A DIAGNOSTIC CRITERION FOR APPROXIMATE FACTOR STRUCTURE

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

First draft: February 2014. This version: January 2017.

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 random coefficient panel models with interactive fixed effects under large cross-section and time-series dimensions. The empirical analysis runs on monthly and quarterly returns for about ten thousand US stocks from January 1968 to December 2011 for several time-invariant and time-varying specifications. For monthly returns, we can choose either among time-invariant specifications with at least four financial factors, or a scaled three-factor specification. For quarterly returns, we cannot select macroeconomic models without the market factor.

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

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

[a]Università della Svizzera Italiana (USI Lugano) and Swiss Finance Institute, [b] European Commission, Joint Research Centre, [c]University of Geneva 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 are grateful to Lu Zhang and Alex Horenstein for sharing their data with us. We thank L. Mancini, E. Sentana, M. Weidmer, participants at ESEM 2014, COMPSTAT 2014, EWM 2015, IMS-China conference 2015, Bristol Econometric Study Group 2015, ESWC 2015, Workshop “Pietro Balestra” on panel data, INFER workshop, DagStat 2016, CORE@50 Conference, SoFiE Annual Meeting 2016, ESEM 2016, FMA 2016 meeting Las Vegas, ICEEE 2017, and participants at seminars at the University of Chicago, University of Orléans, University of St Gallen, Groningen, Tilburg, EPFL, ULB, University of Venice, University College of London, CEMFI, Pittsburgh University, PennState University, Federal Reserve Board of Governors, University of Surrey, and University of Cergy Pontoise. *Disclaimer: The content of this article does not reflect the official opinion of the European Commission. Responsibility for the information and views expressed therein lies entirely with the authors.

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 (1996)). 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. If the set of observable factors is not correctly specified, the no-arbitrage restrictions derived from APT will not hold, and the risk premia estimated by the two-pass regression approach will be meaningless (see Appendix H in Gagliardini, Ossola, Scaillet (2016, GOS) for a discussion of misspecification in the two-pass methodology). Giglio and Xiu (2016) have proposed a three-pass methodology allowing consistent estimation by exploiting an invariance property in time invariant models with omitted factors in balanced panels. Given the large menu of factors available in the literature (the factor zoo of Cochrane (2011), see also Harvey et al. (2016), Harvey and Liu (2016)), we need a simple diagnostic criterion to decide whether we can feel comfortable with the chosen set of observable factors before proceeding further in the empirical analysis of large cross sectional equity data sets under the APT setting.

For models with unobservable (latent) factors only, 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) and Jan and Otter (2008) for a selection procedure in the generalized dynamic factor model of Forni et al. (2000)) has extended the toolkit available to researchers. A first strand of that literature focuses on consistent estimation procedures for the number of factors. Bai and Ng (2002) introduce a penalized least-squares strategy to estimate the number of factors, at least one (see Amengual and Watson (2007) to include dynamic factors). Onatski (2010) looks at the behavior of the adjacent eigenvalues to determine the number of factors when the cross-sectional dimension (n) and the time-series dimension (T) are comparable. Ahn and Horenstein (2013) opt for the same strategy and cover the possibility of zero factors. Caner and Han (2014) propose an estimator with a group bridge penalization to determine the number of unobservable factors. A second strand of that literature develops inference procedures for hypotheses on the number of latent factors. Onatski (2009) deploys a characterization of the largest eigenvalues of a Wishart-distributed covariance matrix with large dimensions in terms of the Tracy-Widom Law. To get a Wishart distribution, Onatski (2009) assumes either Gaussian errors or T much larger than n. Kapetanios (2010) uses subsampling to estimate the limit distribution of the adjacent eigenvalues. Harding (2013) uses free probability theory to derive analytic expressions for the limiting moments of the spectral distribution.

Our paper expands the first strand of the literature by developing a consistent estimation procedure for the number of latent factors in the error terms in a model with observable factors when the cross-section dimension can be much larger than the time series dimension. Concluding for zero omitted factors means weakly cross-sectionally correlated errors. We require $n = O(T^{\bar{\gamma}})$, $\bar{\gamma} > 0$, and $T = O(n^{\gamma})$, $\gamma \in (0, 1]$, which is equivalent to $C_1 n^{1/\bar{\gamma}} \leq T \leq C_2 n^{\gamma}$ for some positive constants C_1, C_2. The case $\gamma < 1$ implies $T/n = o(1)$, namely n is much larger than T, and the case $\bar{\gamma} = \gamma = 1$ implies that n and T are comparable. In our empirical application, we have monthly and quarterly returns for about ten thousand US stocks from January 1968 to December 2011, and this explains why we also investigate the setting $T/n = o(1)$. The asymptotic distribution of the eigenvalues is degenerate under the usual standardisation of the $T \times T$ covariance matrix by n^{-1} when the ratio T/n goes to zero as $T, n \to \infty$. In such a setting, we cannot exploit well-defined limiting characterizations (Marchenko-Pastur distribution, Tracy-Widom distribution) obtained when T/n converges to a strictly positive constant. Without such distributional characterizations, we do not see hope for testing procedures as developed by Onatski (2009). However, a key theoretical result of our paper is that we can still have an asymptotically valid selection procedure for the number of latent factors even in the presence of a degenerate distribution of the eigenvalues of the sample covariance matrix of the errors. We show that this extends to sample covariance matrices of residuals of an estimated linear model with observable factors in unbalanced panels. An extension to residuals instead of true errors is not trivial since we need to cope with a projection matrix in the estimated errors, and there are little results about the analysis of the spectrum of matrix products (as opposed to the many results for matrix sums). The unbalanced nature makes things worse since we also have to take care of the matrix of observability indicators in the product. This further explains why we shy away from putting an additional structure on the errors (Onatski (2010), Ahn and Horenstein (2013)) or estimated errors in unbalanced panels in our assumptions. Most of our assumptions are weaker, and the arguments developed in the proofs of the theorems supporting our extension are new to the literature as further commented below.

For applications of factor models in empirical finance, Bai and Ng (2006) analyze statistics to test whether the observable factors in time-invariant models span the space of unobservable factors (see also Lehmann and Modest (1988) and Connor and Korajczyk (1988)). 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 et al. (forthcoming, 2016) 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. Gonçalves et al. (2015) consider bootstrap prediction intervals for factor models. Factor analysis for large cross-sectional datasets also find applications in studying bond risk premia (Ludvigson and Ng (2007, 2009)) and measuring time-varying macroeconomic uncertainty (Jurado et al. (2015)). Connor et al. (2012) showed that large cross sections exploit data more efficiently in a semiparametric characteristic-based factor model of stock returns. Recent papers (Fan et al. (2015), Pelger (2015), Ait-Sahalia and Xiu (forthcoming, 2016)) have also investigated large-dimensional factor modeling with in-fill asymptotics for high-frequency data.

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 main theoretical contribution shows that step 3) yields asymptotically the correct model selection. The mechanics of the selection are easy to grasp. If we have an approximate factor structure, we expect a vanishing largest eigenvalue because of a lack of a common signal in the error terms. So, if we take a penalizing term with a slower rate towards zero, a negative criterion points to weak cross-sectional correlation. On the contrary, the largest eigenvalue remains bounded from below away from zero if we face omitted factors. We have at least one non vanishing eigenvalue because of a common signal due to omitted factors. The positive largest eigenvalue dominates the vanishing penalizing term, and this explains why we conclude against weak cross sectional correlation when the criterion is positive. We also propose a general version of the diagnostic criterion that determines the number of omitted common factors. As shown below, the criterion coincides with the penalized least-squares approach of Bai and Ng (2002) applied on residuals of an unbalanced panel. 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 et al. (1995)). The panel data model is sufficiently general to accommodate both time-invariant and time-varying factor structures. Allowing for time-varying factor loadings presents challenges for finance theory and econometric modelling, but GOS explain how to solve these issues and give empirical evidence of time-varying risk premia. We recast the factor models as generic random coefficient panel models and develop the theory for large cross-section and time-series dimensions with $n = O(T^{\bar{\gamma}})$, $\bar{\gamma} > 0$, and $T = O(n^{\gamma})$, $\gamma \in (0, 1]$. Omitted latent factors are also called interactive fixed effects in the panel literature (Pesaran (2006), Bai (2009), Moon and Weidner (2015), Gobillon and Magnac (2016)). King et al. (1994) use them to capture the correlation between the unanticipated innovations in observable descriptors of economic performance (e.g. industrial production, inflation, etc.) and stock returns.

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 returns from January 1968 to December 2011. We look at a variety of empirical factors and we build factor models popular in the empirical literature to explain monthly and quarterly equity returns. They differ by the choice of the observable factors. We analyze monthly returns using recent financial specifications such as the five factors of Fama and French (2015), the profitability and investment factors of Hou et al. (2015), the quality minus junk and bet against beta factors of Asness et al. (2014) and Frazzini and Pedersen (2014), as well as other specifications described below. We analyze quarterly returns using macroeconomic specifications including consumption growth (CCAPM), market returns and consumption growth (Epstein and Zin (1989)), the three factors in Yogo (2006), the three factors in Li et al. (2006), and the five factors of Chen et al. (1986). We study time-invariant and time-varying versions of the financial 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)). For monthly returns, our diagnostic criterion is met by time-invariant specifications with at least four financial factors, and a scaled three-factor FF time-varying specification. For quarterly returns, we cannot select macroeconomic models without the market factor.

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 explains how to implement the criterion in practice and how to design a simple graphical diagnostic tool related to the well-known scree plot in principal component analysis. Section 6 contains the empirical results. In Appendices 1 and 2, we gather the theoretical assumptions and some proofs. We use high-level assumptions on cross-sectional and serial dependence of error terms, and show in Appendix 3 that we meet them under a block cross-sectional dependence structure 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 include some Monte-Carlo simulation results under a design mimicking our empirical application to show the practical relevance of our selection procedure in finite samples. The additional empirical results, discussed but not reported in the paper, are available on request.

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 $Z_{\underline{t-1}}$ and $Z_{\underline{t-1}}(\gamma)$, 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, $\varepsilon_{i,t} = \varepsilon_t(\gamma_i)$ be the error terms, and $Z_{i,t} = Z_t(\gamma_i)$ be the stock specific instruments. 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 $x_{1,i,t} = \left(vech\,[X_t]', Z'_{t-1} \otimes Z'_{i,t-1}\right)' \in \mathbb{R}^{d_1}$ 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 $d_2 = K(p+q)$. In vector $x_{2,i,t}$, the first components with common instruments take the interpretation of scaled factors (Cochrane (2005)), while the second components do not since they depend on i. The 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, $k, l = 1, \ldots, p$, where $Z_{t,k}$ denotes the kth component of the vector Z_t. 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). The vector of coefficients β_i is a function of asset specific parameters defining the dynamics of $a_{i,t}$ and $b_{i,t}$ detailed in GOS. 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, and the regression coefficient vector is $\beta_i = (a_i, b_i')'$ of dimension $d = K + 1$.

