A DIAGNOSTIC CRITERION FOR APPROXIMATE FACTOR STRUCTURE

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

First draft: February 2014. This version: February 2015.

Abstract

We build a simple diagnostic criterion for approximate factor structure in large cross-sectional equity datasets. Given a model for asset returns with observable factors, the criterion checks whether the error terms are weakly cross-sectionally correlated or share at least one unobservable common factor. It only requires computing the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals of a large unbalanced panel. A general version of this criterion allows us to determine the number of omitted common factors. The panel data model accommodates both time-invariant and time-varying factor structures. The theory applies to generic random coefficient panel models under large cross-section and time-series dimensions. The empirical analysis runs on monthly returns for about ten thousand US stocks from January 1968 to December 2011 for several time-varying specifications. Among several multi-factor time-invariant models proposed in the literature, we cannot select a model with zero factors in the errors. On the opposite, we conclude for no omitted factor structure in the errors for several time-varying specifications.

JEL Classification: C12, C13, C23, C51, C52, C58, G12.

Keywords: large panel, approximate factor model, asset pricing, model selection.

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

*Acknowledgements: The second author gratefully acknowledges the Swiss National Science Foundation (SNSF) for generously funding her research with a Marie Heim-Voegtlin fellowship. We thank participants at ESEM 2014 in Toulouse, COMPSTAT 2014 in Gèneve and a seminar at the University of Orléans.

1 Introduction

Empirical work in asset pricing vastly relies on linear multi-factor models with either time-invariant coefficients (unconditional models) or time-varying coefficients (conditional models). The factor structure is often based on observable variables (empirical factors) and supposed to be rich enough to extract systematic risks while idiosyncratic risk is left over to the error term. Linear factor models are rooted in the Arbitrage Pricing Theory (APT, Ross (1976), Chamberlain and Rothschild (1983)) or come from a loglinearization of nonlinear consumption-based models (Campbell (1993)). Conditional linear factor models aim at capturing the time-varying influence of financial and macroeconomic variables 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 time-invariant estimates of alphas and betas, and therefore asset pricing test conclusions (Jagannathan and Wang (1996), Lewellen and Nagel (2006), Boguth et al. (2011)). Ghysels (1998) discusses the pros and cons of modeling time-varying betas.

A central and practical issue is to determine whether there are one or more factors omitted in the chosen specification. Approximate factor structures with nondiagonal error covariance matrices (Chamberlain and Rothschild (1983)) answer the potential empirical mismatch of exact factor structures with diagonal error covariance matrices underlying the original APT of Ross (1976). If the set of observable factors is correctly specified, the errors are weakly cross-sectionally correlated. Given the large menu of factors available in the literature (the factor zoo of Cochrane (2011), see also Harvey, Liu, and Zhu (2013)), we need a simple diagnostic criterion to decide whether we can feel comfortable with the chosen set of observable factors.

For models with unobservable (latent) factors, Connor and Korajczyk (1993) are the first to develop a test for the number of factors for large balanced panels of individual stock returns in time-invariant models under covariance stationarity and homoskedasticity. Unobservable factors are estimated by the method of asymptotic principal components developed by Connor and Korajczyk (1986) (see also Stock and Watson (2002)). For heteroskedastic settings, the recent literature on large panels with static factors (see Hallin and Liška (2007) for a selection procedure in the generalized dynamic factor model of Forni et al. (2000)) has extended the toolkit available to researchers. Bai and Ng (2002) introduce a penalized least-squares strategy to estimate the number of factors, at least one, without restrictions on the relation between the cross-sectional dimension (n) and the time-series dimension (T). Caner and Han (2014) propose an estimator with a group bridge penalization to determine the number of unobservable factors. Onatski (2009, 2010) looks at the behavior of the adjacent eigenvalues to determine the number of factors when n and T are comparable. Ahn and Horenstein (2013) opt for the same strategy and cover the possibility of zero factors. Kapetanios (2010) uses subsampling to estimate the limit distribution of the adjacent eigenvalues. The asymptotic distribution of the eigenvalues is degenerate when the ratio T/n vanishes asymptotically (Jonsonn (1982)). In our empirical application on monthly returns for about ten thousand US stocks from January 1968 to December 2011, the cross-sectional dimension is much larger than the time series dimension. This explains why we favor the setting $T/n = o(1)$. This impedes us to exploit the Marchenko-Pastur distribution (Marchenko and Pastur (1967)) or other asymptotic characterizations obtained when T/n converges to a strictly positive constant. In the spirit of Lehmann and Modest (1988) and Connor and Korajczyk (1988), Bai and Ng (2006) analyze statistics to test whether the observable factors in time-invariant models span the space of unobservable factors. They do not impose any restriction on n and T. They find that the three factor model of Fama and French (1993, FF) is the most satisfactory proxy for the unobservable factors estimated from balanced panels of portfolio and individual stock returns. Ahn, Horenstein, and Wang (2013) study a rank estimation method to also check whether time-invariant factor models are compatible with a number of unobservable factors. For portfolio returns, they find that the FF model exhibits a full rank beta (factor loading) matrix.

In this paper, we build a simple diagnostic criterion for approximate factor structure in large cross-sectional datasets. The criterion checks whether the error terms in a given model with observable factors are weakly cross-sectionally correlated or share at least one common factor. It only requires computing the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals of a large unbalanced panel and subtracting a penalization term vanishing to zero for large n and T. The steps of the diagnostic are easy: 1) compute the largest eigenvalue, 2) subtract a penalty, 3) conclude to validity of the proposed approximate factor structure if the difference is negative, or conclude to at least one omitted factor if the difference is positive. Our theoretical contribution shows that step 3) yields asymptotically the correct model selection. We also propose a general version of the diagnostic criterion that determines the number of omitted common factors. We derive all properties for unbalanced panels in the setting of Connor and Korajczyk (1987) to avoid the survivorship bias inherent to studies restricted to balanced subsets of available stock return databases (Brown, Goetzmann, and Ross (1995)). The panel data model is sufficiently general to accommodate both time-invariant and time-varying factor structures (Gagliardini, Ossola, and Scaillet (2011, GOS)). We recast the factor models as generic random coefficient panel models and develop the theory for large cross-section and time-series dimensions with $T/n = o(1)$. Omitted latent factors are also viewed as interactive fixed effects in the panel literature (Pesaran (2006), Bai (2009), Gobillon and Magnac (2014), Moon and Weidner (forthcoming, 2015)). As shown below, the criterion is related to the penalized least-squares approach of Bai and Ng (2002) for model selection with unobservable factors.

For our empirical contribution, 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 January 1968 to December 2011. We look at fifteen empirical factors and we build thirteen factor models popular in the empirical finance literature to explain monthly equity returns. They differ by the choice of the observable factors. We analyze monthly returns using the three factors of FF; the five factors of Chen, Roll, and Ross (1986, CRR); the three factor of Jagannathan and Wang (1996, JW); the three liquidity related factors of Pastor and Stambaugh (2002, LIQ), plus the momentum (MOM) factor and the two return reversal (REV) factors (short-term and long-term). We study time-invariant and time-varying versions of the factor models (Shanken (1990), Cochrane (1996), Ferson and Schadt (1996), Ferson and Harvey (1999)). For the latter, we use both macrovariables and firm characteristics as instruments (Avramov and Chordia (2006)). Among the time-invariant multi-factor models, we cannot select a model with zero factors in the errors. On the opposite, we conclude for no omitted factor structure in the errors for several time-varying specifications.

The outline of the paper is as follows. In Section 2, we consider a general framework of conditional linear factor model for asset returns. In Section 3, we present our diagnostic criterion for approximate factor structure in random coefficient panel models. In Section 4, we provide the diagnostic criterion to determine the number of omitted factors. Section 5 contains the empirical results. In the Appendices 1 and 2, we gather the theoretical assumptions and some proofs. We use high-level assumptions to get our results, and show in Appendix 3 that we meet them under a block cross-sectional dependence structure on the error terms in a serially i.i.d. framework. We place all omitted proofs in the online supplementary materials. There we link our approach to the expectation-maximization (EM) algorithm proposed by Stock and Watson (2002) for unbalanced panels. We also includes some Monte-Carlo simulation results under a design mimicking our empirical application to show the practical relevance of our selection procedure in finite samples. We report some additional empirical results and robustness checks.

2 Conditional factor model of asset returns

In this section, we consider a conditional linear factor model with time-varying coefficients. We work in a multi-period economy (Hansen and Richard (1987)) under an approximate factor structure (Chamberlain and Rothschild (1983)) with a continuum of assets as in GOS. Such a construction is close to the setting advocated by Al-Najjar (1995, 1998, 1999a) in a static framework with an exact factor structure. He discusses several key advantages of using a continuum economy in arbitrage pricing and risk decomposition. A key advantage is robustness of factor structures to asset repackaging (Al-Najjar (1999b); see GOS for a proof).

Let $\mathcal{F}_t$, with $t = 1, 2, ...$, be the information available to investors. Without loss of generality, the continuum of assets is represented by the interval $[0, 1]$. The excess returns $R_t(\gamma)$ of asset $\gamma \in [0, 1]$ at dates $t = 1, 2, ...$ satisfy the conditional linear factor model:

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

where vector f_t gathers the values of K observable factors at date t. The intercept $a_t(\gamma)$ and factor sensitivities $b_t(\gamma)$ are $\mathcal{F}_{t-1}$-measurable. The error terms $\varepsilon_t(\gamma)$ have mean zero and are uncorrelated with the factors conditionally on information $\mathcal{F}_{t-1}$. Moreover, we exclude asymptotic arbitrage opportunities in the economy: there are no portfolios that approximate arbitrage opportunities when the number of assets increases. In this setting, GOS show that the following asset pricing restriction holds:

$$a_t(\gamma) = b_t(\gamma)'\nu_t, \text{ for almost all } \gamma \in [0, 1], \tag{2}$$

almost surely in probability, where random vector $\nu_t \in \mathbb{R}^K$ is unique and is $\mathcal{F}_{t-1}$-measurable. The asset pricing restriction (2) is equivalent to $E[R_t(\gamma)|\mathcal{F}_{t-1}] = b_t(\gamma)'\lambda_t$, where $\lambda_t = \nu_t + E[f_t|\mathcal{F}_{t-1}]$ is the vector of the conditional risk premia.

To have a workable version of Equations (1) and (2), we define how the conditioning information is generated and how the model coefficients depend on it via simple functional specifications. The conditioning information $\mathcal{F}_{t-1}$ contains $\underline{Z_{t-1}}$ and $\underline{Z_{t-1}}(\gamma)$, for all $\gamma \in [0,1]$, where the vector of lagged instruments $Z_{t-1} \in \mathbb{R}^p$ is common to all stocks, the vector of lagged instruments $Z_{t-1}(\gamma) \in \mathbb{R}^q$ is specific to stock γ, and $Z_{\underline{t}} = \{Z_t, Z_{t-1}, ...\}$. Vector Z_{t-1} may include the constant and past observations of the factors and some additional variables such as macroeconomic variables. Vector $Z_{t-1}(\gamma)$ may include past observations of firm characteristics and stock returns. To end up with a linear regression model, we assume that: (i) the vector of factor loadings $b_t(\gamma)$ 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)); (ii) the vector of risk premia λ_t is a linear function of lagged instruments Z_{t-1} (Cochrane (1996), Jagannathan and Wang (1996)); (iii) the conditional expectation of f_t given the information $\mathcal{F}_{t-1}$ depends on Z_{t-1} only and is linear (as e.g. if Z_t follows a Vector Autoregressive (VAR) model of order 1).

To ensure that cross-sectional limits exist and are invariant to reordering of the assets, we introduce a sampling scheme as in GOS. We formalize it so that observable assets are random draws from an underlying population (Andrews (2005)). In particular, we rely on a sample of n assets by randomly drawing i.i.d. indices γ_i from the population according to a probability distribution G on $[0,1]$. For any $n, T \in \mathbb{N}$, the excess returns are $R_{i,t} = R_t(\gamma_i)$. Similarly, let $a_{i,t} = a_t(\gamma_i)$ and $b_{i,t} = b_t(\gamma_i)$ be the coefficients, and $\varepsilon_{i,t} = \varepsilon_t(\gamma_i)$ be the error terms. By random sampling, we get a random coefficient panel model (e.g. Hsiao (2003), Chapter 6). In available datasets, we do not observe asset returns for all firms at all dates. Thus, we account for the unbalanced nature of the panel through a collection of indicator variables $I_{i,t}$, for any asset i at time t. We define $I_{i,t} = 1$ if the return of asset i is observable at date t, and 0 otherwise (Connor and Korajczyk (1987)).

Through appropriate redefinitions of the regressors and coefficients, GOS show that we can rewrite the model for Equations (1) and (2) as a generic random coefficient panel model:

$$R_{i,t} = x'_{i,t}\beta_i + \varepsilon_{i,t}, \tag{3}$$

where the regressor $x_{i,t} = \left(x'_{1,i,t}, x'_{2,i,t}\right)'$ has dimension $d = d_1 + d_2$ and includes vectors $x_{1,i,t} = \left(vech\left[X_t\right]', Z'_{t-1} \otimes Z'_{i,t-1}\right)' \in \mathbb{R}^{d_1}$ and $x_{2,i,t} = \left(f'_t \otimes Z'_{t-1}, f'_t \otimes Z'_{i,t-1}\right)' \in \mathbb{R}^{d_2}$ with $d_1 = p(p+1)/2 + pq$ and $d_2 = K(p+q)$. 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 $vech\,[\cdot]$ stacks the elements of the lower triangular part of a $p \times p$ matrix as a $p\,(p+1)\,/2 \times 1$ vector (see Chapter 2 in Magnus and Neudecker (2007) for properties of this matrix tool). In matrix notation, for any asset i, we have

$$R_i = X_i \beta_i + \varepsilon_i, \tag{4}$$

where R_i and ε_i are $T \times 1$ vectors. Regression (3) contains both explanatory variables that are common across assets (scaled factors) and asset-specific regressors. It includes models with time-invariant coefficients as a particular case. In such a case, the regressor reduces to $x_t = (1, f_t')'$ and is common across assets.

In order to build the diagnostic criterion for the set of observable factors, we consider the following rival models:

$\mathcal{M}_1$: the linear regression model (3), where the errors $(\varepsilon_{i,t})$ are weakly cross-sectionally dependent,

and

$\mathcal{M}_2$: the linear regression model (3), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure.

Under model $\mathcal{M}_1$, the observable factors capture the systematic risk, and the error terms do not feature pervasive forms of cross-sectional dependence (see Assumptions A.1 and A.3 in Appendix 1). Under model $\mathcal{M}_2$, the following error factor structure holds

$$\varepsilon_{i,t} = \theta_i' h_t + u_{i,t}, \tag{5}$$

where the $m \times 1$ vector h_t includes unobservable (i.e., latent or hidden) factors, and the $u_{i,t}$ are weakly cross-sectionally correlated. The $m \times 1$ vector θ_i corresponds to the factor loadings, and the number m of common factors is assumed unknown. In vector notation, we have:

$$\varepsilon_i = H\theta_i + u_i, \tag{6}$$

where H is the $T \times m$ matrix of unobservable factor values, and u_i is a $T \times 1$ vector.

Assumption 1 *Under model $\mathcal{M}_2$: (i) Matrix $\frac{1}{T}\sum_t h_t h_t'$ converges in probability to a positive definite matrix Σ_h, as $T \to \infty$. (ii) $\mu_1\left(\frac{1}{n}\sum_i \theta_i \theta_i'\right) \geq C$, w.p.a. 1 as $n \to \infty$, for a constant $C > 0$, where $\mu_1(.)$ denotes the largest eigenvalue of a symmetric matrix.*

Assumption 1 (i) is a standard identification condition on the latent factor (see Assumption A in Bai and Ng (2002)) and matrix Σ_h can be normalized to the identity matrix I_m. Assumption 1 (ii) requires that at least one factor in the error terms is strong. It is satisfied if the second-order matrix of the loadings $\frac{1}{n}\sum_i \theta_i \theta_i'$ converges in probability to a positive definite matrix (see Assumption B in Bai and Ng (2002)).

We work with the condition:

$$E[x_{i,t} h_t'] = 0, \quad \forall i, \tag{7}$$

that is, orthogonality between latent factors and observable regressors for all stocks. This condition allows us to follow a two-step approach: we first regress stock returns on observable regressors to compute residuals, and then search for latent common factors in the panel of residuals (see next section). We can interpret condition (7) in a partitioned regression: $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ as follows. The Frisch-Waugh-Lovell Theorem (Frisch and Frederick (1933), Lovell (1963)) states that the ordinary least squares (OLS) estimate of β_2 is identical to the OLS estimate of β_2 in the regression $M_{X_1}Y = M_{X_1}X_2\beta_2 + \eta$, where $M_{X_1} = I_d - X_1 \left(X_1' X_1\right)^{-1} X_1'$. Condition (7) is similar to the orthogonality condition $X_1' X_2 = 0$ ensuring that we can estimate β_2 from regressing the residuals $M_{X_1}Y$ on X_2 only, instead of the residuals $M_{X_1}X_2$ coming from the regression of X_2 on X_1. When condition (7) is not satisfied, joint estimation of regression coefficients, latent factor betas and factor values is required (see e.g. Bai (2009), Moon and Weidner (forthcoming, 2015) in a model with homogeneous regression coefficients $\beta_i = \beta$ for all i). If the regressors are common across stocks, i.e. $x_{i,t} = x_t$, we can obtain condition (7) by transformation of the latent factors. It simply corresponds to an identification restriction on the latent factors. If the regressors are stock-specific, ensuring orthogonality between the latent factors h_t and the observable regressors $x_{i,t}$ for all i is more than an identification restriction. It requires an additional assumption where we decompose common and stock-specific components in the regressors vector by writing $x_{i,t} = (x_t', \tilde{x}_{i,t-1}')'$, where $x_t := (vec[X_t]', f_t' \otimes Z_{t-1}')'$ and $\tilde{x}_{i,t} := (Z_{t-1}' \otimes Z_{i,t-1}', f_t' \otimes Z_{i,t-1}')'$.

Assumption 2 *The best linear prediction of the unobservable factor $EL(h_t | \{x_{i,t},\ i = 1, 2, ...\})$ is independent of $\{\tilde{x}_{i,t},\ i = 1, 2, ...\}$.*

Assumption 2 amounts to Granger non-causality from the stock-specific regressors to the latent factors, conditionally on the common regressors. Assumption 2 is verified e.g. if the latent factors are independent of the lagged stock-specific instruments, conditional on the observable factors and the lagged common instruments (see the supplementary materials for a derivation). We keep Assumption 2 as a maintained assumption on the factor structure under $\mathcal{M}_2$. Under Assumption 2, $EL(h_t|\{x_{i,t},\ i=1,2,...\}) =: \Psi x_t$ is a linear function of x_t. Therefore, by transformation of the latent factor $h_t \to h_t - \Psi x_t$, we can assume that $EL(h_t|\{x_{i,t},\ i=1,2,...\}) = 0$, without loss of generality. This condition implies (7).

3 Diagnostic criterion

In this section, we provide the diagnostic criterion that checks whether the error terms are weakly cross-sectionally correlated or share at least one common factor. To compute the criterion, we estimate the generic panel model (3) by OLS asset by asset, and we get estimators $\hat{\beta}_i = \hat{Q}_{x,i}^{-1}\frac{1}{T_i}\sum_t I_{i,t}x_{i,t}R_{i,t}$, 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 get the residuals $\hat{\varepsilon}_{i,t} = R_{i,t} - x'_{i,t}\hat{\beta}_i$, where $\hat{\varepsilon}_{i,t}$ is observable only if $I_{i,t} = 1$. 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. To avoid unreliable estimates of β_i, we apply a trimming approach as in GOS. We define $\mathbf{1}_i^{\chi} = \mathbf{1}\left\{CN\left(\hat{Q}_{x,i}\right) \leq \chi_{1,T}, \tau_{i,T} \leq \chi_{2,T}\right\}$, where $CN\left(\hat{Q}_{x,i}\right) = \sqrt{\mu_1\left(\hat{Q}_{x,i}\right)/\mu_d\left(\hat{Q}_{x,i}\right)}$ is the condition number of the $d \times d$ matrix $\hat{Q}_{x,i}$, $\mu_d\left(\hat{Q}_{x,i}\right)$ is its smallest eigenvalue and $\tau_{i,T} = T/T_i$. 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 also use both trimming conditions in the proofs of the asymptotic results.

We consider the following diagnostic criterion:

$$\xi = \mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}'_i\right) - g(n,T), \tag{8}$$

where the vector $\bar{\varepsilon}_i$ of dimension T gathers the values $\bar{\varepsilon}_{i,t} = I_{i,t}\hat{\varepsilon}_{i,t}$, the penalty $g(n,T)$ is such that $g(n,T) \to 0$ and $C_{n,T}^2 g(n,T) \to \infty$, when $n,T \to \infty$, for $C_{n,T}^2 = \min\{n,T\}$. Bai and Ng (2002) consider several simple potential candidates for the penalty $g(n,T)$. We list and implement them in Section

5. In vector $\bar{\varepsilon}_i$, the unavailable residuals are replaced by zeros. The following model selection rule explains our choice of the diagnostic criterion (8) for approximate factor structure in large unbalanced cross-sectional datasets.

