].$$

This null hypothesis is written on the continuum of assets. Under $\mathcal{H}_0$, we have $E\left[(a_i-b_i'\nu)^2\right]=0$. Since we estimate ν 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 $\hat{e}_i=c_{\hat{\nu}}'\hat{\beta}_i$, which is an empirical counterpart of $E\left[w_i\left(a_i-b_i'\nu\right)^2\right]$.

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

Assumption A.5 *For $n, T \to \infty$ we have* $\frac{1}{\sqrt{n}} \sum_i w_i \tau_i^2 \left(Y_{i,T} \otimes Y_{i,T} - vec\left[S_{ii,T}\right]\right) \Rightarrow N\left(0, \Omega\right)$, *where the asymptotic variance matrix is:*

$$\begin{aligned}\Omega &= \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\left[S_{ij}\otimes S_{ij} + \left(S_{ij}\otimes S_{ij}\right) W_{K+1}\right]\right] \\ &= \textit{a.s.-}\lim_{n\to\infty}\frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\left[S_{ij}\otimes S_{ij} + \left(S_{ij}\otimes S_{ij}\right) W_{K+1}\right].\end{aligned}$$

Assumption A.5 is a high-level CLT condition. We can prove this assumption under primitive conditions on the time series and cross-sectional dependence. For instance, we prove in Appendix 4 that Assumption A.5 holds under a cross-sectional block dependence structure for the errors. Intuitively, the expression of the variance-covariance matrix Ω is related to the result that, for random $(K+1)\times 1$ vectors Y_1 and Y_2 which are jointly normal with covariance matrix S, we have $Cov\left(Y_1\otimes Y_1, Y_2\otimes Y_2\right) = S\otimes S + \left(S\otimes S\right)W_{K+1}$.

Let us now introduce the following statistic $\hat{\xi}_{nT} = T\sqrt{n}\left(\hat{Q}_e - \frac{1}{T}\hat{B}_\xi\right)$, where the recentering term simplifies to $\hat{B}_\xi = 1$ thanks to the weighting scheme. Under the null hypothesis $\mathcal{H}_0$, we prove that $\hat{\xi}_{nT} = \left(vec\left[\hat{Q}_x^{-1}c_{\hat{\nu}}c_{\hat{\nu}}'\hat{Q}_x^{-1}\right]\right)'\frac{1}{\sqrt{n}}\sum_i w_i\tau_i^2\left(Y_{iT}\otimes Y_{i,T} - vec\left[S_{ii,T}\right]\right) + o_p(1)$, which implies $\hat{\xi}_{nT} \Rightarrow N\left(0,\Sigma_\xi\right)$, where $\Sigma_\xi = 2\lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i,j} w_i w_j v_{ij}^2\right] = 2$ a.s.-$\lim_{n\to\infty}\frac{1}{n}\sum_{i,j} w_i w_j v_{ij}^2$ as $n, T \to \infty$ (see Appendix A.2.5). Then, a feasible testing procedure exploits the consistent estimator $\tilde{\Sigma}_\xi = 2\frac{1}{n}\sum_{i,j}\hat{w}_i\hat{w}_j\tilde{v}_{ij}^2$ of the asymptotic variance Σ_ξ, where $\tilde{v}_{ij} = \frac{\tau_{i,T}\tau_{j,T}}{\tau_{ij,T}}c_{\hat{\nu}}'\hat{Q}_x^{-1}\tilde{S}_{ij}\hat{Q}_x^{-1}c_{\hat{\nu}}$.

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

In the homoskedastic case, the asymptotic variance of $\hat{\xi}_{nT}$ reduces to $\Sigma_\xi = 2$a.s.-$\lim_{n\to\infty}\frac{1}{n}\sum_{i,j}\frac{\tau_i\tau_j}{\tau_{ij}^2}\frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}$. For fixed n, we can rely on the test statistic $T\hat{Q}_e$, which is asymptotically distributed as $\frac{1}{n}\sum_j eig_j\chi_j^2$ for $j = 1,\ldots,(n-K)$, where the χ_j^2 are independent chi-square variables with 1 degree of freedom, and the coefficients eig_j are the non-zero eigenvalues of matrix $V_n^{1/2}(W_n - W_nB_n(B_n'W_nB_n)^{-1}B_n'W_n)V_n^{1/2}$ (see Kan, Robotti and Shanken (2012)). 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 condition on the relative expansion rate of n and T for the distributional result on the test statistic in Proposition 5 is more restrictive than the condition for feasible asymptotic normality of the estimators in Proposition 4.

The alternative hypothesis is

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

Let us define the pseudo-true value $\nu_\infty = \arg\inf_{\nu \in \mathbb{R}^K} Q_\infty^w(\nu)$, where $Q_\infty^w(\nu) = E\left[w_i\left(a_i - b_i'\nu\right)^2\right]$ (White (1982), Gourieroux et al. (1984)) and population errors $e_i = a_i - b_i'\nu_\infty = c_{\nu_\infty}'\beta_i$, $i = 1, ..., n$, for all n. In the next proposition, we prove consistency of the test, namely that the statistic $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT}$ diverges to $+\infty$ under the alternative hypothesis $\mathcal{H}_1$ for large n and T. The test of the null $\mathcal{H}_0$ against the alternative $\mathcal{H}_1$ is a one-sided test. 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.5 and C.1-C.5, we have: a)* $\sqrt{n}\left(\hat{\nu} - \frac{1}{T}\hat{B}_{\nu_\infty} - \nu_\infty\right) \Rightarrow N\left(0, \Sigma_{\nu_\infty}\right)$, *where* $\hat{B}_{\nu_\infty} = \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i\tau_{i,T}E_2'\hat{Q}_{x,i}^{-1}\hat{S}_{ii}\hat{Q}_{x,i}^{-1}c_{\hat{\nu}}$ *and* $\Sigma_{\nu_\infty} = Q_b^{-1}E[w_i^2 e_i^2 b_i b_i']Q_b^{-1}$, *and b)* $\sqrt{T}\left(\hat{\lambda} - \lambda_\infty\right) \Rightarrow N\left(0, \Sigma_f\right)$, *where* $\lambda_\infty = \nu_\infty + E\left[f_t\right]$, *as* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $1 < \bar{\gamma} < 3$; *c)* $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT} \xrightarrow{p} +\infty$, *as* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $0 < \bar{\gamma} < \min\left\{2, \ \eta\frac{1-q}{2\delta}\right\}$.

Under the alternative hypothesis $\mathcal{H}_1$, the convergence rate of $\hat{\nu}$ is slower than under $\mathcal{H}_0$, while the convergence rate of $\hat{\lambda}$ remains the same. The asymptotic distribution of the bias-adjusted estimator $\hat{\nu} - \frac{1}{T}\hat{B}_{\nu_\infty}$ is the same as the one got from a cross-sectional regression of a_i on b_i. The condition $\bar{\gamma} > 1$ in Propositions 6 a) and b) ensures that cross-sectional estimation of ν has asymptotically no impact on the estimation of λ.

To study the local asymptotic power, we can adopt the local alternative $\mathcal{H}_{1,nT} : \quad \inf_{\nu \in \mathbb{R}^K} Q_\infty^w(\nu) = \frac{\psi}{\sqrt{n}T} > 0$, for a constant $\psi > 0$. Then we can show that $\hat{\xi}_{nT} \Rightarrow N(\psi, \Sigma_\xi)$, 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:

$$\mathcal{H}_0 : a(\gamma) = 0 \text{ for almost all } \gamma \in [0, 1] \quad \Leftrightarrow \quad E[a_i^2] = 0,$$

against the alternative hypothesis

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

We only have to substitute $\hat{a}_i$ for $\hat{e}_i$, and $E_1 = (1, 0')'$ for $c_{\hat{\nu}}$ in Proposition 5. This gives an extension of Gibbons, Ross and Shanken (1989) with double asymptotics. Implementing the original Gibbons, Ross and Shanken (1989) test, which uses a weighting matrix corresponding to an inverted estimated $n \times n$ covariance matrix, becomes quickly problematic. We expect to compensate the potential loss of power induced by a diagonal weighting via the larger number of restrictions. Our Monte Carlo simulations show that the test exhibits good power properties against both risk-based and non risk-based alternatives (e.g. MacKinlay (1995)) already for a thousand assets with a time series dimension similar to the one in the empirical analysis.

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 (see Connor and Korajczyk (1989) for an intertemporal competitive equilibrium version of the APT yielding time-varying risk premia and Ludvigson (2011) for a discussion within scaled consumption-based models). 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 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.

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

$$R_t(\gamma) = a_t(\gamma) + b_t(\gamma)' f_t + \varepsilon_t(\gamma), \tag{8}$$

where $a_t(\gamma, \omega) = a[\gamma, S^{t-1}(\omega)]$ and $b_t(\gamma, \omega) = b[\gamma, S^{t-1}(\omega)]$, for any $\omega \in \Omega$ and $\gamma \in [0,1]$, and random variable $a(\gamma)$ and random vector $b(\gamma)$, for $\gamma \in [0,1]$, are $\mathcal{F}_0$-measurable.

The intercept $a_t(\gamma)$ and factor sensitivity $b_t(\gamma)$ of asset $\gamma \in [0,1]$ at time t are $\mathcal{F}_{t-1}$-measurable, where the information set $\mathcal{F}_t$ is defined by $\mathcal{F}_t = \left\{S^{-t}(A),\ A \in \mathcal{F}_0\right\}$, for $\mathcal{F}_0 \in \mathcal{F}$, as in Section 2.

Assumption APR.7 *The matrix $\int b(\gamma)b(\gamma)'d\gamma$ is positive definite, P-a.s..*

Since transformation S is measure preserving, Assumption APR.7 implies that the matrix $\int b_t(\gamma)b_t(\gamma)'d\gamma$ is positive definite, P-a.s., for any date $t = 1, 2, ...$

Proposition 7 *Under Assumptions APR.3-APR.7, for any date $t = 1, 2, ...$ there exists a unique random vector $\nu_t \in \mathbb{R}^K$ such that ν_t is $\mathcal{F}_{t-1}$-measurable and:*

$$a_t(\gamma) = b_t(\gamma)'\nu_t, \tag{9}$$

P-a.s. and for almost all $\gamma \in [0,1]$.

We can rewrite the asset pricing restriction as

$$E\left[R_t(\gamma)|\mathcal{F}_{t-1}\right] = b_t(\gamma)'\lambda_t, \tag{10}$$

for almost all $\gamma \in [0,1]$, where $\lambda_t = \nu_t + E\left[f_t|\mathcal{F}_{t-1}\right]$ is the vector of the conditional risk premia.

To have a workable version of Equations (8) and (9), we further specify the conditioning information and how coefficients depend on it. The conditioning information is such that instruments $Z \in \mathbb{R}^p$ and $Z(\gamma) \in \mathbb{R}^q$, for $\gamma \in [0,1]$, are $\mathcal{F}_0$-measurable. Then, the information $\mathcal{F}_{t-1}$ contains $Z_{\underline{t-1}}$ and $Z_{\underline{t-1}}(\gamma)$, for $\gamma \in [0,1]$, where we define $Z_t(\omega) = Z[S^t(\omega)]$ and $Z_t(\gamma, \omega) = Z[\gamma, S^t(\omega)]$. 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 $Z_{t-1}(\gamma)$ are specific to stock γ. They may include past observations of firm characteristics and stock returns. To end up with a linear regression model, we specify that the vector of factor sensitivities $b_t(\gamma)$ is a linear function of lagged instruments Z_{t-1} (Shanken (1990), Ferson and Harvey (1991)) and $Z_{t-1}(\gamma)$ (Avramov and Chordia (2006)): $b_t(\gamma) = B(\gamma)Z_{t-1} + C(\gamma)Z_{t-1}(\gamma)$, where $B(\gamma) \in \mathbb{R}^{K\times p}$ and $C(\gamma) \in \mathbb{R}^{K\times q}$, for any $\gamma \in [0,1]$ and $t = 1, 2, ...$ We can account for nonlinearities by including powers of some explanatory variables among the lagged instruments. 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 information $\mathcal{F}_{t-1}$ depends on Z_{t-1} only and is linear, as, for instance, in an exogeneous Vector Autoregressive (VAR) model of order 1. Since f_t is a subvector of Z_t, then $E[f_t|\mathcal{F}_{t-1}] = FZ_{t-1}$, where $F \in \mathbb{R}^{K\times p}$, for any t. Under these functional specifications the asset pricing restriction (9) implies that the intercept $a_t(\gamma)$ is a quadratic form in lagged instruments Z_{t-1} and $Z_{t-1}(\gamma)$, namely:

$$a_t(\gamma) = Z'_{t-1}B(\gamma)'(\Lambda - F)Z_{t-1} + Z_{t-1}(\gamma)'C(\gamma)'(\Lambda - F)Z_{t-1}. \tag{11}$$

This shows that assuming a priori linearity of $a_t(\gamma)$ in the lagged instruments Z_{t-1} and $Z_{t-1}(\gamma)$ is in general not compatible with linearity of $b_t(\gamma)$ and $E[f_t|Z_{t-1}]$.

The sampling scheme is the same as in Section 2.2, and we use the same type of notation, for example $b_{i,t} = b_t(\gamma_i)$, $B_i = B(\gamma_i)$, $C_i = C(\gamma_i)$ and $Z_{i,t-1} = Z_{t-1}(\gamma_i)$. In particular, we allow for potential correlation between parameters B_i, C_i and asset specific instruments $Z_{i,t-1}$ via the random index γ_i. Then, the conditional factor model (8) with asset pricing restriction (11) written for the sample observations becomes

$$R_{i,t} = Z'_{t-1}B'_i(\Lambda - F)Z_{t-1} + Z'_{i,t-1}C'_i(\Lambda - F)Z_{t-1} + Z'_{t-1}B'_if_t + Z'_{i,t-1}C'_if_t + \varepsilon_{i,t}, \tag{12}$$

which is nonlinear in the parameters Λ, F, B_i, and C_i. In order to implement the two-pass methodology in a conditional context, we rewrite model (12) 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 (Cochrane (2005)), while the second components do not since they depend on i. The regressors also include the predetermined variables $x_{1,i,t} = \left(vech[X_t]', Z'_{t-1} \otimes Z'_{i,t-1}\right)' \in \mathbb{R}^{d_1}$ with $d_1 = p(p+1)/2 + pq$, where the symmetric matrix $X_t = [X_{t,k,l}] \in \mathbb{R}^{p\times p}$ is such that $X_{t,k,l} = Z^2_{t-1,k}$, if $k = l$, and $X_{t,k,l} = 2Z_{t-1,k}Z_{t-1,l}$, otherwise, $k, l =$

$1, \ldots, p$. The vector-half operator $vech\left[\cdot\right]$ stacks the lower elements of a $p \times p$ matrix as a $p\left(p+1\right)/2 \times 1$ vector (see Chapter 2 in Magnus and Neudecker (2007) for properties of this matrix tool). To parallel the analysis of the unconditional case, we can express model (12) as in (2) through appropriate redefinitions of the regressors and loadings (see Appendix 3):

$$R_{i,t} = \beta_i' x_{i,t} + \varepsilon_{i,t}, \tag{13}$$

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

$$\beta_{1,i} = \Psi \beta_{2,i}, \qquad \beta_{2,i} = \left(vec\left[B_i'\right]', vec\left[C_i'\right]'\right)', \tag{14}$$

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

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 $p = 1$ and $q = 0$, and model (13) reduces to model (2).

In (14), 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 (11) implies a constraint on the distribution of random vector β_i via its support. The coefficients of the linear transformation depend on matrix $\Lambda - F$. For the purpose of estimating the loading coefficients of the risk premia in matrix Λ, we rewrite the parameter restrictions as (see Appendix 3):

$$\beta_{1,i} = \beta_{3,i} \nu, \qquad \nu = vec\left[\Lambda' - F'\right], \qquad \beta_{3,i} = \left(\left[D_p^+ \left(B_i' \otimes I_p\right)\right]', \left[W_{p,q}\left(C_i' \otimes I_p\right)\right]'\right)'. \tag{15}$$

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

$$vec\left[\beta_{3,i}'\right] = J_a \beta_{2,i}, \tag{16}$$

where the $d_1 pK \times d_2$ 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 $J_{11} = W_{p(p+1)/2,pK}\left(I_K \otimes \left[\left(I_p \otimes D_p^+\right)\left(W_p \otimes I_p\right)\left(I_p \otimes vec\left[I_p\right]\right)\right]\right)$ and $J_{22} = W_{pq,pK}\left(I_K \otimes \left[\left(I_p \otimes W_{p,q}\right)\left(W_{p,q} \otimes I_p\right)\left(I_q \otimes vec\left[I_p\right]\right)\right]\right)$. The link (16) 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, in which the matrix J_a is equal to I_K. Equations (15) and (16) in the conditional setting are the counterparts of restriction (3) in the unconditional setting.

3.2 Asymptotic properties of time-varying risk premium estimation

We consider a two-pass approach building on Equations (13) and (15).

First Pass: The first pass consists in computing time-series OLS estimators $\hat{\beta}_i = (\hat{\beta}'_{1,i}, \hat{\beta}'_{2,i})' = \hat{Q}_{x,i}^{-1} \frac{1}{T_i} \sum_t I_{i,t} x_{i,t} R_{i,t}$, for $i = 1, ..., n$, where $\hat{Q}_{x,i} = \frac{1}{T_i} \sum_t I_{i,t} x_{i,t} x'_{i,t}$. We use the same trimming device as in Section 2.