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 fully capture the systematic risk, and the error terms do not feature pervasive forms of cross-sectional dependence (see Assumption 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 latent factors may include scaled factors to cover latent time-varying factor loadings with common instruments. We cannot allow for latent time-varying factor loadings with stock-specific instruments because of identification issues. In (5), the θ_i's are also called interactive fixed effects in the panel literature. 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 condition in linear latent factor models (see Assumption A in Bai and Ng (2002)) and we can normalize matrix Σ_h to be the identity matrix I_m for identification. 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) via an analogy with the partitioned regression: $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$. 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 $M_{X_1} = I_n - X_1\left(X_1'X_1\right)^{-1}X_1'$. Condition (7) is tantamount 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 (2015) in a model with homogeneous regression coefficients $\beta_i = \beta$ for all i, and Ando and Bai (2015) for heterogeneous β_i in balanced panels). 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, and is then not an assumption. 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 $x_{i,t} = (x_t', \tilde{x}_{i,t}')'$, where $x_t := (vech[X_t]', f_t' \otimes Z_{t-1}')'$ 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 contemporaneous 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 applied 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 (Assumption A.10). 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 et al. (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 discuss them in Section 5. In vector $\bar{\varepsilon}_i$, the unavailable residuals are replaced by zeros. We use the following assumption on n and T.

Assumption 3 *The cross-sectional dimension n and time series dimension T are such that $n = O(T^{\bar{\gamma}})$, $\bar{\gamma} > 0$, and $T = O(n^{\gamma})$, $\gamma \in (0,1]$.*

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: We select $\mathcal{M}_1$ if $\xi < 0$, and we select $\mathcal{M}_2$ if $\xi > 0$, since under Assumptions 1-3 and Assumptions A.1-A.10, (a) $Pr\left(\xi < 0 \mid \mathcal{M}_1\right) \to 1$, and (b) $Pr\left(\xi > 0 \mid \mathcal{M}_2\right) \to 1$, when $n,T \to \infty$.*

Proposition 1 characterizes an asymptotically valid model selection rule, which treats both models symmetrically. The model selection rule is valid since parts (a) and (b) of Proposition 1 imply $Pr\left(\mathcal{M}_1 | \xi < 0\right) = Pr\left(\xi < 0 | \mathcal{M}_1\right) Pr\left(\mathcal{M}_1\right) \left[Pr\left(\xi < 0 | \mathcal{M}_1\right) Pr\left(\mathcal{M}_1\right) + Pr\left(\xi < 0 | \mathcal{M}_2\right) Pr\left(\mathcal{M}_2\right)\right]^{-1} \to 1$, as $n,T \to \infty$, by Bayes Theorem. Similarly, we have $Pr\left(\mathcal{M}_2 | \xi > 0\right) \to 1$. The diagnostic criterion in Proposition 1 is not a testing procedure since we do not use a critical region based on an asymptotic distribution and a chosen significance level. The zero threshold corresponds to an implicit critical value yielding a test size asymptotically equal to zero since $Pr(\xi < 0 | \mathcal{M}_1) \to 1$. The selection procedure is conservative in diagnosing zero factor by construction. We do not allow type I error under $\mathcal{M}_1$ asymptotically, and really want to ensure that there is no omitted factor as required in the APT setting. This also means that we will not suffer from false discoveries related to a multiple testing problem (see e.g. Barras et al. (2010), Harvey et al. (2016)) in our empirical application where we consider a large variety of factor models on monthly and quarterly data. However, a possibility to achieve p-values is to use a randomisation procedure (see Bandi and Corradi (2014) and Corradi and Swanson (2006)). This type of procedure controls for an error of the first type, conditional on the information provided by the sample and under a randomness induced by auxiliary experiments.

The proof of Proposition 1 shows that the largest eigenvalue in (8) vanishes at a faster rate (Lemma 1 in Appendix A.2.1) 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 (Lemma 4 in Appendix A.2.1) 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 $\mathcal{M}_1$. We can rewrite (8) as $\xi = SS_0 - SS_1 - g\left(n,T\right)$, where $SS_0 = \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2$ is the sum of squared errors and $SS_1 = \min \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^{\chi}\left(\bar{\varepsilon}_{i,t} - \theta_i h_t\right)^2$, where the minimization is w.r.t. the vectors $H \in \mathbb{R}^T$ of factor values and $\Theta = (\theta_1, ..., \theta_n)' \in \mathbb{R}^n$ of factor loadings in a one-factor model, subject to the normalization constraint $\frac{H'H}{T} = 1$. Indeed, the largest eigenvalue $\mu_1\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)$ 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, $\xi = PC\left(0\right) - PC\left(1\right)$, where $PC\left(0\right) = SS_0$, and $PC\left(1\right) = SS_1 + g\left(n,T\right)$.

Lemma 1 in Appendix A.2.1 gives an asymptotic upper bound on the largest eigenvalue of a symmetric matrix based on similar arguments as in Geman (1980), Yin et al. (1988), and Bai and Yin (1993) without exploiting distributional results from random matrix theory valid when n is comparable with T. This exemplifies a key difference with the proportional asymptotics used in Onatski (2010) or Ahn and Horenstein (2013) for balanced panel without observable factors. In Proposition 1, when $\gamma < 1$, 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 either monthly, or quarterly, returns). Another key difference w.r.t. the available literature is the handling of unbalanced panels. We need to address explicitly the presence of the observability indicators $I_{i,t}$ and the trimming devices $\mathbf{1}_i^{\chi}$ in the proofs of the asymptotic results.

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 et al. (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 smaller 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 $\theta_i = \bar{\theta}_i/T^c$ with $c > 0$. Let us assume that $\frac{1}{n}\sum_i \bar{\theta}_i^2$ 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 $c_1/T^{2c} \leq \mu_1 \left(\frac{1}{nT}\sum_i \theta_i^2 HH'\right) \leq c_2/T^{2c}$, $w.p.a.$ 1, 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:

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

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^{-2c}, namely $C_{nT}^{-2} + \bar{\chi}_T T^{-1} = o\left(T^{-2c}\right)$ and $g(n,T) = o\left(T^{-2c}\right)$. To conclude $\mathcal{M}_1$, we need that $g(n,T)$ is the dominant term, namely $T^{-2c} = o\left(g(n,T)\right)$ 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 $c < 1/2$ and $\mathcal{M}_1$ if $c > 1/2$. This means that detecting a weak factor loading structure is difficult if c is not sufficiently small. The factor loadings 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 4 *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 $\mathcal{M}_2(k)$ with $k \geq 1$ 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 $\mathcal{M}_2(k)$, with $m \geq k+1$. Assumption 4 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:

$$\xi(k) = \mu_{k+1}\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) - g(n,T). \tag{9}$$

As discussed in Ahn and Horenstein (2013) (see also Onatski (2013)) for balanced panels, we can rewrite (9) as $\xi(k) = SS_k - SS_{k+1} - g(n,T)$ where $SS_k = \min \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^{\chi}\left(\bar{\varepsilon}_{i,t} - \theta_i' h_t\right)^2$ and the minimization is w.r.t. $H \in \mathbb{R}^{T\times k}$ and $\Theta = (\theta_1, ..., \theta_n)' \in \mathbb{R}^{n\times k}$. 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 $PC(k+1) = SS_{k+1} + (k+1)g(n,T)$.

The following model selection rule extends Proposition 1.

Proposition 2 *Model selection rule: We select $\mathcal{M}_1(k)$ if $\xi(k) < 0$, and we select $\mathcal{M}_2(k)$ if $\xi(k) > 0$, since under Assumptions 1(i), 2-4, and Assumptions A.1-A.11, (a) $Pr[\xi(k) < 0|\mathcal{M}_1(k)] \to 1$ and (b) $Pr[\xi(k) > 0|\mathcal{M}_2(k)] \to 1$, when $n, T \to \infty$.*

The proof of Proposition 2 is also more complicated than the proof of Proposition 1. We need additional arguments to derive an asymptotic upper bound when we look at the $(k+1)$th eigenvalue of a symmetric matrix (Lemma 5 in Appendix A.2.2). We rely on the Courant-Fischer min-max theorem and Courant-Fischer formula (see beginning of Appendix 2) which represent eigenvalues as solutions of constrained quadratic optimization problems. We know that the largest eigenvalue $\mu_1(A)$ of a symmetric positive semi-definite matrix A is equal to its operator norm. There is no such norm interpretation for the smaller eigenvalues $\mu_k(A)$, $k \geq 2$. We cannot directly exploit standard inequalities or bounds associated to a norm when we investigate the asymptotic behavior of the spectrum beyond its largest element. We cannot either exploit distributional results from random matrix theory since we also allow for $T/n = o(1)$. The slow convergence rate $\sqrt{T}$ for the individual estimates $\hat{\beta}_i$ also complicates the proof. In the presence of homogeneous regression coefficients $\beta_i = \beta$ for all i, the estimate $\hat{\beta}$ in Bai (2009) and Moon and Weidner (2015) has a fast convergence rate $\sqrt{nT}$. In that case, controlling for the estimation error in $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} + x_{i,t}'(\beta - \hat{\beta})$ is straightforward due to the small asymptotic contribution of $(\beta - \hat{\beta})$. The approach of Onatski (2010) requires the convergence of the upper edge of the spectrum (i.e., the first k largest eigenvalues of the covariance matrix, with $k/T = o(1)$) to a constant, while the approach of Ahn and Horenstein (2013) requires an asymptotic lower bound on the eigenvalues. Extending these approaches for residuals of an unbalanced panels when $T/n = o(1)$ looks challenging.

We can use the results of Proposition 2 in order to estimate the number of unobservable factors. It suffices to choose the minimum k such that $\xi(k) < 0$. The next proposition states the consistency of that estimate even in the presence of a degenerate distribution of the eigenvalues when $T/n = o(1)$.

Proposition 3 *Let $\hat{k} = \min\{k = 0, 1, ..., T-1 \,:\, \xi(k) < 0\}$, where $\hat{k} = T$ if $\xi(k) \geq 0$ for all $k \leq T-1$. Then, under Assumptions 1(i), 2-4, and Assumptions A.1-A.11, and under $\mathcal{M}_1(k_0)$, we have $P[\hat{k} = k_0] \to 1$, as $n, T \to \infty$.*

In Proposition 3, we do not need to give conditions on the growth rate of the maximum possible number $kmax$ of factors as in Onatski (2010) and Ahn and Horenstein (2013). We believe that this is a strong advantage since there are many possible choices for $kmax$ and the estimated number of factors 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 factors with at least 99 percent chances in most Monte Carlo experiments when n is comparable or much larger than T.

5 Implementation and graphical diagnostic tool

In this section, we discuss how we can implement the model selection rule in practice and design a simple graphical diagnostic tool to determine the number of unobservable factors (see Figures 1 and 3 in the next section). Let us first recognize that

$$\hat{\sigma}^2 = \frac{1}{nT}\sum_i \sum_t \mathbf{1}_i^{\chi} \bar{\varepsilon}_{i,t}^2 = tr\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right) = \sum_{j=1}^{T}\mu_j\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right).$$

The ratio $\mu_j\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)/\hat{\sigma}^2$ gauges the contribution of the jth eigenvalue in percentage of the variance $\hat{\sigma}^2$ of the residuals. Similarly, the sum $\sum_{j=1}^{k}\mu_j\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\varepsilon}_i\bar{\varepsilon}_i'\right)/\hat{\sigma}^2$ gauges the cumulated contribution of the k largest eigenvalues in percentage of $\hat{\sigma}^2$. From Proposition 2, when all eigenvalues in that sum are larger than $g(n, T)$, this is equal to the percentage of $\hat{\sigma}^2$ explained by the k unobservable factors.