Proposition 1 *Model selection rule: Under Assumptions 1, 2 and Assumptions A.1-A.9, (a) we select $\mathcal{M}_1$ if $\xi < 0$, since $Pr\left(\xi < 0 \mid \mathcal{M}_1\right) \to 1$, when $n, T \to \infty$, such that $T/n = o\left(1\right)$; (b) we select $\mathcal{M}_2$ if $\xi > 0$, since $Pr\left(\xi > 0 \mid \mathcal{M}_2\right) \to 1$, when $n, T \to \infty$, such that $T/n = o\left(1\right)$.*

In Proposition 1, we have the additional constraint $T/n = o\left(1\right)$ on the relative rate of the cross-sectional dimension w.r.t. the time series dimension. We use $T/n = o\left(1\right)$ to show the compatibility of Assumption A.3 with a block dependence structure in the error terms. This exemplifies a key difference with the proportional asymptotics used in Onatski (2009, 2010) or Ahn and Horenstein (2013) for balanced panel without observable factors. They rely on the asymptotic distribution of the eigenvalues of large dimensional sample covariances matrices when $n/T \to c > 0$ as $n \to \infty$. The condition $T/n = o\left(1\right)$ agrees with the "large n, small T" case that we face in the empirical application (ten thousand individual stocks monitored over forty-five years of monthly returns).

Proposition 1 characterizes an asymptotically valid model selection rule, which treats both models symmetrically. This is not a testing procedure since we do not use a critical region based on an asymptotic distribution and a chosen significance level. The proof of Proposition 1 shows that the largest eigenvalue in (8) vanishes at a faster rate (see Lemma 4 in the proof) than the penalization term under $\mathcal{M}_1$ when n and T go to infinity. Under $\mathcal{M}_1$, we expect a vanishing largest eigenvalue because of a lack of a common signal in the error terms. The negative penalizing term $-g(n, T)$ dominates in (8), and this explains why we select the first model when ξ is negative. On the contrary, the largest eigenvalue remains bounded from below away from zero (see Lemma 4 in the proof) under $\mathcal{M}_2$ when n and T go to infinity. Under $\mathcal{M}_2$, we have at least one non vanishing eigenvalue because of a common signal due to omitted factors. The largest eigenvalue dominates in (8), and this explains why we select the second model when ξ is positive. We can interpret the criterion (8) as the adjusted gain in fit including a single additional (unobservable) factor in model $\mathcal{M}_1$. In the balanced case, where $I_{i,t} = 1$ for all i and t, we can rewrite (8) as $\xi = SS_0 - SS_1 - g\left(n, T\right)$, where $SS_0 = \dfrac{1}{nT}\sum_i \sum_t \hat{\varepsilon}_{i,t}^2$ is the sum of squared errors and $SS_1 = \min \dfrac{1}{nT}\sum_i \sum_t \left(\hat{\varepsilon}_{i,t} - \theta_i h_t\right)^2$, where the minimization is w.r.t. the vectors $H \in \mathbb{R}^T$ of factor values and $\Theta \in \mathbb{R}^n$ of factor loadings in a one-factor model, subject to the normalization constraint $\frac{H'H}{T} = 1$. Indeed, the largest eigenvalue $\mu_1 \left(\frac{1}{nT} \sum_i \hat{\varepsilon}_i \hat{\varepsilon}_i' \right)$ corresponds to the difference between SS_0 and SS_1. Furthermore, the criterion ξ is equal to the difference of the penalized criteria for zero- and one-factor models defined in Bai and Ng (2002) applied on the residuals. Indeed, $\xi = PC\,(0) - PC\,(1)$, where $PC\,(0) = SS_0$, and $PC\,(1) = SS_1 + g\,(n, T)$. Given such an interpretation in terms of sums of squared errors, we can suggest another diagnostic criterion based on a logarithmic transform as in Corollary 2 of Bai and Ng (2002). The second diagnostic criterion is

$$\check{\xi} = \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \hat{\varepsilon}_{i,t}^2 \right) - \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \hat{\varepsilon}_{i,t}^2 - \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \hat{\varepsilon}_i \hat{\varepsilon}_i' \right) \right) - g(n, T). \qquad (9)$$

In the balanced case, we get $\check{\xi} = \ln(SS_0/SS_1) - g(n, T)$ and it is equal to the difference of $IC\,(0)$ and $IC\,(1)$ criteria in Bai and Ng (2002). Then, the model selection rule is the same as in Proposition 1 with $\check{\xi}$ substituted for ξ.

The recent literature on the properties of the two-pass regressions for fixed n and large T shows that the presence of useless factors (Kan and Zhang (1999a,b), Gospodinov, Kan, and Robotti (2014)) or weak factor loadings (Kleibergen (2009)) does not affect the asymptotic distributional properties of factor loading estimates, but alters the ones of the risk premia estimates. Useless factors have zero loadings, and weak loadings drift to zero at rate $1/\sqrt{T}$. The vanishing rate of the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals does not change if we face useless factors or weak factor loadings in the observable factors under $\mathcal{M}_1$. The same remark applies under $\mathcal{M}_2$. Hence the selection rule remains the same since the probability of taking the right decision still approaches 1. If we have a number of useless factors or weak factor loadings strictly lower than the number m of the omitted factors under $\mathcal{M}_2$, this does not impact the asymptotic rate of the diagnostic criterion if Assumption 1 holds. If we only have useless factors in the omitted factors under $\mathcal{M}_2$, we face an identification issue. Assumption 1 (ii) is not satisfied. We cannot distinguish such a specification from $\mathcal{M}_1$ since it corresponds to a particular approximate factor structure. Again the selection rule remains the same since the probability of taking the right decision still approaches 1. Finally, let us study the case of only weak factor loadings under $\mathcal{M}_2$. We consider a simplified setting:

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

where $\varepsilon_{i,t} = \theta_i h_t + u_{i,t}$ has only one factor with a weak factor loading, namely $m = 1$ and $\theta_i = \bar{\theta}_i / T^{\gamma}$ with $\gamma > 0$. Let us assume that $\mu_1 \left(\frac{1}{n} \sum_i \bar{\theta}_i^2 \right)$ is bounded from below away from zero (see Assumption 1 (ii)) and bounded from above. By the properties of the eigenvalues of a scalar multiple of a matrix, we deduce that $c_1 / T^{2\gamma} \leq \mu_1 \left(\frac{1}{n} \sum_i \theta_i^2 \right) \leq c_2 / T^{2\gamma}$, for some constants c_1, c_2 such that $c_2 \geq c_1 > 0$. Hence, by similar arguments as in the proof of Proposition 1, we get:

$$c_1 T^{-2\gamma} - g(n,T) + O_p\left(C_{nT}^{-2} + \bar{\chi}_T T^{-1}\right) \leq \xi \leq c_2 T^{-2\gamma} - g(n,T) + O_p\left(C_{nT}^{-2} + \bar{\chi}_T T^{-1}\right),$$

where we define $\bar{\chi}_T = \chi_{1,T}^4 \chi_{2,T}^2$. To conclude $\mathcal{M}_2$, we need that $C_{nT}^{-2} + \bar{\chi}_T T^{-1}$ and the penalty $g(n,T)$ vanish at a faster rate than $T^{-2\gamma}$, namely $C_{nT}^{-2} + \bar{\chi}_T T^{-1} = o\left(T^{-2\gamma}\right)$ and $g(n,T) = o\left(T^{-2\gamma}\right)$. To conclude $\mathcal{M}_1$, we need that $g(n,T)$ is the dominant term, namely $T^{-2\gamma} = o\left(g(n,T)\right)$ and $C_{nT}^{-2} + \bar{\chi}_T T^{-1} = o\left(g(n,T)\right)$. As an example, let us take $g(n,T) = T^{-1} \log T$ and $n = T^{\bar{\gamma}}$ with $\bar{\gamma} > 1$, and assume that the trimming is such that $\bar{\chi}_T = o(\log T)$. Then, we conclude $\mathcal{M}_2$ if $\gamma < 1/2$ and $\mathcal{M}_1$ if $\gamma > 1/2$. This means that detecting a weak factor loading structure is difficult if gamma is not sufficiently small. The factor loading should drift to zero not too fast to conclude $\mathcal{M}_2$. Otherwise, we cannot distinguish it asymptotically from weak cross-sectional correlation.

4 Determining the number of factors

In the previous section, we have studied a diagnostic criterion to check whether the error terms are weakly cross-sectionally correlated or share at least one unobservable common factor. This section aims at answering: do we have one, two, or more omitted factors? The design of the diagnostic criterion to check whether the error terms share exactly k unobservable common factors or share at least $k+1$ unobservable common factors follows the same mechanics. We consider the following rival models:

$\mathcal{M}_1(k)$: the linear regression model (3), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure with exactly k unobservable factors, and

$\mathcal{M}_2(k)$: the linear regression model (3), where the errors $(\varepsilon_{i,t})$ satisfy a factor structure with at least $k+1$ unobservable factors.

The above definitions yield $\mathcal{M}_1 = \mathcal{M}_1(0)$ and $\mathcal{M}_2 = \mathcal{M}_2(0)$.

Assumption 3 *Under model $\mathcal{M}_2(k)$, we have $\mu_{k+1}\left(\frac{1}{n}\sum_i \theta_i\theta_i'\right) \geq C$, w.p.a. 1 as $n \to \infty$, for a constant $C > 0$, where $\mu_{k+1}(.)$ denotes the $(k+1)$-th largest eigenvalue of a symmetric matrix.*

Models $\mathcal{M}_1(k)$ and $\mathcal{M}_2(k)$ are subsets of model $\mathcal{M}_2$. Hence, Assumption 1 (i) guarantees the convergence of matrix $\frac{1}{T}\sum_t h_t h_t'$ to a positive definite $k \times k$ matrix under $\mathcal{M}_1(k)$, and to a positive definite $m \times m$ matrix under $\mathcal{M}_2(k)$, respectively, with $m \geq k+1$. Assumption 3 requires that there are at least $k+1$ strong factors under $\mathcal{M}_2(k)$.

The diagnostic criterion exploits the $(k+1)$th largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals:

$$\xi(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - g(n,T). \tag{10}$$

As discussed in Ahn and Horenstein (2013) (see also Onatski (2013)), we can rewrite (10) in the balanced case as $\xi(k) = SS_k - SS_{k+1} - g(n,T)$ where SS_k equals the sample mean of the squared residuals from the time series regressions of individual response variables $(\hat{\varepsilon}_{i,t})$ on the first k principal components of $\frac{1}{nT}\sum_i \hat{\varepsilon}_i\hat{\varepsilon}_i'$. The criterion $\xi(k)$ is equal to the difference of the penalized criteria for k and $(k+1)$- factor models defined in Bai and Ng (2002) applied on the residuals. Indeed, $\xi(k) = PC(k) - PC(k+1)$, where $PC(k) = SS_k + kg(n,T)$ and $PC(k+1) = SS_{k+1} + (k+1)g(n,T)$. To determine the number of unobservable factors, we choose the minimum k such that $\xi(k) < 0$. Graphically, we can build a penalized scree plot where we display the penalized eigenvalues associated with each factor in descending order versus the number of the factor, and use the x-axis for the cut-off point. The number m of unobservable factors in (6) is of no use in such a procedure. This avoids the need to prespecify a maximum possible number of factors ($kmax$) as in Bai and Ng (2002), Onatski (2009, 2010), Ahn and Horenstein (2013). We believe that this is a strong advantage of our methodology since there are many possible choices for $kmax$ and the estimated number of factor is sometimes sensitive to the choice of $kmax$ (see the simulation results in those papers). In the online supplementary materials, we show that our procedure selects the right number of factor with 99 percent chances in most cases when n is much larger than T. The following model selection rule extends Proposition 1 to determine the number of factors.

Proposition 2 *Model selection rule: under Assumptions 1(i), 2 and 3, and Assumptions A.1-A.9, (a) we select* $\mathcal{M}_1(k)$ *if* $\xi(k) < 0$, *since* $Pr[\xi(k) < 0|\mathcal{M}_1(k)] \to 1$, *when* $n, T \to \infty$, *such that* $T/n = o(1)$; *(b) we select* $\mathcal{M}_2(k)$ *if* $\xi(k) > 0$, *since* $Pr[\xi(k) > 0|\mathcal{M}_2(k)] \to 1$, *when* $n, T \to \infty$, *such that* $T/n = o(1)$.

The proof of Proposition 2 is also more complicated than the proof of Proposition 1. The proof of the latter exploits the asymptotic bound on the largest eigenvalue of a symmetric matrix (Lemma 1). We need additional arguments to derive such a bound when we look at the $(k+1)$th eigenvalue (Lemma 5).

5 Empirical results

5.1 Factor models and data description

We consider fifteen non-repetitive empirical factors as in Ahn, Horenstein, and Wang (2013). The three factors of Fama and French (1993) are the monthly excess return on CRSP NYSE/AMEX/Nasdaq value-weighted market portfolio over the risk free rate $r_{m,t}$, and the monthly returns on zero-investment factor-mimicking portfolios for size and book-to-market, denoted by $r_{smb,t}$ and $r_{hml,t}$ respectively. The monthly returns on portfolio for momentum is denoted by $r_{mom,t}$. Two reversal factors are monthly returns on portfolio for short $r_{str,t}$, and long term $r_{ltr,t}$. We have downloaded the time series of these factors from the website of Kenneth French. We consider the five factors of Chen, Roll, and Ross (1986) available from Laura Xiaolei Liu's webpage. The monthly CRR factors are the growth rate of industrial production mp_t, the unexpected inflation ui_t, the term spread uts_t, proxied by the difference between yields on 10-year Treasury and 3-month T-bill, and the default premia upr_t, proxied by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds. Moreover, we consider the three liquidity-related factors of Pastor and Stambaugh (2002) that consist of the monthly liquidity level al_t, traded liquidity tl_t and the innovation in aggregate liquidity il_t. We have downloaded the LIQ factors from the website of Lubos Pastor. Finally, we build the monthly growth rate of labor income lab_t from the Bureau of Economic Analysis's webpage. We proxy the risk free rate with the monthly 30-day T-bill beginning-of-month yield. To account for time-varying coefficients, we use two conditional specifications 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 either (i) the term spread and the default spread, or (ii) the monthly 30-day T-bill and the dividend yields. 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 parsimony explains why we have not included e.g. the size of firm i as an additional stock specific instrument.

Table 1 reports the thirteen linear factor models that we estimate in order to computed the diagnostic criteria. For each model, we specify the empirical factors involved and the number K of observable factors. We look at factor models popular in the empirical finance. We also consider nested models built from the fifteen empirical factors.

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 $n = 10,442$ stocks, and covers the period from January 1968 to December 2011 with $T = 528$ months.

5.2 Diagnostic results

In this section, we compute the diagnostic criteria in Equations (8) and (9) assuming time-invariant and time-varying specifications of the linear factor models listed in Table 1. We need to define the specification for the penalty $g\left(n, T\right)$. Bai and Ng (2002) propose three choices for the penalty function in Equation (8), leading to the following criteria:

1. $\xi_1 = \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) - \hat{\sigma}^2 \left(\frac{n+T}{nT} \right) \ln \left(\frac{nT}{n+T} \right)$;

2. $\xi_2 = \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) - \hat{\sigma}^2 \left(\frac{n+T}{nT} \right) \ln C_{nT}^2;$

3. $\xi_3 = \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) - \hat{\sigma}^2 \left(\frac{\ln C_{nT}^2}{C_{nT}^2} \right),$

where $\hat{\sigma}^2 = \frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{1,i,t}^2$, and $\bar{\varepsilon}_{1,i,t}$ is the fitted residual of the time-varying linear factor model built on the FF, MOM, REV observable factors and a latent factor. Similarly, we get the following logarithmic criteria based on Equation (9). We get the following logarithmic criteria:

1. $\check{\xi}_1 = \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 \right) - \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 - \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) \right) - \left(\frac{n+T}{nT} \right) \ln \left(\frac{nT}{n+T} \right);$

2. $\check{\xi}_2 = \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 \right) - \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 - \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) \right) - \left(\frac{n+T}{nT} \right) \ln C_{nT}^2;$

3. $\check{\xi}_3 = \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 \right) - \ln \left(\frac{1}{nT} \sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 - \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) \right) - \left(\frac{\ln C_{nT}^2}{C_{nT}^2} \right),$

Each time-series is demeaned and standardized to have unit variance before computing the eigenvalues. This ensures that all series have a common scale of measurement and improves the stability of the information extracted from the multivariate time series (see Pena and Poncela (2006)). We fix $\chi_{1,T} = 15$ as advocated by Greene (2008), and $\chi_{2,T} = 546/12$ for the time-invariant estimation and $\chi_{1,T} = 20$ and $\chi_{2,T} = 546/60$ for the time-varying estimation. In Table 2, we report the size of trimmed cross-sectional dimension n^{χ} that comes from the trimming procedure applied in the estimation approach. In some time-varying specifications, we face severe multicollinearity problems due to the correlations within the vector of regressors $x_{i,t}$, that involves cross product of factors f_t and instruments Z_{t-1} (e.g., in the JW and CRR models), and the large dimension of vector $x_{i,t}$ (e.g., the number of parameter to estimate is larger than 40 in models 11-13).

For the time-invariant specifications of (1)-(13) models, we plot the values of the diagnostic criteria ξ_1, ξ_2 and ξ_3 in Figure 1, and $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ in Figure 2. For the time-varying specifications, Figures 3 and 4 plot the values of the diagnostic criteria computed with the common instruments (i). Figures 5 and 6 plot the results by using the second set of common instruments. Since the penalty function is proportional to $\frac{1}{T} \ln T$, the numerical value of criteria ξ_s and $\check{\xi}_s$, with $s = 1, 2, 3$, do not differ much from each other. For the majority of the models, the selected model remains the same when we rely on (8) or (9). In particular, we cannot select a time-invariant model with zero factors in the errors. We conclude for no omitted factor in the error terms when we estimate the time-varying linear factor models based on FF and REV factors. In general, focusing on nested models, when the number of factor increases the diagnostic criteria decreases. Finally, in many cases, the diagnostic criteria is smaller for the time-varying specifications than for the time-invariant models.

In Tables 3-6, we compare the descriptive statistics of four measures of missing factor impact: (i) the estimated time-series coefficient of determination $\hat{\rho}_i^2 = \frac{ESS_i}{TSS_i}$, where $ESS_i = \sum_t I_{i,t}\left(\hat{R}_{i,t} - \bar{\hat{R}}_i\right)^2$, with $\hat{R}_{i,t} = \hat{\beta}_i' x_{i,t}$ and $\bar{\hat{R}}_i = \frac{1}{T_i}\sum_t I_{i,t}\hat{R}_{i,t}$, and $TSS_i = \sum_t I_{i,t}\left(R_{i,t} - \bar{R}_i\right)^2$, with $\bar{R}_i = \frac{1}{T_i}\sum_t I_{i,t} R_{i,t}$; (ii) the estimated adjusted R^2 defined by $\hat{\rho}_{ad,i}^2 = 1 - \frac{(T_i - 1)}{(T_i - d)}\left(1 - \hat{\rho}_i^2\right)$; (iii) the idiosyncratic risk $IdiVol_i = \sqrt{\frac{RSS_i}{T_i}}$, with $RSS_i = \sum_t I_{i,t}\hat{\varepsilon}_{i,t}^2$; (iv) the systematic risk $SysRisk_i = \sqrt{\frac{ESS_i}{T_i}}$, for the time-invariant and time-varying specifications. We consider those estimates as measures of missing factor impact (see Ang, Liu and Schwarz (2008)). The time-series (adjusted) coefficient of determination tend to be a bit larger in the time-varying model than in the time-invariant specifications. The $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$ and $SysRisk_i$ admit large values for the models that introduced the FF, MOM and/or REV factors in their specification. For these linear specifications, we observe that the diagnostic criteria ξ and $\check{\xi}$ admit small values.

5.3 The number of factors

In this section, we compute the diagnostic criteria $\xi(k)$ in (10) that exploit the $(k+1)$-th largest eigenvalue of the empirical cross-sectional covariance matrix of the errors. We compute the diagnostic criteria for the first five eigenvalues, and we use the penalty function $g(n, T)$ defined in the previous section. For each linear factor specification, we build a penalized scree plot. Figures 7 and 8 show the results for the time-invariant specifications. We observe that diagnostic criteria change signs when we consider the time-invariant specifications based on the FF factors. In particular, the diagnostic criteria become negative when $k = 4$ for the FF and Carhart (1997) models. The number of omitted unobservable common factors k is 3 for the time-invariant model that accounts for more than 8 observable factors (e.g., models (11)-(13)). However, the three FF factors alone do not fully explain systematic risk in the excess returns for stocks. Let us consider the results for the time-varying specifications in Figures 9 and 10. In both figures, the cut-off point is smaller than for the time-invariant specifications. Thus, the time-varying specifications capture more properties of excess returns than the corresponding time-invariant models. Indeed, the number of omitted factors is smaller for the time-varying models than for the time-invariant cases. Moreover, the set of common instruments involving the monthly 30-day T-bill and the dividend yields seems to capture in a better way the characteristics of returns of individual stocks.