Second Pass: The second pass consists in computing a cross-sectional estimator of ν 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 $w_i = (diag\,[v_i])^{-1}$, where the v_i are the asymptotic variances of the standardized errors $\sqrt{T}\left(\hat{\beta}_{1,i} - \hat{\beta}_{3,i}\nu\right)$ in the cross-sectional regression for large T. We have $v_i = \tau_i C'_\nu Q_{x,i}^{-1} S_{ii} Q_{x,i}^{-1} C_\nu$, where $Q_{x,i} = E\left[x_{i,t} x'_{i,t} | \gamma_i\right]$, $S_{ii} = \underset{T\to\infty}{\text{plim}} \frac{1}{T} \sum_t \sigma_{ii,t} x_{i,t} x'_{i,t} = E\left[\varepsilon_{i,t}^2 x_{i,t} x'_{i,t} | \gamma_i\right]$, $\sigma_{ii,t} = E\left[\varepsilon_{i,t}^2 | x_{i,\underline{t}}, \gamma_i\right]$, and $C_\nu = \left(E'_1 - \left(I_{d_1} \otimes \nu'\right) J_a E'_2\right)'$, 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 $\hat{v}_i = \tau_{i,T} C'_{\hat{\nu}_1} \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} C_{\hat{\nu}_1}$, where $\hat{S}_{ii} = \frac{1}{T_i} \sum_t I_{i,t} \hat{\varepsilon}_{i,t}^2 x_{i,t} x'_{i,t}$, $\hat{\varepsilon}_{i,t} = R_{i,t} - \hat{\beta}'_i x_{i,t}$ 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_ν, we use the OLS estimator $\hat{\nu}_1 = \left(\sum_i \mathbf{1}_i^\chi \hat{\beta}'_{3,i} \hat{\beta}_{3,i}\right)^{-1} \sum_i \mathbf{1}_i^\chi \hat{\beta}'_{3,i} \hat{\beta}_{1,i}$, i.e., a first-step estimator with unit weights. The WLS estimator is:

$$\hat{\nu} = \hat{Q}_{\beta_3}^{-1} \frac{1}{n} \sum_i \hat{\beta}'_{3,i} \hat{w}_i \hat{\beta}_{1,i}, \tag{17}$$

where $\hat{Q}_{\beta_3} = \frac{1}{n} \sum_i \hat{\beta}'_{3,i} \hat{w}_i \hat{\beta}_{3,i}$ and $\hat{w}_i = \mathbf{1}_i^\chi \left(diag\,[\hat{v}_i]\right)^{-1}$. The final estimator of the risk premia is $\hat{\lambda}_t = \hat{\Lambda} Z_{t-1}$, where we deduce $\hat{\Lambda}$ from the relationship $vec\left[\hat{\Lambda}'\right] = \hat{\nu} + vec\left[\hat{F}'\right]$ with the estimator $\hat{F}$ obtained by a SUR regression of factors f_t on lagged instruments Z_{t-1}: $\hat{F} = \sum_t f_t Z'_{t-1} \left(\sum_t Z_{t-1} Z'_{t-1}\right)^{-1}$.

The next assumption is similar to Assumption A.1.

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

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

Proposition 8 *Under Assumptions APR.3-APR.7,SC.1-SC.2,B.1b) and C.1,C.4-C.6, we get a)* $\|\hat{\nu}-\nu\| = o_p(1)$, *b)* $\left\|\hat{\Lambda}-\Lambda\right\| = o_p(1)$, *when* $n, T \to \infty$ *such that* $n = O(T^{\bar{\gamma}})$ *for* $\bar{\gamma} > 0$.

Part b) implies $\sup_t \left\|\hat{\lambda}_t - \lambda_t\right\| = o_p(1)$ under boundedness of process Z_t (Assumption C.4 written for the conditional model).

Proposition 9 below gives the large-sample distributions under the double asymptotics $n, T \to \infty$. It extends Proposition 3 to the conditional case through adequate use of selection matrices. The following assumptions are similar to Assumptions A.2 and A.3. We make use of $Q_{\beta_3} = E\left[\beta'_{3,i} w_i \beta_{3,i}\right]$, $Q_z = E\left[Z_t Z'_t\right]$, $S_{ij} = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_t \sigma_{ij,t} x_{i,t} x'_{j,t} = E[\varepsilon_{i,t}\varepsilon_{j,t} x_{i,t} x'_{j,t} | \gamma_i, \gamma_j]$ and $S_{Q,ij} = Q^{-1}_{x,i} S_{ij} Q^{-1}_{x,j}$, otherwise, we keep the same notations as in Section 2.

Assumption B.2 *As* $n, T \to \infty$, *a)* $\frac{1}{\sqrt{n}}\sum_i \tau_i \left[(Q^{-1}_{x,i} Y_{i,T}) \otimes v_{3,i}\right] \Rightarrow N(0, S_{v_3})$, *with* $Y_{i,T} = \frac{1}{\sqrt{T}}\sum_t I_{i,t} x_{i,t}\varepsilon_{i,t}$, $v_{3,i} = vec[\beta'_{3,i} w_i]$ *and* $S_{v_3} = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i,j}\frac{\tau_i\tau_j}{\tau_{ij}} S_{Q,ij} \otimes v_{3,i} v'_{3,j}\right] =$ *a.s.-* $\lim_{n\to\infty}\frac{1}{n}\sum_{i,j}\frac{\tau_i\tau_j}{\tau_{ij}}[S_{Q,ij} \otimes v_{3,i} v'_{3,j}]$; *b)* $\frac{1}{\sqrt{T}}\sum_t u_t \otimes Z_{t-1} \Rightarrow N(0, \Sigma_u)$, *where* $\Sigma_u = E\left[u_t u'_t \otimes Z_{t-1} Z'_{t-1}\right]$ *and* $u_t = f_t - F Z_{t-1}$.

Assumption B.3 *For any* $1 \le t, s \le T$, $T \in \mathbb{N}$ *and* $\gamma \in [0,1]$, *we have* $E\left[\varepsilon_t(\gamma)^2 \varepsilon_s(\gamma) | Z_{\underline{T}}, Z_{\underline{T}}(\gamma)\right] = 0$.

Proposition 9 *Under Assumptions APR.3-APR.7,SC.1-SC.2,B.1-B.3 and C.1-C.6, we have a)* $\sqrt{nT}\left(\hat{\nu} - \nu - \frac{1}{T}\hat{B}_\nu\right) \Rightarrow N(0, \Sigma_\nu)$ *where* $\hat{B}_\nu = \hat{Q}^{-1}_{\beta_3} J_b \frac{1}{n}\sum_i \tau_{i,T} vec\left[E'_2 \hat{Q}^{-1}_{x,i}\hat{S}_{ii}\hat{Q}^{-1}_{x,i} C_{\hat{\nu}}\hat{w}_i\right]$ *and* $\Sigma_\nu = \left(vec\left[C'_\nu\right] \otimes Q^{-1}_{\beta_3}\right)' S_{v_3}\left(vec\left[C'_\nu\right] \otimes Q^{-1}_{\beta_3}\right)$, *with* $J_b = \left(vec\left[I_{d_1}\right]' \otimes I_{Kp}\right)\left(I_{d_1} \otimes J_a\right)$ *and* $C_{\hat{\nu}} = \left(E'_1 - \left(I_{d_1} \otimes \hat{\nu}'\right) J_a E'_2\right)'$; *b)* $\sqrt{T} vec\left[\hat{\Lambda}' - \Lambda'\right] \Rightarrow N(0, \Sigma_\Lambda)$ *where* $\Sigma_\Lambda = \left(I_K \otimes Q^{-1}_z\right)\Sigma_u\left(I_K \otimes Q^{-1}_z\right)$, *when* $n, T \to \infty$ *such that* $n = O(T^{\bar{\gamma}})$ *for* $0 < \bar{\gamma} < 3$.

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

We can use Proposition 9 to build confidence intervals. It suffices to replace the unknown quantities Q_x, Q_z, Q_{β_3}, Σ_u, and ν by their empirical counterparts. For matrix S_{v_3}, 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.3, A.4 and C.1-C.6.

3.3 Tests of conditional asset pricing restrictions

Since the equations in (15) correspond to the asset pricing restriction (3), the null hypothesis of correct specification of the conditional model is

$$\mathcal{H}_0: \text{ there exists } \nu \in \mathbb{R}^{pK} \text{ such that } \beta_1(\gamma) = \beta_3(\gamma)\nu, \text{ for almost all } \gamma \in [0,1],$$

where $\beta_1(\gamma)$ and $\beta_3(\gamma)$ are defined as $\beta_{1,i}$ and $\beta_{3,i}$ in Equations (14) and (15) replacing $B(\gamma)$ and $C(\gamma)$ for B_i and C_i. Under $\mathcal{H}_0$, we have $E\left[(\beta_{1,i} - \beta_{3,i}\nu)'(\beta_{1,i} - \beta_{3,i}\nu)\right] = 0$. The alternative hypothesis is

$$\mathcal{H}_1: \quad \inf_{\nu \in \mathbb{R}^{pK}} E\left[(\beta_{1,i} - \beta_{3,i}\nu)'(\beta_{1,i} - \beta_{3,i}\nu)\right] > 0.$$

As in Section 2.5, we build the SSR $\hat{Q}_e = \frac{1}{n}\sum_i \hat{e}_i'\hat{w}_i\hat{e}_i$, with $\hat{e}_i = \hat{\beta}_{1,i} - \hat{\beta}_{3,i}\hat{\nu} = C_{\hat{\nu}}'\hat{\beta}_i$ and the statistic $\hat{\xi}_{nT} = T\sqrt{n}\left(\hat{Q}_e - \frac{1}{T}\hat{B}_\xi\right)$, where $\hat{B}_\xi = d_1$.

Assumption B.4 *For $n, T \to \infty$, we have $\frac{1}{\sqrt{n}}\sum_i \tau_i^2 \left[\left(Q_{x,i}^{-1} \otimes Q_{x,i}^{-1}\right)(Y_{i,T} \otimes Y_{i,T} - vec\left[S_{ii,T}\right])\right] \otimes vec[w_i]$ $\Rightarrow N(0,\Omega)$, where the asymptotic variance matrix is:*

$$\begin{aligned}
\Omega &= \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i,j}\frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\left[S_{Q,ij} \otimes S_{Q,ij} + (S_{Q,ij} \otimes S_{Q,ij})W_d\right] \otimes \left(vec[w_i]vec[w_j]'\right)\right] \\
&= \textit{a.s.-}\lim_{n\to\infty}\frac{1}{n}\sum_{i,j}\frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\left[S_{Q,ij} \otimes S_{Q,ij} + (S_{Q,ij} \otimes S_{Q,ij})W_d\right] \otimes \left(vec[w_i]vec[w_j]'\right).
\end{aligned}$$

Proposition 10 *Under $\mathcal{H}_0$ and Assumptions APR.3-APR.7, SC.1-SC.2, B.1-B.4, A.4 and C.1-C.6, we have $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT} \Rightarrow N(0,1)$, where $\tilde{\Sigma}_\xi = 2\frac{1}{n}\sum_{i,j}\frac{\tau_{i,T}^2\tau_{j,T}^2}{\tau_{ij,T}^2} tr\left[\hat{w}_i\left(C_{\hat{\nu}}'\hat{Q}_{x,i}^{-1}\tilde{S}_{ij}\hat{Q}_{x,j}^{-1}C_{\hat{\nu}}\right)\hat{w}_j\left(C_{\hat{\nu}}'\hat{Q}_{x,j}^{-1}\tilde{S}_{ji}\hat{Q}_{x,i}^{-1}C_{\hat{\nu}}\right)\right]$ as $n, T \to \infty$ such that $n = O\left(T^{\bar{\gamma}}\right)$ for $0 < \bar{\gamma} < \min\left\{2, \eta\frac{1-q}{2\delta}\right\}$.*

Under $\mathcal{H}_1$, we have $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT} \xrightarrow{p} +\infty$, as in Proposition 6.

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

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

against the alternative hypothesis

$$\mathcal{H}_1 : \quad E\left[\beta_{1,i}'\beta_{1,i}\right] > 0.$$

We only have to substitute $\hat{Q}_a = \frac{1}{n}\sum_i \hat{\beta}_{1,i}'\hat{w}_i\hat{\beta}_{1,i}$ for $\hat{Q}_e$, and $E_1 = (I_{d_1} : 0_{d_1\times d_2})'$ for $C_{\hat{\nu}}$. This gives an extension of Gibbons, Ross and Shanken (1989) to the conditional case with double asymptotics. The implementation of the original Gibbons, Ross and Shanken (1989) test is unfeasible here because of the large number nd_1 of restrictions; each $\beta_{1,i}$ is of dimension $d_1 \times 1$, and the estimated covariance matrix to invert is of dimension $nd_1 \times nd_1$.

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, 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)). We proxy the risk free rate with the monthly 30-day T-bill beginning-of-month yield. To account for time-varying alphas, betas and risk premia, we use a conditional specification based on two common variables and a firm-level variable. We take the instruments $Z_t = (1,\ Z_t^{*\prime})'$, 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 a scalar $Z_{i,t}$ corresponding to 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 $x_{i,t}$ has dimension $d = 25$, and parsimony explains why we have not included e.g. the size of firm i as an additional stock specific instrument. We report robustness checks with other conditional specifications in the supplementary materials.

We compute the firm characteristics from Compustat as in the appendix of Fama and French (2008). The CRSP database provides the monthly stock returns data 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 $T = 546$ months. 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, portfolio returns, 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 $\chi_{1,T} = 15$ as advocated by Greene (2008), together with $\chi_{2,T} = 546/12$ for the unconditional estimation and $\chi_{2,T} = 546/60$ for the conditional estimation. In the results reported for each model, we denote by n^{χ} the dimension of the cross-section after trimming. We compute confidence intervals with a data-driven threshold selected by cross-validation as in Bickel and Levina (2008). Table 1 gathers the estimated annual risk premia, with the corresponding confidence intervals at 95% level, for the following unconditional models: the four-factor model, the Fama-French model, and the CAPM. For the Fama-French model and the CAPM, the trimming level $\chi_{1,T}$ is not binding when $\chi_{2,T} = 546/12$. In Table 2, we display the estimates of the components of ν. For individual stocks, we use bias-corrected estimates for λ and ν. For portfolios, 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. For the four-factor model and the individual stocks, the size factor is positively remunerated (2.86%) and it is not significantly different from zero. The value factor commands a significant negative reward (-4.60%). Phalippou (2007) obtains a similar growth premium for portfolios built on stocks with a high institutional ownership. The momentum factor is largely remunerated (7.16%) 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 ν_{mom} (25.40% and 8.66%) that is much larger and less accurate than the estimates for ν_m, ν_{smb} and ν_{hml} (0.85%, -0.26%, 0.03%, and 0.55%, 0.01%, 0.33%). Moreover, while the estimates of ν_m, ν_{smb} and ν_{hml} are statistically not significant for portfolios, the estimates of ν_m and ν_{hml} are statistically different from zero for individual stocks. In particular, the estimate of ν_{hml} is large and negative, which explains the negative estimate on the value premium displayed in Table 1. The size, value and momentum factors are tradable in theory. In practice, their implementation faces transaction costs due to rebalancing and short selling. A non zero ν might capture these market imperfections (Cremers, Petajisto, and Zitzewitz (2010)).

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. As already discussed in Lewellen, Nagel and Shanken (2010), 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. For λ_m, λ_{smb}, and λ_{hml}, we obtain similar inferential results when we consider the Fama-French model. Our point estimates for λ_m, λ_{smb} and λ_{hml}, for large n agree with Ang, Liu and Schwarz (2008). Our point estimates and confidence intervals for λ_m, λ_{smb} and λ_{hml}, agree with the results reported by Shanken and Zhou (2007) for the 25 portfolios.

Let us now consider the conditional four-factor specification. Figure 1 plots the estimated time-varying path of the four risk premia from the individual stocks. For comparison purpose, we also plot the unconditional estimates and the average lambda over time. A well-known bias coming from market-timing and volatility-timing (Jagannathan and Wang (1996), Lewellen and Nagel (2006), Boguth, Carlson, Fisher and Simutin (2011)) explains the discrepancy between the unconditional estimate and the average over time. After trimming, we compute the risk premia on $n^{\chi} = 3,900$ individual assets in the four-factor model. 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 counter-cyclical behavior in size and book-to-market effects. On the contrary, the risk premium for the momentum factor is pro-cyclical. Furthermore, conditional estimates of the value premium are often negative and take positive values mostly in recessions. The conditional estimates of the size premium are most of the time slightly positive.

Figure 2 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 $\lambda_{mom,t}$

are typically smaller than the unconditional estimate in Table 1. Finally, by comparing Figures 1 and 2, 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 2 and 3, we observe that increasing the number of portfolios to 100 does not help in reconciling the discrepancy.

4.3 Results on testing the asset pricing restrictions

As already discussed in Lewellen, Nagel and Shanken (2010), the 25 FF portfolios have four-factor market and momentum betas close to one and zero, respectively. 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 excess 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 3 gathers the results for the tests of the asset pricing restrictions in 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 statistics $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ based on $\hat{Q}_e$ and $\hat{Q}_a$ as well as their associated one-side p-value. Our Monte Carlo simulations show that we need to set a stronger trimming level $\chi_{2,T}$ to compute the test statistic than to estimate the risk premium. We use $\chi_{2,T} = 546/240$. For the 25 and 100 FF portfolios, we compute weighted test statistics (Gibbons, Ross and Shanken (1989)) as well as their associated p-values. For individual stocks, the test statistics reject both null hypotheses $\mathcal{H}_0 : \ a(\gamma) = b(\gamma)' \nu$ and $\mathcal{H}_0 : \ a(\gamma) = 0$ for the three specifications at 5% level. Instead, the null hypothesis $\mathcal{H}_0 : \ a(\gamma) = b(\gamma)' \nu$ is not rejected for the four-factor specification at 1% level. Similar conclusions are obtained when using the two sets of Fama-French portfolios as base assets. Table 4 gathers the results for tests of the asset pricing restrictions in conditional specifications. Contrary to the unconditional case, we do not report the values of the weighted test statistics (Gibbons, Ross and Shanken (1989)) computed for portfolios because of the numerical instability in the inversion of the covariance matrix. The latter has dimension $2,500 \times 2,500$ for the conditional four-factor specification with the 100FF portfolios. Instead, we report the values of the test statistics $T\hat{Q}_e$ and $T\hat{Q}_a$. For individual stocks, the test statistics reject both null hypotheses $\mathcal{H}_0: \ \beta_1(\gamma) = \beta_3(\gamma)\nu$ and $\mathcal{H}_0: \ \beta_1(\gamma) = 0$ for the three specifications at 5% level, but not for the conditional CAPM at 1% level. For portfolios, the two null hypotheses are not rejected under the conditional CAPM even at 5% level.