Therefore, we suggest to work in practice with rescaled eigenvalues which are more informative. We can easily build a scree plot where we display the rescaled eigenvalues $\mu_j \left(\frac{1}{nT} \sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) / \hat{\sigma}^2$ in descending order versus the number of omitted factors k, and use the horizontal line set at $g(n,T)/\hat{\sigma}^2$ as the cut-off point to determine the number of omitted factors. This yields exactly the same choice as the one in Proposition 3. The asymptotic validity of the selection rule in unaffected since $\hat{\sigma}^2$ converges to a strictly positive constant when $n, T \to \infty$. Such a scree plot helps to visually assess which unobservable factors, if needed, explain most of the variability in the residuals. We can set $g(n,T)/\hat{\sigma}^2 = \left(\frac{n+T}{nT} \right) \ln \left(\frac{nT}{n+T} \right)$ following a suggestion in Bai and Ng (2002). Those authors propose two other potential choices $\left(\frac{n+T}{nT} \right) \ln C_{nT}^2$ and $\left(\frac{\ln C_{nT}^2}{C_{nT}^2} \right)$. In our empirical application, n is much larger than T, and they yield identical results.

In Section 3, we saw that $\xi = SS_0 - SS_1 - g(n,T)$. Given such an interpretation in terms of sums of squared errors, we can think about another diagnostic criterion based on a logarithmic version $\check{\xi}$ as in Corollary 2 of Bai and Ng (2002). The second diagnostic criterion is

$$\check{\xi} = \ln\left(\hat{\sigma}^2\right) - \ln\left(\hat{\sigma}^2 - \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) \right) - g(n,T). \tag{10}$$

We get $\hat{\sigma}^2 = SS_0$, and $\check{\xi} = \ln(SS_0/SS_1) - g(n,T)$ 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. For the logarithmic version, Bai and Ng (2002) suggest to use the penalty $g(n,T) = \left(\frac{n+T}{nT} \right) \ln \left(\frac{nT}{n+T} \right)$ since the scaling by $\hat{\sigma}^2$ is implicitly performed by the logarithmic transformation of SS_0 and SS_1. Since, from Equation (10) $\check{\xi} = \ln \left(1 / \left(1 - \mu_1 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i' \right) / \hat{\sigma}^2 \right) \right)$ $-g(n,T)$ and x is close to $\ln(1/(1-x))$ for a small x, we see that a rule based on the rescaled criterion $\xi/\hat{\sigma}^2$ is closely related to the logarithmic version when the rescaled eigenvalue is small. This further explains why we are in favour of working in practice with rescaled eigenvalues.

Prior to computation of the eigenvalues, Bai and Ng (2002) advocate each series to be demeaned and standardize to have unit variance (see also Section 4 in King et al. (1994)). In our setting, each time series of residuals $\bar{\varepsilon}_{i,t}$ have zero mean by construction, and we also standardize them to have unit variance over the sample of T observations before computing the eigenvalues. Working with $\bar{\bar{\varepsilon}}_{i,t} = \bar{\varepsilon}_{i,t} / \sqrt{\frac{1}{T} \sum_t \bar{\varepsilon}_{i,t}^2}$

ensures that all series of residuals have a common scale of measurement and improves the stability of the information extracted from the multivariate time series (see e.g. Pena and Poncela (2006)). Since $tr\left(\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) = n^\chi T$ with $n^\chi = \sum_i \mathbf{1}_i^\chi$, we suggest to work with the normalised matrix $\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'$, so that the variance $\frac{1}{n^\chi T}\sum_i \sum_t \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_{i,t}^2$ of the scaled residuals is 1 by construction, and we can interpret $\mu_j\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$ directly as percentage of the variance of the normalised residuals.

From Johnstone (2001), we know that for a matrix of residuals, all of whose entries are independent standard Gaussian variates in a balanced panel, the distribution of the largest eigenvalue of the corresponding Wishart variable suitably normalized approaches the Tracy-Widom law of order 1 under proportional asymptotics. That result implies that, for such standard Gaussian residuals, the largest eigenvalue that we compute should be approximately $1/T$ if T is smaller than n (see also Geman (1980)) without the need to rely on a scaling by an estimated variance $\hat{\sigma}^2$. This further explains why we are in favor of working with standardised residuals, so that we are as close as possible to a standardized Gaussian reference model. This is akin to use the standard rule of thumb based on a Gaussian reference model in nonparametric density estimation (Silverman (1986)). We know the rate of convergence of the kernel density estimate but need an idea of the constant to use that information for practical bandwidth choice. In our setting, we can set the constant to one, when we face independent standard Gaussian residuals. The Gaussian reference model also suggests to use the penalisation $g(n,T) = \frac{\left(\sqrt{n}+\sqrt{T}\right)^2}{nT} \ln\left(\frac{nT}{\left(\sqrt{n}+\sqrt{T}\right)^2}\right)$, and this is our choice in the empirical section with n^χ substituted for n.

Finally, we can also investigate the ratio $\mu_j^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^T \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$ and the cumulated contribution $\sum_{j=1}^k \mu_j^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^T \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$. The denominator corresponds to the square of the Frobenius (or Hilbert-Schmidt) norm of the matrix $\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\varepsilon}_i \bar{\varepsilon}_i'$ since the sum of the squared eigenvalues of a positive semidefinite symmetric matrix $A = (a_{ij})$ corresponds to $tr(A'A) = \sum_{i,j} a_{ij}^2$. Those quantities measure the contributions of the omitted factors in terms of the off-diagonal terms (correlation part) in addition to the diagonal terms (residual variance). Here we follow Fiorentini and Sentana

(2015, pages 158-159) who prefer to look at the fraction of the Frobenius norm instead of the usual fraction of the trace of the sample covariance matrix to judge the representativeness of principal components. King et al. (1994) use the Frobenius norm to decompose the sample covariance of residuals of a Vector AutoRegressive (VAR) model and obtain starting values for maximum likelihood estimation of the parameters of a factor model for the error terms. A selection rule based on the squared eigenvalues being above or below the squared penalty delivers the same diagnostic, but helps to gauge the impact on correlation explanation by the omitted factors.

6 Empirical results

In this section, we compute the diagnostic criteria and the number of omitted factors using a large variety of combinations of financial and macroeconomic factors. We estimate linear factor models using monthly and quarterly data from January 1968 to December 2011.

6.1 Factor models and data description

We consider several linear factor models that involve financial and macroeconomic variables. Let us start with the financial specifications listed in Table 1. We estimate these linear specifications using monthly data. We proxy the risk free rate with the monthly 30-day T-bill beginning-of-month yield. The three factors of Fama and French (1993) are the monthly excess return $r_{m,t}$ on CRSP NYSE/AMEX/Nasdaq value-weighted market portfolio over the risk free rate, and the monthly returns on zero-investment factor-mimicking portfolios for size and book-to-market, denoted by $r_{smb,t}$ and $r_{hml,t}$. The monthly returns on portfolio for momentum is denoted by $r_{mom,t}$ (Carhart (1997)). The two operative profitability factors of Fama and French (2015) are the difference between monthly returns on diversified portfolios with robust and weak profitability and investments, and with low and high investment stocks, denoted by $r_{rmw,t}$ and $r_{cma,t}$. We have downloaded the time series of these factors from the website of Kenneth French. We denote the monthly returns of size, investment, and profitability portfolios introduced by Hou et al. (2015) by $r_{me,t}$, $r_{I/A,t}$ and $r_{ROE,t}$ (see also Hou et al. (2014)). Furthermore, we include quality minus junk (qmj_t) and bet against beta (bab_t) factors as described in Asness et al. (2014) and Frazzini and Pedersen (2014). The factor return qmj_t is the average return on the two high quality portfolios minus the average return on the two low quality (junk) portfolios. The bet against beta factor is a portfolio that is long low-beta securities and short high-beta securities. We have downloaded these data from the website of AQR.

As additional specifications, we consider the two reversal factors which are monthly returns on portfolios for short-term and long-term reversals from the website of Kenneth French. Besides, the monthly returns of industry-adjusted value, momentum and profitability factors are available from the website of Robert Novy-Marx (see Novy-Marx (2013)). We also include the three liquidity-related factors of Pastor and Stambaugh (2002) that consist of monthly liquidity level, traded liquidity, and the innovation in aggregate liquidity. We have downloaded them from the website of Lubos Pastor.

In Table 2, we list the linear factor specifications that involve financial and macroeconomic variables. We estimate these specifications using quarterly data. We consider the aggregate consumption growth cg_t for the CCAPM (Lucas (1978), Breeden (1979)) and the Epstein and Zin (1989) model (see also Epstein and Zin (1991)), the durable and nondurable-consumption growth rate introduced by Yogo (2006) and denoted by dcg_t and $ndcg_t$. The investment factors used in Li et al. (2006) track the changes in the gross private investment for households, for non-financial corporate and for non-financial non-corporate firms, and are denoted by dhh_t, $dcorp_t$, and $dncorp_t$. Finally, we consider the five factors of Chen et al. (1986) available from the website of Laura Xiaolei Liu. Those factors are the growth rate of industrial production mp_t, the unexpected inflation ui_t, the change in the expected inflation dei_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.

To account for time-varying coefficients, we consider two conditional specifications: (i) $Z_{t-1} = (1, divY_{t-1})'$ and (ii) $Z_{t-1} = (1, divY_{t-1})'$, $Z_{i,t-1} = bm_{i,t-1}$, where $divY_{t-1}$ is the lagged dividend yield and the asset specific instrument $bm_{i,t-1}$ corresponds to the lagged book-to-market equity of firm i. We compute the firm characteristic from Compustat as in the appendix of Fama and French (2008). We refer to Avramov and Chordia (2006) for convincing theoretical and empirical arguments in favor of the chosen conditional specifications. The parsimony and the empirical results below explain why we have not included an additional firm characteristic such as the size of firm i.

As additional specifications, we consider the lagged default spread, term spread, monthly 30-day T-bill, aggregate consumption-to-wealth ratio (Lettau and Ludvigson (2001)), and labour-to-consumption ratio (Santos and Veronesi (2006)) as common instruments.

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 $T = 546$ months. We constructed the quarterly stock returns from the monthly data and $T = 176$. In order to account for the unbalanced characteristic, if the monthly observability indicators $I_{i,t}, I_{i,t+1}$ and $I_{i,t+2}$ are observed, we built the returns of the quarter $s = 1, 2, 3, 4$ as the average of the three monthly returns at time $t, t+1$ and $t+2$. Otherwise, the observability indicator of the quarter s takes value zero.

6.2 Results for financial models

In this section, we compute the diagnostic criteria for the linear factor models listed in Table 1. We fix $\chi_{1,T} = 15$ as advocated by Greene (2008) and $\chi_{2,T} = 546/60$, i.e., at least 60 months of return observations as in Bai and Ng (2002). In Table 3, we report the trimmed cross-sectional dimension n^{χ}. 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-products of factors f_t and instruments Z_{t-1}. These problems explain why we shrink from $n^{\chi} = 6,775$ for time-invariant models to around three thousand assets for time-varying models.