Table 1: Linear factor models

Model	Empirical factors	K
(1) CAPM	$r_{m,t}$	1
(2) FF model	$r_{m,t}, r_{smb,t}, r_{hml,t}$	3
(3) LIQ model	al_t, tl_t, il_t	3
(4) JW model	$r_{m,t}, lab_{t,}, upr_t$	3
(5) MOM and REV factors	$r_{mom,t}, r_{str,t}, r_{ltr,t}$	3
(6) Carhart (1997) model	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}$	4
(7) CRR model	$mp_t, ui_t, dei_t, uts_t, upr_t$	5
(8) FF and REV factors	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{str,t}, r_{ltr,t}$	5
(9) FF and JW factors	$r_{m,t}, r_{smb,t}, r_{hml,t}, lab_{t,}, upr_t$	5
(10) FF, MOM and REV factors	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}, r_{str,t}, r_{ltr,t}$	6
(11) FF and CRR factors	$r_{m,t}, r_{smb,t}, r_{hml,t}, mp_t, ui_t, dei_t$ uts_t, upr_t	8
(12) FF, CRR and JW factors	$r_{m,t}, r_{smb,t}, r_{hml,t}, mp_t, ui_t, dei_t,$ uts_t, upr_t, lab_t	9
(13) FF, MOM, REV, CRR and JW factors	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}, r_{str,t}, r_{ltr,t},$ $mp_t, ui_t, dei_t, uts_t, upr_t, lab_t$	12

The table lists the linear factor models that we estimate in order to compute the diagnostic criteria. For each model, we give the empirical factors which are involved. K is the number of observable factors. FF, CRR, MOM, REV, LIQ and JW refer to the three Fama-French factors, the five Chen-Roll-Ross macroeconomic factors, the momentum factor, the reversal factors, the three liquidity factors of Pastor and Stambaugh (2002), and the three Jagannathan and Wang (1996) factors, respectively.

Table 2: Trimmed cross-sectional dimensions n^χ and number of parameter to estimate d

Model	time-invariant spec.	time-varying spec.	(i)	(ii)
	n^χ	d	n^χ	n^χ
(1) CAPM	10,410	13	5,046	1,661
(2) FF model	10,410	21	4,476	1,476
(3) LIQ model	10,410	21	3,393	1,008
(4) JW model	7,578	21	-	-
(5) MOM and REV factors	10,410	21	4,568	1,471
(6) Carhart (1997) model	10,410	25	4,020	1,354
(7) CRR model	7,171	29	-	-
(8) FF and REV factors	10,396	29	3,828	1,076
(9) FF and JW factors	5,271	29	-	-
(10) FF, MOM and REV factors	7,461	33	3,217	960
(11) FF and CRR factors	6,786	41	-	-
(12) FF, CRR and JW factors	6,110	45	-	-
(13) FF, MOM, REV, CRR and JW factors	5,572	57	-	-

For each linear factor model, the table reports the trimmed cross-sectional dimension n^χ that comes from the estimation procedure. For the time-varying specifications, n^χ is given for the two sets of instruments (i) and (ii) described in Section 5.1. Moreover, the dimension of vector $x_{i,t}$, denoted by d, is also specified. For the time-invariant specifications, the number of regressors corresponds to the number of observable factors K (see Table 1).

Table 3: Summary statistics of $\hat{\rho}_i^2$ and $\hat{\rho}_{ad,i}^2$ for the time-invariant specifications

Model	1	2	3	4	5	6	7	8	9	10	11	12	13
							$\hat{\rho}_i^2$						
Min	0.0000	0.0013	0.0012	0.0009	0.0005	0.0050	0.0019	0.0025	0.0060	0.0090	0.0083	0.0130	0.0305
Quantile 25%	0.0948	0.1475	0.0429	0.1071	0.0521	0.1618	0.0338	0.1746	0.1534	0.1857	0.1730	0.1754	0.1972
Median	0.1872	0.2509	0.0889	0.1882	0.1107	0.2671	0.0652	0.2803	0.2454	0.2950	0.2596	0.2617	0.2823
Mean	0.2399	0.2948	0.1761	0.2198	0.1974	0.3111	0.1069	0.3239	0.2678	0.3383	0.2774	0.2782	0.2995
Quantile 75%	0.3172	0.3856	0.2110	0.2954	0.2576	0.4051	0.1274	0.4181	0.3525	0.4374	0.3623	0.3617	0.3827
Max	0.9828	0.9849	0.9863	0.9514	0.9868	0.9916	0.9535	0.9933	0.9574	0.9971	0.9582	0.9473	0.8934
Std	0.2003	0.2020	0.2105	0.1541	0.2143	0.2043	0.1232	0.2044	0.1561	0.2072	0.1420	0.1379	0.1389
							$\hat{\rho}_{ad,i}^2$						
Min	-0.0689	-0.2114	-0.2900	-0.1845	-0.2223	-0.3304	-0.1898	-0.3737	-0.3639	-0.5507	-0.1401	-0.1761	-0.2287
Quantile 25%	0.0845	0.1164	0.0188	0.0841	0.0282	0.1219	0.005	0.1253	0.1140	0.1274	0.1158	0.1203	0.1302
Median	0.1778	0.2220	0.0571	0.1664	0.0781	0.2276	0.0275	0.2319	0.2093	0.2379	0.2103	0.2115	0.2219
Mean	0.2285	0.2621	0.1388	0.1955	0.1614	0.2679	0.0627	0.2701	0.2298	0.2744	0.2234	0.2239	0.2332
Quantile 75%	0.3067	0.3549	0.1587	0.2735	0.2074	0.3618	0.0740	0.3652	0.3175	0.3723	0.3117	0.3106	0.3232
Max	0.9815	0.9808	0.9811	0.9417	0.9826	0.9878	0.9324	0.9884	0.9440	0.9937	0.9344	0.9176	0.8182
Std	0.2004	0.2026	0.2053	0.1534	0.2095	0.2049	0.1122	0.2044	0.1561	0.2071	0.1429	0.1383	0.1395

The table contains the descriptive statistics (cross-sectional minimum, 25% and 75% quantiles, median, mean, maximum and standard deviation) of the estimated coefficient of determination ($\hat{\rho}_i^2$), the estimated adjusted coefficients of determination ($\hat{\rho}_{ad,i}^2$) for the time-invariant linear factor models.

Table 4: Summary statistics of $IdiVol_i$ and $SysRisk_i$ for the time-invariant specifications

Model	1	2	3	4	5	6	7	8	9	10	11	12	13
							$IdiVol_i$						
Min	0.0109	0.0104	0.0103	0.0105	0.0106	0.0104	0.0085	0.0101	0.0102	0.0101	0.0081	0.0081	0.0260
Quantile 25%	0.1121	0.1069	0.1183	0.1046	0.1165	0.1057	0.1128	0.1042	0.0997	0.1029	0.0964	0.0947	0.0912
Median	0.1558	0.1488	0.1616	0.1432	0.1591	0.1466	0.1521	0.1453	0.1376	0.1432	0.1322	0.1283	0.1229
Mean	0.1781	0.1707	0.1829	0.1613	0.1797	0.1679	0.1684	0.1656	0.1555	0.1630	0.1482	0.1426	0.1366
Quantile 75%	0.2158	0.2072	0.2196	0.1951	0.2166	0.2037	0.2016	0.2015	0.1888	0.1986	0.1797	0.1731	0.1661
Max	3.0700	3.0206	3.2080	2.6677	2.8717	2.7810	3.2114	2.7925	2.3288	2.5306	2.0782	1.3228	1.3010
Std	0.1097	0.1045	0.1071	0.0904	0.1029	0.1013	0.0923	0.0996	0.0867	0.0965	0.0804	0.0711	0.0680
							$SysRisk_i$						
Min	0.001	0.0039	0.0022	0.0041	0.0022	0.0056	0.0045	0.0065	0.0059	0.0075	0.0083	0.0084	0.0100
Quantile 25%	0.0471	0.0572	0.0308	0.0469	0.0338	0.0589	0.0258	0.0608	0.0548	0.0623	0.0566	0.0560	0.0575
Median	0.0702	0.0825	0.0516	0.0660	0.0585	0.0854	0.0410	0.0880	0.0757	0.0904	0.0772	0.0758	0.0777
Mean	0.1030	0.1162	0.0904	0.0831	0.0970	0.1202	0.0570	0.1232	0.0924	0.1264	0.0883	0.0846	0.0857
Quantile 75%	0.1108	0.1268	0.0960	0.0949	0.1071	0.1312	0.0663	0.1359	0.1064	0.1396	0.1062	0.1028	0.1046
Max	3.4809	3.5239	3.3542	3.7981	3.6462	3.7159	3.3509	3.7073	4.0148	3.8908	4.1501	1.0911	1.1171
Std	0.1190	0.1223	0.1241	0.0812	0.1274	0.1249	0.0702	0.1257	0.0837	0.1283	0.0700	0.0454	0.0437

The table contains the descriptive statistics (cross-sectional minimum, 25% and 75% quantiles, median, mean, maximum and standard deviation) of the idiosyncratic risks ($IdiVol_i$), and the systematic risks ($SysRisk_i$) for the time-invariant linear factor models.

Table 5: Summary statistics of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$ and $SysRisk_i$ for the time-varying specifications (i)

Model	1	2	3	5	6	8	10
				$\hat{\rho}_i^2$			
Min	0.0218	0.0465	0.0242	0.0325	0.0514	0.0547	0.0559
Quantile 25%	0.1569	0.2294	0.1037	0.1414	0.2517	0.2673	0.2790
Median	0.2305	0.3112	0.1434	0.2024	0.3331	0.3471	0.3598
Mean	0.2440	0.3225	0.1612	0.2316	0.3435	0.3566	0.3680
Quantile 75%	0.3191	0.4069	0.1984	0.2938	0.4261	0.4370	0.4468
Max	0.7052	0.8512	0.7069	0.8822	0.8628	0.9032	0.9051
Std	0.1147	0.1265	0.0832	0.1228	0.1272	0.1229	0.1221
				$\hat{\rho}_{ad,i}^2$			
Min	-0.1268	-0.2463	-0.1287	-0.2671	-0.3089	-0.2981	-0.3427
Quantile 25%	0.0884	0.1306	0.0214	0.0532	0.1435	0.1461	0.1519
Median	0.1666	0.2240	0.0563	0.1052	0.2358	0.2429	0.2497
Mean	0.1794	0.2296	0.0656	0.1270	0.2404	0.2433	0.2479
Quantile 75%	0.2575	0.3223	0.0983	0.1764	0.3335	0.3374	0.3423
Max	0.6764	0.7835	0.5897	0.8287	0.7804	0.8279	0.8101
Std	0.1201	0.1362	0.0708	0.1132	0.1380	0.1365	0.1369
				$IdiVol_i$			
Min	0.0358	0.0315	0.0387	0.0334	0.0312	0.0322	0.0311
Quantile 25%	0.0948	0.0868	0.0980	0.0965	0.0847	0.0838	0.0809
Median	0.1283	0.1171	0.1338	0.1274	0.1134	0.1125	0.1083
Mean	0.1421	0.1310	0.1460	0.1393	0.1268	0.1255	0.1209
Quantile 75%	0.1730	0.1603	0.1774	0.1679	0.1545	0.1537	0.1472
Max	0.7487	0.6984	0.7015	0.6770	0.6842	0.6529	0.6236
Std	0.0683	0.0630	0.0682	0.0627	0.0611	0.0601	0.0574
				$SysRisk_i$			
Min	0.0091	0.0131	0.0102	0.0106	0.0155	0.0150	0.0185
Quantile 25%	0.0523	0.0597	0.0387	0.0463	0.0609	0.0623	0.0617
Median	0.0698	0.0805	0.0555	0.0667	0.0820	0.0834	0.0823
Mean	0.0766	0.0874	0.0638	0.0765	0.0894	0.0917	0.0909
Quantile 75%	0.0931	0.1068	0.0803	0.0954	0.1096	0.1115	0.1096
Max	0.4210	0.6166	0.5218	0.6278	0.6208	0.6352	0.6359
Std	0.0360	0.0407	0.0379	0.0442	0.0422	0.0438	0.0442

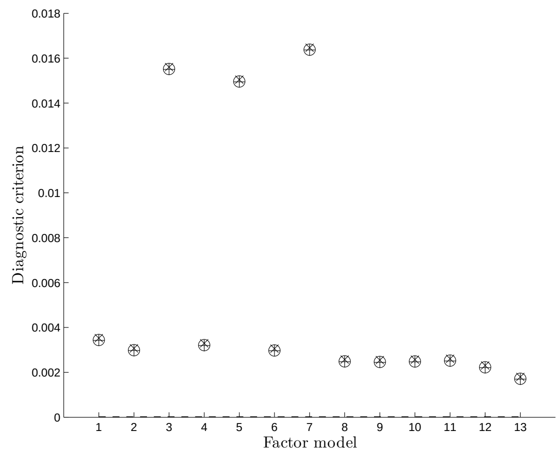
The table contains the descriptive statistics (cross-sectional minimum, 25% and 75% quantiles, median, mean, maximum and standard deviation) of the estimated coefficient of determination ($\hat{\rho}_i^2$), the estimated adjusted coefficients of determination ($\hat{\rho}_{ad,i}^2$), the idiosyncratic risks ($IdiVol_i$), and the systematic risks ($SysRisk_i$) for the time-varying linear factor models estimated by using the term spread and the default spread as common instruments.

Table 6: Summary statistics of $\hat{\rho}_i^2$, $\hat{\rho}_{ad,i}^2$, $IdiVol_i$ and $SysRisk_i$ for the time-varying specifications (ii)

Model	1	2	3	5	6	8	10
				$\hat{\rho}_i^2$			
Min	0.0210	0.0490	0.0200	0.0306	0.0562	0.0688	0.0730
Quantile 25%	0.1368	0.2034	0.0827	0.1096	0.2240	0.2366	0.2508
Median	0.2012	0.2795	0.1128	0.1545	0.3044	0.3013	0.3162
Mean	0.2200	0.3010	0.1349	0.1908	0.3237	0.3100	0.3225
Quantile 75%	0.2893	0.3778	0.1638	0.2423	0.4034	0.3761	0.3876
Max	0.7134	0.7780	0.5545	0.7180	0.7885	0.8491	0.7868
Std	0.1094	0.1306	0.0783	0.1157	0.1336	0.1070	0.1049
				$\hat{\rho}_{ad,i}^2$			
Min	-0.0934	-0.1017	-0.1479	-0.2031	-0.1336	-0.1400	-0.1738
Quantile 25%	0.0765	0.1174	0.0170	0.0425	0.1292	0.1263	0.1313
Median	0.1487	0.2095	0.0472	0.0784	0.2190	0.2111	0.2188
Mean	0.1633	0.2176	0.0544	0.0994	0.2284	0.2142	0.2199
Quantile 75%	0.2396	0.3080	0.0811	0.1333	0.3263	0.3016	0.3042
Max	0.6497	0.7218	0.3925	0.5954	0.7208	0.7573	0.6318
Std	0.1135	0.1333	0.0592	0.0949	0.1358	0.1221	0.1239
				$IdiVol_i$			
Min	0.0377	0.0347	0.0385	0.0313	0.0316	0.0311	0.0289
Quantile 25%	0.0878	0.0808	0.0885	0.0899	0.0787	0.0768	0.0749
Median	0.1200	0.1098	0.1219	0.1180	0.1073	0.1091	0.1048
Mean	0.1357	0.1260	0.1379	0.1333	0.1233	0.1247	0.1222
Quantile 75%	0.1651	0.1535	0.1680	0.1612	0.1506	0.1552	0.1537
Max	0.7620	0.7186	0.7430	0.6825	0.7141	0.6540	0.6430
Std	0.0718	0.0659	0.0713	0.0645	0.0652	0.0670	0.0660
				$SysRisk_i$			
Min	0.0111	0.0164	0.0088	0.0101	0.0167	0.0188	0.0190
Quantile 25%	0.0468	0.0541	0.0310	0.0365	0.0557	0.0544	0.0550
Median	0.0618	0.0727	0.0454	0.0565	0.0756	0.0728	0.0733
Mean	0.0666	0.0786	0.0548	0.0640	0.0817	0.0809	0.0817
Quantile 75%	0.0813	0.0964	0.0707	0.0805	0.1002	0.0989	0.0999
Max	0.3502	0.3626	0.3189	0.4540	0.3713	0.5322	0.5439
Std	0.0306	0.0361	0.0358	0.0386	0.0386	0.0419	0.0421

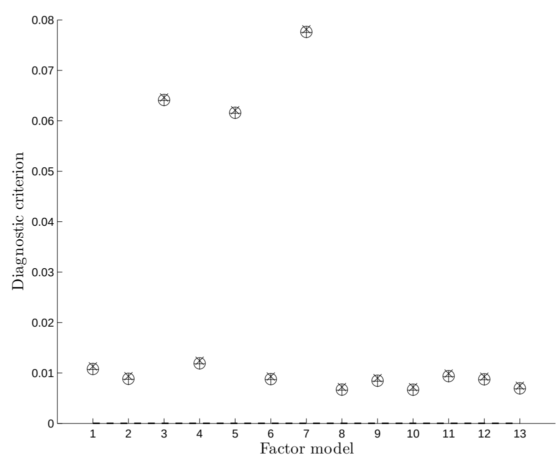
The table contains the descriptive statistics (cross-sectional minimum, 25% and 75% quantiles, median, mean, maximum and standard deviation) of the estimated coefficient of determination ($\hat{\rho}_i^2$), the estimated adjusted coefficients of determination ($\hat{\rho}_{ad,i}^2$), the idiosyncratic risks ($IdiVol_i$), and the systematic risks ($SysRisk_i$) for the time-varying linear factor models estimated by using the monthly 30-day T-bill and the dividend yields as common instruments.

Figure 1: **Values of the diagnostic criteria ξ_1, ξ_2 and ξ_3 for the time-invariant models**



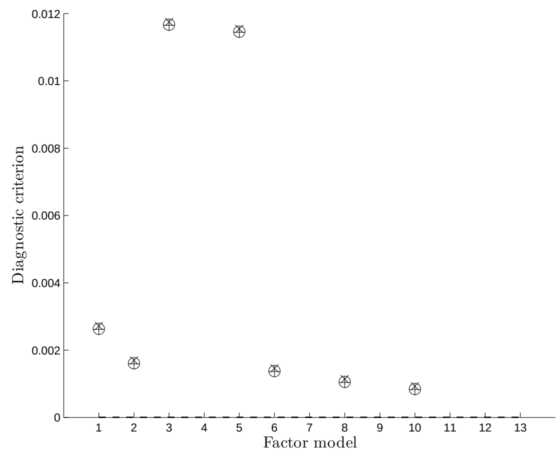
The figure plots the values of the diagnostic criteria ξ_1(circle), ξ_2 (plus sign) and ξ_3 (cross) for the time-invariant specifications. We also report the zero axis (dashed horizontal line).

Figure 2: **Estimated values of the diagnostic criteria $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ for the time-invariant models**



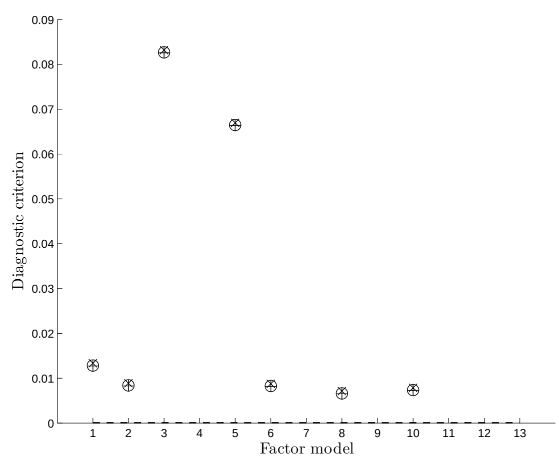
The figure plots the values of the logarithmic diagnostic criteria $\check{\xi}_1$(circle), $\check{\xi}_2$ (plus sign) and $\check{\xi}_3$ (cross) for the time-invariant specifications. We also report the zero axis (dashed horizontal line).