For individual stocks, the rejection of the asset pricing restriction using a conditional multi-factor specification (at 1% level), and the non rejection under an unconditional specification, might seem counterintuitive. Indeed, for a given choice of the factors and instruments, the set of unconditional specifications satisfying the no-arbitrage restriction $a(\gamma) = b(\gamma)'\nu$, is a strict subset of the collection of conditional specifications with $a_t(\gamma) = b_t(\gamma)'\nu_t$. However, what we are testing here is whether the projection of the DGP on a given conditional or unconditional factor specification is compatible with no-arbitrage. The set of unconditional factor models is included in the set of conditional factor models, and it may well be the case that the projection of the DGP on the former set satisfies the no-arbitrage restrictions, while the projection on the latter does not. Therefore, the results in Tables 3 and 4 for individual stocks are not incompatible with each other. A similar argument might explain why in Table 4 we fail to reject the asset pricing restriction $\mathcal{H}_0: \ \beta_1(\gamma) = \beta_3(\gamma)\nu$ under the conditional CAPM (at level 1% for individual assets, and 5% for portfolios), while this restriction is rejected under the three- and four-factor specifications.

The analysis of the validity of the asset pricing restrictions could be completed by an analysis of correct specification of the different conditional and unconditional factor models. A specification test would assess whether the proposed set of linear factors captures the systematic risk component in equity returns, and clearly differs from the test of the no-arbitrage restrictions introduced above. Developing a test of correct specification of conditional factor models with an unbalanced panel and double asymptotics is beyond the scope of the paper. We leave this interesting topic for future research.

References

N. Al-Najjar. Decomposition and characterization of risk with a continuum of random variables. *Econometrica*, 63(5):1195–1224, 1995.

N. Al-Najjar. Factor analysis and arbitrage pricing in large asset economies. *Journal of Economic Theory*, 78(2):231–262, 1998.

N. Al-Najjar. On the robustness of factor structures to asset repackaging. *Journal of Mathematical Economics*, 31(3):309–320, 1999.

D. W. K. Andrews. Cross-section regression with common shocks. *Econometrica*, 73(5):1551–1585, 2005.

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

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

M. Arellano and S. Bonhomme. Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies*, 79(3):987–1020, 2012.

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

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

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

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

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

D.A. Belsley, E. Kuh, and R.E. Welsch. *Regression diagnostics - Identifying influential data and sources of collinearity*. John Wiley & Sons, New York, 2004.

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

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

P. Billingsley. *Probability and measure, 3rd Edition*. John Wiley & Sons, New York, 1995.

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

O. Boguth, M. Carlson, A. Fisher, and M. Simutin. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics*, 102 (2):363–389, 2011.

S. Bonhomme. Functional differencing. *Econometrica*, 80(4):1337–1385, 2012.

D. Bosq. *Nonparametric statistics for stochastic processes*. Springer-Verlag, New York, 1998.

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

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

G. Chamberlain. Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5): 1305–1323, 1983.

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

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

J. H. Cochrane. *Asset pricing*. Princeton University Press, Princeton, 2005.

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

G. Connor and R. A. Korajczyk. An intertemporal equilibrium beta pricing model. *The Review of Financial Studies*, 2(3):373–392, 1989.

G. Connor, M. Hagmann, and O. Linton. Efficient semiparametric estimation of the Fama-French model and extensions. *Econometrica*, 80(2):713–754, 2012.

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

M. Cremers, A. Petajisto, and E. Zitzewitz. Should benchmark indices have alpha? Revisiting performance evaluation. *Working paper*, 2010.

J. Davidson. *Stochastic limit theory*. Advanced texts in Econometrics. Oxford University Press, Oxford, 1994.

J. Doob. *Stochastic processes*. John Wiley & Sons, New York, 1953.

D. Duffie. *Dynamic asset pricing theory, 3rd Edition*. Princeton University Press, Princeton, 2001.

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

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

E. F. Fama and K. R. French. Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193, 1997.

E. F. Fama and K. R. French. Dissecting anomalies. *The Jounal of Finance*, 63(4):1653–1678, 2008.

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

J. Fan, Y. Liao, and M. Mincheva. High dimensional covariance matrix estimation in approximate factor structure. *Annals of Statistics*, 39(6):3320–3356, 2011.

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

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

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

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

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

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

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

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

J. Gomes, L. Kogan, and L. Zhang. Equilibrium cross section of returns. *Journal of Political Economy*, 111 (4):693–732, 2003.

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

W. Greene. *Econometric Analysis, 6th ed.* Prentice Hall, 2008.

J. Hahn and G. Kuersteiner. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, 70(4):1639–1657, 2002.

J. Hahn and W. Newey. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319, 2004.

L. P. Hansen and S. F. Richard. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 55(3):587–613, 1987.

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

C. Hsiao. *Analysis of Panel Data*. Econometric Society Monographs, 2nd edition, Cambridge University Press, 2003.

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

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

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

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91, 1993.

R. Kan, C. Robotti, and J. Shanken. Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, forthcoming, 2012.

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

T. Lancaster. The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413, 2000.

M. Lettau and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849, 2001.

J. Lewellen and S. Nagel. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289–314, 2006.

J. Lewellen, S. Nagel, and J. Shanken. A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics*, 96(2):175–194, 2010.

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

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

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

S. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. *Handbook of the Economics of Finance vol. 2*, edited by George Constantinides, Milton Harris and Rene Stulz, forthcoming, 2011.

S. Ludvigson and S. Ng. The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, 83:171–222, 2007.

S. Ludvigson and S. Ng. Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12): 5027–5067, 2009.

C. MacKinlay. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, 38(1):3–28, 1995.

J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2007.

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

J. Neyman and E. L. Scott. Consistent estimation from partially consistent observations. *Econometrica*, 16 (1):1–32, 1948.

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

M. H. Pesaran and T. Yamagata. Testing slope homogeneity in large panels. *Journal of Econometrics*, 142 (1):50–93, 2008.

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

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

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

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

S. A. Ross. A simple approach to the valuation of risky streams. *Journal of Business*, 51(3):453–476, 1978.

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

J. Shanken. The arbitrage pricing theory: Is it testable? *Journal of Finance*, 37(5):1129–1140, 1982.

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

J. Shanken. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2): 99–120, 1990.

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

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

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

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

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

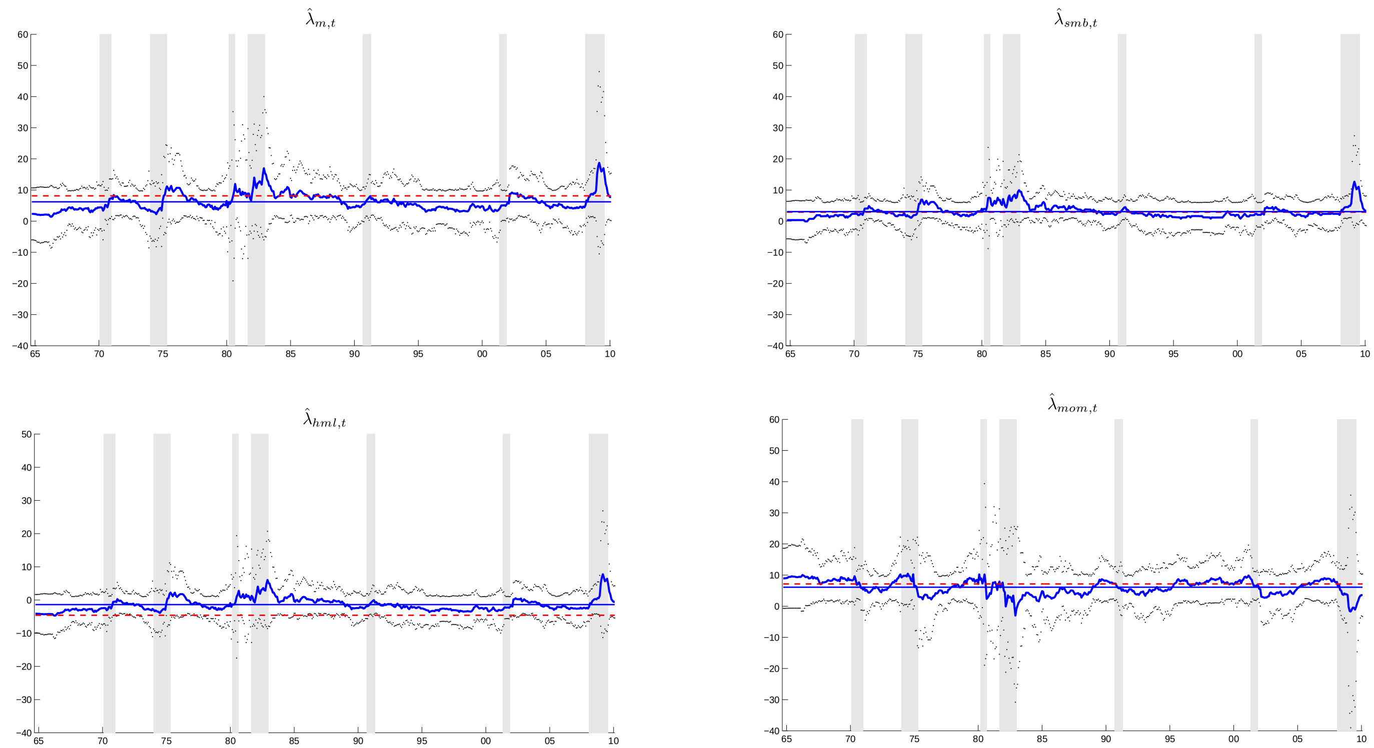
W. Stout. *Almost sure convergence*. Academic Press, New York, 1974.

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

H. White. *Asymptotic theory for econometricians*. Academic Press, 2001.

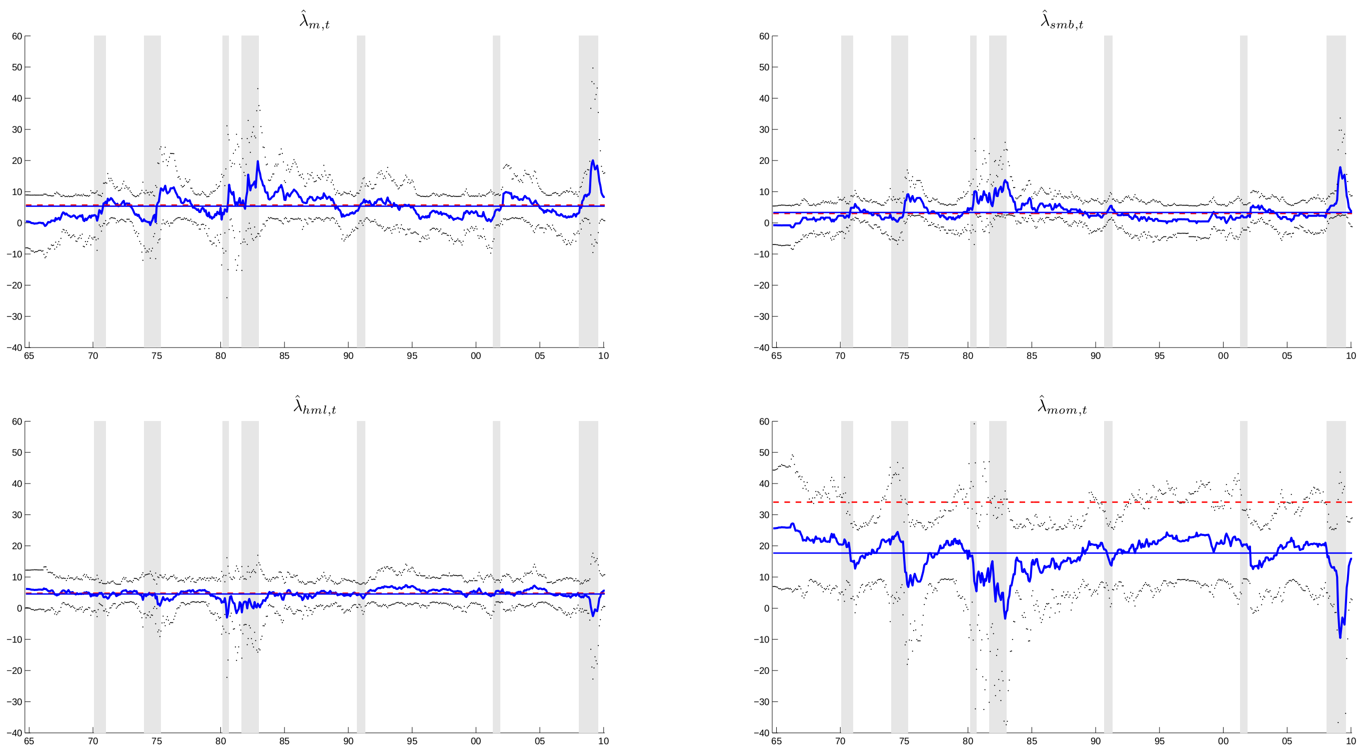
L. Zhang. The value premium. *Jounal of Finance*, 60(1):67–103, 2005.

Figure 1: Path of estimated annualized risk premia with $n = 9,936$



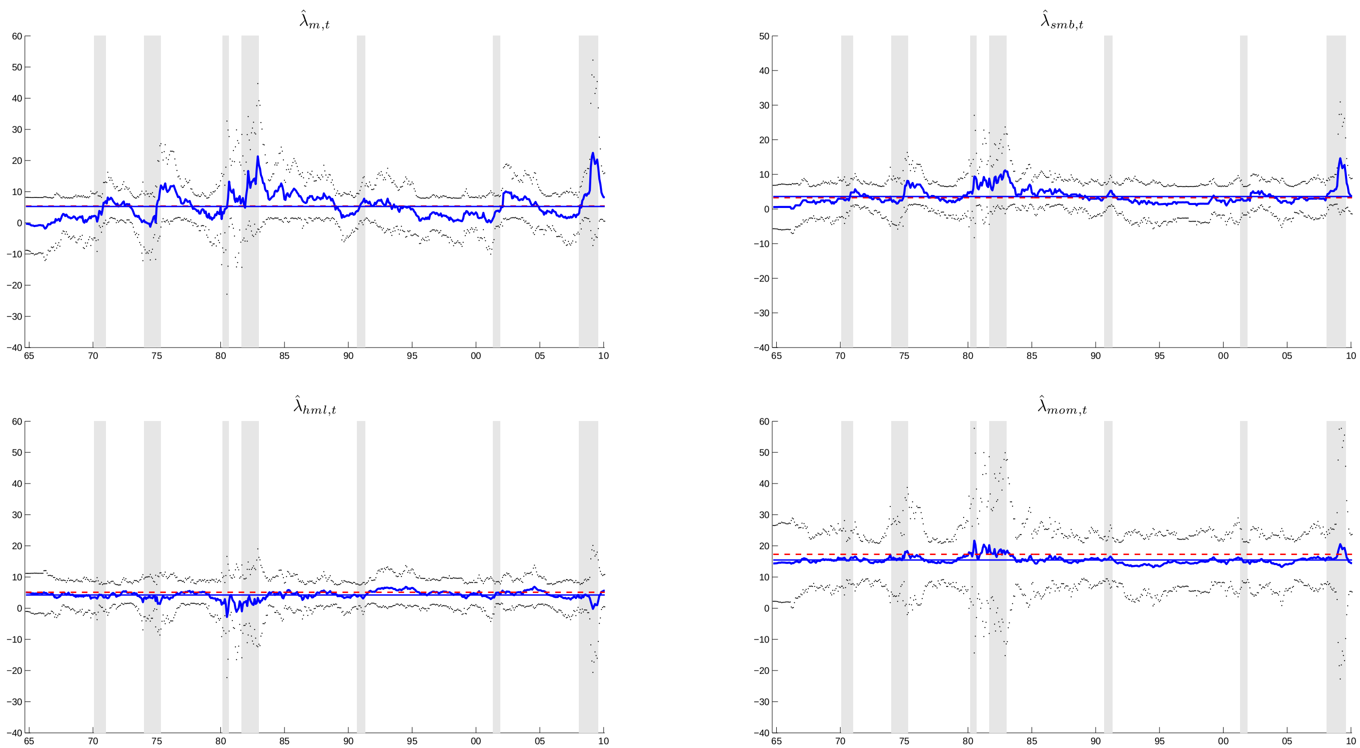
The figure plots the path of estimated annualized risk premia $\hat{\lambda}_{m,t}$, $\hat{\lambda}_{smb,t}$, $\hat{\lambda}_{hml,t}$ and $\hat{\lambda}_{mom,t}$ and their pointwise confidence intervals at 95% probability level. We also report the unconditional estimate (dashed horizontal line) and the average conditional estimate (solid horizontal line). We consider all stocks as base assets ($n = 9,936$ and $n^{\chi} = 3,900$). The vertical shaded areas denote recessions determined by the National Bureau of Economic Research (NBER). The recessions start at the peak of a business cycle and end at the trough.

Figure 2: Path of estimated annualized risk premia with $n = 25$



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 use the returns of the 25 Fama-French portfolios. 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).

Figure 3: Path of estimated annualized risk premia with $n = 100$



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 use the returns of the 100 Fama-French portfolios. 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

	Stocks ($n = 9,936$)		Portfolios ($n = 25$)		Portfolios ($n = 100$)	
	bias corrected estimate (%)	95% conf. interval	point estimate (%)	95% conf. interval	point estimate (%)	95% conf. interval
Four-factor model						
($n^{\chi} = 9,902$)						
λ_m	8.14	(3.26, 13.02)	5.70	(0.73, 10.67)	5.41	(0.42, 10.39)
λ_{smb}	2.86	(-0.50, 6.22)	3.02	(-0.48, 6.51)	3.28	(-0.27, 6.83)
λ_{hml}	-4.60	(-8.06, -1.14)	4.81	(1.21, 8.41)	5.11	(1.52, 8.71)
λ_{mom}	7.16	(2.56, 11.76)	34.03	(9.98, 58.07)	17.29	(8.55, 26.03)
Fama-French model						
($n^{\chi} = 9,904$)						
λ_m	7.77	(2.89, 12.65)	5.04	(0.11, 9.97)	4.88	(-0.08, 9.83)
λ_{smb}	2.64	(-0.72, 5.99)	3.00	(-0.42, 6.42)	3.35	(-0.13, 6.83)
λ_{hml}	-5.18	(-8.65, -1.72)	5.20	(1.66, 8.74)	5.20	(1.63, 8.77)
CAPM						
($n^{\chi} = 9,904$)						
λ_m	7.42	(2.54, 12.31)	6.98	(1.93, 12.02)	7.16	(2.06, 12.25)

The table contains the estimated annualized risk premia for the market (λ_m), size (λ_{smb}), book-to-market (λ_{hml}) and momentum (λ_{mom}) factors. We report the bias corrected estimates $\hat{\lambda}_B$ of λ for individual stocks ($n = 9,936$). In order to build the confidence intervals for $n = 9,936$, we use the HAC estimator $\hat{\Sigma}_f$ defined in Section 2.4. When we consider 25 and 100 portfolios as base assets, we compute an estimate of the variance-covariance matrix $\Sigma_{\lambda,n}$ defined in Section 2.4.