Table 4 reports the contribution in percentage of the first eigenvalue μ_1 with respect to the variance of normalized residuals $\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'$, that is equal to one by construction under our variance scaling to one for each time series of residuals. We also report the number of omitted factors k, the contribution of the first k eigenvalues, i.e., $\sum_{j=1}^{k}\mu_j$, and the incremental contribution of the $k+1$ eigenvalue μ_{k+1}. For each model, we also specify the numerical value of the penalisation function $g\left(n^{\chi}, T\right)$, as defined in Section 5.

Let us start with the results for the time-invariant specifications. The number k of omitted factors is larger than one for the most popular financial models, e.g., the CAPM (Sharpe (1964)), the three-factor Fama-French model (FF) and the four-factor Carhart (1997) model (CAR). On the contrary, for the recent proposals based on profitability and investment (5FF, HXZ), quality minus junk (QMJ), and bet against beta

(BAB) factors, we find no omitted latent factor. We observe that adding observable factors helps to reduce the contribution of the first eigenvalue μ_1 to the variance of residuals. However, when we face latent factors, the omitted systematic contribution $\sum_{j=1}^{k} \mu_j$ only accounts for a small proportion of the residual variance. For instance, we find $k = 2$ omitted factors in the CAPM. Those two latent factors only contribute to $\mu_1 + \mu_2 = 4.06\%$ of the residual variance. Figure 1 summarizes this information graphically by displaying the penalized scree plots and the plots of cumulated eigenvalues for the CAPM. For instance, $\mu_3 = 1.47\%$ lies below the horizontal line $g(n^\chi, T) = 1.79\%$ in Panel A for the time-invariant CAPM, so that $k = 2$. In Panel B for the time-invariant CAPM, the vertical bar $\mu_1 + \mu_2 = 4.06\%$ is divided into the contribution of $\mu_1 = 2.16\%$ (light grey area) and that of $\mu_2 = 1.90\%$ (dark grey area). Figure 2 Panel A displays the scree plots of squared eigenvalues for the CAPM and the square $g^2\left(n^\chi, T\right)$ of the penalisation function relative to the squared Frobenius norm $\sum_{l=1}^{T} \mu_l^2 \left(\frac{1}{nT} \sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i' \right)$. By construction, the conclusion of the number of omitted factor is the same as for the scree plot shown in Figure 1. From the plot of cumulated squared eigenvalues in Figure 2 Panel B, we conclude that the two omitted factors contribute more to the relative explanation of the correlation part than of the residual variance. For example, we get that the sum of the square of the two first eigenvalues accounts for 22.51% of the square of the Frobenius norm for the time-invariant CAPM. Thus, the two latent factors are much more representative of the off-diagonal components. We conclude similarly for the time-invariant FF model, even if the correlation explanation provided by the single omitted factor is lower.

For the time-varying specifications (i) and (ii) of Table 4, we still find one omitted factor for the CAPM. We see that the scaled three-factor FF model with $Z_{t-1} = (1, divY_{t-1})'$ passes the diagnostic criterion. The largest eigenvalue $\mu_1 = 1.37\%$ lies below the horizontal line $g(n^\chi, T) = 2.05\%$ in Figure 3 Panel B, and its square μ_1^2 only contributes to 5.80% of the square of the Frobenius norm in Figure 4 Panel B for the scaled three-factor FF model, so that $k = 0$. The additional stock specific instrument $Z_{i,t-1} = bm_{i,t-1}$ is not necessary to exhaust the cross-sectional dependence. Hence, the empirical message of Table 4 is that we can choose either among time-invariant specifications with at least four financial factors, or a scaled FF model. The latter is more parsimonious for the factor space in the conditional sense ($K = 3$ versus $K = 4$), but less parsimonious for the parameter space ($d = 9$ versus $d = 5$). From an econometric point of view, it is not clear which parsimony we should favor to decide between the time-invariant specification (more factors, less parameters) and the time-varying specification (less factors, more parameters). From a finance point of view, the first one is better suited for static (unconditional) investment decisions while the second one is better suited for dynamic (conditional) investment decisions. The choice between the two models should meet the investor needs or answer the empirical research question at hand. For a balanced panel of monthly returns for $4,883$ stocks on the period January 1994 to December 1998 ($T = 60$), Bai and Ng (2002) find only two latent factors.

As observed in GOS, measures of limits-to-arbitrage and missing factor impact (not reported here) like those in Pontiff (2006), Lam and Wei (2011), Ang et al. (2010), Stambaugh et al. (2015) decrease with the number of observable factors.

Concerning the additional factors and instruments mentioned in Section 6.1, none of them allow to reach a more parsimonious factor structure in a time-invariant or time-varying setting. Moreover, neither the time-invariant CAPM, FF and CAR models, nor their time-varying specifications with term spread, default spread and book-to-market equity used in GOS, pass the diagnostic criterion. As conjectured in GOS, this might be one reason for the rejection of the asset pricing restrictions.

6.3 Results for macroeconomic models

In this section, we perform the empirical exercises on the macroeconomic linear factor models listed in Table 2. We fix $\chi_{1,T} = 15$ and $\chi_{2,T} = 176/20$, i.e., at least 20 quarterly return observations. In Table 5, we report the trimmed cross-sectional dimension n^{χ}. The quarterly dataset has $6,707$ stocks with more than twenty quarterly observations. The trimming is driven by the multicollinearity between regressors, when $n^{\chi} < 6,707$. Table 5 further reports the empirical results for the macroeconomic models. The time-invariant specifications which include only macroeconomic variables (CCAPM, NDC and DC, LVX, and CRR) and exclude the market, do not pass the diagnostic criterion. We find $k = 2$ omitted factors. Moreover, $\mu_1 + \mu_2$ is about 15% of the residual variance in Table 5 and $\mu_1^2 + \mu_2^2$ accounts for 58% of the square of the Frobenius norm in Figure 5, in contrast to the 4.06% and 22.51% found for the time-invariant CAPM with monthly returns. The latent factors in the macro economic models are both representative of the residual variance (diagonal values) and the correlation part (off-diagonal values). When we incorporate the market (EZ and YO), we find no omitted latent factors. This is not surprising since, for quarterly data, the CAPM fully captures the systematic risk of individual stocks, with $\mu_1 = 3,15\%$, $g(n^\chi, T) = 3,74\%$, and $n^\chi = 6,707$. We do not report results for time-varying specifications. We have a limited sample size $T = 176$. Because of multicollinearity problems and the parameter dimension being up to $d = 14$, the estimation yields imprecise results. The trimmed sample size n^χ is often below T, which violates our large panel assumption.

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. *Journal of Financial and Quantitative Analysis*, forthcoming, 2016.

Y. Ait-Sahalia and D. Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, forthcoming, 2016.

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. Amengual and M. W. Watson. Consistent estimation of the number of dynamic factors in large N and T panel. *Journal of Business and economic statistics*, 25(1):91–96, 2007.

T. Ando and J. Bai. Asset pricing with a general multifactor structure. *Journal of Financial Econometrics*, 13(3):556–604, 2015.

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

A. Ang, J. Liu, and K. Schwarz. Using stocks or portfolios in tests of factor models. *Working Paper*, 2010.

C. S. Asness, A. Frazzini, and L.H. Pedersen. Quality minus junk. *Working paper*, 2014.

D. Avramov and T. Chordia. Asset pricing models and financial market anomalies. *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.

F. Bandi and V. Corradi. Nonparametric nonstationarity tests. *Econometric Theory*, 30:127–149, 2014.

L. Barras, O. Scaillet, and R. Wermers. False discoveries in mutual fund performance: measuring luck in estimated alphas. *Journal of Finance*, 65:179–216, 2010.

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.

D. Breeden. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7(3):265–296, 1979.

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

J. Y. Campbell. Understanding risk and return. *Journal of Political Economy*, 104(2):298–345, 1996.

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*, 32(3):359–374, 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. *Asset pricing*. Princeton University Press, Princeton, 2005.

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.

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

V. Corradi and N. Swanson. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. *Journal of Econometrics*, 132:195–229, 2006.

L. G. Epstein and S. E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica*, 57(4):937–969, 1989.

L. G. Epstein and S. E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an empirical analysis. *Journal of Political Economy*, 99(2):263–286, 1991.

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.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1): 1–22, 2015.

J. Fan, A. Furger, and D. Xiu. Incorporating global industrial classification standard into portfolio allocation:

A simple factor-based large covariance matrix estimator with high frequency data. *Journal of Business and Economic Statistics*, 34:489–503, 2015.

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.

G. Fiorentini and E. Sentana. *Tests for Serial Dependence in Static, Non-Gaussian Factor Models, in Unobserved Components and Time Series Econometrics*. Oxford University Press, Oxford, 2015.

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

A. Frazzini and L.H. Pedersen. Betting against beta. *Journal of Financial Economics*, 111:1–25, 2014.

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. *Econometrica*, 84(3):985–1046, 2016.

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.

S. Giglio and D. Xiu. Inference on risk premia in the presence of omitted factors. *Booth School of Business working paper*, 2016.

L. Gobillon and T. Magnac. Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics, forthcoming*, 2016.

S. Gonçalves, B. Perron, and A. Djogbenou. Bootstrap prediction intervals for factor models. *Journal of*

Business and Economic Statistics, forthcoming, 2015.

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.

M. Harding. Estimating the number of factors in large dimensional factor models. *Working Paper*, 2013.

C. R. Harvey and Y. Liu. Lucky factors. *Working Paper*, 2016.

C. R. Harvey, Y. Liu, and H. Zhu. ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.

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

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

K. Hou, C. Xue, and L. Zhang. A comparison of new factor models. *Working Paper*, 2014.

K. Hou, C. Xue, and L. Zhang. Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705, 2015.

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.

J. Jan and P. Otter. Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach. *Econometric reviews*, 4-6(27):385–397, 2008.

Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295 – 327, 2001.

K. Jurado, S. Ludvigson, and S. Ng. Measuring uncertainty. *The American Economic Review*, 105(3):

1177–1216, 2015.

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.

M. King, E. Sentana, and S. Wadhwani. Volatility and links between national stock markets. *Econometrica*, 62(4):901–933, 1994.

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

E. Lam and J. Wei. Limits-to-arbitrage, investment frictions, and the asset growth anomaly. *Journal of Financial Economics*, 102(1):127–149, 2011.

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.

Q. Li, M. Vassalou, and Y. Xing. Sector investment growth rates and the cross section of equity returns. *Journal of Business*, 21:2417–2448, 2006.

M. Lovell. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58:993–1010, 1963.

R. E. Lucas. Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445, 1978.

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

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

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

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

H. Moon and M. Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579, 2015.

R. Novy-Marx. The other side of value: the gross profitability premium. *Journal of Financial Economics*, 108:1–28, 2013.

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.

M. Pelger. Large-dimensional factor modeling based on high-frequency observations. *Working Paper, Stanford University*, 2015.

D. Pena and P. Poncela. Dimension reduction in multivariate time series. 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.

J. Pontiff. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42

(1-2):35–52, 2006.

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.

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

T. Santos and P. Veronesi. Labor income and predictable stock returns. *Review of Financial Studies*, 19: 1–44, 2006.

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

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

B. W. Silverman. *Density Estimation in Statistics and Data Analysis*. Chapman and Hall, London, 1986.

R. F. Stambaugh, J. Yu, and Y. Yuan. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance*, 70(5):1903–1948, 2015.

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 inequalities 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.