Figure 3: Values of the diagnostic criteria ξ_1, ξ_2 and ξ_3 for the time-varying models (i)



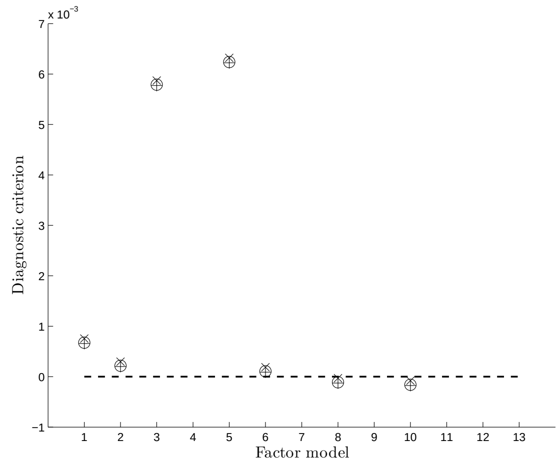
The figure plots the values of the diagnostic criteria ξ_1(circle), ξ_2 (plus sign) and ξ_3 (cross) for the time-varying specifications when Z_t^* includes default and term spreads. The diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (dashed horizontal line).

Figure 4: Values of the diagnostic criteria $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ for the time-varying models (i)



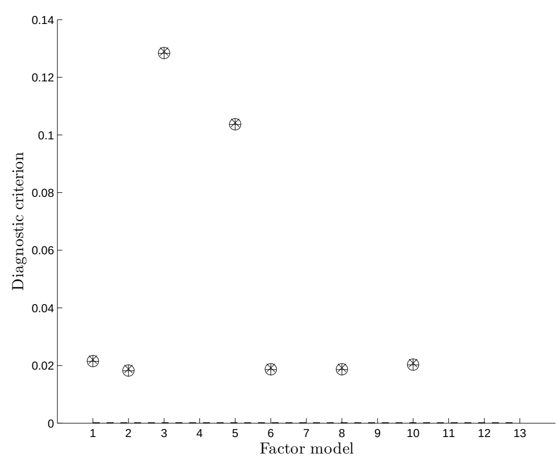
The figure plots the values of the logarithmic diagnostic criteria $\check{\xi}_1$(circle), $\check{\xi}_2$ (plus sign) and $\check{\xi}_3$ (cross) for the time-varying specifications when Z_t^* includes default and term spreads. The logarithmic diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (dashed horizontal line).

Figure 5: Values of the diagnostic criteria ξ_1, ξ_2 and ξ_3 for the time-varying models (ii)



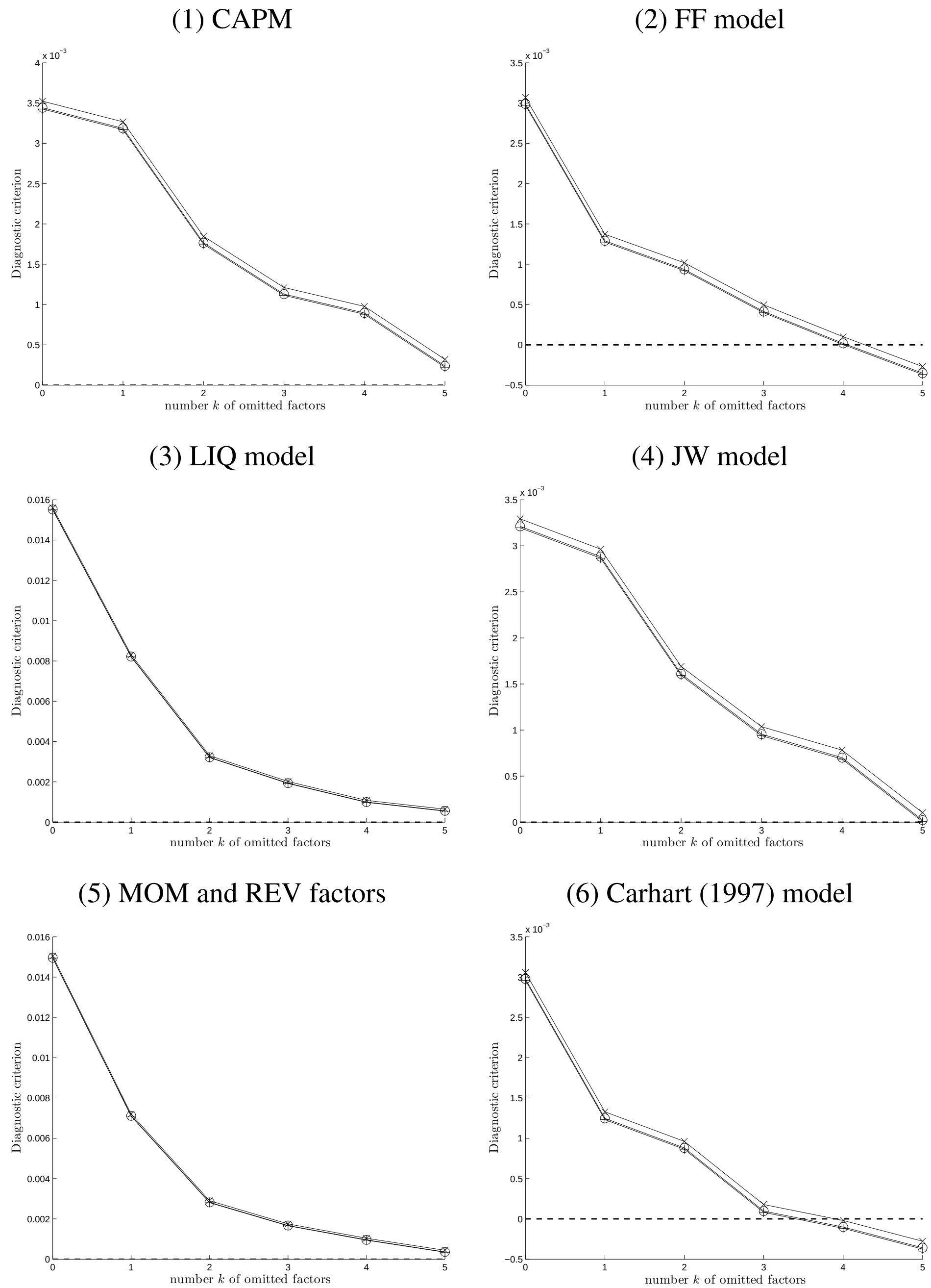
The figure plots the values of the diagnostic criteria ξ_1(circle), ξ_2 (plus sign) and ξ_3 (cross) for the time-varying specifications when Z_t^* includes one-month T-Bill and dividend yield. The diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (dashed horizontal line).

Figure 6: Values of the diagnostic criteria $\check{\xi}_1, \check{\xi}_2$ and $\check{\xi}_3$ for the time-varying models (ii)



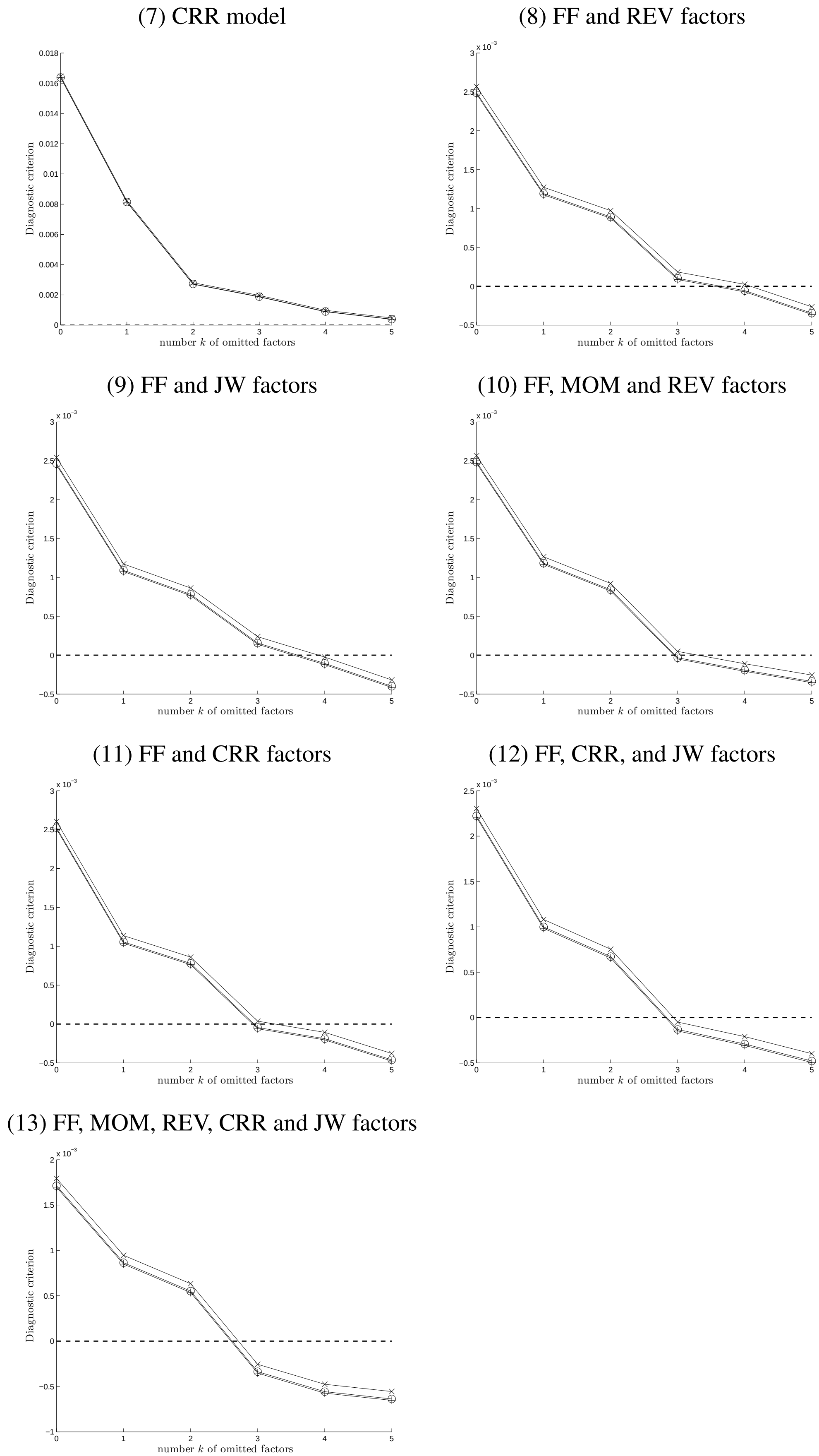
The figure plots the values of the logarithmic diagnostic criteria $\check{\xi}_1$(circle), $\check{\xi}_2$ (plus sign) and $\check{\xi}_3$ (cross) for the time-varying specifications when Z_t^* includes one-month T-Bill and dividend yield. The logarithmic diagnostic criteria cannot be computed for the JW, CRR, (9), (11)-(13) models due to the multicollinearity problems. We also report the zero axis (dashed horizontal line).

Figure 7: Values of criteria $\xi(k)$ for the time-invariant models



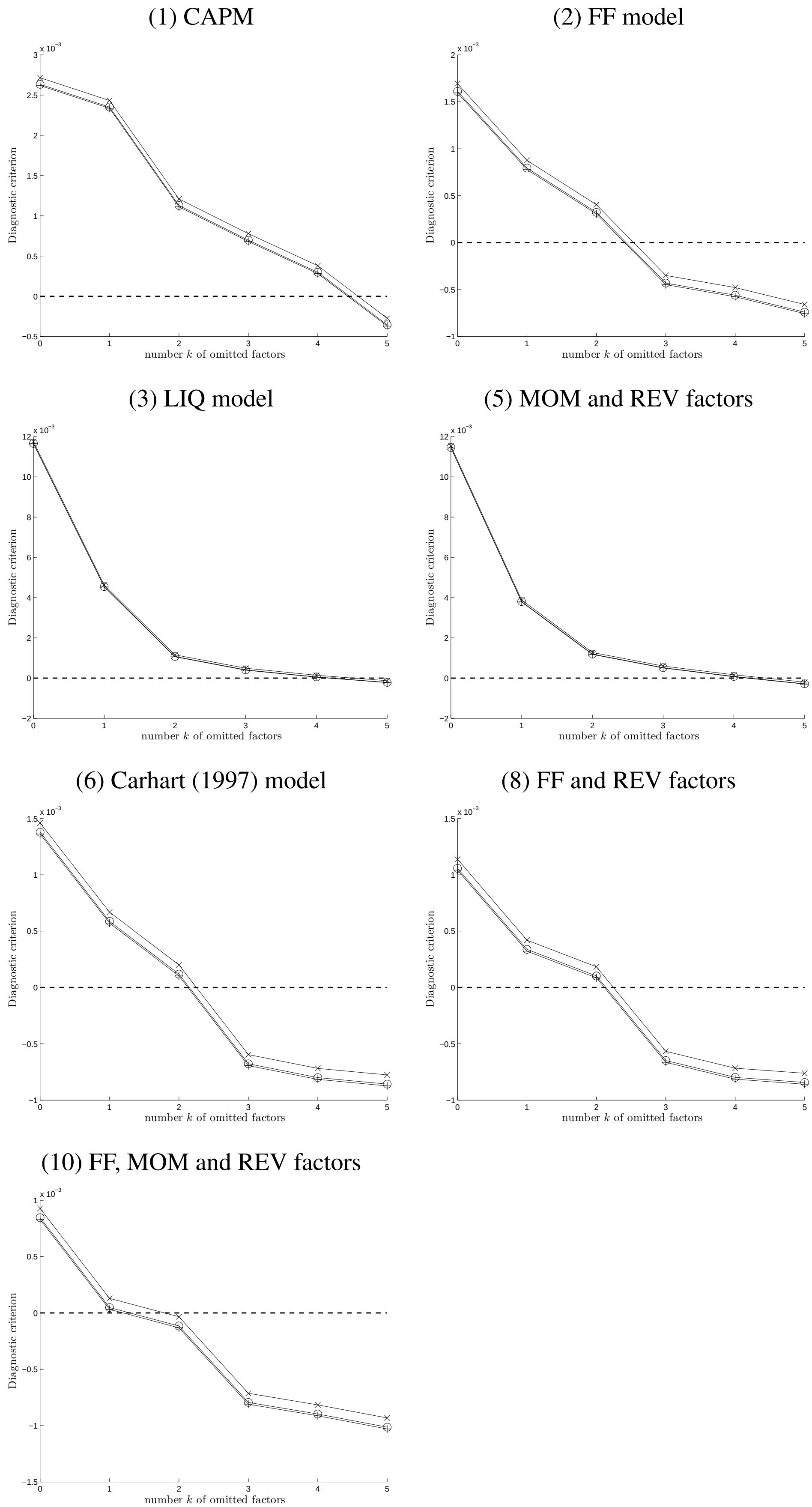
The figure plots the values of the diagnostic criteria $\xi_1(k)$(circle), $\xi_2(k)$ (plus sign) and $\xi_3(k)$ (cross) with $k = 0, 1, ..., 5$, for the time-invariant specifications (1)-(6). We also report the zero axis (dashed horizontal line).

Figure 8: Values of criteria $\xi(k)$ for the time-invariant models



The figure plots the values of the diagnostic criteria $\xi_1(k)$(circle), $\xi_2(k)$ (plus sign) and $\xi_3(k)$ (cross) with $k = 0, 1, ..., 5$, for the time-invariant specifications (7)-(13). We also report the zero axis (dashed horizontal line).

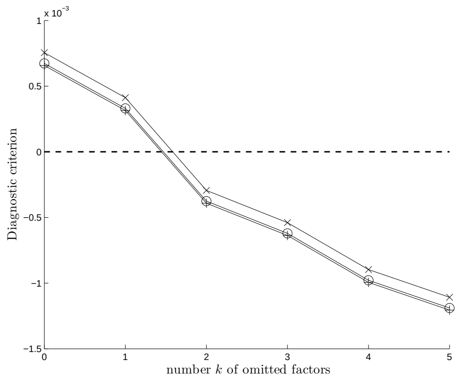
Figure 9: Values of criteria $\xi\,(k)$ for the time-varying models (i)



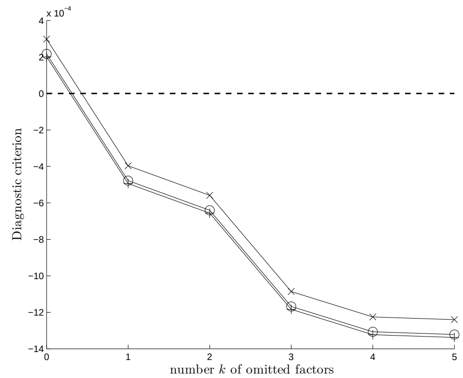
The figure plots the values of the diagnostic criteria $\xi_1\,(k)$(circle), $\xi_2\,(k)$ (plus sign) and $\xi_3\,(k)$ (cross) with $k = 0, 1, ..., 5$, for the time-varying specifications when Z_t^* includes default and term spreads. We also report the zero axis (dashed horizontal line).

Figure 10: Values of criteria $\xi(k)$ for the time-varying models (ii)

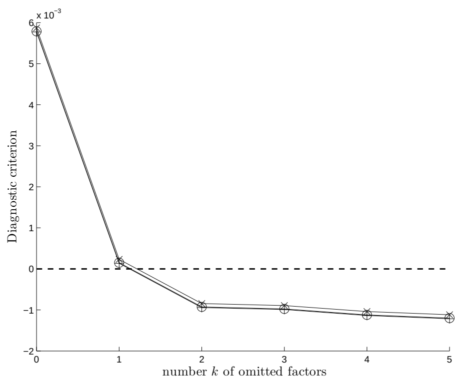
(1) CAPM



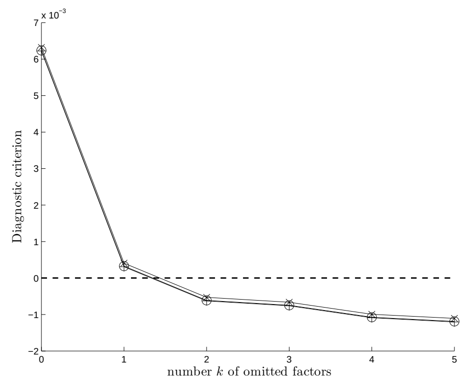
(2) FF model



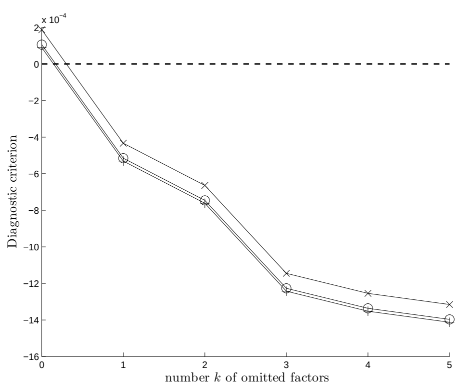
(3) LIQ model



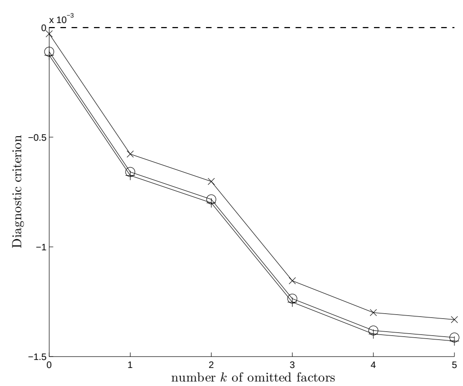
(5) MOM and REV factors



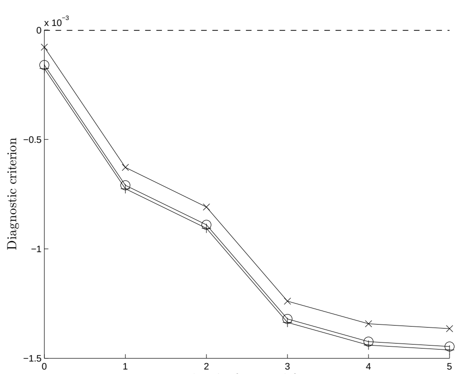
(6) Carhart (1997) model



(8) FF and REV factors



(10) FF, MOM and REV factors



The figure plots the values of the diagnostic criteria $\xi_1(k)$(circle), $\xi_2(k)$ (plus sign) and $\xi_3(k)$ (cross) with $k = 0, 1, ..., 5$, for the time-varying specifications when Z_t^* includes one-month T-Bill and dividend yield. We also report the zero axis (dashed horizontal line).

References

S. C. Ahn and A. R. Horenstein. Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3): 1203–1227, 2013.

S. C. Ahn, A. R. Horenstein, and N. Wang. Beta matrix and common factors in stock returns. *Working paper*, 2013.

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. Decomposition and characterization of risk with a continuum of random variables: Corrigendum. *Econometrica*, 67(4):919–920, 1999a.

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

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

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

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

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.

Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, 21(3):1275–1294, 1993.

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

D.S. Bernstein. *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2009.

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

J. Y. Campbell. Understanding risk and return. *NBER Working Papers, 4554*, 1993.

M. Caner and X. Han. Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimator. *Journal of Business and Economics Statistics*, forthcoming, 2014.

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

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

N. Chen, R. Roll, and S. Ross. Economic forces and the stock market. *Journal of Business*, 59:563–594, 1986.

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. Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108, 2011.

G. Connor and R. A. Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3):373–394, 1986.

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. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21:255–289, 1988.