Table 2: Estimated annualized ν for the unconditional models

	Stocks ($n = 9,936$)		Portfolios ($n = 25$)		Portfolios ($n = 100$)	
	bias corrected estimate (%)	95% conf. interval	point estimate (%)	95% conf. interval	point estimate (%)	95% conf. interval
Four-factor model						
($n^{\chi} = 9,902$)						
ν_m	3.29	(2.88, 3.69)	0.85	(-0.10, 1.79)	0.55	(-0.46, 1.57)
ν_{smb}	-0.41	(-0.95, 0.13)	-0.26	(-1.24, 0.72)	0.01	(-1.14, 1.16)
ν_{hml}	-9.38	(-10.12, -8.64)	0.03	(-0.95, 1.01)	0.33	(-0.63, 1.30)
ν_{mom}	-1.47	(-2.86, -0.08)	25.40	(1.80, 49.00)	8.66	(1.23, 16.10)
Fama-French model						
($n^{\chi} = 9,904$)						
ν_m	2.92	(2.48, 3.35)	0.18	(-0.51, 0.87)	0.02	(-0.84, 0.88)
ν_{smb}	-0.63	(-1.11, -0.15)	-0.27	(-0.93, 0.40)	0.08	(-0.85, 1.01)
ν_{hml}	-9.96	(-10.62, -9.31)	0.41	(-0.32, 1.15)	0.42	(-0.44, 1.28)
CAPM						
($n^{\chi} = 9,904$)						
ν_m	2.57	(2.17, 2.97)	2.12	(0.85, 3.40)	2.30	(0.84, 3.77)

The table contains the annualized estimates of the components of vector ν for the market (ν_m), size (ν_{smb}), book-to-market (ν_{hml}) and momentum (ν_{mom}) factors. We report the bias corrected estimates $\hat{\nu}_B$ of ν 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 variance-covariance matrix $\Sigma_{\nu,n}$ defined in Section 2.4.

Table 3: Test results for asset pricing restrictions in the unconditional models

	Test of the null hypothesis $\mathcal{H}_0: \ a(\gamma) = b(\gamma)' \nu$			Test of the null hypothesis $\mathcal{H}_0: \ a(\gamma) = 0$		
	Stocks	Portfolios ($n = 25$)	Portfolios ($n = 100$)	Stocks	Portfolios ($n = 25$)	Portfolios ($n = 100$)
			Four-factor model			
	$(n^{\chi} = 1,400)$			$(n^{\chi} = 1,400)$		
Test statistic	2.0088	35.2231	253.2575	19.1803	74.9100	263.3395
p-value	0.0223	0.0267	0.0000	0.0000	0.0000	0.0000
			Fama-French model			
	$(n^{\chi} = 1,400)$			$(n^{\chi} = 1,400)$		
Test statistic	2.9593	83.6846	253.9652	28.0328	87.3767	270.7899
p-value	0.0015	0.0000	0.0000	0.0000	0.0000	0.0000
			CAPM			
	$(n^{\chi} = 1,400)$			$(n^{\chi} = 1,400)$		
Test statistic	8.2576	110.8368	276.3679	11.5882	111.6735	278.3949
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

We compute the statistics $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT}$ based on $\hat{Q}_e$ and $\hat{Q}_a$ defined in Proposition 5 for the individual stocks to test the null hypotheses $\mathcal{H}_0: \ a(\gamma) = b(\gamma)' \nu$ and $\mathcal{H}_0: \ a(\gamma) = 0$, respectively. The trimming levels are $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/240$. For $n = 25$ and $n = 100$, we compute the weighted statistics $T\hat{e}'\hat{V}^{-1}\hat{e}$ and $T\hat{a}'\hat{V}_a^{-1}\hat{a}$ (Gibbons, Ross and Shanken (1989)), where $\hat{e}$ and $\hat{a}$ are $n \times 1$ vectors with elements $\hat{e}_i$ and $\hat{a}_i$, and $\hat{V} = (\hat{v}_{ij})$ and $\hat{V}_a = (\hat{v}_{a,ij})$ are $n \times n$ matrices with elements $\hat{v}_{ij} = \hat{c}_{\nu}'\hat{Q}_x^{-1}\hat{S}_{ij}\hat{Q}_x^{-1}\hat{c}_{\nu}$, and $\hat{v}_{a,ij} = E_1'\hat{Q}_x^{-1}\hat{S}_{ij}\hat{Q}_x^{-1}E_1$. The table reports the p-values of the statistics.

Table 4: Test results for the asset pricing restrictions in the conditional models

	Test of the null hypothesis $\mathcal{H}_0 :\ \beta_1(\gamma) = \beta_3(\gamma)\nu$			Test of the null hypothesis $\mathcal{H}_0 :\ \beta_1(\gamma) = 0$		
	Stocks	Portfolios ($n = 25$)	Portfolios ($n = 100$)	Stocks	Portfolios ($n = 25$)	Portfolios ($n = 100$)
Four-factor model						
	$(n^\chi = 1,373)$			$(n^\chi = 1,373)$		
Test statistic	3.2514	13.4815	11.6389	3.8683	14.3080	12.2098
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Fama-French model						
	$(n^\chi = 1,393)$			$(n^\chi = 1,393)$		
Test statistic	3.1253	15.7895	12.8938	3.8136	15.9038	13.0349
p-value	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000
CAPM						
	$(n^\chi = 1,395)$			$(n^\chi = 1,395)$		
Test statistic	1.7322	9.2934	9.5153	1.7381	9.6680	9.8007
p-value	0.0416	0.2076	0.1825	0.0411	0.0000	0.0000

We compute the statistics $\tilde{\Sigma}_\xi^{-1/2}\hat{\xi}_{nT}$ based on $\hat{Q}_e$ and $\hat{Q}_a$ defined in Proposition 5 for the individual stocks to test the null hypotheses $\mathcal{H}_0 :\ \beta_1(\gamma) = \beta_3(\gamma)\nu$ and $\mathcal{H}_0 :\ \beta_1(\gamma) = 0$, respectively. The trimming levels are $\chi_{1,T} = 15$ and $\chi_{2,T} = 546/240$. For $n = 25$ and $n = 100$, we compute the test statistics $T\hat{Q}_e$ and $T\hat{Q}_a$. The table reports the p-values of the statistics.

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 exist 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:*

a) $\mathbb{P}\left[\left\|\frac{1}{T}\sum_t \left(x_t x_t' - E\left[x_t x_t'\right]\right)\right\| \geq \delta\right] \leq C_1 T \exp\left\{-C_2 \delta^2 T^{\eta}\right\} + C_3 \delta^{-1} \exp\left\{-C_4 T^{\bar{\eta}}\right\}$.

Furthermore, for all $\delta > 0$, $T \in \mathbb{N}$, *and* $1 \leq k, l, m \leq K+1$, *the same upper bound holds for:*

b) $\sup_{\gamma \in [0,1]} \mathbb{P}\left[\left\|\frac{1}{T}\sum_t I_t(\gamma)\left(x_t x_t' - E\left[x_t x_t'\right]\right)\right\| \geq \delta\right]$; *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) $\sup_{\gamma, \gamma' \in [0,1]} \mathbb{P}\left[\left|\frac{1}{T}\sum_t (I_t(\gamma) I_t(\gamma') - E[I_t(\gamma) I_t(\gamma')])\right| \geq \delta\right]$;

e) $\sup_{\gamma, \gamma' \in [0,1]} \mathbb{P}\left[\left\|\frac{1}{T}\sum_t I_t(\gamma) I_t(\gamma')\left(\varepsilon_t(\gamma)\varepsilon_t(\gamma') x_t x_t' - E\left[\varepsilon_t(\gamma)\varepsilon_t(\gamma') x_t x_t'\right]\right)\right\| \geq \delta\right]$;

f) $\sup_{\gamma, \gamma' \in [0,1]} \mathbb{P}\left[\left|\frac{1}{T}\sum_t I_t(\gamma) I_t(\gamma') x_{t,k} x_{t,l} x_{t,m} \varepsilon_t(\gamma)\right| \geq \delta\right]$.

Assumption C.2 *There exists a constant* $M > 0$ *such that for all* $T \in \mathbb{N}$ *we have:*

$$\sup_{\gamma \in [0,1]} E\left[\frac{1}{T}\sum_{t_1, t_2, t_3} |cov(\varepsilon_{t_1}^2(\gamma), \varepsilon_{t_2}(\gamma)\varepsilon_{t_3}(\gamma) | x_{\underline{T}})|\right] \leq M.$$

Assumption C.3 *There exists a constant* $M > 0$ *such that for all* $n, T \in \mathbb{N}$ *we have:*

a) $E\left[\frac{1}{nT}\sum_{i,j}\sum_{t_1, t_2} E\left[\left|cov\left(\varepsilon_{i,t_1}^2, \varepsilon_{j,t_2}^2 | x_{\underline{T}}, \gamma_i, \gamma_j\right)\right|^2 | \gamma_i, \gamma_j\right]^{1/2}\right] \leq M$.

b) $E\left[\frac{1}{nT^2}\sum_{i,j}\sum_{t_1, t_2, t_3, t_4} E\left[\left|cov\left(\varepsilon_{i,t_1}\varepsilon_{i,t_2}, \varepsilon_{j,t_3}\varepsilon_{j,t_4} | x_{\underline{T}}, \gamma_i, \gamma_j\right)\right|^2 | \gamma_i, \gamma_j\right]^{1/2}\right] \leq M$;

c) $E\left[\frac{1}{nT^2}\sum_{i,j}\sum_{t_1, t_2, t_3, t_4} E\left[\left|cov\left(\eta_{i,t_1}\varepsilon_{i,t_2}, \eta_{j,t_3}\varepsilon_{j,t_4} | x_{\underline{T}}, \gamma_i, \gamma_j\right)\right|^2 | \gamma_i, \gamma_j\right]^{1/2}\right] \leq M$, *where* $\eta_{i,t} := \varepsilon_{i,t}^2 - \sigma_{ii,t}$;

d) $E\left[\frac{1}{nT^2}\sum_{i,j}\sum_{t_1, t_2, t_3, t_4} E\left[\left|cov\left(\eta_{i,t_1}\eta_{i,t_2}, \eta_{j,t_3}\eta_{j,t_4} | x_{\underline{T}}, \gamma_i, \gamma_j\right)\right|^2 | \gamma_i, \gamma_j\right]^{1/2}\right] \leq M$;

e) $E\left[\frac{1}{nT^3}\sum_{i,j}\sum_{t_1,\ldots,t_6} E\left[\left|cov\left(\varepsilon_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{i,t_3}, \varepsilon_{j,t_4}\varepsilon_{j,t_5}\varepsilon_{j,t_6}|x_{\underline{T}},\gamma_i,\gamma_j\right)\right|^2|\gamma_i,\gamma_j\right]^{1/2}\right] \leq M$*;*

f) $E\left[\frac{1}{nT^3}\sum_{i,j}\sum_{t_1,\ldots,t_6} E\left[\left|cov\left(\eta_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{i,t_3}, \eta_{j,t_4}\varepsilon_{j,t_5}\varepsilon_{j,t_6}|x_{\underline{T}},\gamma_i,\gamma_j\right)\right|^2|\gamma_i,\gamma_j\right]^{1/2}\right] \leq M.$

Assumption C.4 *a) There exists a constant* $M > 0$ *such that* $\|x_t\| \leq M$*, P-a.s.. Moreover, b)* $\sup_{\gamma\in[0,1]} \|\beta(\gamma)\| < \infty$ *and c)* $\inf_{\gamma\in[0,1]} E[I_t(\gamma)] > 0$*.*

Assumption C.5 *The trimming constants satisfy* $\chi_{1,T} = O\left((\log T)^{\kappa_1}\right)$ *and* $\chi_{2,T} = O\left((\log T)^{\kappa_2}\right)$*, with* κ_1*,* $\kappa_2 > 0$*.*

Assumptions C.1 and C.2 restrict the serial dependence of the factors and the individual processes of observability indicators and error terms. Specifically, Assumption C.1 a) gives an upper bound for large-deviation probabilities of the sample average of random matrices $x_t x_t'$. It implies that the first two sample moments of the factor vector converge in probability to the corresponding population moments at a rate $O_p(T^{-\eta/2}(\log T)^c)$, for some $c > 0$. Assumptions C.1 b)-f) give similar upper bounds for large-deviation probabilities of sample averages of processes involving factors, observability indicators and error terms, uniformly w.r.t. $\gamma \in [0,1]$. We use these assumptions to prove the convergence of time series averages uniformly across assets. Assumption C.2 involves conditional covariances of products of error terms. Assumptions C.1 and C.2 are satisfied e.g. when the factors and the individual processes of observability indicators and error terms feature mixing serial dependence, with mixing coefficients uniformly bounded w.r.t. $\gamma \in [0,1]$ (see e.g. Bosq (1998), Theorems 1.3 and 1.4). Assumptions C.3 a)-f) restrict both serial and cross-sectional dependence of the error terms. They involve conditional covariances between products of error terms $\varepsilon_{i,t}$ and innovations $\eta_{i,t} = \varepsilon_{i,t}^2 - \sigma_{ii,t}$ for different assets and dates. These assumptions can be satisfied under weak serial and cross-sectional dependence of the errors, such as temporal mixing and block dependence across assets. Assumptions C.4 a) and b) require uniform upper bounds on factor values, factor loadings and intercepts. Assumption C.4 c) implies that asymptotically the fraction of the time period in which an asset return is observed is bounded away from zero uniformly across assets. Assumptions C.4 a)-c) ease the proofs. Assumption C.5 gives an upper bound on the divergence rate of the trimming constants. The slow logarithmic divergence rate allows to control the first-pass estimation error in the second-pass regression.

For conditional models, we use Assumptions C.1-C.5 with x_t replaced by the extended vector of common and firm-specific regressors as defined in Section 3.1. More precisely, for Assumption C.1a) we replace x_t by $x_t(\gamma) := \left(vech(X_t)', Z_{t-1}' \otimes Z_{t-1}(\gamma)', f_t' \otimes Z_{t-1}', f_t' \otimes Z_{t-1}(\gamma)'\right)'$, and require the bound to be valid uniformly w.r.t. $\gamma \in [0,1]$. For Assumptions C.1 b)-f) we replace x_t by $x_t(\gamma)$. For Assumptions C.2 and C.3 we replace $x_{\underline{T}}$ by $x_{\underline{T}}(\gamma)$, and by $x_{\underline{T}}(\gamma_i), x_{\underline{T}}(\gamma_j)$, respectively. For Assumption C.4a) we replace the bound on $\|x_t\|$ with bounds on $\|Z_t\|$, and on $\|Z_t(\gamma)\|$ uniformly w.r.t. $\gamma \in [0,1]$. Furthermore, we use:

Assumption C.6 *There exists a constant $M > 0$ such that $\left\|E\left[u_t u_t' | Z_{t-1}\right]\right\| \leq M$ for all t, where $u_t = f_t - E[f_t|\mathcal{F}_{t-1}]$.*

Assumption C.6 requires a bounded conditional variance-covariance matrix for the linear innovation u_t associated with the factor process. We use this assumption to prove that we can consistently estimate matrix F of the coefficients of the linear projection of factor f_t on variables Z_{t-1} by a SUR regression.

Appendix 2: Unconditional factor model

A.2.1 Proof of Proposition 1 and link with Chamberlain and Rothschild (1983)

To ease notations, we assume w.l.o.g. that the continuous distribution G is uniform on $[0,1]$. For a given countable collection of assets $\gamma_1, \gamma_2, ...$ in $[0,1]$, let $\mu_n = A_n + B_n E[f_1|\mathcal{F}_0]$ and $\Sigma_n = B_n V[f_1|\mathcal{F}_0] B_n'$ $+\Sigma_{\varepsilon,1,n}$, for $n \in \mathbb{N}$, be the mean vector and the covariance matrix of asset excess returns $\left(R_1(\gamma_1), ..., R_1(\gamma_n)\right)'$ conditional on $\mathcal{F}_0$, where $A_n = [a(\gamma_1), ..., a(\gamma_n)]'$, and $B_n = [b(\gamma_1), ..., b(\gamma_n)]'$. Let $e_n = \mu_n - B_n \left(B_n' B_n\right)^{-1} B_n' \mu_n = A_n - B_n \left(B_n' B_n\right)^{-1} B_n' A_n$ be the residual of the orthogonal projection of μ_n (and A_n) onto the columns of B_n. Furthermore, let $\mathcal{P}_n$ denote the set of portfolios p_n that invest in the risk-free asset and risky assets $\gamma_1, ..., \gamma_n$, for $n \in \mathbb{N}$, with portfolio shares measurable w.r.t. $\mathcal{F}_0$, and let $\mathcal{P}$ denote the set of portfolio sequences (p_n), with $p_n \in \mathcal{P}_n$. For portfolio $p_n \in \mathcal{P}_n$, the cost, the conditional expected return, and the conditional variance are given by $C(p_n) = \alpha_{0,n} + \alpha_n' \iota_n$, $E\left[p_n|\mathcal{F}_0\right] = R_0 C(p_n) + \alpha_n' \mu_n$, and $V\left[p_n|\mathcal{F}_0\right] = \alpha_n' \Sigma_n \alpha_n$, where $\iota_n = (1, ..., 1)'$ and $\alpha_n = (\alpha_{1,n}, ..., \alpha_{n,n})'$. Moreover, let $\rho = \sup_p E[p|\mathcal{F}_0]/V[p|\mathcal{F}_0]^{1/2}$, where the sup is w.r.t. portfolios $p \in \bigcup_{n \in \mathbb{N}} \mathcal{P}_n$ with $C(p) = 0$

and $p \neq 0$, be the maximal Sharpe ratio of zero-cost portfolios. For expository purpose, we do not make explicit the dependence of μ_n, Σ_n, e_n, $\mathcal{P}_n$, and ρ on the collection of assets (γ_i).