M. Yogo. A consumption-based explanation of expected stock returns. *Journal of Finance*, 612(2):539–580, 2006.

Table 1: Financial linear factor models

Model	Factors	K
CAPM	$r_{m,t}$	1
FF	$r_{m,t}, r_{smb,t}, r_{hml,t}$	3
CAR	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{mom,t}$	4
5FF	$r_{m,t}, r_{smb,t}, r_{hml,t}, r_{rmw,t}, r_{cma,t}$	5
HXZ	$r_{m,t}, r_{me,t}, r_{I/A,t}, r_{ROE,t}$	4
FF and QMJ	$r_{m,t}, r_{smb,t}, r_{hml,t}, qmj_t$	4
FF and BAB	$r_{m,t}, r_{smb,t}, r_{hml,t}, bab_t$	4

The table lists the linear factor models based on financial variables. We estimate these specifications by using monthly data. For each model, we report the factors labeling and their number K. FF, CAR, 5FF, HXZ, QMJ and BAB refer to the three Fama-French factors, the four Carhart factors, the five Fama-French factors, the four Hou-Xue-Zhang factors, quality minus junk factor, and bet against beta factor.

Table 2: Macroeconomic linear factor models

Model	Factors	K
CCAPM	cg_t	1
EZ	$r_{m,t}, cg_t$	2
NDC and DC	$ndcg_t, dcg_t$	2
YO	$r_{m,t}, ndcg_t, dcg_t$	3
LVX	$dhh_t, dcorp_t, dncorp_t$	3
CRR	$mp_t, ui_t, dei_t, uts_t, upr_t$	5

The table lists the linear factor models based on macroeconomic variables and the market. We estimate these specifications by using quarterly data. For each model, we report the factors labeling and their number K. EZ, NDC and DC, YO, LVX and CRR refer to the two Epstein-Zin factors, the two nondurable and durable consumption growth factors, the three Yogo factors, the three Li-Vassalou-Xing factors, and the five Chen-Roll-Ross factors.

Table 3: Trimmed cross-sectional dimensions n^{χ} and number d of parameters to estimate for financial models

Financial model	Time-invariant	Time-varying					
	n^{χ}	(i)	d	n^{χ}	(ii)	d	n^{χ}
CAPM	6,775		5	3,766		8	3,004
FF	6,775		9	3,536		14	2,780
CAR	6,775		11	3,468		17	2,608
5 FF	6,775		13	2,957		20	1,991
HXZ	6,775		11	3,344		17	2,612
FF and QMJ	6,775		11	3,365		17	2,423
FF and BAB	6,775		11	3,224		17	2,441

For each financial model of Table 1, we report the trimmed cross-sectional dimension n^{χ} for estimation from monthly data. For the time-varying specifications, we give the dimension d of vector $x_{i,t}$ and n^{χ} for two sets of instruments: (i) $Z_{t-1} = (1, divY_{t-1})'$ and (ii) $Z_{t-1} = (1, divY_{t-1})'$, $Z_{i,t-1} = bm_{i,t-1}$. For the time-invariant specifications, we have $d = K + 1$ (see Table 1).

Table 4: Results for time-invariant and time-varying financial models

Financial model	Panel A - *Time-invariant* μ_1	k	$\sum_{j=1}^{k}\mu_j$	μ_{k+1}	$g(n^{\chi},T)$	Panel B - *Time-varying*	μ_1	k	$\sum_{j=1}^{k}\mu_j$	μ_{k+1}	$g(n^{\chi},T)$
CAPM	2.16%	2	4.06%	1.47%	1.79%	(i)	2.87%	1	2.87%	1.79%	2.02%
						(ii)	3.00%	1	3.00%	1.98%	2.13%
FF	2.03%	1	2.03%	1.16%	1.79%	(i)	1.37%	0	0.00%	1.37%	2.05%
						(ii)	1.53%	0	0.00%	1.53%	2.17%
CAR	2.03%	1	2.03%	1.12%	1.79%	(i)	1.34%	0	0.00%	1.34%	2.05%
						(ii)	1.51%	0	0.00%	1.51%	2.20%
5FF	1.42%	0	0.00%	1.42%	1.79%	(i)	1.45%	0	0.00%	1.45%	2.13%
						(ii)	1.81%	0	0.00%	1.81%	2.37%
HXZ	1.43%	0	0.00%	1.43%	1.79%	(i)	1.35%	0	0.00%	1.35%	2.07%
						(ii)	1.54%	0	0.00%	1.54%	2.20%
FF and QMJ	1.39%	0	0.00%	1.39%	1.79%	(i)	1.33%	0	0.00%	1.33%	2.07%
						(ii)	1.60%	0	0.00%	1.60%	2.24%
FF and BAB	1.64%	0	0.00%	1.64%	1.79%	(i)	1.40%	0	0.00%	1.40%	2.09%
						(ii)	1.58%	0	0.00%	1.58%	2.24%

The table shows the contribution of the first eigenvalue μ_1 to the variance of normalised residuals, the number of omitted factors k, the contributions of the first k and of the $k+1$ eigenvalues, and the penalty term $g(n^{\chi},T)$. Panels A and B report empirical results for time-invariant and time-varying financial models estimated from monthly data, respectively. The time-varying specifications use two sets of instruments: (i) $Z_{t-1} = (1, divY_{t-1})'$ and (ii) $Z_{t-1} = (1, divY_{t-1})'$, $Z_{i,t-1} = bm_{i,t-1}$.

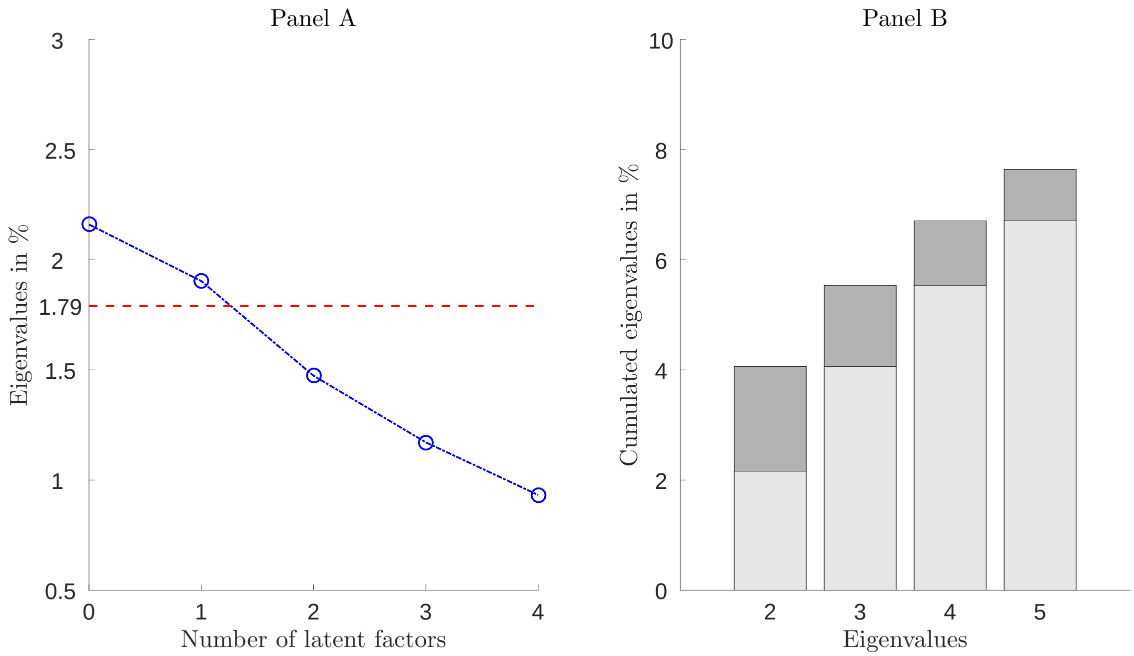
Table 5: Results for the macroeconomic models

Macroeconomic model	n^{χ}	μ_1	k	$\sum_{j}^{k} \mu_j$	μ_{k+1}	$g(n^{\chi}, T)$
CCAPM	$6,707$	8.12%	2	14.36%	3.29%	3.74%
EZ	$6,707$	3.07%	0	0.00%	3.07%	3.74%
NDC and DC	$6,306$	8.07%	2	14.21%	3.36%	3.76%
YO	$6,270$	3.38%	0	0.00%	3.38%	3.76%
LVX	$6,707$	7.96%	2	14.04%	3.34%	3.74%
CRR	$6,153$	6.42%	4	15.74%	1.13%	1.82%

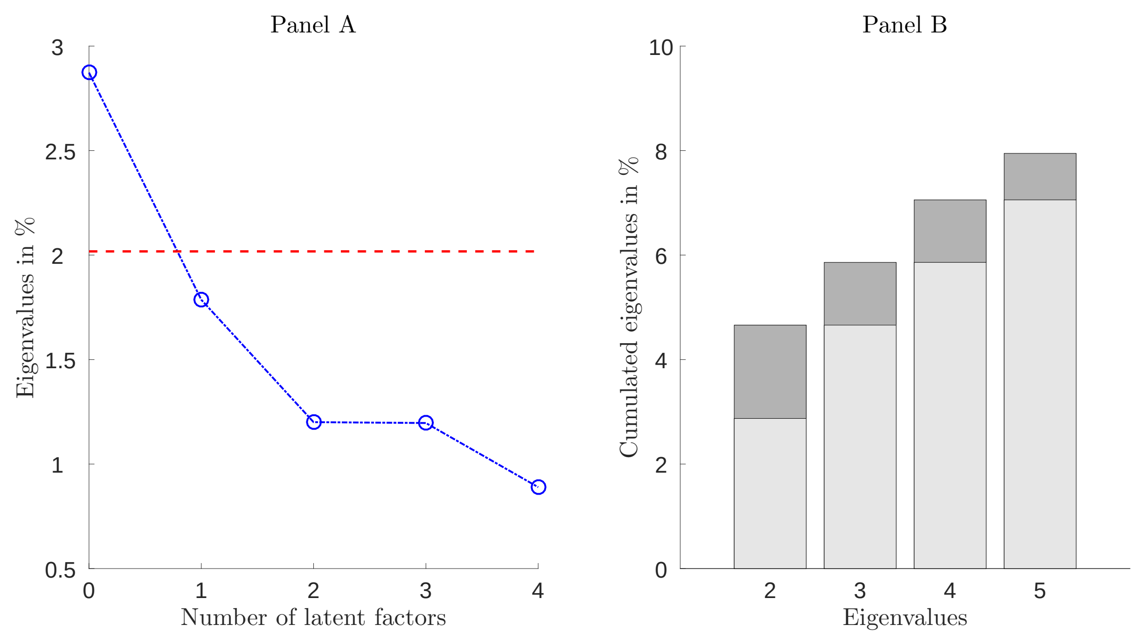
For each macroeconomic model of Table 2, we report the trimmed cross-sectional dimension n^{χ} for time-invariant specifications estimated from quarterly data. We further show the contribution of the first eigenvalue μ_1 to the variance of normalised residuals, the number of omitted factors k, the contributions of the first k and of the $k+1$ eigenvalues, and the penalty term $g(n^{\chi}, T)$.