G. Connor and R. A. Korajczyk. A test for the number of factors in an approximate factor model. *Journal of Finance*, 48(4):1263–1291, 1993.

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. Dissecting anomalies. *Journal of Finance*, 63(4):1653–1678, 2008.

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.

R. Frisch and V. Frederick. Partial time regressions as compared with individual trends. *Econometrica*, 1 (4):387–401, 1933.

P. Gagliardini, E. Ossola, and O. Scaillet. Time-varying risk premium in large cross-sectional equity datasets. *Working paper*, 2011.

S. Geman. A limit theorem for the norm of random matrices. *The Annals of Probability*, 8(2):252–261, 1980.

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.

L. Gobillon and T. Magnac. Regional policy evaluation: Interactive fixed effects and synthetic controls. *Working Paper TSE*, 2014.

N. Gospodinov, R. Kan, and C. Robotti. Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies*, 27(7):2139–2170, 2014.

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

M. Hallin and R. Liška. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617, 2007.

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.

C. R. Harvey, Y. Liu, and H. Zhu. ... and the cross-section of expected returns. *Working Paper*, 2013.

R. Horn and C. Johnson. *Matrix analysis*. Cambridge University Press, 1985.

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.

D. Jonsonn. Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12:1–38, 1982.

R. Kan and C. Zhang. Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, 54 (1):203–235, 1999a.

R. Kan and C. Zhang. GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics*, 54(1):103–127, 1999b.

G. Kapetanios. A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business and Economic Statistics*, 28(3):397–409, 2010.

F. Kleibergen. Test of risk premia in linear factor models. *Journal of Econometrics*, 149(2):149–173, 2009.

B. N. Lehmann and D. M. Modest. The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, 21:213–254, 1988.

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.

M. Lovell. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 304(304):993–1010, 1963.

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

V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik a:4*, pages 457–483, 1967.

C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Siam, 2000.

H. Moon and M. Weidner. Linear regression for panel with unknown. *Econometrica*, forthcoming, 2015.

A. Onatski. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5): 1447–1479, 2009.

A. Onatski. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4):1004–1016, 2010.

A. Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2013.

L. Pastor and R.F. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63:315–349, 2002.

D. Pena and P. Poncela. Dimension reduction in multivariate time series advances in distribution theory, order statistics, and inference. In N. Balakrishnan, E. Castillo, and J. M. Sarabia, editors, *Advances in distribution theory, order statistics, and inference*, pages 433–458. Birkhauser, 2006.

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

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

B. Rennie and A. Dobson. On Stirling numbers of the second kind. *Journal of Combinatorial Theory*, 7: 116–121, 1969.

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

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

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

J. von Neumann. Some matrix inequalieties and metrization of matrix space. *Tomsk Universitet Review*, 1: 286–300, 1937.

Y.Q. Yin, Z.D. Bai, and P.R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory*, 78:509–521, 1988.

Appendix 1 Regularity conditions

In this appendix, we list and comment additional assumptions used in the proofs in Appendix 2. The error terms $(\varepsilon_{i,t})$ are $\varepsilon_{i,t} = u_{i,t}$ under model $\mathcal{M}_1$, and $\varepsilon_{i,t} = \theta_i' h_t + u_{i,t}$ under model $\mathcal{M}_2$ (see Equation (6)). Since models $\mathcal{M}_1(k)$ and $\mathcal{M}_2(k)$ are subsets of model $\mathcal{M}_2$, the assumptions stated for $\mathcal{M}_2$ also hold for $\mathcal{M}_1(k)$ and $\mathcal{M}_2(k)$, for any $k \geq 1$. We use M as a generic constant in the assumptions.

Assumption A.1 *For a constant $M > 0$ and for all $n, T \in \mathbb{N}$, we have:*

$$\frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}\left|E\left[u_{i,t_1}u_{i,t_2}u_{j,t_3}u_{j,t_4}\left|x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j\right.\right]\right| \leq M.$$

Assumption A.2 *We have $E[|u_{i,t}|^q] \leq M$, for all i, t, and some constants $q \geq 8$ and $M > 0$.*

Assumption A.3 *Let $\delta = \delta_n \uparrow \infty$ be a diverging sequence such that $\sqrt{T}/\delta^{q-1} = o(1)$ and $\delta \geq n^{\beta}$, for $\beta > 2/q$. Let $e_{i,t} = u_{i,t}1\{|u_{i,t}| \leq \delta\} - E[u_{i,t}1\{|u_{i,t}| \leq \delta\}]$. Then:*

$$\frac{1}{n^k}\sum_{i_1,\dots,i_k}\sum_{t_1,\dots,t_k}|E[e_{i_1,t_k}e_{i_1,t_1}e_{i_2,t_1}e_{i_2,t_2}e_{i_3,t_2}\cdots e_{i_{k-1},t_{k-1}}e_{i_k,t_{k-1}}e_{i_k,t_k}]| \leq M^k,$$

for a sequence of integers $k = k_n \uparrow \infty$ and a constant $M > 0$, where indices $i_1, ..., i_k$ run from 1 to n, and indices $t_1, ..., t_k$ from 1 to T.

Assumption A.4 *There exists a constant $M > 0$ such that $\|x_{i,t}\| \leq M$, P-a.s., for any i and t.*

Assumption A.5 *Under model $\mathcal{M}_2$, a) there exists a constant $M > 0$ such that $\|h_t\| \leq M$, P-a.s., for all t. Moreover, b) $\|\theta_i\| < M$, for all i.*

Assumption A.6 *Under model $\mathcal{M}_2$, for a constant $M > 0$ and for all $n, T \in \mathbb{N}$, we have:*

$$\frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}\|E[(x_{i,t_1}h_{t_1}')(x_{i,t_2}h_{t_2}')'(x_{j,t_3}h_{t_3}')(x_{j,t_4}h_{t_4}')'|\gamma_i,\gamma_j]\| \leq M.$$

Assumption A.7 *Variables $(I_{i,t})$ and $(\varepsilon_{i,t})$ are independent.*

Assumption A.8 *Under model $\mathcal{M}_2(k)$, for any $l = 1, ..., k$ we have $\mu_1(W^{(l)}) = O_p(C_{n,T}^{-2})$, where $W^{(l)} = [w_{t,s}^{(l)}]$ is the $T \times T$ matrix with elements $w_{t,s}^{(l)} = \dfrac{1}{nT}\sum_i (I_{i,t} - \tau_{i,T}^{-1})(I_{i,s} - \tau_{i,T}^{-1})\theta_{i,l}^2$.*

Assumption A.9 *The trimming constants $\chi_{1,T}$ and $\chi_{2,T}$ are such that $\chi_{1,T}^2\chi_{2,T} = o\left(Tg\left(n, T\right)\right)$.*

Appendix 2 Proofs

We start by listing several results known from matrix theory. They are used several times in the proofs.

(i) Weyl inequality: The singular-value version states that if A and B are $T \times n$ matrices, then $\mu_{i+j-1}[(A+B)(A+B)']^{1/2} \leq \mu_i(AA')^{1/2} + \mu_j(BB')^{1/2}$, for any $1 \leq i, j \leq \min\{n, T\}$ such that $1 \leq i+j \leq \min\{n, T\}+1$ (see Theorem 3.3.16 of Horn and Johnson (1985)). The Weyl inequality for $i = k+1$ and $j = 1$ yields:

$$\mu_{k+1}[(A+B)(A+B)']^{1/2} \leq \mu_{k+1}(AA')^{1/2} + \mu_1(BB')^{1/2}, \tag{11}$$

$$\mu_{k+1}[(A+B)(A+B)']^{1/2} \geq \mu_{k+1}(AA')^{1/2} - \mu_1(BB')^{1/2}, \tag{12}$$

for any $T \times n$ matrices A and B and integer k such that $0 \leq k \leq \min\{n, T\}-1$. With $k = 0$, $\mu_1[(A+B)(A+B)']^{1/2} \leq \mu_1(AA')^{1/2} + \mu_1(BB')^{1/2}$, and $\mu_1[(A+B)(A+B)']^{1/2} \geq \mu_1(AA')^{1/2} - \mu_1(BB')^{1/2}$. We also use Weyl inequality for eigenvalues: for any $T \times T$ symmetric matrices A and B we have: $\mu_{i+j-1}(A+B) \leq \mu_i(A) + \mu_j(B)$, for any $1 \leq i, j \leq T$ such that $i+j \leq T+1$ (see Theorem 8.4.11 in Bernstein (2009)).

(ii) Equality between largest eigenvalue and operator norm: The largest eigenvalue $\mu_1(A)$ of a symmetric positive semi-definite matrix A is equal to its operator norm $\|A\|_{op} = \max\limits_{x:\|x\|=1} \|Ax\|$. Besides, $\|A\|_{op} \leq \|A\|$ for any square matrix A, where $\|\cdot\|$ is the Frobenius norm (see e.g. Meyer (2000)).

(iii) Inequalities fot the eigenvalues of matrix products: if A and B are $m \times m$ positive semidefinite and positive deinite matirces, respectively,

$$\lambda_k(A)\,\lambda_m(B) \leq \lambda_k(AB) \leq \lambda_k(A)\,\lambda_1(B), \tag{13}$$

for $k = 1, 2, ..., m$ (see Fact 8.19.17 in Bernstein (2009)).

(iv) Courant-Fischer min-max Theorem: If A is a $T \times T$ symmetric matrix, for $k = 1, ..., T$,

$$\mu_k(A) = \min_{\mathcal{G}:dim(\mathcal{G})=T-k+1} \max_{x\in\mathcal{G}:\|x\|=1} x'Ax, \tag{14}$$

where the minimization is w.r.t. the $(T-k+1)$-dimensional linear subspace $\mathcal{G}$ of $\mathbb{R}^T$ (see e.g. Bernstein (2009)). Tha max-min formulation states:

$$\mu_k(A) = \max_{\mathcal{G}:dim(\mathcal{G})=T-k+1} \min_{x\in\mathcal{G}:\|x\|=1} x'Ax, \tag{15}$$

where the minimization is w.r.t. the k-dimensional linear subspace $\mathcal{G}$ of $\mathbb{R}^T$.

(v) Courant-Fischer formula: If A is a $T \times T$ symmetric matrix, for $k = 1, ..., T$,

$$\mu_k(A) = \max_{x \in \mathcal{F}_{k-1}^{\perp} : \|x\|=1} x' A x, \tag{16}$$

where $\mathcal{F}_k^{\perp}$ is the orthogonal complement of $\mathcal{F}_k$ with $\mathcal{F}_k$ being the linear space spanned by the eigenvectors associated to the k largest eigenvalues of matrix A.

A.2.1 Proof of Proposition 1

a) The OLS estimator of β_i in matrix notation is $\hat{\beta}_i = \left(\tilde{X}_i' \tilde{X}_i\right)^{-1} \tilde{X}_i' \tilde{R}_i$, with $\tilde{R}_i = \boldsymbol{I}_i \odot R_i$, where $\boldsymbol{I}_i$ is the $T \times 1$ vector of indicators $I_{i,t}$ for asset i, and $\odot$ is the Hadamard product. We get the vector of residuals $\hat{\varepsilon}_i = R_i - X_i \left(\tilde{X}_i' \tilde{X}_i\right)^{-1} \tilde{X}_i' \tilde{R}_i$. Then, we have $\bar{\varepsilon}_i = \boldsymbol{I}_i \odot \hat{\varepsilon}_i = M_{\tilde{X}_i} \tilde{R}_i = M_{\tilde{X}_i} \tilde{\varepsilon}_i$, where $\tilde{\varepsilon}_i = \boldsymbol{I}_i \odot \varepsilon_i$ and $M_{\tilde{X}_i} = I_T - P_{\tilde{X}_i}$, with $P_{\tilde{X}_i} = \tilde{X}_i \left(\tilde{X}_i' \tilde{X}_i\right)^{-1} \tilde{X}_i'$. Thus, under $\mathcal{M}_1$, we have the decomposition $\mathbf{1}_i^{\chi} \bar{\varepsilon}_i = \tilde{\varepsilon}_i - (1 - \mathbf{1}_i^{\chi}) \tilde{\varepsilon}_i - \mathbf{1}_i^{\chi} P_{\tilde{X}_i} \tilde{\varepsilon}_i$. From Weyl inequality (11) with $k = 0$, and the inequality between matrix norms, we get:

$$\mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right)^{1/2} \leq \mu_1 \left(\frac{1}{nT} \sum_i \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \right)^{1/2} + I_1^{1/2} + I_2^{1/2}, \tag{17}$$

where:

$$I_1 := \| \frac{1}{nT} \sum_i (1 - \mathbf{1}_i^{\chi}) \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \|, \qquad I_2 := \| \frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} P_{\tilde{X}_i} \tilde{\varepsilon}_i \tilde{\varepsilon}_i' P_{\tilde{X}_i} \|. \tag{18}$$

We bound the largest eigenvalue of matrix $\frac{1}{nT} \sum_i \tilde{\varepsilon}_i \tilde{\varepsilon}_i'$ and the remainder terms I_1 and I_2 in the next two lemmas.

Lemma 1 *Under model $\mathcal{M}_1$ and Assumptions A.2, A.3, A.7, as $n, T \to \infty$ such that $T/n = o(1)$, we have $\mu_1 \left(\frac{1}{nT} \sum_i \tilde{\varepsilon}_i \tilde{\varepsilon}_i' \right) = O_p(C_{n,T}^{-2})$.*

Lemma 2 *Under model $\mathcal{M}_1$ and Assumptions A.1, A.4, as $n, T \to \infty$ such that $T/n = o(1)$, we have: (i) $I_1 = O_p(T^{-\bar{b}})$, for any $\bar{b} > 0$; (ii) $I_2 = O_p(\chi_{1,T}^4 \chi_{2,T}^2 / T)$.*

From Inequality (17) and Lemmas 1 and 2, we get $\xi = O_p(C_{n,T}^{-2}) + O_p(\frac{\chi_{1,T}^4 \chi_{2,T}^2}{T}) - g(n,T)$. Then, from Assumption A.9 on the trimming constants and the properties of penalty function $g(n,T)$, Proposition 1(a) follows.

b) Let us now consider the case $\mathcal{M}_2$. We have $\bar{\varepsilon}_i = M_{\tilde{X}_i}\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i = \tilde{H}_i\theta_i + \tilde{u}_i$, where $\tilde{H}_i = \mathbf{I}_i \odot H$ and H is the $T \times m$ matrix of latent factor values, with $m \geq 1$. Hence, we have the decomposition $\mathbf{1}_i^\chi \bar{\varepsilon}_i = \tilde{H}_i\theta_i + \tilde{u}_i - (1 - \mathbf{1}_i^\chi)\tilde{\varepsilon}_i - \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{H}_i\theta_i - \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{u}_i$. By using Weyl inequalities (11) and (12) with $k = 0$, and the inequality between matrix norms, we get:

$$\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i\bar{\varepsilon}_i'\right)^{1/2} \geq \mu_1\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right)^{1/2} - \mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right)^{1/2} - I^{1/2}, \qquad (19)$$

where $I^{1/2} = I_1^{1/2} + I_3^{1/2} + I_4^{1/2}$, term I_1 is defined as in (18), and

$$I_3^{1/2} := \|\frac{1}{nT}\sum_i \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i'P_{\tilde{X}_i}\|^{1/2}, \qquad I_4^{1/2} := \|\frac{1}{nT}\sum_i \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{u}_i\tilde{u}_i'P_{\tilde{X}_i}\|^{1/2}. \qquad (20)$$

By Lemma 1 applied on $\tilde{u}_i$ instead of $\tilde{\varepsilon}_i$, we have $\mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right) = O_p(C_{n,T}^{-2})$. Moreover, from the next Lemma 3 and Assumption A.9 on the trimming constants, we get $I = O_p(C_{n,T}^{-2})$ under $\mathcal{M}_2$.

Lemma 3 *Under model $\mathcal{M}_2$ and Assumptions A.1, A.4, A.6, as $n, T \to \infty$ such that $T/n = o(1)$, we have: (i) $I_1 = O_p(T^{-\bar{b}})$, for any $\bar{b} > 0$; (ii) $I_3 = O_p(\chi_{1,T}^4\chi_{2,T}^2/T)$; (iii) $I_4 = O_p(\chi_{1,T}^4\chi_{2,T}^2/T)$.*

The next Lemma 4 provides a lower bound for the first term in the r.h.s. of Inequality (19).

Lemma 4 *Under model $\mathcal{M}_2$ and Assumptions 1, 2 and ..., we have $\mu_1\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) \geq C$, w.p.a. 1, for a constant $C > 0$.*

Then, from Inequality (19) and Lemma 4, we get $\xi \geq C/2$, w.p.a. 1, and Proposition 1(b) follows.

A.2.2 Proof of Proposition 2

We prove Proposition 2 along similar lines as Proposition 1 by exploiting the Weyl inequalities (11) and (12) for a generic k.

a) Let us first consider the case $\mathcal{M}_1(k)$. We have $\bar{\varepsilon}_i = M_{\tilde{X}_i}\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i = \tilde{H}_i\theta_i + \tilde{u}_i$, where $\tilde{H}_i = \mathbf{I}_i \odot H$ and H is the $T \times k$ matrix of latent factor values. Then, $\mathbf{1}_i^\chi \bar{\varepsilon}_i = \tilde{H}_i\theta_i + \tilde{u}_i - (1-\mathbf{1}_i^\chi)\tilde{\varepsilon}_i - \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{H}_i\theta_i - \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{u}_i$. From Weyl inequalities (11) and (12), and the inequality between matrix norms, we get:

$$\mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i\bar{\varepsilon}_i'\right)^{1/2} \leq \mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right)^{1/2} + \mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right)^{1/2} + I^{1/2}, \qquad (21)$$

where $I^{1/2} = I_1^{1/2} + I_3^{1/2} + I_4^{1/2}$ and terms I_1, I_3 and I_4 are defined as in the proof of Proposition 1. Since model $\mathcal{M}_1(k)$ is included in model $\mathcal{M}_2$ for any $k \geq 1$, we get $I = O_p(C_{n,T}^{-2})$, from Lemma 3 and Assumption A.9 on the trimming constants. Moreover, $\mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right) = O_p(C_{n,T}^{-2})$ by Lemma 1 with $\tilde{u}_i$ replacing $\tilde{\varepsilon}_i$. The first term in the r.h.s. of (21) is bounded by the next lemma.

Lemma 5 *Under model $\mathcal{M}_1(k)$ and Assumptions A.5 a), A.8, we have* $\mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) = O_p(C_{n,T}^{-2})$.

From Inequality (21) and Lemma 5, we get $\xi = O_p(C_{n,T}^{-2}) - g(n,T)$. Then, by the properties of $g(n,T)$, Proposition 2*a*) follows.

b)Let us now consider the case $\mathcal{M}_2(k)$. We have $\bar{\varepsilon}_i = M_{\tilde{X}_i}\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i = \tilde{H}_i\theta_i + \tilde{u}_i$, where $\tilde{H}_i = \mathbf{I}_i \odot H$ and H is the $T \times m$ matrix of latent factor values, with $m \geq k+1$. By similar arguments as in part **a)**, using Weyl inequalities (11) and (12), and the inequality between matrix norms, we get:

$$\mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i\bar{\varepsilon}_i'\right)^{1/2} \geq \mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right)^{1/2} - \mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right)^{1/2} - I^{1/2}. \qquad (22)$$

As in part **a)** we have $\mu_1\left(\frac{1}{nT}\sum_i \tilde{u}_i\tilde{u}_i'\right) = O_p(C_{n,T}^{-2})$ and $I = O_p(\chi_{1,T}^4\chi_{2,T}^2/T) = O_p(C_{n,T}^{-2})$.

Lemma 6 *Under model $\mathcal{M}_2(k)$ and Assumptions 1(i), 2 and 3, we have* $\mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) \geq C$, *w.p.a.* 1, *for a constant $C > 0$.*

Then, from Inequality (22) and Lemma 6, we get $\xi \geq C/2$, w.p.a. 1, and Proposition 2(b) follows.

A.2.3 Proof of Lemma 1

We prove:

$$\limsup_{n,T\to\infty} \mu_1 \left(\frac{1}{n}\tilde{\mathcal{E}}\tilde{\mathcal{E}}'\right) \leq C,\ a.s., \tag{23}$$

for a constant $C < \infty$, where $\tilde{\mathcal{E}}$ is the $T \times n$ matrix with elements $\tilde{\varepsilon}_{i,t} = I_{i,t}\varepsilon_{i,t}$. Then, since $T/n = o(1)$, the statement of Lemma 1 follows. To show (23), we follow similar arguments as in Geman (1980), Yin, Bai, and Krishnaiah (1988), and Bai and Yin (1993).