The statement of Proposition 1 is proved by contradiction. Suppose that $\inf_{\nu \in \mathbb{R}^K} \int [a(\gamma) - b(\gamma)'\nu]^2 d\gamma = \int [a(\gamma) - b(\gamma)'\nu_\infty]^2 d\gamma > 0$, where $\nu_\infty = \left(\int b(\gamma) b(\gamma)' d\gamma \right)^{-1} \int b(\gamma) a(\gamma) d\gamma$. By the strong LLN and Assumption APR.2, we have that:

$$\frac{1}{n}\|e_n\|^2 = \inf_{\nu \in \mathbb{R}^K} \frac{1}{n} \sum_{i=1}^{n} [a(\gamma_i) - b(\gamma_i)'\nu]^2 \to \int [a(\gamma) - b(\gamma)'\nu_\infty]^2 d\gamma, \tag{18}$$

as $n \to \infty$, for any sequence (γ_i) in a set $\mathcal{J}_1 \subset \Gamma$, with measure $\mu_\Gamma(\mathcal{J}_1) = 1$. Let us now show that an asymptotic arbitrage portfolio exists based on any sequence in $\mathcal{J}_1 \bigcap \mathcal{J}$, where set $\mathcal{J}$ is defined in Assumption APR.4 (i). Define the portfolio sequence (q_n) with investments $\alpha_n = \frac{1}{\|e_n\|^2} e_n$ and $\alpha_{0,n} = -\iota_n' \alpha_n$. This static portfolio has zero cost, i.e., $C(q_n) = 0$, while $E[q_n|\mathcal{F}_0] = 1$ and $V[q_n|\mathcal{F}_0] \leq eig_{\max}(\Sigma_{\varepsilon,1,n})\|e_n\|^{-2}$. Moreover, we have $V[q_n|\mathcal{F}_0] = E\left[(q_n - E[q_n|\mathcal{F}_0])^2 |\mathcal{F}_0\right] \geq E\left[(q_n - E[q_n|\mathcal{F}_0])^2 |\mathcal{F}_0, q_n \leq 0\right] P[q_n \leq 0|\mathcal{F}_0] \geq P[q_n \leq 0|\mathcal{F}_0]$. Hence, we get: $P[q_n > 0|\mathcal{F}_0] \geq 1 - V[q_n|\mathcal{F}_0] \geq 1 - eig_{\max}(\Sigma_{\varepsilon,1,n})\|e_n\|^{-2}$. Thus, by using $eig_{\max}(\Sigma_{\varepsilon,1,n}) = o(n)$ from Assumption APR.4 (i) and $\|e_n\|^{-2} = O(1/n)$ from Equation (18), we get $P[q_n > 0|\mathcal{F}_0] \to 1$, P-a.s.. By using the Law of Iterated Expectation and the Lebesgue dominated convergence theorem, $P[q_n > 0] \to 1$. Hence, portfolio (q_n) is an asymptotic arbitrage opportunity. Since asymptotic arbitrage portfolios are ruled out by Assumption APR.5, it follows that we must have $\int [a(\gamma) - b(\gamma)'\nu_\infty]^2 d\gamma = 0$, that is, $a(\gamma) = b(\gamma)'\nu$, for $\nu = \nu_\infty$ and almost all $\gamma \in [0, 1]$. Such vector ν is unique by Assumption APR.2, and Proposition 1 follows.

Let us now establish the link between the no-arbitrage conditions and asset pricing restrictions in CR on the one hand, and the asset pricing restriction (3) in the other hand. Let $\mathcal{J}^* \subset \Gamma$ be the set of countable collections of assets (γ_i) such that $\mathbb{P}$ [Conditions (i) and (ii) hold for any static portfolio sequence (p_n) in $\mathcal{P}] = 1$, where Conditions (i) and (ii) are: (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 $\delta \geq 0$. Condition (i) means that, if the conditional variability and cost vanish, so does the conditional expected return. Condition (ii) means that, if the conditional variability vanishes and the cost is positive, the conditional expected return is non-negative. They correspond to Conditions A.1 (i) and (ii) in CR written conditionally on $\mathcal{F}_0$ and for a given countable collection of assets (γ_i). Hence, the set $\mathcal{J}^*$ is the set permitting no asymptotic arbitrage opportunities in the sense of CR almost surely (see also Chamberlain (1983)).

Proposition APR: *Under Assumptions APR.1-APR.4, either* $\mu_\Gamma\left(\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty\right)=\mu_\Gamma(\mathcal{J}^*)=1$, *or* $\mu_\Gamma\left(\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty\right)=\mu_\Gamma(\mathcal{J}^*)=0$. *The former case occurs if, and only if, the asset pricing restriction (3) holds.*

The fact that $\mu_\Gamma\left(\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty\right)$ is either $=1$, or $=0$, is a consequence of the Kolmogorov zero-one law (e.g., Billingsley (1995)). Indeed, $\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty$ if, and only if, $\inf_{\nu\in\mathbb{R}^K}\sum_{i=n}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty$, for any $n\in\mathbb{N}$. Thus, the zero-one law applies since the event $\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty$ belongs to the tail sigma-field $\mathcal{T}=\bigcap_{n=1}^{\infty}\sigma(\gamma_i,i=n,n+1,...)$, and the variables γ_i are i.i.d. under measure μ_Γ.

Proof of Proposition APR: The proof involves four steps.

STEP 1: If $\mu_\Gamma\left(\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty\right)>0$, then the asset pricing restriction (3) holds. This step is proved by contradiction. Suppose that the asset pricing restriction (3) does not hold, and thus $\int[a(\gamma)-b(\gamma)'\nu_\infty]^2d\gamma>0$. Then, we get $\mu_\Gamma\left(\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty\right)=0$, by the convergence in (18).

STEP 2: If the asset pricing restriction (3) holds, then $\mu_\Gamma\left(\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty\right)=1$. Indeed, $\mu_\Gamma\left(\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2=0\right)=1$, if the asset pricing restriction (3) holds for some vector $\nu\in\mathbb{R}^K$.

STEP 3: If $\mu_\Gamma(\mathcal{J}^*)>0$, then the asset pricing restriction (3) holds. By following the same arguments as in CR on p. 1295-1296, we have $\rho^2\geq\mu_n'\Sigma_{\varepsilon,1,n}^{-1}\mu_n$ and $\Sigma_{\varepsilon,1,n}^{-1}\geq eig_{\max}(\Sigma_{\varepsilon,1,n})^{-1}[I_n-B_n(B_n'B_n)^{-1}B_n']$, for any (γ_i) in $\mathcal{J}^*$. Thus, we get: $\rho^2 eig_{\max}(\Sigma_{\varepsilon,1,n})\geq\mu_n'\left(I_n-B_n(B_n'B_n)^{-1}B_n'\right)\mu_n=\min_{\lambda\in\mathbb{R}^K}\|\mu_n-B_n\lambda\|^2=\min_{\nu\in\mathbb{R}^K}\|A_n-B_n\nu\|^2=\min_{\nu\in\mathbb{R}^K}\sum_{i=1}^{n}[a(\gamma_i)-b(\gamma_i)'\nu]^2$, for any $n\in\mathbb{N}$, P-a.s.. Hence, we deduce

$$\min_{\nu\in\mathbb{R}^K}\frac{1}{n}\sum_{i=1}^{n}[a(\gamma_i)-b(\gamma_i)'\nu]^2\leq\rho^2\frac{1}{n}eig_{\max}(\Sigma_{\varepsilon,1,n}),\tag{19}$$

for any n, P-a.s., and for any sequence (γ_i) in $\mathcal{J}^*$. Moreover, $\rho < \infty$, P-a.s., by the same arguments as in CR, Corollary 1, and by using that the condition in CR, footnote 6, is implied by our Assumption APR.4 (ii). Then, by the convergence in (18), the LHS of Inequality (19) converges to $\int [a(\gamma) - b(\gamma)'\nu_\infty]^2 d\gamma$, for μ_Γ-almost every sequence (γ_i) in $\mathcal{J}^*$. From Assumption APR.4 (i), the RHS is $o(1)$, P-a.s., for μ_Γ-almost every sequence (γ_i) in Γ. Since $\mu_\Gamma(\mathcal{J}^*) > 0$, it follows that $\int [a(\gamma) - b(\gamma)'\nu_\infty]^2 d\gamma = 0$, i.e., $a(\gamma) = b(\gamma)'\nu$, for $\nu = \nu_\infty$ and almost all $\gamma \in [0, 1]$.

STEP 4: If the asset pricing restriction (3) holds, then $\mu_\Gamma(\mathcal{J}^*) = 1$. If (3) holds, it follows that $e_n = 0$ and $\mu_n = B_n(B_n'B_n)^{-1}B_n'\mu_n$, for all n, for μ_Γ-almost all sequences (γ_i). Then, we get $E[p_n|\mathcal{F}_0] = R_0 C(p_n)$ $+\alpha_n' B_n(B_n'B_n/n)^{-1}B_n'\mu_n/n$. Moreover, we have: $V[p_n|\mathcal{F}_0] = (B_n^{'}\alpha_n)'V[f_1|\mathcal{F}_0](B_n^{'}\alpha_n) + \alpha_n^{'}\Sigma_{\varepsilon,1,n}\alpha_n \geq eig_{\min}(V[f_1|\mathcal{F}_0]) \left\| B_n^{'}\alpha_n \right\|^2$, where $eig_{\min}(V[f_1|\mathcal{F}_0]) > 0$, P-a.s. (Assumption APR.4 (iii)). Since $B_n'B_n/n$ converges to a positive definite matrix and $B_n'\mu_n/n$ is bounded, for μ_Γ-almost any sequence (γ_i), Conditions (i) and (ii) in the definition of set $\mathcal{J}^*$ follow, for μ_Γ-almost any sequence (γ_i), that is, $\mu_\Gamma(\mathcal{J}^*) = 1$.

A.2.2 Proof of Proposition 2

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

$$\hat{\nu} - \nu = \hat{Q}_b^{-1} \frac{1}{n} \sum_i \hat{w}_i \hat{b}_i c_\nu' \left(\hat{\beta}_i - \beta_i \right). \tag{20}$$

The consistency of $\hat{\nu}$ follows from the next Lemma, which is proved in Section A.2.2 c) below. The notation $I_{n,T} = O_{p,log}(a_{n,T})$ means that $I_{n,T}/a_{n,T}$ is bounded in probability by some power of the logarithmic term $\log(T)$ as $n, T \to \infty$.

Lemma 1 *Under Assumptions A.1 b), SC.1-SC.2, C.1, C.4 and C.5, we have: (i)* $\sup_i \mathbf{1}_i^\chi \|\hat{\beta}_i - \beta_i\| = O_{p,log}\left(T^{-\eta/2}\right)$; *(ii)* $\sup_i w_i = O(1)$; *(iii)* $\frac{1}{n}\sum_i |\hat{w}_i - w_i| = o_p(1)$; *(iv)* $\hat{Q}_b - Q_b = o_p(1)$, *when* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $\bar{\gamma} > 0$.

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

c) Proof of Lemma 1: (i) We use $\hat{\beta}_i - \beta_i = \frac{\tau_{i,T}}{\sqrt{T}} \hat{Q}_{x,i}^{-1} Y_{i,T}$ and $\mathbf{1}_i^{\chi} \tau_{i,T} \leq \chi_{2,T}$. Moreover, $\|\hat{Q}_{x,i}^{-1}\|^2 = Tr\left(\hat{Q}_{x,i}^{-2}\right) = \sum_{k=1}^{K+1} \lambda_{k,i}^{-2} \leq (K+1) CN\left(\hat{Q}_{x,i}\right)^2$, where the $\lambda_{k,i}$ are the eigenvalues of matrix $\hat{Q}_{x,i}$ and we use $eig_{max}\left(\hat{Q}_{x,i}\right) \geq 1$, which implies $\mathbf{1}_i^{\chi} \|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}$. Thus, $\sup_i \mathbf{1}_i^{\chi} \|\hat{\beta}_i - \beta_i\| = O_{p,log}\left(T^{-1/2} \sup_i \|Y_{i,T}\|\right)$ from Assumption C.5. Now let $\delta_T := T^{-\eta/2} (\log T)^{(1+\bar{\gamma})/(2C_2)}$, where $\eta, C_2 > 0$ are as in Assumption C.1 and $\bar{\gamma} > 0$ is such that $n = O(T^{\bar{\gamma}})$. We have:

$$\mathbb{P}\left[T^{-1/2} \sup_i \|Y_{i,T}\| \geq \delta_T\right] \leq n\mathbb{P}\left[T^{-1/2}\|Y_{i,T}\| \geq \delta_T\right] = nE\left[\mathbb{P}\left(T^{-1/2}\|Y_{i,T}\| \geq \delta_T | \gamma_i\right)\right]$$

$$\leq n \sup_{\gamma \in [0,1]} \mathbb{P}\left[\left\|\frac{1}{T}\sum_t I_t(\gamma) x_t \varepsilon_t(\gamma)\right\| \geq \delta_T\right] \leq n\left(C_1 T \exp\left\{-C_2 \delta_T^2 T^{\eta}\right\} + C_3 \delta_T^{-1} \exp\left\{-C_4 T^{\bar{\eta}}\right\}\right) = O(1),$$

from Assumption C.1 c). Part (i) follows. By using $w_i = v_i^{-1}$, $\tau_i \geq 1$ and $eig_{min}(S_{ii}) \geq M^{-1} eig_{min}(Q_x)$ from Assumption A.1 b), part (ii) follows. Part (iii) is proved in the supplementary materials by using Assumptions C.1, C.4 and C.5. Finally, part (iv) follows from $\hat{Q}_b - Q_b = \frac{1}{n}\sum_i (\hat{w}_i \hat{b}_i \hat{b}_i' - w_i b_i b_i')$ $+ \frac{1}{n}\sum_i w_i b_i b_i' - Q_b$, by using parts (i)-(iii) and the LLN.

A.2.3 Proof of Proposition 3

a) Asymptotic normality of $\hat{\nu}$. From Equation (20) and by using $\hat{\beta}_i - \beta_i = \frac{\tau_{i,T}}{\sqrt{T}} \hat{Q}_{x,i}^{-1} Y_{i,T}$ we get:

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

Let $I_1 := \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i Y_{i,T}' \hat{Q}_{x,i}^{-1} c_\nu$. Then, from Equation (21) and the definition of $\hat{B}_\nu$, we get:

$$\begin{aligned} \sqrt{nT}\left(\hat{\nu} - \frac{1}{T}\hat{B}_\nu - \nu\right) &= \hat{Q}_b^{-1} I_1 + \frac{1}{\sqrt{T}} \hat{Q}_b^{-1} E_2' \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T}^2 \left(\hat{Q}_{x,i}^{-1} Y_{i,T} Y_{i,T}' \hat{Q}_{x,i}^{-1} c_\nu - \tau_{i,T}^{-1} \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} c_{\hat{\nu}}\right) \\ &=: \hat{Q}_b^{-1} I_1 + \frac{1}{\sqrt{T}} \hat{Q}_b^{-1} E_2' I_2. \quad (22) \end{aligned}$$

Let us first show that $\hat{Q}_b^{-1} I_1$ is asymptotically normal. We use the next Lemma, which is proved below in Subsection A.2.3 c).

Lemma 2 *Under Assumptions A.1, A.3, SC.1-SC.2 and C.1, C.3-C.5, we have* $I_1 = \frac{1}{\sqrt{n}} \sum_i w_i \tau_i b_i Y'_{i,T} \hat{Q}_x^{-1} c_\nu + o_p(1)$*, when* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $\bar{\gamma} > 0$.

From Lemmas 1 (iv) and 2, and using $vec\left[ABC\right] = \left[C' \otimes A\right] vec\left[B\right]$ (MN Theorem 2, p. 35), we have:

$$\hat{Q}_b^{-1} I_1 = \hat{Q}_b^{-1} \left(\frac{1}{\sqrt{n}} \sum_i w_i \tau_i b_i Y'_{i,T} \right) \hat{Q}_x^{-1} c_\nu + o_p(1) = \left(c'_\nu \hat{Q}_x^{-1} \otimes \hat{Q}_b^{-1} \right) \frac{1}{\sqrt{n}} \sum_i w_i \tau_i \left(Y_{i,T} \otimes b_i \right) + o_p(1).$$

Then, we deduce $\hat{Q}_b^{-1} I_1 \Rightarrow N\left(0, \Sigma_\nu\right)$, by Assumptions A.2a) and C.1a) and Lemma 1 (iv).

Let us now show that $\frac{1}{\sqrt{T}} I_2 = o_p(1)$. We have:

$$\begin{aligned} I_2 &= \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T}^2 \hat{Q}_{x,i}^{-1} \left(Y_{i,T} Y_{i,T}^{'} - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_\nu - \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T}^2 \hat{Q}_{x,i}^{-1} \left(\tau_{i,T}^{-1} \hat{S}_{ii}^0 - S_{ii,T} \right) \hat{Q}_{x,i}^{-1} c_\nu \\ &\quad - \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} \hat{Q}_{x,i}^{-1} \left(\hat{S}_{ii} - \hat{S}_{ii}^0 \right) \hat{Q}_{x,i}^{-1} c_\nu - \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} \left(c_{\hat{\nu}} - c_\nu \right) \\ &=: \left(I_{21} - I_{22} - I_{23} \right) c_\nu - I_{24} \left(c_{\hat{\nu}} - c_\nu \right), \end{aligned} \tag{23}$$

where $\hat{S}_{ii}^0 := \frac{1}{T_i} \sum_t I_{i,t} \varepsilon_{i,t}^2 x_t x_t^{'}$ and $S_{ii,T} = \frac{1}{T} \sum_t I_{i,t} \sigma_{ii,t} x_t x'_t$. The various terms are bounded in the next Lemma.

Lemma 3 *Under Assumptions A.1, A.3, SC.1-SC.2, C.1-C.5, (i)* $I_{21} = \frac{1}{\sqrt{n}} \sum_i w_i \tau_i^2 \hat{Q}_x^{-1} \left(Y_{i,T} Y_{i,T}^{'} - S_{ii,T} \right) \hat{Q}_x^{-1}$ $+ O_{p,log}\left(\frac{\sqrt{n}}{T} \right) = O_p(1) + O_{p,log}\left(\frac{\sqrt{n}}{T} \right)$, *(ii)* $I_{22} = O_{p,log}\left(\frac{1}{\sqrt{T}} + \frac{\sqrt{n}}{T} \right)$, *(iii)* $I_{23} = O_{p,log}\left(\frac{\sqrt{n}}{T} \right)$ *(iv)* $I_{24} = O_{p,log}\left(\sqrt{n} \right)$ *and (v)* $c_{\hat{\nu}} - c_\nu = O_{p,log}\left(\frac{1}{\sqrt{nT}} + \frac{1}{T} \right)$*, when* $n, T \to \infty$ *such that* $n = O\left(T^{\bar{\gamma}}\right)$ *for* $\bar{\gamma} > 0$.