Figure 1: Number of omitted factors and cumulated eigenvalues for the CAPM

Time-invariant CAPM



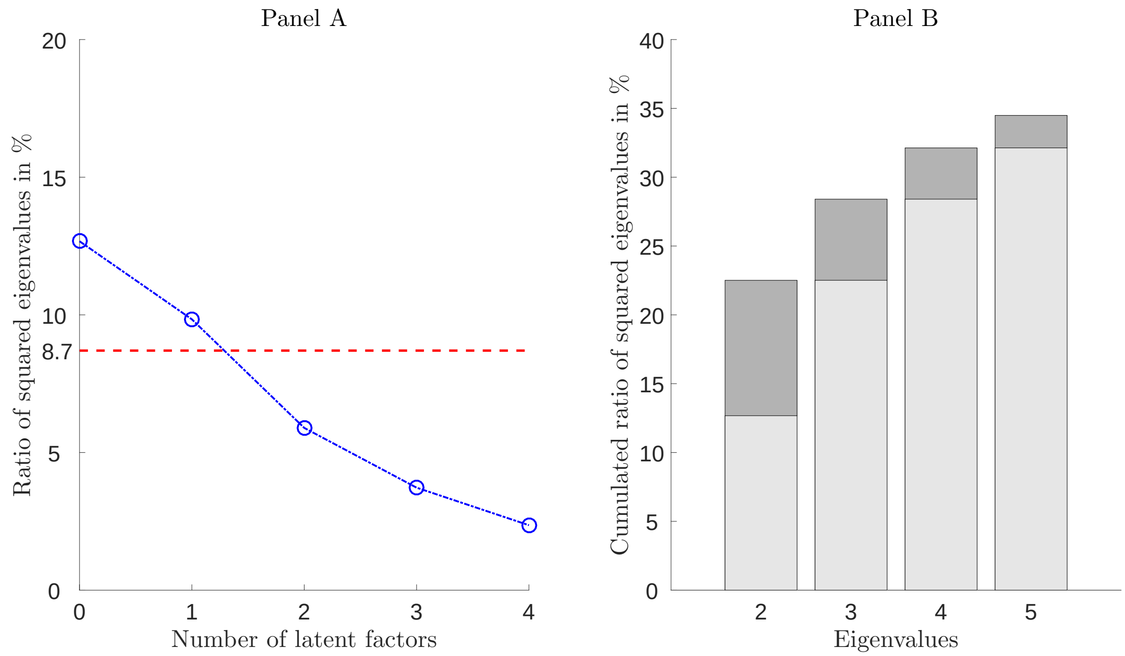
Time-varying CAPM



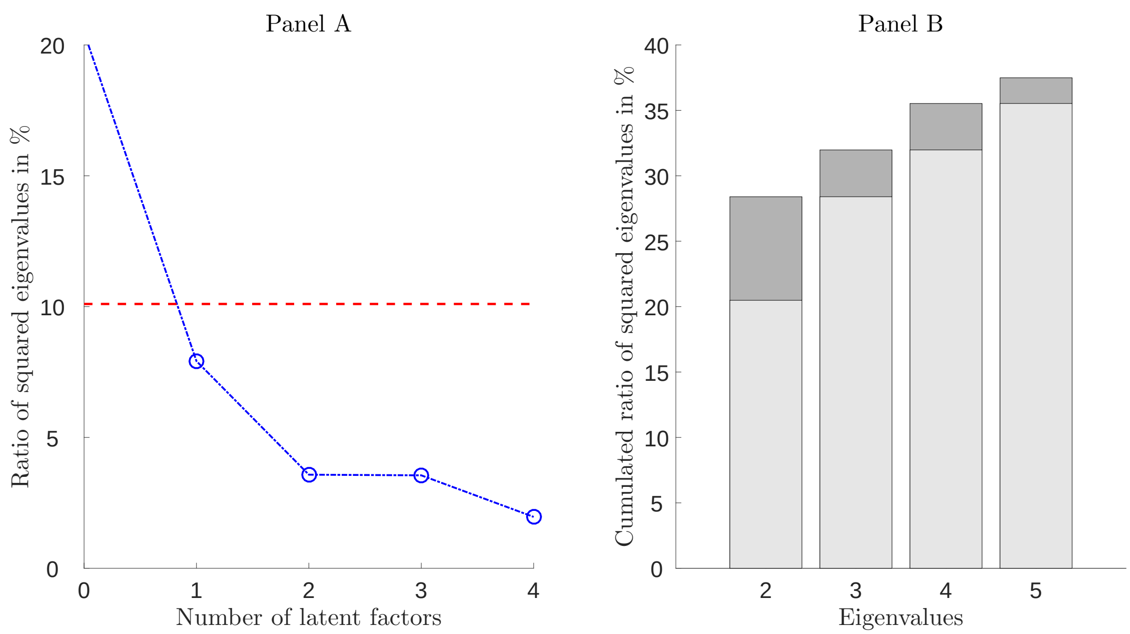
Panel A plots the scree-plot of the values of the first five eigenvalues in percentage, i.e. $\mu_j\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^{\chi}, T\right)$. Panel B plots the cumulated eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1}\mu_l\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$, the dark grey is the contribution of the jth eigenvalue in percentage. The figure reports results for the CAPM for the time-invariant and time-varying specifications with $Z_{t-1} = (1, divY_{t-1})'$.

Figure 2: Number of omitted factors and cumulated squared eigenvalues for the CAPM

Time-invariant CAPM



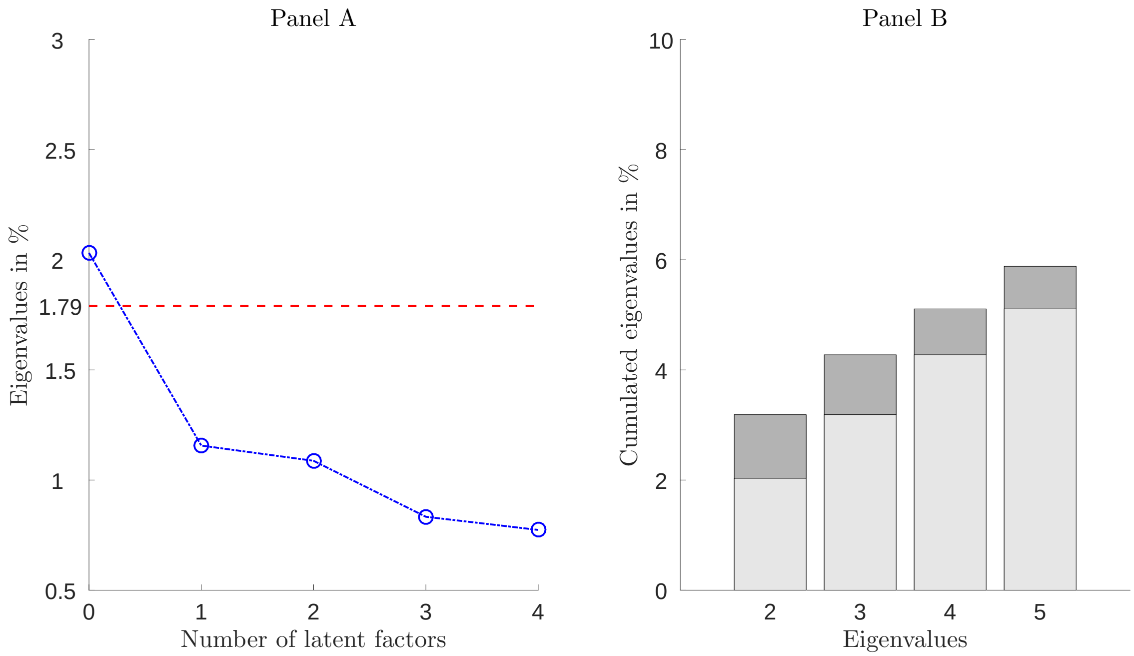
Time-varying CAPM



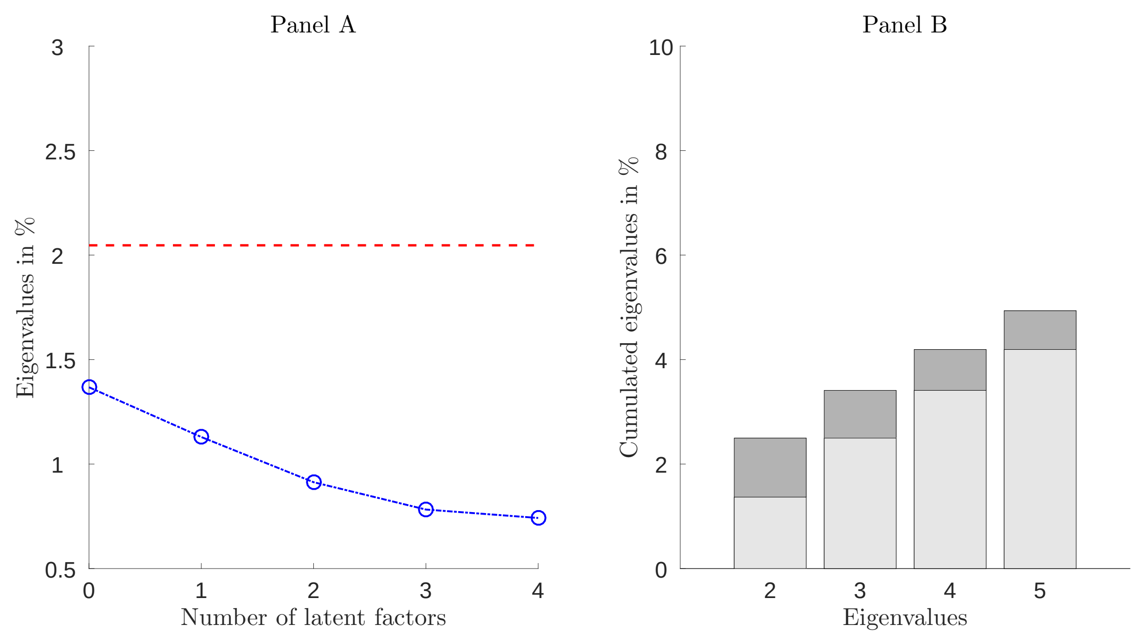
Panel A plots the scree-plot of the values of the first five squared eigenvalues in percentage, i.e. $\mu_j^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^\chi, T\right)^2 / \sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$. Panel B plots the cumulated squared eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1}\mu_l^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$, the dark grey is the contribution of the jth squared eigenvalue in percentage. The figure reports results for the CAPM for the time-invariant and time-varying specifications with $Z_{t-1} = (1, divY_{t-1})'$.