We first establish suitable versions of the so-called truncation and centralization lemmas. We denote by Ξ and E the $T \times n$ matrices with elements $(\xi_{i,t})$ and $(e_{i,t})$, respectively, where $\xi_{i,t} = \varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}$ and $e_{i,t} = \xi_{i,t} - E[\xi_{i,t}]$, and $\delta = \delta_n \uparrow \infty$ is a diverging sequence as in Assumption A.3. Let us define matrices $\tilde{E} = (I_{i,t}e_{i,t})$ and $\tilde{\Xi} = (I_{i,t}\xi_{i,t})$ by analogy to $\tilde{\mathcal{E}}$. Lemma 7 shows that we can substitute the truncated $\xi_{i,t}$ and $I_{i,t}\xi_{i,t}$ for $\varepsilon_{i,t}$ and $I_{i,t}\varepsilon_{i,t}$, and Lemma 8 shows that we can substitute the centered $I_{i,t}e_{i,t}$ for the $I_{i,t}\xi_{i,t}$ to show boundedness of the largest eigenvalue in (23). Lemmas 7 and 8 are proved in the supplementary material.

Lemma 7 *Under Assumption A.2, if $\delta = \delta_n$ is such that $\delta \geq n^\beta$ for $\beta > 2/q$, then: (i) $P\left(\mathcal{E} \neq \Xi\ i.o.\right) = 0$, and (ii) $P\left(\tilde{\mathcal{E}} \neq \tilde{\Xi}\ i.o.\right) = 0$, where $i.o.$ means infinitely often for $n = 1, 2, ...$*

Lemma 8 *Under Assumption A.2, if $\delta = \delta_n \uparrow \infty$ such that $\sqrt{T}/\delta^{q-1} = o(1)$, then:*

$$\mu_1\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) = \mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right) + o(1), \quad a.s.$$

From Lemma 7(ii) and Lemma 8, condition (23) is implied by:

$$\limsup_{n,T\to\infty} \mu_1 \left(\frac{1}{n}\tilde{E}\tilde{E}'\right) \leq C,\ a.s., \tag{24}$$

for a constant $C < \infty$.

Now, we use that the upper bound (24) is implied by the condition:

$$\sum_{n=1}^{\infty} E\left[\left(\mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)/C\right)^k\right] < \infty, \tag{25}$$

for an increasing sequence of integers $k = k_n \uparrow \infty$. To prove the validity of condition (25), we use that:

$$\begin{aligned}\mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)^k &\leq Tr\left[\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)^k\right] \\ &= \frac{1}{n^k}\sum_{i_1,\ldots,i_k}\sum_{t_1,\ldots,t_k}\tilde{e}_{i_1,t_1}\tilde{e}_{i_2,t_1}\tilde{e}_{i_2,t_2}\tilde{e}_{i_3,t_2}\cdots\tilde{e}_{i_{k-1},t_{k-1}}\tilde{e}_{i_k,t_{k-1}}\tilde{e}_{i_k,t_k}\tilde{e}_{i_1,t_k},\end{aligned}$$

for any integer k, where in the summation the indices i_1, ..., i_k run from 1 to n, and indices t_1, ..., t_k run from 1 to T. Therefore, from Assumption A.7:

$$E\left[\mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)^k\right] \leq \frac{1}{n^k}\sum_{i_1,\ldots,i_k}\sum_{t_1,\ldots,t_k}|E[e_{i_1,t_k}e_{i_1,t_1}e_{i_2,t_1}e_{i_2,t_2}e_{i_3,t_2}\cdots e_{i_{k-1},t_{k-1}}e_{i_k,t_{k-1}}e_{i_k,t_k}]|.$$

Then, we get $E\left[\mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)^k\right] \leq C_1^k$, for the sequence $k = k_n$ defined in Assumption A.3. Condition (25) holds for any $C > C_1$, and the conclusion follows.

A.2.4 Proof of Lemma 2

i) We have:

$$\begin{aligned}I_1^2 &= \|\frac{1}{nT}\sum_i(1-\mathbf{1}_i^\chi)\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\|^2 \\ &= \frac{1}{n^2T^2}\sum_{i,j}(1-\mathbf{1}_i^\chi)(1-\mathbf{1}_j^\chi)(\tilde{\varepsilon}_i'\tilde{\varepsilon}_j)^2 \\ &= \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2}(1-\mathbf{1}_i^\chi)(1-\mathbf{1}_j^\chi)I_{i,t_1}I_{j,t_1}I_{i,t_2}I_{j,t_2}\varepsilon_{i,t_1}\varepsilon_{j,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_2}.\end{aligned}$$

By the Cauchy-Schwarz inequality:

$$E[I_1^2] \leq \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2}E[1-\mathbf{1}_i^\chi]^{1/4}E[1-\mathbf{1}_j^\chi]^{1/4}E[\varepsilon_{i,t_1}^8]^{1/8}E[\varepsilon_{j,t_1}^8]^{1/8}E[\varepsilon_{i,t_2}^8]^{1/8}E[\varepsilon_{j,t_2}^8]^{1/8}.$$

Now, we have $E[\varepsilon_{i,t}^8] \leq M$ from Assumption A.2 and $E[1-\mathbf{1}_i^\chi] = P[\mathbf{1}_i^\chi = 0] = O(T^{-\bar{b}})$ for any $\bar{b} > 0$, uniformly in i and t (see GOS). Then, $I_1 = O_p(T^{-\bar{b}})$ for any $\bar{b} > 0$.

ii) We have:

$$
\begin{aligned}
I_2^2 &= \|\frac{1}{nT}\sum_i \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i' P_{\tilde{X}_i}\|^2 \\
&= \frac{1}{n^2T^2}\sum_{i,j}\mathbf{1}_i^\chi\mathbf{1}_j^\chi Tr\left[P_{\tilde{X}_i}\tilde{\varepsilon}_i\tilde{\varepsilon}_i' P_{\tilde{X}_i}P_{\tilde{X}_j}\tilde{\varepsilon}_j\tilde{\varepsilon}_j' P_{\tilde{X}_j}\right] \\
&= \frac{1}{n^2T^2}\sum_{i,j}\mathbf{1}_i^\chi\mathbf{1}_j^\chi \frac{\tau_{i,T}^2\tau_{j,T}^2}{\tau_{T,ij}^2} Tr\left[\hat{Q}_{x,i}^{-1}\left(\frac{\tilde{X}_i'\tilde{\varepsilon}_i}{\sqrt{T}}\right)\left(\frac{\tilde{\varepsilon}_i'\tilde{X}_i}{\sqrt{T}}\right)\hat{Q}_{x,i}^{-1}\hat{Q}_{x,ij}\hat{Q}_{x,j}^{-1}\left(\frac{\tilde{X}_j'\tilde{\varepsilon}_j}{\sqrt{T}}\right)\right. \\
&\quad \left.\left(\frac{\tilde{\varepsilon}_j'\tilde{X}_j}{\sqrt{T}}\right)\hat{Q}_{x,j}^{-1}\hat{Q}_{x,ji}\right],
\end{aligned}
$$

where $\hat{Q}_{x,ij} = \frac{1}{T_{i,j}}\sum_t I_{i,t}I_{j,t}x_{i,t}x_{j,t}'$ and $\tau_{ij,T} = T/T_{ij}$. By using $Tr(AB') \leq \|A\|\|B\|$, $\mathbf{1}_i^\chi\|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$, $\mathbf{1}_i^\chi\tau_{i,T} \leq \chi_{2,T}$, $\|x_{i,t}\| \leq M$ (Assumption A.4), $\tau_{ij,T} \geq 1$, for all i and t, we get:

$$
\begin{aligned}
I_2^2 &\leq \frac{C\chi_{1,T}^8\chi_{2,T}^4}{n^2T^2}\sum_{i,j}\|\frac{\tilde{\varepsilon}_i'\tilde{X}_i}{\sqrt{T}}\|^2\|\frac{\tilde{\varepsilon}_j'\tilde{X}_j}{\sqrt{T}}\|^2 \\
&= \frac{C\chi_{1,T}^8\chi_{2,T}^4}{n^2T^4}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4} I_{i,t_1}I_{i,t_2}I_{j,t_3}I_{j,t_4}\varepsilon_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_3}\varepsilon_{j,t_4}x_{i,t_1}'x_{i,t_2}x_{j,t_3}'x_{j,t_4}.
\end{aligned}
$$

Thus:

$$
\begin{aligned}
& E[I_2^2|I_{i,\underline{T}}, I_{j,\underline{T}}, x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j] \\
\leq\ & \frac{C\chi_{1,T}^8\chi_{2,T}^4}{n^2T^4}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}\|x_{i,t_1}\|\|x_{i,t_2}\|\|x_{j,t_3}\|\|x_{j,t_4}\||E[\varepsilon_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_3}\varepsilon_{j,t_4}|x_{i,\underline{T}}, x_{j,\underline{T}}, \gamma_i, \gamma_j]| \\
\leq\ & \frac{CM^5\chi_{1,T}^8\chi_{2,T}^4}{T^2},
\end{aligned}
$$

from Assumptions A.1 and A.4. It follows $E[I_2^2] = O(\frac{\chi_{1,T}^8\chi_{2,T}^4}{T^2})$, which implies $I_2 = O_p(\frac{\chi_{1,T}^4\chi_{2,T}^2}{T})$.

A.2.5 Proof of Lemma 3

i) The proof of Lemma 3(i) is the same as that of Lemma 2(i), since the bound $E[|\varepsilon_{i,t}|^8] \leq M$ applies under $\mathcal{M}_2$ as well (Assumptions A.2 and A.5).

ii) The proof of Lemma 3(ii) is similar to that of Lemma 2(ii), by replacing $\tilde{\varepsilon}_i$ with $\tilde{H}_i\theta_i$ and using

Assumption A.6. We have:

$$\begin{aligned}
I_2^2 &= \|\frac{1}{nT}\sum_i \mathbf{1}_i^\chi P_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i' P_{\tilde{X}_i}\|^2 \\
&= \frac{1}{n^2T^2}\sum_{i,j}\mathbf{1}_i^\chi\mathbf{1}_j^\chi Tr\left[P_{\tilde{X}_i}\tilde{H}_i\theta_i\theta_i'\tilde{H}_i'P_{\tilde{X}_i}P_{\tilde{X}_j}\tilde{H}_j\theta_j\theta_j'\tilde{H}_j'P_{\tilde{X}_j}\right] \\
&= \frac{1}{n^2T^2}\sum_{i,j}\mathbf{1}_i^\chi\mathbf{1}_j^\chi\frac{\tau_{i,T}^2\tau_{j,T}^2}{\tau_{T,ij}^2}Tr\left[\hat{Q}_{x,i}^{-1}\left(\frac{\tilde{X}_i'\tilde{H}_i}{\sqrt{T}}\right)\theta_i\theta_i'\left(\frac{\tilde{H}_i'\tilde{X}_i}{\sqrt{T}}\right)\hat{Q}_{x,i}^{-1}\hat{Q}_{x,ij}\hat{Q}_{x,j}^{-1}\left(\frac{\tilde{X}_j'\tilde{H}_j}{\sqrt{T}}\right)\right. \\
&\quad \left.\theta_j\theta_j'\left(\frac{\tilde{H}_j'\tilde{X}_j}{\sqrt{T}}\right)\hat{Q}_{x,j}^{-1}\hat{Q}_{x,ji}\right].
\end{aligned}$$

By using $Tr(AB') \leq \|A\|\|B\|$, $\mathbf{1}_i^\chi\|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$, $\mathbf{1}_i^\chi\tau_{i,T} \leq \chi_{2,T}$, $\|\theta_i\| \leq M$, $\|x_{i,t}\| \leq M$, $\tau_{ij,T} \geq 1$, for all i and t, we get:

$$\begin{aligned}
I_2^2 &\leq \frac{C\chi_{1,T}^8\chi_{2,T}^4}{n^2T^2}\sum_{i,j}\|\frac{\tilde{H}_i'\tilde{X}_i}{\sqrt{T}}\|^2\|\frac{\tilde{H}_j'\tilde{X}_j}{\sqrt{T}}\|^2 \\
&= \frac{C\chi_{1,T}^8\chi_{2,T}^4}{n^2T^4}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4} I_{i,t_1}I_{i,t_2}I_{j,t_3}I_{j,t_4}h_{t_1}'h_{t_2}x_{i,t_1}'x_{i,t_2}h_{t_3}'h_{t_4}x_{j,t_3}'x_{j,t_4}.
\end{aligned}$$

Thus:

$$\begin{aligned}
E[I_2^2|I_{\underline{T},i}, I_{\underline{T},j}, \gamma_i, \gamma_j] &\leq \frac{C\chi_{1,T}^8\chi_{2,T}^4}{n^2T^4}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4}|E[h_{t_1}'h_{t_2}x_{i,t_1}'x_{i,t_2}h_{t_3}'h_{t_4}x_{j,t_3}'x_{j,t_4}|\gamma_i,\gamma_j]| \\
&\leq \frac{CM\chi_{1,T}^8\chi_{2,T}^4}{T^2},
\end{aligned}$$

where the first inequality comes from Assumption 2, i.e., $EL\left[h_t|\left\{x_{i,t}, i=1,2,...\right\}\right]=0$, and the second inequality comes from Assumption A.6. It follows $E[I_2^2] = O(\frac{\chi_{1,T}^8\chi_{2,T}^4}{T^2})$, which implies $I_2 = O_p(\frac{\chi_{1,T}^4\chi_{2,T}^2}{T})$.

iii) The proof of Lemma 3(iii) is the same as that of Lemma 2(ii), by replacing $\tilde{\varepsilon}_i$ with $\tilde{u}_i$.

A.2.6 Proof of Lemma 4

We have:

$$\mu_1\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) = \max_{x\in\mathbb{R}^T:\|x\|=1} x'\left(\frac{1}{nT}\sum_i\tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right)x.$$

From Assumption 1 (i), matrix $\frac{1}{T}H'H = \frac{1}{T}\sum_t h_t h_t'$ is positive definite w.p.a. 1. Thus, for any $a \in \mathbb{R}^m$ with $\|a\| = 1$, the vector $x(a) \in \mathbb{R}^T$ defined by $x(a) = \frac{1}{\sqrt{T}} Ha[a'(H'H/T)a]^{-1/2}$ is such that $\|x(a)\| = 1$, w.p.a. 1. Therefore:

$$
\begin{aligned}
\mu_1\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) &\geq \max_{a\in\mathbb{R}^m:\|a\|=1} x(a)'\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right)x(a) \\
&= \max_{a\in\mathbb{R}^m:\|a\|=1} \frac{a'\left[\frac{1}{n}\sum_i (H'\tilde{H}_i/T)\theta_i\theta_i'(\tilde{H}_i'H/T)\right]a}{a'(H'H/T)a} \\
&= \max_{a\in\mathbb{R}^m:\|a\|=1} \frac{a'\left[\frac{1}{n}\sum_i \tau_{i,T}^{-2}\left(\frac{1}{T_i}\sum_t I_{i,t}h_t h_t'\right)\theta_i\theta_i'\left(\frac{1}{T_i}\sum_t I_{i,t}h_t h_t'\right)\right]a}{a'\left(\frac{1}{T}\sum_t h_t h_t'\right)a}.
\end{aligned}
$$

We have $a'\left(\frac{1}{T}\sum_t h_t h_t'\right)a \leq \mu_1\left(\frac{1}{T}\sum_t h_t h_t'\right)$, for any $a \in \mathbb{R}^m$ such that $\|a\| = 1$, and from Assumption 1 (i), we have $\mu_1\left(\frac{1}{T}\sum_t h_t h_t'\right) \leq 2\mu_1(\Sigma_h)$ w.p.a. 1. Moreover, from GOS, under Assumptions ... we have $\sup_{1\leq i\leq n} \|\frac{1}{T_i}\sum_t I_{i,t}h_t h_t' - \Sigma_h\| = o_p(1)$, $\sup_{1\leq i\leq n} |\tau_{i,T} - \tau_i| = o_p(1)$, and $1 \leq \tau_i \leq M$, for all i. It follows:

$$
\mu_1\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) \geq C \max_{a\in\mathbb{R}^m:\|a\|=1} a'\Sigma_h\left(\frac{1}{n}\sum_i \theta_i\theta_i'\right)\Sigma_h a = C\mu_1\left(\Sigma_h\left(\frac{1}{n}\sum_i \theta_i\theta_i'\right)\Sigma_h\right),
$$

for a constant $C > 0$. From the inequality (13) for the eigenvalues of a matrix product applied twice, we have $\mu_1\left(\Sigma_h\left(\frac{1}{n}\sum_i \theta_i\theta_i'\right)\Sigma_h\right) \geq \mu_1\left(\frac{1}{n}\sum_i \theta_i\theta_i'\right)\mu_m(\Sigma_h)^2$. From Assumption 1 (ii), the conclusion follows.

A.2.7 Proof of Lemma 5

We start with the case $k = 1$, and then extend the arguments to the case $k \geq 2$.

a) When $k = 1$, let us consider matrix $\tilde{A} = \frac{1}{nT}\sum_i \theta_i^2 \tilde{H}_i\tilde{H}_i' = (\tilde{a}_{t,s})$ with elements

$$
\tilde{a}_{t,s} = \frac{1}{nT}\sum_i I_{i,t}I_{i,s}\theta_i^2 h_t h_s =: a_{t,s}h_t h_s.
$$

Further, define matrices $A = (a_{t,s})$ and $D = diag(h_t : t = 1, ..., T)$. Then $\tilde{A} = DAD$, and both $\tilde{A}$ and A are positive semidefinite matrices. In the first step of the proof, we show that:

$$\mu_2(\tilde{A}) \leq M^2 \mu_2(A), \tag{26}$$

where M is the constant in Assumption A.5 a).

Let $\mathcal{G}$ be a linear subspace of $\mathbb{R}^T$ and consider the maximization problem:

$$\max_{x \in \mathcal{G}: \|x\|=1} x' \tilde{A} x = \max_{x \in \mathcal{G}: \|x\|=1} x' DAD x.$$

For $x \in \mathcal{G}$ such that $\|x\| = 1$, define $y = Dx$. Then, $y \in D(\mathcal{G})$ (the image of space $\mathcal{G}$ under the linear mapping defined by matrix D) and

$$\|y\|^2 \leq \|h\|^2_{\infty,T} \|x\|^2 = \|h\|^2_{\infty,T} \leq M^2,$$

where $\|h\|_{\infty,T} = \max\limits_{t=1,...,T} |h_t| \leq M$ under Assumption A.5 a). Then:

$$\max_{x \in \mathcal{G}: \|x\|=1} x' \tilde{A} x \leq \max_{y \in D(\mathcal{G}): \|y\| \leq M} y' A y = M^2 \max_{y \in D(\mathcal{G}): \|y\|=1} y' A y. \tag{27}$$

Suppose that $h_t \neq 0$ for all $t = 1, ..., T$ (an event of probability 1). Then D corresponds to a one-to-one linear mapping. Let $\mathcal{F}_1$ be the eigenspace associated to the largest eigenvalue of matrix A, and define $\mathcal{G} = D^{-1}(\mathcal{F}_1^{\perp})$, which is a linear space of dimension $T - 1$. Then, from Inequality (27) we get:

$$\max_{x \in D^{-1}(\mathcal{F}_1^{\perp}): \|x\|=1} x' \tilde{A} x \leq M^2 \max_{y \in \mathcal{F}_1^{\perp}: \|y\|=1} y' A y. \tag{28}$$

From the Courant min-max theorem (14), we have:

$$\mu_2(\tilde{A}) \leq \max_{x \in D^{-1}(\mathcal{F}_1^{\perp}): \|x\|=1} x' \tilde{A} x,$$

and, from the Courant-Fisher formula (16), we have:

$$\mu_2(A) = \max_{y \in \mathcal{F}_1^{\perp}: \|y\|=1} y' A y. \tag{29}$$

Then, Inequality (28) implies bound (26).

Finally, let us bound $\mu_2(A)$. By writing $A = \frac{1}{nT}(B+C)(B+C)'$, where B and C are $T \times n$ matrices with elements $b_{i,t} = \theta_i \tau_{i,T}^{-1}$ and $c_{i,t} = \theta_i(I_{i,t} - \tau_{i,T}^{-1})$, respectively, the Weyl's inequality (12) implies:

$$\mu_2(A)^{1/2} \leq \mu_2\left(\frac{1}{nT}BB'\right)^{1/2} + \mu_1\left(\frac{1}{nT}CC'\right)^{1/2} = \mu_1\left(W\right)^{1/2},$$

where matrix BB' has rank 1, and the elements of matrix $W = \frac{1}{nT}CC'$ are $w_{t,s} = \frac{1}{nT}\sum_i (I_{i,t} - \tau_{i,T}^{-1})(I_{i,s} - \tau_{i,T}^{-1})\theta_i^2$. Thus, from Assumption A.8 we get $\mu_2(A) = O_p(C_{n,T}^{-2})$. From bound (26) the conclusion follows.

b) Let us now consider the case $k \geq 1$. Consider the matrix $\tilde{A} = \frac{1}{nT}\sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' = (\tilde{a}_{t,s})$ with elements

$$\begin{aligned}
\tilde{a}_{t,s} &= \frac{1}{nT}\sum_i I_{i,t}I_{i,s}\theta_i' h_t \theta_i' h_s \\
&= \sum_{m,l}\left(\frac{1}{nT}\sum_i I_{i,t}I_{i,s}\theta_{i,m}\theta_{i,l}\right) h_{t,m}h_{s,l} =: \sum_{m,l} a_{t,s}^{(m,l)} h_{t,m}h_{s,l},
\end{aligned}$$

where summation w.r.t. m, l is from 1 to k. Then, we have:

$$\tilde{A} = \sum_{m,l} D^{(m)} A^{(m,l)} D^{(l)} = DAD, \tag{30}$$

where $A^{(m,l)} = [a_{t,s}^{(m,l)}]$, $D^{(m)} = diag(h_{t,m} : t = 1, ..., T)$, the $T \times (Tk)$ matrix D is defined by $D = [D^{(1)} : ... : D^{(k)}]$ and A is $(Tk) \times (Tk)$ block matrix with blocks $A^{(m,l)}$.