From Equation (23) and Lemma 3 we get $\frac{1}{\sqrt{T}} I_2 = o_p(1) + O_{p,log}\left(\frac{\sqrt{n}}{T\sqrt{T}} \right)$. From $n = O(T^{\bar{\gamma}})$ with $\bar{\gamma} < 3$, we get $\frac{1}{\sqrt{T}} I_2 = o_p(1)$ and the conclusion follows.

b) Asymptotic normality of $\hat{\lambda}$. We have $\sqrt{T}\left(\hat{\lambda} - \lambda \right) = \frac{1}{\sqrt{T}} \sum_t \left(f_t - E\left[f_t\right] \right) + \sqrt{T}\left(\hat{\nu} - \nu \right)$. By using $\sqrt{T}\left(\hat{\nu} - \nu \right) = O_p\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{T}} \right) = o_p\left(1\right)$, the conclusion follows from Assumption A.2b).

c) Proof of Lemma 2: Write:

$$I_1 = \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i Y'_{i,T} \hat{Q}_x^{-1} c_\nu + \frac{1}{\sqrt{n}} \sum_i \hat{w}_i \tau_{i,T} b_i Y'_{i,T} \left(\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1} \right) c_\nu \qquad =: I_{11} \hat{Q}_x^{-1} c_\nu + I_{12} c_\nu.$$

Let us decompose I_{11} as:

$$\begin{aligned} I_{11} &= \frac{1}{\sqrt{n}}\sum_i w_i\tau_i b_i Y'_{i,T} + \frac{1}{\sqrt{n}}\sum_i \left(\mathbf{1}_i^{\chi}-1\right) w_i\tau_i b_i Y'_{i,T} + \frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi} w_i \left(\tau_{i,T}-\tau_i\right) b_i Y'_{i,T} \\ &\quad + \frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi}\left(\hat{v}_i^{-1}-v_i^{-1}\right)\tau_{i,T} b_i Y'_{i,T} \qquad =: I_{111}+I_{112}+I_{113}+I_{114}. \end{aligned}$$

Similarly, for I_{12} we have:

$$\begin{aligned} I_{12} &= \frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi} v_i^{-1}\tau_{i,T} b_i Y'_{i,T}\left(\hat{Q}_{x,i}^{-1}-\hat{Q}_x^{-1}\right) \\ &\quad + \frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi}\left(\hat{v}_i^{-1}-v_i^{-1}\right)\tau_{i,T} b_i Y'_{i,T}\left(\hat{Q}_{x,i}^{-1}-\hat{Q}_x^{-1}\right) \qquad =: I_{121}+I_{122}. \end{aligned}$$

The conclusion follows by proving that terms I_{112}, I_{113}, I_{114}, I_{121} and I_{122} are $o_p(1)$.

Proof that $I_{112} = o_p(1)$. We use the next Lemma.

Lemma 4 *Under Assumptions SC.1-SC.2, C.1 b), d) and C.4 a), c):* $\mathbb{P}\left[\mathbf{1}_i^{\chi}=0\right] = O(T^{-\bar{b}})$*, for any* $\bar{b}>0$.

In Lemma 4, the unconditional probability $\mathbb{P}\left[\mathbf{1}_i^{\chi}=0\right]$ is independent of i since the indices (γ_i) are i.i.d. By using the bound $\|I_{112}\| \leq \frac{C}{\sqrt{n}}\sum_i (1-\mathbf{1}_i^{\chi})\|Y_{i,T}\|$ from Assumptions C.4 b) and c) and Lemma 1 (ii), the bound $\sup_i E[\|Y_{i,T}\| | x_{\underline{T}}, I_{\underline{T}}, \{\gamma_i\}] \leq C$ from Assumptions A.1 a) and b), and Lemma 4, it follows $I_{112} = O_p(\sqrt{n}T^{-\bar{b}})$, for any $\bar{b}>0$. Since $n = O\left(T^{\bar{\gamma}}\right)$, with $\bar{\gamma}>0$, we get $I_{112} = o_p(1)$.

Proof that $I_{113} = o_p(1)$. We have $E\left[\|I_{113}\|^2 | x_{\underline{T}}, I_{\underline{T}}, \{\gamma_i\}\right] \leq \frac{C}{nT}\sum_{i,j}\sum_t \mathbf{1}_i^{\chi}\mathbf{1}_j^{\chi}|\tau_{i,T}-\tau_i||\tau_{j,T}-\tau_j||\sigma_{ij,t}|$ from Assumption A.1 a). By Cauchy-Schwarz inequality and Assumption A.1 c), we get $E\left[\|I_{113}\|^2|\{\gamma_i\}\right] \leq CM\sup_{\gamma\in[0,1]} E\left[\mathbf{1}_i^{\chi}|\tau_{i,T}-\tau_i|^4|\gamma_i=\gamma\right]^{1/2}$. By using $\tau_{i,T}-\tau_i = -\tau_{i,T}\tau_i\frac{1}{T}\sum_t\left(I_{i,t}-E[I_{i,t}|\gamma_i]\right)$ and $\mathbf{1}_i^{\chi}\tau_{i,T} \leq \chi_{2,T}$, we get $\sup_{\gamma\in[0,1]} E\left[\mathbf{1}_i^{\chi}|\tau_{i,T}-\tau_i|^4|\gamma_i=\gamma\right] \leq C\chi_{2,T}^4 \sup_{\gamma\in[0,1]} E\left[\left|\frac{1}{T}\sum_t\left(I_t(\gamma)-E[I_t(\gamma)]\right)\right|^4\right] = o(1)$ from Assumption C.5 and the next Lemma.

Lemma 5 *Under Assumption C.1 d):* $\sup_{\gamma\in[0,1]} E\left[\left|\frac{1}{T}\sum_t\left(I_t(\gamma)-E[I_t(\gamma)]\right)\right|^4\right] = O(T^{-c})$*, for some* $c>0$.

Then, $I_{113} = o_p(1)$.

Proof that $I_{114} = o_p(1)$. From $\hat{v}_i^{-1}-v_i^{-1} = -v_i^{-2}\left(\hat{v}_i-v_i\right)+\hat{v}_i^{-1}v_i^{-2}\left(\hat{v}_i-v_i\right)^2$, we get:

$$I_{114} = -\frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\left(\hat{v}_i-v_i\right)\tau_{i,T} b_i Y'_{i,T} + \frac{1}{\sqrt{n}}\sum_i \mathbf{1}_i^{\chi}\hat{v}_i^{-1}v_i^{-2}\left(\hat{v}_i-v_i\right)^2\tau_{i,T} b_i Y'_{i,T} =: I_{1141}+I_{1142}.$$

Let us first consider I_{1141}. We have:

$$\begin{aligned}\hat{v}_i - v_i &= \tau_{i,T} c'_{\hat{\nu}_1} \hat{Q}_{x,i}^{-1} \left(\hat{S}_{ii} - S_{ii}\right) \hat{Q}_{x,i}^{-1} c_{\hat{\nu}_1} + 2\tau_{i,T}(c_{\hat{\nu}_1} - c_\nu)' \hat{Q}_{x,i}^{-1} S_{ii} \hat{Q}_{x,i}^{-1} c_{\hat{\nu}_1} \\ &\quad + \tau_{i,T}(c_{\hat{\nu}_1} - c_\nu)' \hat{Q}_{x,i}^{-1} S_{ii} \hat{Q}_{x,i}^{-1} (c_{\hat{\nu}_1} - c_\nu) + 2\tau_{i,T} c'_\nu \left(\hat{Q}_{x,i}^{-1} - Q_x^{-1}\right) S_{ii} \hat{Q}_{x,i}^{-1} c_\nu \\ &\quad + \tau_{i,T} c'_\nu \left(\hat{Q}_{x,i}^{-1} - Q_x^{-1}\right) S_{ii} \left(\hat{Q}_{x,i}^{-1} - Q_x^{-1}\right) c_\nu + (\tau_{i,T} - \tau_i) c'_\nu Q_x^{-1} S_{ii} Q_x^{-1} c_\nu. \end{aligned} \tag{24}$$

The contribution of the first two terms to I_{1141} is:

$$\begin{aligned} I_{11411} &= -\frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^2 c'_{\hat{\nu}_1} \hat{Q}_{x,i}^{-1} \left(\hat{S}_{ii} - S_{ii}\right) \hat{Q}_{x,i}^{-1} c_{\hat{\nu}_1} b_i Y'_{i,T}, \\ I_{11412} &= -\frac{2}{\sqrt{n}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^2 (c_{\hat{\nu}_1} - c_\nu)' \hat{Q}_{x,i}^{-1} S_{ii} \hat{Q}_{x,i}^{-1} c_{\hat{\nu}_1} b_i Y'_{i,T}. \end{aligned}$$

We first show $I_{11412} = o_p(1)$. For this purpose, it is enough to show that $c_{\hat{\nu}_1} - c_\nu = O_p(T^{-c})$, for some $c > 0$, and $\frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^2 \left(\hat{Q}_{x,i}^{-1} S_{ii} \hat{Q}_{x,i}^{-1}\right)_{kl} b_i Y'_{i,T} = O_p\left(\chi_{1,T}^2 \chi_{2,T}^2\right)$, for any $k, l = 1, ..., K+1$. The first statement follows from the proof of Proposition 2 but with known weights equal to 1. To prove the second statement, we use bounds $\mathbf{1}_i^\chi \tau_{i,T} \leq \chi_{2,T}$ and $\mathbf{1}_i^\chi \|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}$ and Assumption A.1 c). Let us now prove that $I_{11411} = o_p(1)$. For this purpose, it is enough to show that

$$J_1 := \frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^2 \left(\hat{Q}_{x,i}^{-1} \left(\hat{S}_{ii} - S_{ii}\right) \hat{Q}_{x,i}^{-1}\right)_{kl} b_i Y'_{i,T} = o_p(1), \tag{25}$$

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

$$\begin{aligned} \hat{S}_{ii} - S_{ii} &= \frac{1}{T_i} \sum_t I_{i,t} \left(\varepsilon_{it}^2 x_t x'_t - S_{ii}\right) + \frac{1}{T_i} \sum_t I_{i,t} \left(\hat{\varepsilon}_{i,t}^2 - \varepsilon_{it}^2\right) x_t x'_t \\ &= \frac{\tau_{i,T}}{\sqrt{T}} W_{1,i,T} + \frac{\tau_{i,T}}{\sqrt{T}} W_{2,i,T} - \frac{2\tau_{i,T}^2}{T} W_{3,i,T} \hat{Q}_{x,i}^{-1} Y_{i,T} + \frac{\tau_{i,T}^3}{T} \hat{Q}_{x,i}^{(4)} \hat{Q}_{x,i}^{-1} Y_{i,T} Y'_{i,T} \hat{Q}_{x,i}^{-1}, \end{aligned} \tag{26}$$

where $W_{1,i,T} := \frac{1}{\sqrt{T}} \sum_t I_{i,t} x_t^2 \eta_{i,t}$, $\eta_{i,t} = \varepsilon_{it}^2 - \sigma_{ii,t}$, $W_{2,i,T} := \frac{1}{\sqrt{T}} \sum_t I_{i,t} \zeta_{i,t}$, $\zeta_{i,t} := \sigma_{ii,t} x_t^2 - S_{ii}$, $W_{3,i,T} := \frac{1}{\sqrt{T}} \sum_t I_{i,t} \varepsilon_{i,t} x_t^3$, $\hat{Q}_{x,i}^{(4)} := \frac{1}{T} \sum_t I_{i,t} x_t^4$ and x_t is treated as a scalar to ease notation. Then:

$$\begin{aligned} J_1 &= \frac{1}{\sqrt{nT}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^3 \hat{Q}_{x,i}^{-1} W_{1,i,T} \hat{Q}_{x,i}^{-1} b_i Y'_{i,T} + \frac{1}{\sqrt{nT}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^3 \hat{Q}_{x,i}^{-1} W_{2,i,T} \hat{Q}_{x,i}^{-1} b_i Y'_{i,T} \\ &\quad - \frac{2}{\sqrt{n}T} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^4 \hat{Q}_{x,i}^{-1} W_{3,i,T} \hat{Q}_{x,i}^{-1} Y_{i,T} \hat{Q}_{x,i}^{-1} b_i Y'_{i,T} + \frac{1}{\sqrt{n}T} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^5 \hat{Q}_{x,i}^{-1} \hat{Q}_{x,i}^{(4)} \hat{Q}_{x,i}^{-1} Y_{i,T} Y'_{i,T} \hat{Q}_{x,i}^{-2} b_i Y'_{i,T} \\ &=: J_{11} + J_{12} + J_{13} + J_{14}. \end{aligned}$$

Let us consider J_{11}. We have:

$$E\left[J_{11}|x_{\underline{T}}, I_{\underline{T}}, \{\gamma_i\}\right] = \frac{1}{\sqrt{nT^3}}\sum_i\sum_{t,s}\mathbf{1}_i^{\chi}v_i^{-2}\tau_{i,T}^3\hat{Q}_{x,i}^{-2}b_i x_t^2 x_s E\left[\varepsilon_{i,t}^2\varepsilon_{i,s}|x_{\underline{T}},\gamma_i\right] = 0,$$

from Assumption A.3. Moreover, from Assumption C.4:

$$V\left[J_{11}|x_{\underline{T}}, I_{\underline{T}}, \{\gamma_i\}\right] \leq \frac{C}{nT^3}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}\mathbf{1}_i^{\chi}\mathbf{1}_j^{\chi}\tau_{i,T}^3\tau_{j,T}^3\|\hat{Q}_{x,i}^{-1}\|^2\|\hat{Q}_{x,j}^{-1}\|^2|cov\left(\eta_{i,t_1}\varepsilon_{i,t_2},\eta_{j,t_3}\varepsilon_{j,t_4}|x_{\underline{T}},\gamma_i,\gamma_j\right)|.$$

By using $\mathbf{1}_i^{\chi}\|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}$, $\mathbf{1}_i^{\chi}\tau_{i,T} \leq \chi_{2,T}$, the Law of Iterated Expectations and Assumptions C.3 c) and C.5, we get $E\left[J_{11}\right] = 0$ and $V[J_{11}] = o(1)$. Thus $J_{11} = o_p(1)$. By similar arguments and using Assumptions A.1 c) and C.3 e), we get $J_{12} = o_p(1)$, $J_{13} = o_p(1)$ and $J_{14} = o_p(1)$. Hence the bound in Equation (25) follows, and $I_{11411} = o_p(1)$. Paralleling the detailed arguments provided above, we can show that all other remaining terms making I_{114} are also $o_p(1)$.

Proof that $I_{121} = o_p(1)$. From:

$$\hat{Q}_{x,i}^{-1} - \hat{Q}_x^{-1} = -\hat{Q}_{x,i}^{-1}\left(\frac{1}{T_i}\sum_t I_{i,t}x_t x_t' - \hat{Q}_x\right)\hat{Q}_x^{-1} = -\tau_{i,T}\hat{Q}_{x,i}^{-1}W_{i,T}\hat{Q}_x^{-1} + \hat{Q}_{x,i}^{-1}W_T\hat{Q}_x^{-1}, \qquad (27)$$

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

$$\begin{aligned} I_{121} &= \left(-\frac{1}{\sqrt{n}}\sum_i\mathbf{1}_i^{\chi}v_i^{-1}\tau_{i,T}^2 b_i Y_{i,T}'\hat{Q}_{x,i}^{-1}W_{i,T} + \frac{1}{\sqrt{n}}\sum_i\mathbf{1}_i^{\chi}v_i^{-1}\tau_{i,T}b_i Y_{i,T}'\hat{Q}_{x,i}^{-1}W_T\right)\hat{Q}_x^{-1} \\ &=: \left(I_{1211} + I_{1212}\right)\hat{Q}_x^{-1}. \end{aligned}$$

Let us consider term I_{1211}. From Assumption C.4, $\mathbf{1}_i^{\chi}\|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}$ and $\mathbf{1}_i^{\chi}\tau_{i,T} \leq \chi_{2,T}$, we have:

$$E\left[\|I_{1211}\|^2|x_{\underline{T}}, I_{\underline{T}}, \{\gamma_i\}\right] \leq \frac{C\chi_{1,T}^2\chi_{2,T}^4}{nT}\sum_{i,j}\sum_t|\sigma_{ij,t}|\|W_{i,T}\|\|W_{j,T}\|.$$

Then, from Cauchy-Schwarz inequality, we get $E\left[\|I_{1211}\|^2|\{\gamma_i\}\right] \leq C\chi_{1,T}^2\chi_{2,T}^4\frac{1}{n}\sum_{i,j}E[\sigma_{ij,t}^2|\gamma_i,\gamma_j]^{1/2}$

$\sup_i E\left[\|W_{i,T}\|^4|\gamma_i\right]^{1/2}$, where $\sup_i E\left[\|W_{i,T}\|^4|\gamma_i\right] \leq \sup_{\gamma\in[0,1]} E\left[\left\|\frac{1}{T}\sum_t I_t(\gamma)\left(x_t x_t' - Q_x\right)\right\|^4\right] = O(T^{-c})$ from Assumptions C.1 b) and C.4 a). Then, from Assumptions A.1 c) and C.5 it follows $E[\|I_{1211}\|^2] = o(1)$ and thus $I_{1211} = o_p(1)$. Similarly we can show $I_{1212} = o_p(1)$, and then $I_{121} = o_p(1)$.

Proof that $I_{122} = o_p(1)$. The statement follows by combining arguments similar as for I_{114} and I_{121}.