Figure 3: Number of omitted factors and cumulated eigenvalues for the FF model

Time-invariant FF

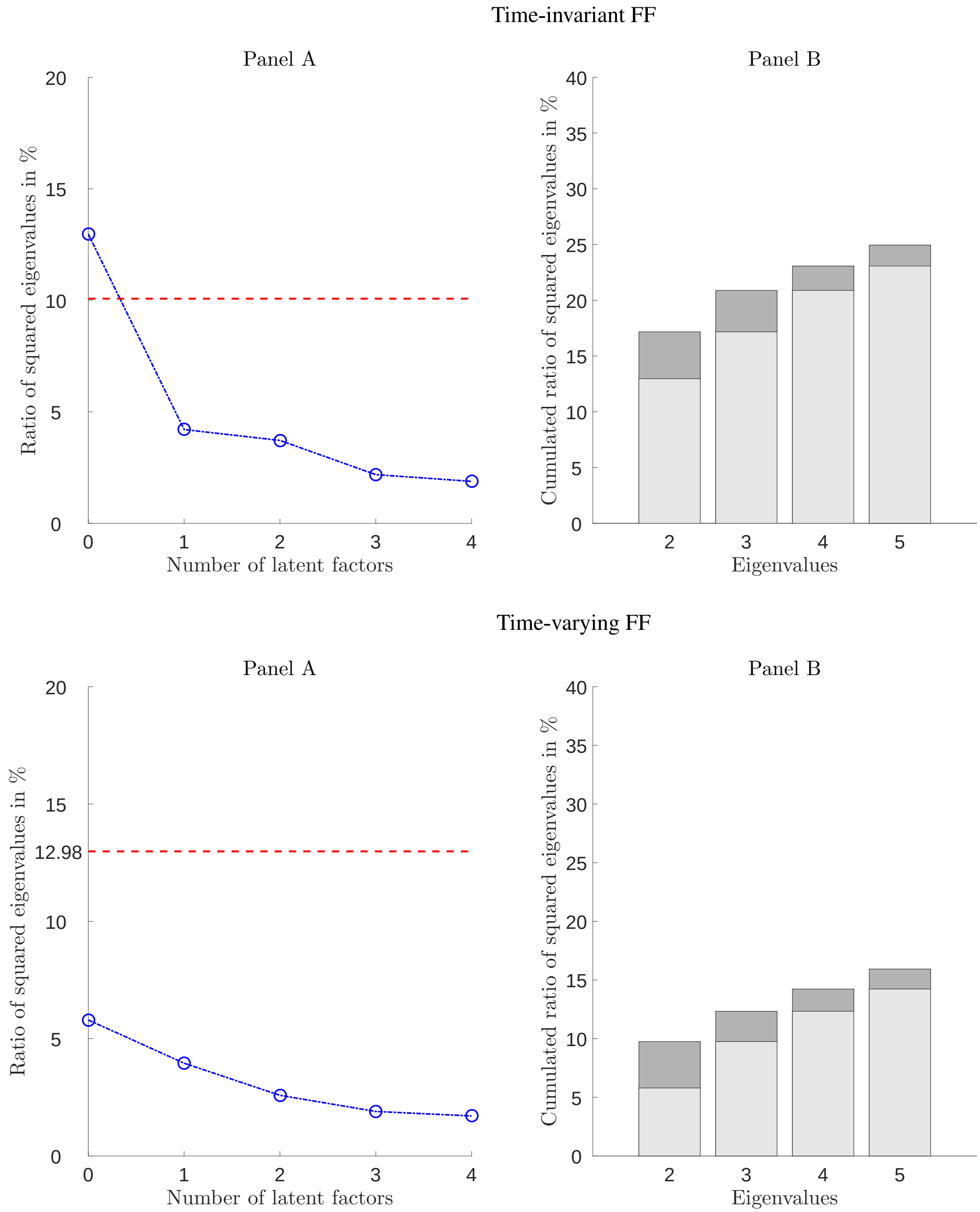


Time-varying FF



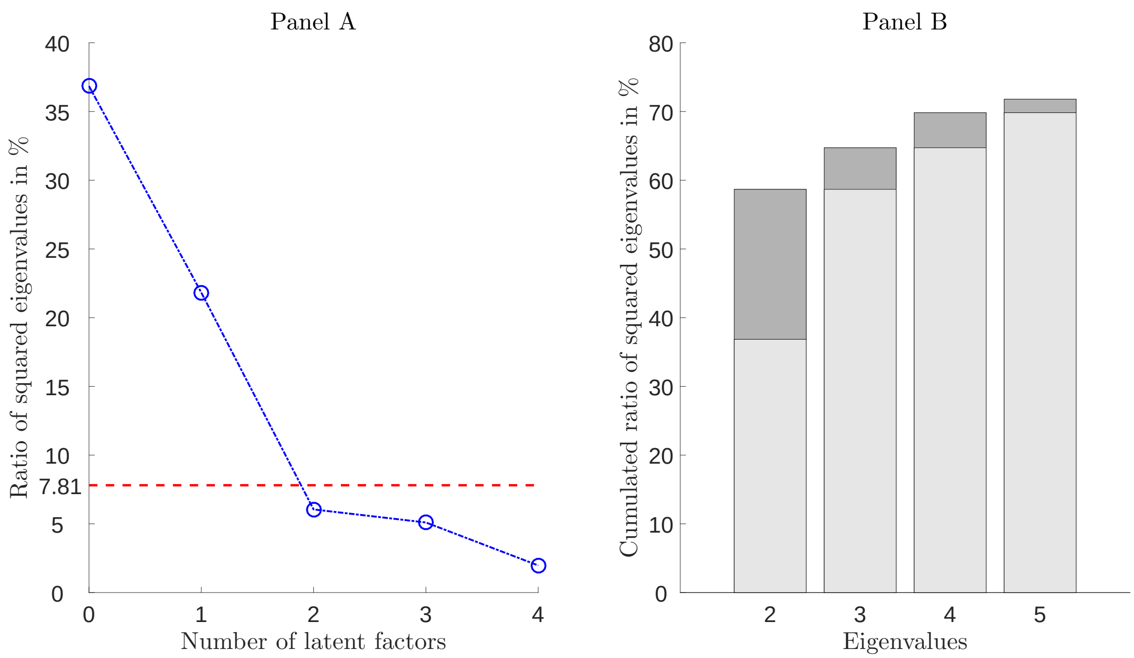
Panel A plots the scree-plot of the values of the first five eigenvalues in percentage, i.e. $\mu_j\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$ with $j=1,...,5$. The horizonal line corresponds to the penalty function $g\left(n^{\chi},T\right)$. Panel B plots the cumulated eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1}\mu_l\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$, the dark grey is the contribution of the jth eigenvalue in percentage. The figure reports results for the FF model for the time-invariant and time-varying specifications with $Z_{t-1}=(1,divY_{t-1})'$.

Figure 4: Number of omitted factors and cumulated squared eigenvalues for the FF model



Panel A plots the scree-plot of the values of the first five squared eigenvalues in percentage, i.e. $\mu_j^2\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)/\sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$ with $j=1,...,5$. The horizonal line corresponds to the penalty function $g\left(n^{\chi},T\right)^2/\sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$. Panel B plots the cumulated squared eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1}\mu_l^2\left(\frac{1}{n^{\chi}T}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)/\sum_{l=1}^{T}\mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^{\chi}\bar{\bar{\varepsilon}}_i\bar{\bar{\varepsilon}}_i'\right)$, the dark grey is the contribution of the jth squared eigenvalue in percentage. The figure reports results for the FF model for the time-invariant and time-varying specifications with $Z_{t-1}=(1, divY_{t-1})'$.

Figure 5: Number of omitted factors and cumulated squared eigenvalues for the CCAPM model



Panel A plots the scree-plot of the values of the first five squared eigenvalues in percentage, i.e. $\mu_j^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T} \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$ with $j = 1, ..., 5$. The horizonal line corresponds to the penalty function $g\left(n^\chi, T\right)^2 / \sum_{l=1}^{T} \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$. Panel B plots the cumulated squared eigenvalues in percentage. The light grey area corresponds to $\sum_{l=1}^{j-1} \mu_l^2\left(\frac{1}{n^\chi T}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right) / \sum_{l=1}^{T} \mu_l^2\left(\frac{1}{nT}\sum_i \mathbf{1}_i^\chi \bar{\bar{\varepsilon}}_i \bar{\bar{\varepsilon}}_i'\right)$, the dark grey is the contribution of the jth sqaured eigenvalue in percentage.

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:*

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

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 $e_{i,t} = u_{i,t} 1\{|u_{i,t}| \leq \delta\} - E[u_{i,t} 1\{|u_{i,t}| \leq \delta\}|\gamma_i]$. Then:*

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

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

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

Assumption A.7 *The processes $(I_t(\gamma))$ and $(\varepsilon_t(\gamma))$ are independent.*

Assumption A.8 *There exist constants η, $\bar{\eta} \in (0, 1]$ and $C_1, C_2, C_3, C_4 > 0$ such that, for all $\delta > 0$ and $n, T \in \mathbb{N}$, we have:*

a) $\sup_{1\leq i\leq n} P\left[\|\frac{1}{T}\sum_t I_{i,t}\left(h_t h_t' - \Sigma_h\right)\| \geq \delta|\gamma_i\right] \leq C_1 T \exp\{-C_2\delta^2 T^\eta\} + C_3\delta^{-1}\exp\{-C_4 T^{\bar{\eta}}\}.$
Furthermore the same upper bound holds for
b) $\sup_{1\leq i\leq n} P\left[|\frac{1}{T}\sum_t I_{i,t} - E[I_{i,t}|\gamma_i]| \geq \delta|\gamma_i\right],$
c) $\sup_{1\leq i\leq n} P\left[|\frac{1}{T}\sum_t \left(x_{i,t}x_{i,t}' - E[x_{i,t}x_{i,t}'|\gamma_i]\right)| \geq \delta|\gamma_i\right].$

Assumption A.9 $\inf_{1\leq i\leq n} E[I_{i,t}|\gamma_i] \geq M^{-1}$, *for all* $n \in \mathbb{N}$ *and a constant* $M > 0$.

Assumption A.10 *The trimming constants* $\chi_{1,T}$ *and* $\chi_{2,T}$ *are such that* $\chi_{1,T}^4\chi_{2,T}^2 = o\left(Tg\left(n,T\right)\right)$.

Assumption A.11 *We have* $\mu_1(W) = O_p(C_{n,T}^{-2})$, *where* $W = [w_{t,s}]$ *is the* $T\times T$ *matrix with elements* $w_{t,s} = \frac{1}{nT}\sum_i (I_{i,t}-\bar{I}_t)(I_{i,s}-\bar{I}_s)$, *and* $\bar{I}_t = \frac{1}{n}\sum_i I_{i,t}$.