Lemma 9 *Let* $\begin{pmatrix} A & B \\ B' & D \end{pmatrix}$ *be a positive definite (or semi-definite) block matrix. Then:*

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix} \leq 2 \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix},$$

where the inequality is w.r.t. the ranking of symmetric matrices.

By repeated application of Lemma 9 we get:

$$A \leq 2^{k-1} \begin{pmatrix} A^{(1,1)} & & \\ & \ddots & \\ & & A^{(k,k)} \end{pmatrix}.$$

This implies:

$$\tilde{A} \leq 2^{k-1} \sum_m D^{(m)} A^{(m,m)} D^{(m)}. \tag{31}$$

Since two symmetric matrices are ranked if, and only if, their corresponding eigenvalues are ranked, we get:

$$\mu_{k+1}(\tilde{A}) \leq 2^{k-1} \mu_{k+1} \left(\sum_m D^{(m)} A^{(m,m)} D^{(m)} \right). \tag{32}$$

Moreover, we use the next lemma.

Lemma 10 *For k symmetric matrices A_1, A_2, ... A_k:*

$$\mu_{k+1}(A_1 + ... + A_k) \leq \mu_2(A_1) + ... + \mu_2(A_k).$$

From Inequality (32) and Lemma 10 we get:

$$\mu_{k+1}(\tilde{A}) \leq 2^{k-1} \sum_m \mu_2 \left(D^{(m)} A^{(m,m)} D^{(m)} \right).$$

By using the arguments deployed for the case $k = 1$ in part **a**), we have:

$$\mu_2(D^{(m)} A^{(m,m)} D^{(m)}) \leq M^2 \mu_2(A^{(m,m)}). \tag{33}$$

Therefore, we get:

$$\mu_{k+1}(\tilde{A}) \leq 2^{k-1} M^2 \sum_m \mu_2(A^{(m,m)}).$$

As in part **a**), the Weyl's inequality and Assumption A.8 imply $\mu_2(A^{(m,m)}) \leq \mu_1(W^{(m)}) = O_p(C_{n,T}^{-2})$. Thus $\mu_{k+1}(\tilde{A}) = O_p(C_{n,T}^{-2})$.

A.2.8 Proof of Lemma 6

From the Courant-Fisher max-min Theorem (15), we have:

$$\mu_{k+1} \left(\frac{1}{nT} \sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' \right) = \max_{\mathcal{G}:\dim(\mathcal{G})=k+1} \min_{x \in \mathcal{G}:\|x\|=1} x' \left(\frac{1}{nT} \sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' \right) x, \tag{34}$$

where the maximization is w.r.t. the linear $(k+1)$-dimensional subspace $\mathcal{G}$ of $\mathbb{R}^T$. From Assumption 1 (i), under model $\mathcal{M}_2(k)$ matrix $H/\sqrt{T}$ has full column-rank equal to m, w.p.a. 1, with $m \geq k+1$. Thus, for any linear subspace $\mathbb{A}$ of $\mathbb{R}^m$ with dimension $k+1$, the set $\mathcal{G}_{\mathbb{A}} := \left\{ x \in \mathbb{R}^T : \, x = \frac{1}{\sqrt{T}} H a, \, a \in \mathbb{A} \right\}$ is a linear subspace of $\mathbb{R}^T$ of dimension $k+1$. We deduce from (34):

$$\begin{aligned} \mu_{k+1}\left(\frac{1}{nT} \sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' \right) &\geq \max_{\mathbb{A}:\dim(\mathbb{A})=k+1} \min_{x \in \mathcal{G}_{\mathbb{A}}:\|x\|=1} x' \left(\frac{1}{nT} \sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' \right) x \\ &= \max_{\mathbb{A}:\dim(\mathbb{A})=k+1} \min_{a \in \mathbb{A}:\|a\|=1} \frac{a' \left(\frac{1}{n} \sum_i \frac{H' \tilde{H}_i}{T} \theta_i \theta_i' \frac{\tilde{H}_i' H}{T} \right) a}{a' \left(\frac{1}{T} H' H \right) a}. \end{aligned}$$

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

$$\mu_{k+1}\left(\frac{1}{nT} \sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' \right) \geq C \max_{\mathbb{A}:\dim(\mathbb{A})=k+1} \min_{a \in \mathbb{A}:\|a\|=1} a' \Sigma_h \left(\frac{1}{n} \sum_i \theta_i \theta_i' \right) \Sigma_h a,$$

w.p.a. 1. By the max-min Theorem, the r.h.s. is such that:

$$\max_{\mathbb{A}:\dim(\mathbb{A})=k+1} \min_{a \in \mathbb{A}:\|a\|=1} a' \Sigma_h \left(\frac{1}{n} \sum_i \theta_i \theta_i' \right) \Sigma_h a = \mu_{k+1}\left(\Sigma_h \left(\frac{1}{n} \sum_i \theta_i \theta_i' \right) \Sigma_h \right).$$

Moreover, from the inequality (13) for the eigenvalues of product matrices applied twice, we have $\mu_{k+1}\left(\Sigma_h \left(\frac{1}{n} \sum_i \theta_i \theta_i' \right) \Sigma_h \right) \geq \mu_{k+1}\left(\left(\frac{1}{n} \sum_i \theta_i \theta_i' \right) \right) \mu_m(\Sigma_h)^2$. Then, from Assumptions 1 (i) and 3, the conclusion follows.

Appendix 3 Check of Assumption A.3 under block dependence

In this appendix, we verify that Assumption A.3 is satisfied under a block-dependence structure in a serially i.i.d. framework. Here, $\varepsilon_{i,t}$ and $\varepsilon_{j,s}$ are independent if either i and j belong to different blocks, or $t \neq s$. There are $b = b_n$ blocks of approximate size $d = d_n$, where $bd = O(n)$.

1) Let $\omega > 0$ be such that $E[\varepsilon_{i,t}^2] \leq \omega^2$, for all i and t, and define $\phi_{i,t} = e_{i,t}/\omega$. The scaled $\phi_{i,t}$ are such that $E[\phi_{i,t}] = 0$, $E[\phi_{i,t}^2] \leq 1$, and $E[|\phi_{i,t}|^{r-2}] = O(\delta^{r-2})$, for all $r \geq 3$, uniformly in i and t. Note that $\phi_{i,t}$ is a (nonlinear) transformation of $\varepsilon_{i,t}$. Hence, the variables $\phi_{i,t}$ have the same block dependence structure

as the variables $\varepsilon_{i,t}$. Moreover:

$$\begin{aligned}
& \frac{1}{n^k}\sum_{i_1,...,i_k}\sum_{t_1,...,t_k} |E[e_{i_1,t_k}e_{i_1,t_1}e_{i_2,t_1}\varepsilon_{i_2,t_2}e_{i_3,t_2}\cdots e_{i_{k-1},t_{k-1}}e_{i_k,t_{k-1}}e_{i_k,t_k}]| \\
\leq \quad & \omega^{2k}\frac{1}{n^k}\sum_{i_1,...,i_k}\sum_{t_1,...,t_k} |E[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\phi_{i_3,t_2}\cdots \phi_{i_{k-1},t_{k-1}}\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}] \\
=: \quad & \omega^{2k} I_k. \qquad (35)
\end{aligned}$$

Let us now bound I_k.

2) For $m = 1, ..., k$, let $\mathcal{C}_m$ denote the set of k-tuples $(i_1, ..., i_k)$ such that indices i_1, ..., i_k belong to m different blocks. Let N_m denote the number of different $2k$-tuples $(i_1, ..., i_k)$, $(t_1, ..., t_k)$ such that $(i_1, ..., i_k) \in \mathcal{C}_m$ and the expectation $E[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\phi_{i_3,t_2}\cdots \phi_{i_{k-1},t_{k-1}}\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}]$ does not vanish. Moreover, let Q_m be an upper bound for such a non vanishing expectation. Then:

$$I_k \leq \frac{1}{n^k}\sum_{m=1}^{k} N_m Q_m. \qquad (36)$$

3) We need upper bounds for N_m and Q_m, for $m = 1, 2, ..., k$, and any integer k.

- $m = 1$: The number of k-tuples $(i_1, ..., i_k)$ with all indices in the same block is $O(bd^k)$. Indeed, we can select the block among b alternatives, and we have $O(d^k)$ possibilities to select the indices within the block. Then, $N_1 = O(bd^kT^k)$. Moreover, by the Cauchy-Schwarz inequality,

  $$E\left[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\phi_{i_3,t_2}\cdots \phi_{i_{k-1},t_{k-1}}\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}\right] \leq \sup_{i,t} E[|\phi_{i,t}|^{2k}] = O(\delta^{2k-2}).$$

  Thus, $Q_1 = O(\delta^{2(k-1)})$.

- $m = k$: The number of k-tuples $(i_1, ..., i_k)$ with indices in k different blocks is $O(b^kd^k)$. For such a k-tuple:

  $$E\left[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\cdots \phi_{i_k,t_{k-1}}\phi_{i_k,t_k}\right] = E\left[\phi_{i_1,t_k}\phi_{i_1,t_1}\right]E\left[\phi_{i_2,t_1}\phi_{i_2,t_2}\right]\cdots E\left[\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}\right].$$

  Hence, the indices t_1, ... t_k must be all equal for this expectation not to vanish. Then, $N_k = O(b^kd^kT)$ and $Q_k \leq 1$. [1]

[1]For $k > b$, there are no k-tuples $(i_1, ..., i_k)$ with indices in k different blocks, and $N_k = 0$. The upper bound $N_k = O(b^kd^kT)$ trivially holds also in this case. However, this case will not occur with our choice of sequence k, since (42) implies $k = o(b)$, see below.

- $m=2$: The number N_2 is $O(b^2) \times \left\{ \begin{matrix} k \\ 2 \end{matrix} \right\} \times O(d^k) \times O(T^{k-1})$, where $\left\{ \begin{matrix} k \\ 2 \end{matrix} \right\} = 2^{k-1}-1$ is the number of different ways in which we can divide k objects into two (non-empty) groups (a Stirling number of the second kind). Indeed, $O(b^2)$ is a bound for the number of different ways to select the two distinct blocks. Then, for each $j=1,...,k$ we select whether index i_j is in the first or the second block; we have $\left\{ \begin{matrix} k \\ 2 \end{matrix} \right\}$ different possibilities. Once we have fixed the blocks, we have $O(d^k)$ alternatives to select the indices. By block dependence, the expectation $E[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\cdots\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}]$ can be splitted into two expectations, and at least a pair of indices in the k-tuple $(t_1,...,t_k)$ must be equal for the expectation not to vanish. Hence the term $O(T^{k-1})$.

  Suppose the expectation $E[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\cdots\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}]$ is splitted into two expectations, with r_1 indices i_j in the first block, and r_2 indices in the second block, $r_1+r_2=k$. Then, $E[\phi_{i_1,t_k}\phi_{i_1,t_1}\phi_{i_2,t_1}\phi_{i_2,t_2}\cdots\phi_{i_k,t_{k-1}}\phi_{i_k,t_k}] = O(\delta^{2(r_1-1)}) \times O(\delta^{2(r_1-1)}) = O(\delta^{2(k-2)})$. Hence, $Q_2 = O(\delta^{2(k-2)})$.

- Generic m: We have

$$\begin{aligned} N_m &= O(b^m) \times \left\{ \begin{matrix} k \\ m \end{matrix} \right\} \times O(d^k) \times O(T^{k-m+1}), & (37) \\ Q_m &= = (\delta^{2(k-m)}), & (38) \end{aligned}$$

  where the Stirling number of the second kind $\left\{ \begin{matrix} k \\ m \end{matrix} \right\} = \frac{1}{m!}\sum_{j=0}^{m}(-1)^{m-j}\binom{m}{j} j^k$ gives the number of different ways in which we can divide k objects into m (non-empty) groups (see e.g. Rennie and Dobson (1969)) and $\binom{k}{m}$ is a binomial coefficient.

From bounds (36), (37) and (38), and using $d = O(n/b)$, we get:

$$\begin{aligned} I_k &\leq \frac{const}{n^k}\sum_{m=1}^{k} b^m d^k \left\{ \begin{matrix} k \\ m \end{matrix} \right\} T^{k-m+1}\delta^{2(k-m)} \\ &= const \times T \sum_{m=1}^{k} \left\{ \begin{matrix} k \\ m \end{matrix} \right\} (\delta^2 T/b)^{k-m}. & (39) \end{aligned}$$

4) We exploit the following upper bound for the Stirling numbers of the second kind (see Rennie and Dobson (1969), Theorem 3):

$$\left\{ \begin{matrix} k \\ m \end{matrix} \right\} \leq \frac{1}{2}\binom{k}{m} m^{k-m}. \qquad (40)$$

Then, we get:

$$\begin{aligned}\sum_{m=1}^{k}\left\{{k \atop m}\right\}(\delta^2T/b)^{k-m} &\leq \frac{1}{2}\sum_{m=1}^{k}\binom{k}{m}(\delta^2T/b)^{k-m}m^{k-m}\\ &\leq \frac{1}{2}\sum_{m=0}^{k}\binom{k}{m}(k\delta^2T/b)^{k-m}\\ &= \frac{1}{2}(1+k\delta^2T/b)^k,\end{aligned}$$

from the binomial theorem. Thus, from (39), we get:

$$I_k \leq constT(1+k\delta^2T/b)^k. \tag{41}$$

5) Assume that the sequence $k = k_n \uparrow \infty$ is such that:

$$k\delta^2T/b = o(1), \quad T = O(e^k). \tag{42}$$

From (41) and (42), we get $I_k \leq (2e)^k$. Then, from (35):

$$\frac{1}{n^k}\sum_{i_1,\dots,i_k}\sum_{t_1,\dots,t_k}|E[e_{i_1,t_k}e_{i_1,t_1}e_{i_2,t_1}\varepsilon_{i_2,t_2}e_{i_3,t_2}\cdots e_{i_{k-1},t_{k-1}}e_{i_k,t_{k-1}}e_{i_k,t_k}]| \leq (2e\omega)^k,$$

i.e. the bound in Assumption A.3 holds with $C = 2e\omega$.

6) Let us now verify compatibility of the different rates, i.e., that we can choose sequences $\delta = n^\beta$ and $k = c\log(n)$, $\beta, c > 0$, such that $\sqrt{T}/\delta^{q-1} = o(1)$ and they match conditions (42). Let $n \geq T^{\bar{\gamma}}$ and $b \geq n^\alpha$, with $\bar{\gamma} > 1$ and $\alpha \in (0,1)$. Condition $T = O(e^k)$ is satisfied if $c \geq 1/\bar{\gamma}$. Condition $k\delta^2T/b = o(1)$ implies:

$$\beta < \frac{1}{2}(\alpha - 1/\bar{\gamma}). \tag{43}$$

Condition $\sqrt{T}/\delta^{q-1} = o(1)$ implies:

$$\beta > \frac{1}{2\bar{\gamma}(q-1)}. \tag{44}$$

Then, there exists a power $\beta > 0$ satisfying conditions (43) and (44) if, and only if, $\frac{1}{2}(\alpha - 1/\bar{\gamma}) > \frac{1}{2\bar{\gamma}(q-1)}$, i.e.

$$\bar{\gamma} > \frac{1}{\alpha}\frac{q}{q-1}. \tag{45}$$

This condition provides a restriction on the relative growth rate of the cross-sectional and time series dimensions in terms of: (i) the strength of cross-sectional dependence, and (ii) the existence of higher-order moments of the error terms. We can have $\bar{\gamma}$ (arbitrarily) close to 1, if cross-sectional dependence is sufficiently weak and the tails of the errors are sufficiently thin. Condition (45) clarifies the link between the behaviour of expectations of products of error terms and the assumption of a bounded largest eigenvalue used for example in Chamberlain and Rothschild (1983) p. 1294 for arbitrage pricing theory.

Table 7: Selection probabilities, unbalanced case

trimming level	$\chi_{2,T} = \frac{T}{12}$					$\chi_{2,T} = \frac{T}{60}$				
n	500	1,000	3,000	6,000	9,000	500	1,000	3,000	6,000	9,000
$Pr(\xi_1 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_2 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_3 < 0 \mid \mathcal{M}_1)$	0.0000	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_1 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_2 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_3 < 0 \mid \mathcal{M}_1)$	0.0020	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_1 > 0 \mid \mathcal{M}_2)$	0.9520	1.0000	1.0000	1.0000	1.0000	0.9960	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_2 > 0 \mid \mathcal{M}_2)$	0.6180	0.9980	1.0000	1.0000	1.0000	0.9380	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_3 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_1 > 0 \mid \mathcal{M}_2)$	0.9480	1.0000	1.0000	1.0000	1.0000	0.9960	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_2 > 0 \mid \mathcal{M}_2)$	0.6140	0.9980	1.0000	1.0000	1.0000	0.9360	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_3 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

trimming level	$\chi_{2,T} = \frac{T}{120}$					$\chi_{2,T} = \frac{T}{240}$				
n	500	1,000	3,000	6,000	9,000	500	1,000	3,000	6,000	9,000
$Pr(\xi_1 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0040	1.0000	1.0000	1.0000
$Pr(\xi_2 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.1140	1.0000	1.0000	1.0000
$Pr(\xi_3 < 0 \mid \mathcal{M}_1)$	0.0000	0.9660	1.0000	1.0000	1.0000	0.0000	0.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_1 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0040	1.0000	1.0000	1.0000
$Pr(\check{\xi}_2 < 0 \mid \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.1480	1.0000	1.0000	1.0000
$Pr(\check{\xi}_3 < 0 \mid \mathcal{M}_1)$	0.0000	0.9800	1.0000	1.0000	1.0000	0.0000	0.0000	1.0000	1.0000	1.0000
$Pr(\xi_1 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_2 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_3 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_1 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_2 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_3 > 0 \mid \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

SUPPLEMENTARY MATERIALS

A diagnostic criterion for approximate factor structure

Patrick Gagliardini, Elisa Ossola and Olivier Scaillet

Appendix 4 Proofs of technical Lemmas

A.4.1 Proof of Lemma Lemma 7

We follow the arguments in the proof of Lemma 2.2 in Yin, Bai, and Krishnaiah (1988). From the conditions $\delta \geq n^{\beta}, T/n = o(1)$, we have:

$$\begin{aligned} P(\mathcal{E} \neq \Xi \; i.o.) &\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} P\left(\bigcup_{2^{m-1}\leq n<2^m} \bigcup_{i=1}^{n}\bigcup_{t=1}^{T} \{|\varepsilon_{i,t}| > \delta\} \right) \\ &\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} P\left(\bigcup_{i=1}^{2^m}\bigcup_{t=1}^{2^m} \{|\varepsilon_{i,t}| > 2^{(m-1)\beta}\} \right) \\ &\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} 2^{2m} P\left(|\varepsilon_{i,t}| > 2^{(m-1)\beta} \right). \end{aligned}$$

Thus, part (i) follows from the summability condition:

$$\sum_{m=1}^{\infty} 2^{2m} P\left(|\varepsilon_{i,t}| > 2^{(m-1)\beta} \right) < \infty. \tag{46}$$

To prove the summability condition (46). We use the Chebyshev inequality and Assumption A.2. We have $P\left(|\varepsilon_{i,t}| > 2^{(m-1)\beta} \right) \leq E[|\varepsilon_{i,t}|^q]/2^{(m-1)\beta q} \leq M/2^{(m-1)\beta q}$. Therefore, we get:

$$\sum_{m=1}^{\infty} 2^{2m} P\left(|\varepsilon_{i,t}| > 2^{(m-1)\beta} \right) \leq M \sum_{m=1}^{\infty} \frac{2^{2m}}{2^{(m-1)\beta q}} = M 2^{\beta q} \sum_{m=1}^{\infty} \frac{1}{2^{(\beta q-2)m}} < \infty,$$

since $q\beta > 2$.

Part (ii) is a straightforward consequence of part (i), since $P(\tilde{\mathcal{E}} \neq \tilde{\Xi} \; i.o.) \leq P(\mathcal{E} \neq \Xi \; i.o.)$.