A.2.4 Proof of Proposition 4

From Proposition 3, we have to show that $\tilde{\Sigma}_\nu - \Sigma_\nu = o_p(1)$. By $\Sigma_\nu = \left(c_\nu' Q_x^{-1} \otimes Q_b^{-1}\right) S_b \left(Q_x^{-1} c_\nu \otimes Q_b^{-1}\right)$ and $\tilde{\Sigma}_\nu = \left(c_{\hat{\nu}}' \hat{Q}_x^{-1} \otimes \hat{Q}_b^{-1}\right) \tilde{S}_b \left(\hat{Q}_x^{-1} c_{\hat{\nu}} \otimes \hat{Q}_b^{-1}\right)$, where $\tilde{S}_b = \frac{1}{n}\sum_{i,j} \hat{w}_i \hat{w}_j \frac{\tau_{i,T}\tau_{j,T}}{\tau_{ij,T}} \tilde{S}_{ij} \otimes \hat{b}_i \hat{b}_j'$, and the consistency of $\hat{Q}_x$ and $\hat{Q}_b$, the statement follows if $\tilde{S}_b - S_b = o_p(1)$. The leading terms in $\tilde{S}_b - S_b$ are given by $I_3 := \frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} \left(\tilde{S}_{ij} - S_{ij}\right) \otimes b_i b_j'$ and $I_4 := \frac{1}{n}\sum_{i} w_i w_j \tau_i \tau_j (\tau_{ij,T}^{-1} - \tau_{ij}^{-1}) S_{ij} \otimes b_i b_j'$, while the other ones can be shown to be $o_p(1)$ by arguments similar to the proofs of Propositions 2 and 3.

Proof of $I_3 = o_p(1)$. By using that $\tau_i \leq M$, $\tau_{ij} \geq 1$, $w_i \leq M$ and $\|b_i\| \leq M$, $I_3 = o_p(1)$ follows if we show: $\frac{1}{n}\sum_{i,j} \left\|\tilde{S}_{ij} - S_{ij}\right\| = o_p(1)$. For this purpose, we introduce the following Lemmas 6 and 7 that extend results in Bickel and Levina (2008) from the i.i.d. case to the time series case including random individual effects.

Lemma 6 *Let* $\psi_{nT} := \max_{i,j} \left\|\hat{S}_{ij} - S_{ij}\right\|$, *and* $\Psi_{nT}(\xi) := \max_{i,j} \mathbb{P}\left[\left\|\hat{S}_{ij} - S_{ij}\right\| \geq \xi\right]$, *for* $\xi > 0$. *Under Assumptions SC.1, SC.2, A.4,* $\frac{1}{n}\sum_{i,j} \left\|\tilde{S}_{ij} - S_{ij}\right\| = O_p\left(\psi_{nT} n^\delta \kappa^{-q} + n^\delta \kappa^{1-q} + \psi_{nT} n^2 \Psi_{nT}((1-v)\kappa)\right)$, *for any* $v \in (0,1)$.

Lemma 7 *Under Assumptions SC.1, SC.2, C.1, C.4 and C.5, 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* $\psi_{nT} = O_p\left(\sqrt{\frac{\log n}{T^\eta}}\right)$, *when* $n, T \to \infty$ *such that* $n = O(T^{\bar{\gamma}})$ *for* $\bar{\gamma} > 0$.

In Lemma 6, the probability $\mathbb{P}\left[\left\|\hat{S}_{ij} - S_{ij}\right\| \geq \xi\right]$ is the same for all pairs (i,j) with $i = j$, and for all pairs with $i \neq j$, since this probability is marginal w.r.t. the individual random effects. From Lemmas 6 and 7, it follows $\frac{1}{n}\sum_{i,j} \left\|\tilde{S}_{ij} - S_{ij}\right\| = O_p\left(\left(\frac{\log n}{T^\eta}\right)^{(1-q)/2} n^\delta\right) = o_p(1)$, since $n = O(T^{\bar{\gamma}})$ with $\bar{\gamma} < \eta\frac{1-q}{2\delta}$.

Proof of $I_4 = o_p(1)$. From $w_i \leq M$, $\tau_i \leq M$ and $b_i \leq M$, we have $E[\|I_4\||\{\gamma_i\}] \leq C \sup_{i,j} E[|\tau_{ij,T}^{-1} - \tau_{ij}^{-1}| \gamma_i, \gamma_j] \frac{1}{n}\sum_{i,j} \|S_{ij}\|$. By using the inequalities $\sup_{i,j} E[|\tau_{ij,T}^{-1} - \tau_{ij}^{-1}| \gamma_i, \gamma_j]$ $\leq \sup_{\gamma,\gamma' \in [0,1]} E\left[\left\|\frac{1}{T}\sum_t (I_t(\gamma) I_t(\gamma') - E[I_t(\gamma) I_t(\gamma')])\right\|\right]$ and $\|S_{ij}\| \leq E[|\sigma_{ij,t}||\gamma_i, \gamma_j]$, from Assumptions A.1 c) and C.1 d) we get $E[\|I_4\|] = o(1)$, which implies $I_4 = o_p(1)$.

A.2.5 Proof of Proposition 5

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

Lemma 8 *Under $\mathcal{H}_0$ and Assumptions APR.1-APR.5, SC.1-SC.2, A.1-A.3 and C.1-C.5, we have* $\hat{Q}_e = \frac{1}{n}\sum_i \hat{w}_i \left[c'_{\hat{\nu}}\left(\hat{\beta}_i - \beta_i\right)\right]^2 + O_{p,log}\left(\frac{1}{nT} + \frac{1}{T^2}\right)$*, when $n, T \to \infty$ such that $n = O\left(T^{\bar{\gamma}}\right)$ for $\bar{\gamma} > 0$.*

From Lemma 4 and $n = O(T^{\bar{\gamma}})$ for $0 < \bar{\gamma} < 2$, it follows $\hat{\xi}_{nT} = \frac{1}{\sqrt{n}}\sum_i \hat{w}_i \left\{\left[c'_{\hat{\nu}}\sqrt{T}\left(\hat{\beta}_i - \beta_i\right)\right]^2 + \right.$ $\left. -\tau_{i,T} c'_{\hat{\nu}} \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} c_{\hat{\nu}}\right\} + o_p(1)$. By using $\sqrt{T}\left(\hat{\beta}_i - \beta_i\right) = \tau_{i,T}\hat{Q}_{x,i}^{-1} Y_{i,T}$, we get

$$\begin{aligned}
\hat{\xi}_{nT} &= \frac{1}{\sqrt{n}}\sum_i \hat{w}_i \tau_{i,T}^2 c'_{\hat{\nu}} \hat{Q}_{x,i}^{-1}\left(Y_{i,T}Y'_{i,T} - \tau_{i,T}^{-1}\hat{S}_{ii}\right)\hat{Q}_{x,i}^{-1} c_{\hat{\nu}} + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_i \hat{w}_i \tau_{i,T}^2 c'_{\hat{\nu}} \hat{Q}_{x,i}^{-1}\left(Y_{i,T}Y'_{i,T} - S_{ii,T}\right)\hat{Q}_{x,i}^{-1} c_{\hat{\nu}} - \frac{1}{\sqrt{n}}\sum_i \hat{w}_i \tau_{i,T}^2 c'_{\hat{\nu}} \hat{Q}_{x,i}^{-1}\left(\tau_{i,T}^{-1}\hat{S}_{ii} - S_{ii,T}\right)\hat{Q}_{x,i}^{-1} c_{\hat{\nu}} \\
&\quad + o_p(1) \qquad =: c'_{\hat{\nu}}(I_{21} - I_{22} - I_{23})c_{\hat{\nu}} + o_p(1),
\end{aligned}$$

where I_{21}, I_{22} and I_{23} are defined in (23). By Lemma 3 (i)-(iii), and the consistency of $\hat{\nu}$, we have $\hat{\xi}_{nT} = \frac{1}{\sqrt{n}}\sum_i w_i \tau_i^2 c'_{\hat{\nu}} \hat{Q}_x^{-1}\left(Y_{i,T}Y'_{i,T} - S_{ii,T}\right)\hat{Q}_x^{-1} c_{\hat{\nu}} + O_{p,log}\left(\frac{\sqrt{n}}{T}\right) + o_p(1)$. Moreover, from $n = O(T^{\bar{\gamma}})$ with $\bar{\gamma} < 2$, the remainder term $O_{p,log}(\sqrt{n}/T)$ is $o_p(1)$. Then, by using $tr\left[A'B\right] = vec\left[A\right]' vec\left[B\right]$, and $vec\left[YY'\right] = (Y \otimes Y)$ for a vector Y, we get

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

By using Assumption A.5, and by consistency of $\hat{\nu}$ and $\hat{Q}_x$, we get $\hat{\xi}_{nT} \Rightarrow N(0, \Sigma_\xi)$, where $\Sigma_\xi = \left(vec\left[Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right]\right)' \Omega \left(vec\left[Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right]\right)$. By using MN Theorem 3 Chapter 2, we have

$$\begin{aligned}
vec\left[Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right]' (S_{ij} \otimes S_{ij})\, vec\left[Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right] &= tr\left[S_{ij} Q_x^{-1} c_\nu c'_\nu Q_x^{-1} S_{ij} Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right] \\
&= \left(c'_\nu Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)^2, \qquad (28)
\end{aligned}$$

and

$$vec\left[Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right]' (S_{ij} \otimes S_{ij})\, W_{K+1} vec\left[Q_x^{-1} c_\nu c'_\nu Q_x^{-1}\right] = \left(c'_\nu Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)^2. \qquad (29)$$

Then, from the definition of Ω in Assumption A.5 and Equations (28) and (29), we deduce $\Sigma_\xi = 2 \text{ a.s.-}\lim_{n\to\infty} \frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\left(c_\nu' Q_x^{-1} S_{ij} Q_x^{-1} c_\nu\right)^2$. Finally, $\tilde{\Sigma}_\xi = \Sigma_\xi + o_p(1)$ follows from $\frac{1}{n}\sum_{i,j}\|\tilde{S}_{ij} - S_{ij}\| = o_p(1)$ and $\frac{1}{n}\sum_{i,j}\|\tilde{S}_{ij} - S_{ij}\|^2 = o_p(1)$.

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

$$\begin{aligned}
\hat{\nu} - \nu_\infty &= \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i \hat{b}_i c_{\nu_\infty}' \hat{\beta}_i = \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i \hat{b}_i c_{\nu_\infty}'\left(\hat{\beta}_i - \beta_i\right) + \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i \hat{b}_i e_i \qquad (30)\\
&= \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i b_i\left(\hat{\beta}_i - \beta_i\right)' c_{\nu_\infty} + \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i E_2'\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_i - \beta_i\right)' c_{\nu_\infty}\\
&\quad + \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i b_i e_i + \hat{Q}_b^{-1}\frac{1}{n}\sum_i \hat{w}_i\left(\hat{b}_i - b_i\right) e_i.
\end{aligned}$$

Equation (30) is the analogue of Equation (20), and the consistency of $\hat{\nu}$ for ν_∞ follows as in the proof of Proposition 2 and by using $E\left[w_i b_i e_i\right] = 0$. Thus, we get:

$$\begin{aligned}
&\sqrt{n}\left(\hat{\nu} - \frac{1}{T}\hat{B}_{\nu_\infty} - \nu_\infty\right)\\
=\; & \hat{Q}_b^{-1}\frac{1}{\sqrt{nT}}\sum_i \hat{w}_i \tau_{i,T} b_i Y_{i,T}' \hat{Q}_{x,i}^{-1} c_{\nu_\infty} + \frac{1}{T}\hat{Q}_b^{-1} E_2' \frac{1}{\sqrt{n}}\sum_i \hat{w}_i \tau_{i,T}^2\left(\hat{Q}_{x,i}^{-1} Y_{i,T} Y_{i,T}' \hat{Q}_{x,i}^{-1} c_{\nu_\infty} - \tau_{i,T}^{-1}\hat{Q}_{x,i}^{-1}\hat{S}_{ii}\hat{Q}_{x,i}^{-1} c_{\hat{\nu}}\right)\\
& + \hat{Q}_b^{-1}\frac{1}{\sqrt{n}}\sum_i w_i b_i e_i + \hat{Q}_b^{-1}\frac{1}{\sqrt{n}}\sum_i (\hat{w}_i - w_i) b_i e_i + \hat{Q}_b^{-1}\frac{1}{\sqrt{nT}}\sum_i \hat{w}_i \tau_{i,T} e_i Y_{i,T}' \hat{Q}_{x,i}^{-1} E_2\\
& =: I_{51} + I_{52} + I_{53} + I_{54} + I_{55}.
\end{aligned}$$

From Assumption SC.2 and $E\left[w_i b_i e_i\right] = 0$, we get $\frac{1}{\sqrt{n}}\sum_i w_i b_i e_i \Rightarrow N\left(0, E\left[w_i^2 e_i^2 b_i b_i'\right]\right)$ by the CLT. Thus, $I_{53} \Rightarrow N\left(0, Q_b^{-1} E\left[w_i^2 e_i^2 b_i b_i'\right] Q_b^{-1}\right)$. Then, the asymptotic distribution of $\hat{\nu}$ follows if terms I_{51}, I_{52}, I_{54} and I_{55} are $o_p(1)$. From similar arguments as for term I_1 in the proof of Proposition 3, we have $\frac{1}{\sqrt{n}}\sum_i \hat{w}_i \tau_{i,T} b_i Y_{i,T}' \hat{Q}_{x,i}^{-1} c_{\nu_\infty} = O_p(1)$ and $\frac{1}{\sqrt{n}}\sum_i \hat{w}_i \tau_{i,T} e_i Y_{i,T}' \hat{Q}_{x,i}^{-1} E_2 = O_p(1)$. Thus $I_{51} = o_p(1)$ and $I_{55} = o_p(1)$. From similar arguments as for term I_2 in the proof of Proposition 3, we have $I_{52} = o_p(1)$. Moreover, term $I_{54} = o_p(1)$ from similar arguments as for I_{112} and I_{114}.

b) Asymptotic normality of $\hat{\lambda}$. We have $\sqrt{T}\left(\hat{\lambda} - \lambda_\infty\right) = \sqrt{T}\left(\hat{\nu} - \nu_\infty\right) + \frac{1}{\sqrt{T}}\sum_t \left(f_t - E\left[f_t\right]\right)$. By using $\bar{\gamma} > 1$ and $\sqrt{T}\left(\hat{\nu}-\nu_{\infty}\right) = O_p\left(\sqrt{\frac{T}{n}}+\frac{1}{\sqrt{T}}\right) = o_p(1)$, the conclusion follows.

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

Lemma 9 *Under $\mathcal{H}_1$ and Assumptions SC.1, SC.2, A.1-A.3, C.1-C.5, we have* $\hat{Q}_e = \frac{1}{n}\sum_i \hat{w}_i \left[c_{\hat{\nu}}'\left(\hat{\beta}_i-\beta_i\right)\right]^2$ $+\frac{1}{n}\sum_i w_i e_i^2 + O_{p,log}\left(\frac{1}{n}+\frac{1}{\sqrt{nT}}+\frac{1}{\sqrt{T^3}}\right)$, *when $n, T \to \infty$ such that $n = O\left(T^{\bar{\gamma}}\right)$ for $\bar{\gamma} > 0$.*

By similar arguments as in the proof of Proposition 5 and using $\bar{\gamma} < 2$, we get:

$$
\begin{aligned}
\hat{\xi}_{nT} &= \frac{1}{\sqrt{n}}\sum_i w_i \tau_i^2 c_{\hat{\nu}}' \hat{Q}_x^{-1}\left(Y_{i,T}Y_{i,T}'-S_{ii,T}\right)\hat{Q}_x^{-1}c_{\hat{\nu}} + T\frac{1}{\sqrt{n}}\sum_i w_i e_i^2 + O_{p,log}\left(\frac{T}{\sqrt{n}}+\sqrt{T}\right) + o_p(1) \\
&= T\sqrt{n}E\left[w_i\left(a_i - b_i'\nu_{\infty}\right)^2\right] + O_p(T).
\end{aligned}
$$

Under $\mathcal{H}_1$, we have $E\left[w_i\left(a_i-b_i'\nu_{\infty}\right)^2\right] > 0$, since $w_i > 0$ and $\left(a_i - b_i\nu_{\infty}\right)^2 > 0$, P-a.s. Moreover, $\tilde{\Sigma}_{\xi} = \Sigma_{\xi} + o_p(1)$. Thus, $\tilde{\Sigma}_{\xi}^{-1/2}\hat{\xi}_{nT} = T\sqrt{n}\left(\Sigma_{\xi}^{-1/2}E\left[w_i\left(a_i-b_i'\nu_{\infty}\right)^2\right]+o_p(1)\right)$.

Appendix 3: Conditional factor model

A.3.1 Proof of Proposition 7

Proposition 7 is proved along similar lines as Proposition 1. Hence we only highlight the slight differences. We can work at $t = 1$ because of stationarity, and use that $a_1(\gamma)$, $b_1(\gamma)$, for $\gamma \in [0, 1]$, are $\mathcal{F}_0$-measurable. Then, the proof by contradiction uses the strong LLN applied conditionally on $\mathcal{F}_0$ and Assumption APR.7 as in the proof of Proposition 1. A result similar to Proposition APR also holds true with straightforward modifications to accommodate the conditional case.

A.3.2 Derivation of Equations (13) and (14)

From Equation (12) and by using $vec[ABC] = \left[C' \otimes A\right] vec[B]$ (MN Theorem 2, p. 35), we get $Z_{t-1}'B_i'f_t = vec\left[Z_{t-1}'B_i'f_t\right] = \left[f_t' \otimes Z_{t-1}'\right] vec\left[B_i'\right]$, and $Z_{i,t-1}'C_i'f_t = \left[f_t' \otimes Z_{i,t-1}'\right] vec\left[C_i'\right]$, which gives $Z_{t-1}'B_i'f_t + Z_{i,t-1}'C_i'f_t = x_{2,i,t}'\beta_{2,i}$.

Let us now consider the first two terms in the RHS of Equation (12).

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

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

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

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

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

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

b) By the properties of the tr and vec operators, we have

$$\begin{aligned} Z'_{i,t-1}C'_i(\Lambda-F)Z_{t-1} &= tr\left[Z_{t-1}Z'_{i,t-1}C'_i(\Lambda-F)\right]=vec\left[Z_{i,t-1}Z'_{t-1}\right]'vec\left[C'_i(\Lambda-F)\right] \\ &= (Z_{t-1}\otimes Z_{i,t-1})'\left[(\Lambda-F)'\otimes I_q\right]vec\left[C'_i\right]. \end{aligned}$$

By combining a) and b), we get $Z'_{t-1}B'_i(\Lambda-F)Z_{t-1}+Z'_{i,t-1}C'_i(\Lambda-F)Z_{t-1}=x'_{1,i,t}\beta_{1,i}$ and $\beta_{1,i}=\Psi\beta_{2,i}$.

A.3.3 Derivation of Equation (15)

We use $\beta_{1,i}=\left(\left(\frac{1}{2}D_p^+\left[vec\left[B'_i(\Lambda-F)\right]+vec\left[(\Lambda-F)'B_i\right]\right]\right)',\left(vec\left[C'_i(\Lambda-F)\right]\right)'\right)'$ from Section A.3.2. a) From the properties of the vec operator, we get

$$vec\left[B'_i(\Lambda-F)\right]+vec\left[(\Lambda-F)'B_i\right]=\left(I_p\otimes B'_i\right)vec\left[\Lambda-F\right]+\left(B'_i\otimes I_p\right)vec\left[\Lambda'-F'\right].$$

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

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

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

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

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

A.3.4 Derivation of Equation (16)

We use $vec\left[\beta'_{3,i}\right] = \left(vec\left[\{D_p^+\left(B_i' \otimes I_p\right)\}'\right]', vec\left[\{W_{p,q}\left(C_i' \otimes I_p\right)\}'\right]'\right)'$.

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

$$\begin{aligned} vec\left[D_p^+\left(B_i' \otimes I_p\right)\right] &= \left(I_{pK} \otimes D_p^+\right) vec\left[B_i' \otimes I_p\right] \\ &= \left(I_{pK} \otimes D_p^+\right)\{I_K \otimes \left[\left(W_p \otimes I_p\right)\left(I_p \otimes vec\left[I_p\right]\right)\right]\} vec\left[B_i'\right] \\ &= \left\{I_K \otimes \left[\left(I_p \otimes D_p^+\right)\left(W_p \otimes I_p\right)\left(I_p \otimes vec\left[I_p\right]\right)\right]\right\} vec\left[B_i'\right]. \end{aligned}$$

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

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

By combining a) and b) the conclusion follows.

A.3.5 Proof of Proposition 8

a) Consistency of $\hat{\nu}$. By definition of $\hat{\nu}$, we have: $\hat{\nu} - \nu = \hat{Q}_{\beta_3}^{-1} \frac{1}{n} \sum_i \hat{\beta}'_{3,i} \hat{w}_i \left(\hat{\beta}_{1,i} - \hat{\beta}_{3,i}\nu\right)$. From Equation (16) and MN Theorem 2 p. 35, we get $\hat{\beta}_{3,i}\nu = vec[\nu'\hat{\beta}'_{3,i}] = \left(I_{d_1} \otimes \nu'\right) vec[\hat{\beta}'_{3,i}] = \left(I_{d_1} \otimes \nu'\right) J_a \hat{\beta}_{2,i}$. Moreover, by using matrices E_1 and E_2, we obtain $\left(\hat{\beta}_{1,i} - \hat{\beta}_{3,i}\nu\right) = \left[E_1' - \left(I_{d_1} \otimes \nu'\right) J_a E_2'\right] \hat{\beta}_i = C_\nu' \hat{\beta}_i = C_\nu' \left(\hat{\beta}_i - \beta_i\right)$, from Equation (15). It follows that

$$\hat{\nu} - \nu = \hat{Q}_{\beta_3}^{-1} \frac{1}{n} \sum_i \hat{\beta}'_{3,i} \hat{w}_i C_\nu' \left(\hat{\beta}_i - \beta_i\right). \tag{31}$$

By comparing with Equation (20) and by using the same arguments as in the proof of Proposition 2 applied to $\beta'_{3,i}$ instead of b_i, the result follows.

b) Consistency of $\hat{\Lambda}$. By definition of $\hat{\Lambda}$, we deduce $\left\|vec\left[\hat{\Lambda}' - \Lambda'\right]\right\| \leq \|\hat{\nu} - \nu\| + \left\|vec\left[\hat{F}' - F'\right]\right\|$. By part a), $\|\hat{\nu} - \nu\| = o_p(1)$. By the LLN and Assumptions C.1a), C.4a) 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 get that $\left\|vec\left[\hat{F}' - F'\right]\right\| = o_p(1)$. The result follows.

A.3.6 Proof of Proposition 9

a) Asymptotic normality of $\hat{\nu}$. From Equation (31) 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

$$\begin{aligned}\sqrt{nT}\left(\hat{\nu}-\nu\right) &= \hat{Q}_{\beta_3}^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_{i,T}\hat{\beta}_{3,i}'\hat{w}_i C_\nu'\hat{Q}_{x,i}^{-1}Y_{i,T} = \hat{Q}_{\beta_3}^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_{i,T}\beta_{3,i}'\hat{w}_i C_\nu'\hat{Q}_{x,i}^{-1}Y_{i,T} \\ &\quad + \hat{Q}_{\beta_3}^{-1}\frac{1}{\sqrt{n}}\sum_i \tau_{i,T}\left(\hat{\beta}_{3,i}-\beta_{3,i}\right)'\hat{w}_i C_\nu'\hat{Q}_{x,i}^{-1}Y_{i,T} \quad =: \hat{Q}_{\beta_3}^{-1}I_{61} + I_{62}.\end{aligned}$$

Term I_{61} is the analogue of term I_1 in the proof of Proposition 3. To analyse I_{62}, we use the following lemma.

Lemma 10 *Let A be a $m \times n$ matrix and b be a $n \times 1$ vector. Then,* $Ab = \left(vec\left[I_n\right]' \otimes I_m\right) vec\left[vec\left[A\right]b'\right]$.

By Lemma 10, Equation (16) and $\sqrt{T}vec\left[\left(\hat{\beta}_{3,i}-\beta_{3,i}\right)'\right] = \tau_{i,T}J_a E_2'\hat{Q}_{x,i}^{-1}Y_{i,T}$, we have

$$\begin{aligned} I_{62} &= \hat{Q}_{\beta_3}^{-1}\frac{1}{\sqrt{nT}}\sum_i \tau_{i,T}^2\left(vec\left[I_{d_1}\right]'\otimes I_{Kp}\right)vec\left[J_a E_2'\hat{Q}_{x,i}^{-1}Y_{i,T}Y_{i,T}'\hat{Q}_{x,i}^{-1}C_\nu\hat{w}_i\right] \\ &= \hat{Q}_{\beta_3}^{-1}\frac{1}{\sqrt{nT}}\sum_i \tau_{i,T}^2 J_b vec\left[E_2'\hat{Q}_{x,i}^{-1}Y_{i,T}Y_{i,T}'\hat{Q}_{x,i}^{-1}C_\nu\hat{w}_i\right] = \sqrt{\frac{n}{T}}\hat{B}_\nu + \frac{1}{\sqrt{T}}\hat{Q}_{\beta_3}^{-1}I_{63},\end{aligned}$$

where $I_{63} := \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2 J_b vec\left[E_2'\left(\hat{Q}_{x,i}^{-1}Y_{i,T}Y_{i,T}'\hat{Q}_{x,i}^{-1}C_\nu - \tau_{i,T}^{-1}\hat{Q}_{x,i}^{-1}\hat{S}_{ii}\hat{Q}_{x,i}^{-1}C_{\hat{\nu}}\right)\hat{w}_i\right]$. We get:

$$\sqrt{nT}\left(\hat{\nu}-\frac{1}{T}\hat{B}_\nu-\nu\right) = \hat{Q}_{\beta_3}^{-1}I_{61} + \frac{1}{\sqrt{T}}\hat{Q}_{\beta_3}^{-1}I_{63}, \tag{32}$$

which is the analogue of Equation (22) in the proof of Proposition 3. Let us now derive the asymptotic behaviour of the terms in the RHS of (32). By MN Theorem 2 p. 35, we have $I_{61} = \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}\left[(Y_{i,T}'\hat{Q}_{x,i}^{-1})\otimes(\beta_{3,i}'\hat{w}_i)\right]vec\left[C_\nu'\right]$. Similarly as in Lemma 2, we have $I_{61} = \frac{1}{\sqrt{n}}\sum_i \tau_i\left[(Y_{i,T}'Q_{x,i}^{-1})\otimes(\beta_{3,i}'w_i)\right]vec\left[C_\nu'\right] + o_p(1)$. Then, by the properties of the vec operator, we get $\hat{Q}_{\beta_3}^{-1}I_{61} = \left(vec\left[C_\nu'\right]'\otimes\hat{Q}_{\beta_3}^{-1}\right)\frac{1}{\sqrt{n}}\sum_i \tau_i vec\left[(Y_{i,T}'Q_{x,i}^{-1})\otimes(\beta_{3,i}'w_i)\right] + o_p(1)$. Moreover, by using the equality $vec\left[(Y_{i,T}'Q_{x,i}^{-1})\otimes(\beta_{3,i}'w_i)\right] = (Q_{x,i}^{-1}Y_{i,T})\otimes vec\left[\beta_{3,i}'w_i\right]$ (see MN Theorem 10 p. 55), we get $\hat{Q}_{\beta_3}^{-1}I_{61} = \left(vec\left[C_\nu'\right]'\otimes\hat{Q}_{\beta_3}^{-1}\right)\frac{1}{\sqrt{n}}\sum_i \tau_i\left[(Q_{x,i}^{-1}Y_{i,T})\otimes v_{3,i}\right] + o_p(1)$. Then $\hat{Q}_{\beta_3}^{-1}I_{61} \Rightarrow N\left(0,\Sigma_\nu\right)$ follows from Assumption B.2 a). Let us now consider I_{63}. By similar arguments as in the proof of Proposition 3 (control of term I_2), $\frac{1}{\sqrt{T}} I_{63} = o_p(1)$. The conclusion follows.

b) Asymptotic normality of $vec\left(\hat{\Lambda}'\right)$**.** We have $\sqrt{T} vec\left[\hat{\Lambda}' - \Lambda'\right] = \sqrt{T} vec\left[\hat{F}' - F'\right] + \sqrt{T}(\hat{\nu} - \nu)$. By using $\sqrt{T} vec\left[\hat{F}' - F'\right] = \left[I_K \otimes \left(\frac{1}{T}\sum_t Z_{t-1} Z_{t-1}'\right)^{-1}\right] \frac{1}{\sqrt{T}} \sum_t u_t \otimes Z_{t-1}$ 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.3.7 Proof of Proposition 10

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

$$\begin{aligned} \hat{Q}_e &= \frac{1}{n}\sum_i \left(\hat{\beta}_i - \beta_i\right)' C_{\hat{\nu}} \hat{w}_i C_{\hat{\nu}}' \left(\hat{\beta}_i - \beta_i\right) + O_{p,log}\left(\frac{1}{nT} + \frac{1}{T^2}\right) \\ &= \frac{1}{nT}\sum_i \tau_{i,T}^2 tr\left[C_{\hat{\nu}}' \hat{Q}_{x,i}^{-1} Y_{i,T} Y_{i,T}' \hat{Q}_{x,i}^{-1} C_{\hat{\nu}} \hat{w}_i\right] + O_{p,log}\left(\frac{1}{nT} + \frac{1}{T^2}\right). \end{aligned}$$

By using that $\tau_{i,T} tr\left[C_{\hat{\nu}}' \hat{Q}_{x,i}^{-1} \hat{S}_{ii} \hat{Q}_{x,i}^{-1} C_{\hat{\nu}} \hat{w}_i\right] = \mathbf{1}_i^{\chi} d_1$, Lemma 4 in the conditional case and $n = O(T^{\bar{\gamma}})$ with $\bar{\gamma} < 2$, we get:

$$\begin{aligned} \hat{\xi}_{nT} &= \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2 tr\left[C_{\hat{\nu}}' \hat{Q}_{x,i}^{-1}\left(Y_{i,T} Y_{i,T}' - \tau_{i,T}^{-1} \hat{S}_{ii}\right) \hat{Q}_{x,i}^{-1} C_{\hat{\nu}} \hat{w}_i\right] + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_i \tau_i^2 tr\left[C_{\hat{\nu}}' Q_{x,i}^{-1}\left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) Q_{x,i}^{-1} C_{\hat{\nu}} w_i\right] + o_p(1). \end{aligned}$$

Now, by using $tr(ABCD) = vec(D')'(C' \otimes A) vec(B)$ (MN Theorem 3, p. 35) and $vec(ABC) = (C' \otimes A) vec(B)$ for conformable matrices, we have:

$$\begin{aligned} & tr\left[C_{\hat{\nu}}' Q_{x,i}^{-1}\left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) Q_{x,i}^{-1} C_{\hat{\nu}} w_i\right] = vec[w_i]'\left(C_{\hat{\nu}}' \otimes C_{\hat{\nu}}'\right) vec\left[Q_{x,i}^{-1}\left(Y_{i,T} Y_{i,T}' - S_{ii,T}\right) Q_{x,i}^{-1}\right] \\ =\ & vec[w_i]'\left(C_{\hat{\nu}}' \otimes C_{\hat{\nu}}'\right)\left(Q_{x,i}^{-1} \otimes Q_{x,i}^{-1}\right) vec\left[Y_{i,T} Y_{i,T}' - S_{ii,T}\right] \\ =\ & vec[w_i]'\left(C_{\hat{\nu}}' \otimes C_{\hat{\nu}}'\right)\left(Q_{x,i}^{-1} \otimes Q_{x,i}^{-1}\right)\left(Y_{i,T} \otimes Y_{i,T} - vec[S_{ii,T}]\right) \\ =\ & vec\left[C_{\hat{\nu}}' \otimes C_{\hat{\nu}}'\right]'\left\{\left[\left(Q_{x,i}^{-1} \otimes Q_{x,i}^{-1}\right)\left(Y_{i,T} \otimes Y_{i,T} - vec[S_{ii,T}]\right)\right] \otimes vec[w_i]\right\}. \end{aligned}$$

Thus, we get $\hat{\xi}_{nT} = vec\left[C_{\hat{\nu}}' \otimes C_{\hat{\nu}}'\right]' \frac{1}{\sqrt{n}}\sum_i \tau_i^2\left[\left(Q_{x,i}^{-1} \otimes Q_{x,i}^{-1}\right)\left(Y_{i,T} \otimes Y_{i,T} - vec\left[S_{ii,T}\right]\right)\right] \otimes vec[w_i]$. From Assumption B.4, we get $\hat{\xi}_{nT} \Rightarrow N(0, \Sigma_{\xi})$, where $\Sigma_{\xi} = vec\left[C_{\nu}' \otimes C_{\nu}'\right]' \Omega vec\left[C_{\nu}' \otimes C_{\nu}'\right]$. Now, by using that $tr(ABCD) = vec(D)'(A \otimes C')vec(B')$ we have:

$$
\begin{aligned}
& vec\left[C'_\nu \otimes C'_\nu\right]' \left[(S_{Q,ij} \otimes S_{Q,ij}) \otimes vec[w_i]vec[w_j]'\right] vec\left[C'_\nu \otimes C'_\nu\right] \\
= \; & tr\left[(S_{Q,ij} \otimes S_{Q,ij})\left(C_\nu \otimes C_\nu\right) vec[w_j]vec[w_i]' \left(C'_\nu \otimes C'_\nu\right)\right] \\
= \; & vec[w_i]' \left[\left(C'_\nu S_{Q,ij} C_\nu\right) \otimes \left(C'_\nu S_{Q,ij} C_\nu\right)\right] vec[w_j] = tr\left[\left(C'_\nu S_{Q,ij} C_\nu\right) w_j \left(C'_\nu S_{Q,ji} C_\nu\right) w_i\right] \\
= \; & tr\left[\left(C'_\nu Q_{x,i}^{-1} S_{ij} Q_{x,j}^{-1} C_\nu\right) w_j \left(C'_\nu Q_{x,j}^{-1} S_{ji} Q_{x,i}^{-1} C_\nu\right) w_i\right],
\end{aligned}
$$

and similarly we have $vec\left[C'_\nu \otimes C'_\nu\right]' \left[(S_{Q,ij} \otimes S_{Q,ij})W_d \otimes vec[w_i]vec[w_j]'\right] vec\left[C'_\nu \otimes C'_\nu\right]$ $= tr\left[\left(C'_\nu Q_{x,i}^{-1} S_{ij} Q_{x,j}^{-1} C_\nu\right) w_j \left(C'_\nu Q_{x,j}^{-1} S_{ji} Q_{x,i}^{-1} C_\nu\right) w_i\right]$. Thus, we get the asymptotic variance matrix $\Sigma_\xi = 2 \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i,j} \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2} tr\left[\left(C'_\nu Q_{x,i}^{-1} S_{ij} Q_{x,j}^{-1} C_\nu\right) w_j \left(C'_\nu Q_{x,j}^{-1} S_{ji} Q_{x,i}^{-1} C_\nu\right) w_i\right]\right]$. From $\tilde{\Sigma}_\xi = \Sigma_\xi + o_p(1)$, the conclusion follows.

Appendix 4: Check of assumptions under block dependence

In this appendix, we verify that the eigenvalue condition in Assumption APR.4 (i), and the cross-sectional/time-series dependence and CLT conditions in Assumptions A.1-A.5, 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 $E[\varepsilon_t(\gamma)] = 0$ and $E[\varepsilon_t(\gamma)^3] = 0$, 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 γ and γ' 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) and the indicators $(I_t(\gamma))$, $\gamma \in [0,1]$, are i.i.d. over time, mutually independent, and independent of the errors $(\varepsilon_t(\gamma))$, $\gamma \in [0,1]$.

BD.4 There exists 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 in 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 requires that the largest block size shrinks with n and that there are not too many large blocks, i.e., the partition in independent blocks is sufficiently fine grained asymptotically. Within blocks, covariances do not need to vanish asymptotically.

Lemma 11 *Let Assumptions BD.1-4 on block dependence and Assumptions SC.1-SC.2 on random sampling hold. Then, Assumptions APR.4 (i), A.1, A.2, A.3, A.4 (with any $q \in (0,1)$ and $\delta \in (1/2,1)$) and A.5 are satisfied.*

The proof of Lemma 11 uses a result on almost sure convergence in Stout (1974), a large deviation theorem based on the Hoeffding's inequality in Bosq (1998), and CLTs for martingale difference arrays in Davidson (1994) and White (2001).

Instead of a block structure, we can also assume that the covariance matrix is full, but with off-diagonal elements vanishing asymptotically. In that setting, we can carry out similar checks.