A.4.2 Proof of Lemma Lemma 8

We follow the arguments in Bai and Yin (1993), p. 1278. We use the von Neumann inequality (von Neumann (1937)): for any $n \times T$ matrices A and B,

$$tr(A'B) \leq \sum_{k=1}^{T} \mu_k(A'A)^{1/2} \mu_k(B'B)^{1/2}. \tag{47}$$

We have:

$$\begin{aligned}
\left[\mu_1^{1/2}\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) - \mu_1^{1/2}\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)\right]^2 &\leq \sum_{k=1}^{T}\left[\mu_k^{1/2}\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) - \mu_k^{1/2}\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)\right]^2 \\
&= tr\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) + tr\left(\frac{1}{n}\tilde{E}\tilde{E}'\right) - 2\sum_{k=1}^{T}\mu_k^{1/2}\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right)\mu_k^{1/2}\left(\frac{1}{n}\tilde{E}\tilde{E}'\right).
\end{aligned}$$

The last term in the r.h.s. is bounded by the von Neumann inequality (47):

$$\begin{aligned}
\left[\mu_1^{1/2}\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) - \mu_1^{1/2}\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)\right]^2 &\leq tr\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) + tr\left(\frac{1}{n}\tilde{E}\tilde{E}'\right) - 2\frac{1}{n}tr\left(\tilde{\Xi}\tilde{E}'\right) \\
&= \frac{1}{n}tr\left[(\tilde{\Xi} - \tilde{E})(\tilde{\Xi} - \tilde{E})'\right]. \qquad (48)
\end{aligned}$$

The elements of matrix $\tilde{\Xi} - \tilde{E}$ are $I_{i,t}E[\varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}]$. By the zero-mean property of the errors $\varepsilon_{i,t}$, the Minkowski inequality and Assumption A.2, we have:

$$|E[\varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}]| = |E[\varepsilon_{i,t}1\{|\varepsilon_{i,t}| > \delta\}]| \leq E[|\varepsilon_{i,t}|^q]^{1/q} P[|\varepsilon_{i,t}| > \delta]^{1/\bar{q}},$$

where $1/q + 1/\bar{q} = 1$, with q defined in Assumption A.2. By the Chebyshev inequality and Assumption A.2, we get:

$$E[|\varepsilon_{i,t}|^q]^{1/q} P[|\varepsilon_{i,t}| > \delta]^{1/\bar{q}} \leq = E[|\varepsilon_{i,t}|^q]^{1/q}\left(\frac{E[|\varepsilon_{i,t}|^q]}{\delta^q}\right)^{1/\bar{q}} = \frac{E[|\varepsilon_{i,t}|^q]}{\delta^{q-1}} \leq \frac{M}{\delta^{q-1}}.$$

Thus, we get:

$$\frac{1}{n}tr\left[(\tilde{\Xi} - \tilde{E})(\tilde{\Xi} - \tilde{E})'\right] = \frac{1}{n}\sum_i \sum_t I_{i,t}E[\varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}]^2 \leq T\frac{M^2}{\delta^{2(q-1)}}. \tag{49}$$

From inequalities (48) and (49), we get $|\mu_1^{1/2}\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) - \mu_1^{1/2}\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)| \leq \sqrt{T}\frac{M}{\delta^{q-1}}$. Since the sequence $\delta = \delta_n$ is such that $\sqrt{T}/\delta^{q-1} = o(1)$, the conclusion follows.

A.4.3 Proof of Lemma Lemma 9

We have:

$$2\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} - \begin{pmatrix} A & B \\ B' & D \end{pmatrix} = \begin{pmatrix} A & -B \\ -B' & D \end{pmatrix},$$

and:

$$\begin{pmatrix} x_1' & x_2' \end{pmatrix}\begin{pmatrix} A & -B \\ -B' & D \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1' & -x_2' \end{pmatrix}\begin{pmatrix} A & B \\ B' & D \end{pmatrix}\begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} \geq 0,$$

for all $x = (x_1', x_2')'$.

A.4.4 Proof of Lemma Lemma 10

By repeated application of the Weyl inequality for eigenvalues (see Appendix 2 (i)) we have:

$$\begin{aligned} \mu_{k+1}(A_1 + ... + A_k) &\leq \mu_k(A_1 + ... + A_{k-1}) + \mu_2(A_k) \\ &\leq \mu_{k-1}(A_1 + ... + A_{k-2}) + \mu_2(A_{k-1}) + \mu_2(A_k) \\ &\quad \cdots \\ &\leq \mu_2(A_1) + ... + \mu_2(A_k). \end{aligned}$$

Appendix 5 Verification that conditional independence implies Assumption 2

Let us verify that 2 is true if the latent factors are independen of the lagged stock-specific instruments, conditional on the observable factors and the lagged common instruments.

We have:

$$\begin{aligned} h_t \perp \{Z_{i,t-1}, i = 1, ...\} \mid f_t, Z_{t-1} &\Rightarrow h_t \perp \{\tilde{x}_{i,t-1}, i = 1, ...\} \mid f_t, Z_{t-1} \\ &\Rightarrow h_t \perp \{\tilde{x}_{i,t-1}, i = 1, ...\} \mid x_t \\ &\Rightarrow EL[h_t | x_{i,t-1}, i = 1, ...] = EL[h_t | x_t], \end{aligned}$$

where $A \perp B | C$ denotes independence of A and B conditional on C.

Appendix 6 Link with Stock and Watson (2002)

We consider the EM algorithm proposed by Stock and Watson (2002):

$$\tilde{\tilde{\varepsilon}}_{i,t} = \begin{cases} \hat{\varepsilon}_{i,t}, & \text{if } I_{i,t} = 1, \\ \hat{\theta}_i \hat{h}_t, & \text{if } I_{i,t} = 0. \end{cases}$$

The statistic is $\xi = \mu_1 \left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT} \right) - \frac{1}{nT} \sum_i \sum_t (1 - I_{i,t}) \left(\hat{\theta}_i \hat{h}_t \right)^2 - g(n, T)$. Below we show that ξ is the difference of the EM criteria under the two models. Comparing the two test statistics gives the following link: $\frac{1}{nT} \sum_i \sum_t (1 - I_{i,t}) \left(\hat{\theta}_i \hat{h}_t \right)^2 = \frac{1}{nT} \left\| \hat{\varepsilon} - \tilde{\tilde{\varepsilon}} \right\|^2$.

To study the EM algorithm, we work as if the true error terms $\varepsilon_{i,t}$ are observed when $I_{i,t} = 1$. This error is replaced by the residual $\hat{\varepsilon}_{i,t}$. We consider the jth iteration of the algorithm. Let $\tilde{\zeta} = \left(\tilde{\Theta}, \tilde{H} \right)$ denotes the estimates of Θ and H obtained from the $(j-1)$th iteration, and let $Q\left(\zeta, \tilde{\zeta}\right) = E_{\tilde{\zeta}} \left[\mathcal{L}(\zeta) | \varepsilon \right]$, where $\mathcal{L}(\zeta) = \frac{1}{nT} \sum_i \sum_t \left(\varepsilon^*_{i,t} - \theta_i h_t \right)^2$, and $E_{\tilde{\zeta}}[\cdot|\varepsilon]$ denotes conditional expectation given the panel of observations under parameter $\tilde{\zeta}$. We study $Q\left(\zeta, \tilde{\zeta}\right)$ under the two models. Under both $\mathcal{M}_1$ and $\mathcal{M}_2$, we consider a pseudo model for the innovations such that $u_{i,t} \sim i.i.d. \left(0, \sigma^2_{i,t}\right)$

- Under $\mathcal{M}_1$: we get

$$Q_0\left(\zeta, \tilde{\zeta}\right) = E\left[\frac{1}{nT} \sum_i \sum_t \left(\varepsilon^*_{i,t}\right)^2 | \varepsilon \right] = \frac{1}{nT} \sum_i \sum_t E\left[\left(\varepsilon^*_{i,t}\right)^2 | \varepsilon \right].$$

We have

$$E\left[\varepsilon^*_{i,t} | \varepsilon\right] = \begin{cases} \varepsilon_{i,t}, & \text{if } I_{i,t} = 1, \\ 0, & \text{if } I_{i,t} = 0, \end{cases} \text{ and } V\left[\varepsilon^*_{i,t} | \varepsilon\right] = \begin{cases} 0, & \text{if } I_{i,t} = 1, \\ \sigma^2_{i,t}, & \text{if } I_{i,t} = 0. \end{cases}$$

and $E\left[\left(\varepsilon^*_{i,t}\right)^2 | \varepsilon \right] = I_{i,t}\varepsilon^2_{i,t} + (1 - I_{i,t}) \sigma^2_{i,t}$. Thus,

$$Q_0 = Q_0\left(\zeta, \tilde{\zeta}\right) = \frac{1}{nT} \sum_i \sum_t I_{i,t}\varepsilon^2_{i,t} + \frac{1}{nT} \sum_i \sum_t (1 - I_{i,t}) \sigma^2_{i,t}.$$

- Under $\mathcal{M}_2$: we get

$$
\begin{aligned}
Q_1\left(\zeta, \tilde{\zeta}\right) &= E_{\tilde{\zeta}}\left[\frac{1}{nT}\sum_i\sum_t\left(\varepsilon^*_{i,t}-\theta_i h_t\right)^2|\varepsilon\right] \\
&= \frac{1}{nT}\sum_i\sum_t E_{\tilde{\zeta}}\left[\left(\varepsilon^*_{i,t}-\theta_i h_t\right)^2|\varepsilon\right] \\
&= \frac{1}{nT}\sum_i\sum_t E_{\tilde{\zeta}}\left[\left(\varepsilon^*_{i,t}-E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right]+E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right]-\theta_i h_t\right)^2|\varepsilon\right] \\
&= \frac{1}{nT}\sum_i\sum_t V_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right]+\frac{1}{nT}\sum_i\sum_t\left(E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right]-\theta_i h_t\right)^2.
\end{aligned}
$$

We have

$$
\tilde{\varepsilon}_{i,t} := E_{\tilde{\zeta}}\left[\varepsilon^*_{i,t}|\varepsilon\right] = \begin{cases}\varepsilon_{i,t}, & \text{if } I_{i,t}=1, \\ \tilde{\theta}_i\tilde{h}_t, & \text{if } I_{i,t}=0,\end{cases} \text{ and } V\left[\varepsilon^*_{i,t}|\varepsilon\right] = \begin{cases}0, & \text{if } I_{i,t}=1, \\ \sigma^2_{i,t}, & \text{if } I_{i,t}=0.\end{cases}
$$

Thus, $Q_1\left(\zeta, \tilde{\zeta}\right)=\frac{1}{nT}\sum_i\sum_t\left(\tilde{\tilde{\varepsilon}}_{i,t}-\theta_i h_t\right)^2+\frac{1}{nT}\sum_i\sum_t\left(1-I_{i,t}\right)\sigma^2_{i,t}$, and the values of ζ that minimize $Q_1\left(\zeta, \tilde{\zeta}\right)$ can be calculated by $\min_{\zeta}\frac{1}{nT}\sum_i\sum_t\left(\tilde{\tilde{\varepsilon}}_{i,t}-\theta_i h_t\right)^2$. This minimization problem reduces to the usual PCA on data $\tilde{\tilde{\varepsilon}}$: $\min_{\zeta}\frac{1}{nT}\sum_i\sum_t\left(\tilde{\tilde{\varepsilon}}_{i,t}-\theta_i h_t\right)^2=\frac{1}{nT}\sum_i\sum_t\tilde{\tilde{\varepsilon}}^2_{i,t}-\mu_1\left(\frac{\tilde{\varepsilon}\tilde{\varepsilon}'}{nT}\right)$. Therefore, at convergence with $\hat{\zeta}=\tilde{\zeta}$, we have

$$
\begin{aligned}
Q_1\left(\hat{\zeta}, \tilde{\zeta}\right) &= \frac{1}{nT}\sum_i\sum_t\tilde{\tilde{\varepsilon}}^2_{i,t}-\mu_1\left(\frac{\tilde{\varepsilon}\tilde{\varepsilon}'}{nT}\right)+\frac{1}{nT}\sum_i\sum_t\left(1-I_{i,t}\right)\sigma^2_{i,t} \\
&= \frac{1}{nT}\sum_i\sum_t I_{i,t}\varepsilon^2_{i,t}+\frac{1}{nT}\sum_i\sum_t\left(1-I_{i,t}\right)\left(\hat{\theta}_i\hat{h}_t\right)^2 \\
&\quad -\mu_1\left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right)+\frac{1}{nT}\sum_i\sum_t\left(1-I_{i,t}\right)\sigma^2_{i,t}.
\end{aligned}
$$

Finally, the difference of the two EM criteria is

$$
Q_0-Q_1\left(\hat{\zeta}, \hat{\zeta}\right) = \mu_1\left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right)-\frac{1}{nT}\sum_i\sum_t\left(1-I_{i,t}\right)\left(\hat{\theta}_i\hat{h}_t\right)^2,
$$

which gives the interpretation of the test statistic.

Appendix 7 Monte-Carlo experiments

In this section, we perform simulation exercises on balanced and unbalanced panels in order to study the properties of our diagnostic criterion. We pay particular attention to the probability of diagnosing the correct model and its interaction with n and T in finite samples. The simulation design mimics the empirical features of our data. The balanced case serves as benchmark to understand when T and n are sufficiently large to apply theory. The unbalanced case shows that we can exploit the guidelines found for the balanced case when we substitute the average of the sample sizes of the individual assets, i.e., a kind of operative sample size, for T. To summarize our Monte Carlo findings, we do not face any finite sample distortions for the selection rule under $\mathcal{M}_1$ for most combinations of n and T, since we get estimates of $Pr(\xi < 0|\mathcal{M}_1)$ close to 1, and under $\mathcal{M}_2$ when n is larger than $3,000$, since we get estimates of $Pr(\xi > 0|\mathcal{M}_2)$ close to 1. In light of these results, we do not expect to face significant diagnostic bias in our empirical application.

A.7.1 Balanced panel

Under $\mathcal{M}_1$, we simulate S datasets of excess returns from a one-factor model (CAPM). A simulated dataset includes: a vector of factor loadings $b^s \in \mathbb{R}^n$, and a variance-covariance matrix $\Omega^s \in \mathbb{R}^{n \times n}$. At each simulation $s = 1, ..., S$, we randomly draw $n \leq 10,410$ assets from the sample of our empirical analysis that comprises 10,410 individual stocks with $T_i \geq 12$. The assets are listed by industrial sectors. We use the classification proposed by Ferson and Harvey (1999). The vector b^s is composed by the estimated factor loadings for the n randomly chosen assets. At each simulation, we build a block diagonal matrix Ω^s with blocks matching industrial sectors. The n elements of the main diagonal of Ω^s correspond to the variances of the estimated residuals of the individual assets. The off-diagonal elements of Ω^s are covariances computed by fixing correlations within block equal to the average correlation of the industrial sector computed from the $10,410 \times 10,410$ thresholded variance-covariance matrix of estimated residuals. Hence we get a setting in line with the weak block dependence case shown in GOS to exhibit an approximate factor structure.

Let us define $R^s_{i,t}$ the simulated excess returns of asset i at time t as follows

$$R^s_{i,t} = b^s_i f_t + \varepsilon^s_{i,t}, \text{ for } i = 1, ..., n, \text{ and } t = 1, ..., T, \tag{50}$$

where f_t is the market excess returns and $\varepsilon^s_{i,t}$ is the error term. In Equation (50), we impose the intercepts to be zero to satisfy the no-arbitrage restrictions for tradable factors. The $n \times 1$ error vectors ε_t^s are independent across time and Gaussian with mean zero and variance-covariance matrix Ω_B^s. We apply our diagnostic criterion on every simulated dataset of excess returns. Since the panel is balanced, we do not need to fix $\chi_{2,T}$. We only use $\chi_{1,T} = 15$. However, this trimming level does not affect the number of assets n in the simulations.

In order to study the properties under $\mathcal{M}_2$, we generate data under a three-factor alternative hypothesis, i.e., two omitted factors, and then we estimate a one-factor model to get the residuals. We build the simulated dataset as above except that we use estimated loadings, variance, and covariances for the Fama-French model on the CRSP dataset instead of the CAPM estimates.

In order to understand how our diagnostic criterion works for different finite samples, we perform exercises combining different values of the cross-sectional dimension n and the time dimension T. Table 8 reports estimates of $Pr(\xi < 0|\mathcal{M}_1)$ and $Pr(\xi > 0|\mathcal{M}_2)$, i.e., selection probabilities of the correct model estimated from the simulated datasets.

Table 8: Selection probabilities, balanced case

T	150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500
$Pr(\xi_1 < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_2 < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_3 < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_1 < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_2 < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_3 < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_1 > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_2 > 0 \vert \mathcal{M}_2)$	0.9580	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi_3 > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_1 > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_2 > 0 \vert \mathcal{M}_2)$	0.9640	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\check{\xi}_3 > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

A.7.2 Unbalanced panel

Let us repeat similar exercises as in the previous section, but with unbalanced characteristics for the simulated datasets. We introduce these characteristics through a matrix of observability indicators $I^s \in \mathbb{R}^{n \times T}$. The matrix gathers the indicator vectors for the n randomly chosen assets. We fix the maximal sample size $T = 528$ as in the empirical application.

In the unbalanced setting, the excess returns $R^s_{i,t}$ of asset i at time t under $\mathcal{M}_1$ is:

$$R^s_{i,t} = b^s_i f_t + \varepsilon^s_{i,t}, \text{ if } I^s_{i,t} = 1, \text{ for } i = 1, ..., n, \text{ and } t = 1, ..., T, \tag{51}$$

where $I^s_{i,t}$ is the observability indicator of asset i at time t in simulation s. Under $\mathcal{M}_2$, we again replace CAPM estimates with estimates for the Fama-French model to get a three-factor alternative.

In Tables 9 and 10, we provide the operative cross-sectional and time-series sample sizes in the Monte-Carlo repetitions for trimming $\chi_{1,T} = 15$ and four different levels of trimming $\chi_{2,T}$. More precisely, in Table 9, we report the average number $\bar{n}^{\chi}$ of retained assets across simulations, as well as the minimum $\min(n^{\chi})$ and the maximum $\max(n^{\chi})$ across simulations (rounded). For the lowest level of trimming $\chi_{2,T} = T/12$, all assets are kept in all simulations, while for the level of trimming $\chi_{2,T} = T/60$ on average we keep about two thirds of the assets. In Table 10, we report the average across assets of the $\bar{T}_i$, that are the average time-series size T_i for asset i across simulations, as well as the min and the max of the $\bar{T}_i$. Since the distribution of T_i for an asset i is right-skewed, we also report the average across assets of the median T_i. For trimming level $\chi_{2,T} = T/60$, the average mean time-series size is about 180 months, while the average median time-series size is 140 months.

Table 7 reports estimates of $Pr(\xi < 0|\mathcal{M}_1)$ and $Pr(\xi > 0|\mathcal{M}_2)$. These probabilities are close to 1 for most combinations of cross-sectional sample size n and trimming level $\chi_{2,T}$. The detection probability for model $\mathcal{M}_2$ is low only for trimming level $\chi_{2,T} = T/240$ and cross-sectional sample sizes $n = 500, 1000$. In fact, in Table 9, we see that the operative sample size is too small in such cases (below 100 in all simulations). For $n = 3,000$, or larger, the probabilities $Pr(\xi < 0|\mathcal{M}_1)$ and $Pr(\xi > 0|\mathcal{M}_2)$ are 1 for all trimming levels.

Table 9: Operative cross-sectional sample size

trimming level	$\chi_{2,T} = \frac{T}{12}$					$\chi_{2,T} = \frac{T}{60}$				
n	500	1,000	3,000	6,000	9,000	500	1,000	3,000	6,000	9,000
$\bar{n}^{\chi}$	500	1,000	3,000	6,000	9,000	326	651	1,955	3,905	5,857
$\min(n^{\chi})$	500	1,000	3,000	6,000	9,000	299	613	1,890	3,820	5,823
$\max(n^{\chi})$	500	1,000	3,000	6,000	9,000	359	694	2,018	3,977	5,903
trimming level	$\chi_{2,T} = \frac{T}{120}$					$\chi_{2,T} = \frac{T}{240}$				
n	500	1,000	3,000	6,000	9,000	500	1,000	3,000	6,000	9,000
$\bar{n}^{\chi}$	194	388	1,161	2,325	3,488	65	128	386	772	1,158
$\min(n^{\chi})$	162	348	1,080	2,245	3,437	44	97	338	712	1,123
$\max(n^{\chi})$	223	434	1,223	2,398	3,533	88	162	442	826	1,185

Table 10: Operative time-series sample size

trimming level	$\chi_{2,T} = \frac{T}{12}$	$\chi_{2,T} = \frac{T}{60}$	$\chi_{2,T} = \frac{T}{120}$	$\chi_{2,T} = \frac{T}{240}$
mean$(\bar{T}_i)$	126	175	235	365
$\min(\bar{T}_i)$	113	158	216	331
$\max(\bar{T}_i)$	141	190	260	400
mean(median (T_i))	88	141	198	344