Assumption A.1 restricts serial dependence in the bivariate process of error terms $(u_{i,t}, u_{j,t})$ of any two assets. It involves conditional expectations of products of error terms $u_{i,t}$ for different dates and any pair of assets. That assumption can be satisfied under weak serial dependence of the errors $(u_{i,t}, u_{j,t})$, such as mixing, with mixing size uniformly bounded across pairs (i,j). Assumption A.2 is an upper bound on higher-orders moments of $u_{i,t}$, to control tail thickness. Assumption A.3 is a restriction on both serial and cross-sectional dependence of the error terms and on the growth rates of n and T. We use Assumptions A.2 and A.3 to characterize the asymptotic behavior of the spectrum of the cross-sectional variance-covariance matrix of errors under the rival models. Assumption A.2 yields the so-called truncation and centralization lemmas, which are used together with Assumption A.2 in the proof of Lemma 1 building on Geman (1980), Yin et al. (1988) and Bai and Yin (1993). For those lemmas, we do not need a structure on the error terms based on matrix transformations of i.i.d. random variables as in Onatski (2010) and Ahn and Horenstein (2013). In Appendix 3, we show that Assumptions A.1 and A.3 are satisfied under cross-sectional block-dependence and time-series independence of the errors, provided n grows sufficiently faster than T. Under cross-sectional independence of the errors, the condition $T/n = o(1)$ is enough as discussed at the end of Appendix 3. The arguments in Yin et al. (1988), page 520, show that Assumption A.3 is also satisfied under i.i.d. $u_{i,t}$ and proportional asymptotics. Assumptions A.4 and A.5 require upper bounds on regressor values, latent factors and factor loadings. Assumption A.6 restricts serial dependence of the products of latent factors and regressors. Recall that matrices $x_{i,t}h_t'$ are zero-mean under Assumption 2. In Assumption A.7, we assume a missing-at-random design (Rubin (1976)), that is, independence between unobservability and return generation. Another design would require an explicit modeling of the link between the unobservability mechanism and the return process of the continuum of assets (Heckman (1979)); this would yield a nonlinear factor structure. Assumption A.8 a) restricts the serial dependence of the latent factors and the individual processes of observability indicators. Specifically, Assumption A.8 a) gives an upper bound for large deviation probabilities of the sample average of zero-mean random matrices $h_t h_t' - \Sigma_h$, computed over date with available observations for assets i, uniformly w.r.t. asset i. It implies that the unbalanced sample moment of squared components of the latent factor vector converges in probability to the corresponding population moment at a rate $O_p(T^{-\eta/2}(\log T)^c)$, for some $c > 0$. Assumptions A.8 b) and c) give similar upper bounds for large-deviation probabilities of sample averages of observability indicators and cross-moments of regressors uniformly w.r.t. asset i. We use such assumptions to get the convergence of time-series averages uniformly across assets as in GOS. Assumption A.9 implies that asymptotically the fraction of the time period in which an asset return is observed is bounded away from zero uniformly across assets, so that $\tau_i = \operatorname*{plim}_{T\to\infty} \tau_{i,T} = E[I_{i,t}|\gamma_i]^{-1}$ is bounded uniformly across all assets as in GOS. Assumption A.10 gives an upper bound on the divergence rate of the trimming constants. Assumption A.11 controls the rate at which the largest eigenvalue of the matrix with entries made of cross-sectional empirical covariances of observability indicators vanishes to zero. The matrix gathering those empirical covariances should not be associated to an omitted factor structure.

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 $0 \leq k \leq \min\{n, T\} - 1$. 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 $\|A\|_{op} = \max_{x:\|x\|=1} \|Ax\|$. Besides, $\|A\|_{op} \leq \|A\|$ for any square matrix A, where $\|\cdot\|$ is the Frobenius norm (see e.g. Meyer (2000)).

(iii) Inequalities for the eigenvalues of matrix products: If A and B are $m \times m$ positive semidefinite and positive definite matrices, respectively,

$$\mu_k(A)\,\mu_m(B) \leq \mu_k(AB) \leq \mu_k(A)\,\mu_1(B), \tag{13}$$

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, we have, 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)). The max-min formulation states:

$$\mu_k(A) = \max_{\mathcal{G}:dim(\mathcal{G})=k} \min_{x\in\mathcal{G}:\|x\|=1} x'Ax, \tag{15}$$

where the maximization 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, we have, for $k = 1, ..., T$,

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

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, and $\mathcal{F}_0 \equiv \mathbb{R}^T$.

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 $\tilde{X}_i = \boldsymbol{I}_i \odot X_i$ and $\tilde{R}_i = \boldsymbol{I}_i \odot R_i$, 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 3, A.2, A.3, A.7, as $n, T \to \infty$, 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 3, A.1, A.2, A.4, A.8 b), c) and A.9, as $n, T \to \infty$, 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.10 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 inequality (12) with $k=0$, and the inequality between matrix norms, we get:

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

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

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

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.10 on the trimming constants, we get $I = o_p(g(n,T))$ under $\mathcal{M}_2$.

Lemma 3 *Under model $\mathcal{M}_2$ and Assumptions 3, A.2, A.4, A.5 and A.6, as $n, T \to \infty$, 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, A.8 and A.9, 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:

$$\mu_{k+1} \left(\frac{1}{nT} \sum_i \mathbf{1}_i^{\chi} \bar{\varepsilon}_i \bar{\varepsilon}_i' \right)^{1/2} \leq \mu_{k+1} \left(\frac{1}{nT} \sum_i \tilde{H}_i \theta_i \theta_i' \tilde{H}_i' \right)^{1/2} + \mu_1 \left(\frac{1}{nT} \sum_i \tilde{u}_i \tilde{u}_i' \right)^{1/2} + I^{1/2}, \qquad (20)$$

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 $I = o_p(g(n,T))$, from Lemma 3 and Assumption A.10 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 (20) is bounded by the next lemma.

Lemma 5 *Under model $\mathcal{M}_1(k)$ and Assumptions A.5 and A.11, we have* $\mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) = O_p(C_{n,T}^{-2})$.

The bound in Lemma 5 would be trivial in the case $\tilde{H}_i = H$, i.e. with a balanced panel, because in that case $\mu_{k+1}\left(\frac{1}{nT}\sum_i H_i\theta_i\theta_i'H_i'\right) = 0$ under $\mathcal{M}_1(k)$.

From Inequality (20) and Lemma 5, we get $\xi = O_p(C_{n,T}^{-2}) + o_p(g(n,T)) - g(n,T)$. 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:

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

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 $I = o_p(g(n,T))$.

Lemma 6 *Under model $\mathcal{M}_2(k)$ and Assumptions 1(i), 4, A.8 and A.9, 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 (21) and Lemma 6, we get $\xi \geq C/2$, w.p.a. 1, and Proposition 2(b) follows.

A.2.3 Proof of Proposition 3

Let us define the events $A_k = \{\xi(k) \geq 0\}$, $k = 0,...,k_0-1$, and $A_{k_0} = \{\xi(k_0) < 0\}$. We have $P[\hat{k} = k_0] = P[\{A_0 \cap A_1 \cap ... \cap A_{k_0-1}\} \cap A_{k_0}]$. For generic events B and C, we have $P[B\cap C] = P[B] + P[C] - P[B\cup C]$, and we conclude that $P[B\cap C] \to 1$ if both $P[B]$ and $P[C]$ converge to 1 since $P[B\cup C] \geq P[B]$ and $P[B\cup C] \geq P[C]$. Applying repeatedly this argument to the probability $P[\{A_0 \cap A_1 \cap ... \cap A_{k_0-1}\} \cap A_{k_0}]$ yields $P[\hat{k} = k_0] \to 1$ since $P[A_k] \to 1$, $k = 0,...,k_0-1$, and $P[A_{k_0}] \to 1$, under $\mathcal{M}_1(k_0)$ from Proposition 2.

A.2.4 Proof of Lemma 1

We prove:

$$\limsup_{n,T\to\infty} \mu_1\left(\frac{1}{n}\tilde{\mathcal{E}}\tilde{\mathcal{E}}'\right) \leq C, \ a.s., \tag{22}$$

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, the statement of Lemma 1 follows. To show (22), we follow similar arguments as in Geman (1980), Yin et al. (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 $(e_{i,t})$, where $\xi_{i,t} = \varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}$ and $e_{i,t} = \xi_{i,t} - E[\xi_{i,t}|\gamma_i]$, and $\delta = \delta_n \uparrow \infty$ is a diverging sequence as in Assumption A.3. Let us define matrices $\tilde{E}$ and $\tilde{\Xi}$ with elements $(I_{i,t}e_{i,t})$ and $(I_{i,t}\xi_{i,t})$ 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 (22). We prove Lemmas 7 and 8 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) $P(\mathcal{E} \neq \Xi \ i.o.) = 0$, and (ii) $P\left(\tilde{\mathcal{E}} \neq \tilde{\Xi} \ i.o.\right) = 0$, where $i.o.$ means infinitely often for $n = 1, 2, \ldots$.*

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 (22) is implied by:

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

for a constant $C < \infty$. Now, we use that the upper bound (23) is implied by the condition:

$$\sum_{n=1}^{\infty} E\left[\left(\mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)/C\right)^k\right] < \infty, \tag{24}$$

for an increasing sequence of integers $k = k_n \uparrow \infty$. To prove the validity of condition (24), we use that:

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

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:

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

Then, we get $E\left[\mu_1\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)^k\right] \leq M^k$, for the sequence $k = k_n$ defined in Assumption A.3. Condition (24) holds for any $C > M$, and the conclusion follows.

A.2.5 Proof of Lemma 2

i) We have $I_1^2 = \| \frac{1}{nT}\sum_i(1-\mathbf{1}_i^\chi)\tilde{\varepsilon}_i\tilde{\varepsilon}_i'\|^2 = \frac{1}{n^2T^2}\sum_{i,j}(1-\mathbf{1}_i^\chi)(1-\mathbf{1}_j^\chi)(\tilde{\varepsilon}_i'\tilde{\varepsilon}_j)^2 = \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2}(1-\mathbf{1}_i^\chi)(1-\mathbf{1}_j^\chi)I_{i,t_1}I_{j,t_1}I_{i,t_2}I_{j,t_2}\varepsilon_{i,t_1}\varepsilon_{j,t_1}\varepsilon_{i,t_2}\varepsilon_{j,t_2}$. 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 from Assumptions A.4, A.8c) and A.9 (see Lemma 7 in GOS). Then, $I_1 = O_p(T^{-\bar{b}})$ for any $\bar{b} > 0$.

ii) We have:

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

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:

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

Hence $E[I_2^2] \leq \frac{CM^5\chi_{1,T}^8\chi_{2,T}^4}{T^2}$, 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.2.6 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:

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

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:

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

Hence $E[I_2^2] \leq \frac{CM\chi_{1,T}^8\chi_{2,T}^4}{T^2}$, 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.2.7 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 the proof of Lemma 3 in GOS, under Assumptions A.8 and A.9, and $n = O\left(T^{\bar{\gamma}}\right)$, $\bar{\gamma} > 0$, 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:

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

w.p.a. 1, for a constant $C > 0$. From 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.2.8 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:

$$\mu_2(\tilde{A}) \leq M^2 \mu_2(A), \tag{25}$$

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 $\max_{x \in \mathcal{G}: \|x\|=1} x'\tilde{A}x = \max_{x \in \mathcal{G}: \|x\|=1} x'DADx$. 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 $\|y\|^2 \leq \|h\|_{\infty,T}^2 \|x\|^2 = \|h\|_{\infty,T}^2 \leq M^2$, where $\|h\|_{\infty,T} = \max_{t=1,\ldots,T} |h_t| \leq M$ under Assumption A.5 a). Then:

$$\max_{x \in \mathcal{G}: \|x\|=1} x'\tilde{A}x \leq \max_{y \in D(\mathcal{G}): \|y\| \leq M} y'Ay = M^2 \max_{y \in D(\mathcal{G}): \|y\|=1} y'Ay. \tag{26}$$

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 subspace of $\mathbb{R}^T$ with dimension $T-1$. Then, from Inequality (26) we get:

$$\max_{x \in D^{-1}(\mathcal{F}_1^{\perp}): \|x\|=1} x'\tilde{A}x \leq M^2 \max_{y \in \mathcal{F}_1^{\perp}: \|y\|=1} y'Ay. \tag{27}$$

From the Courant-Fisher 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: $\mu_2(A) = \max_{y \in \mathcal{F}_1^{\perp}: \|y\|=1} y'Ay$. Then, Inequality (27) implies bound (25).

Finally, let us bound $\mu_2(A)$. By writing $A = \frac{1}{nT}(B+C)(B+C)'$, where $B = (b_{t,i})$ and $C = (c_{t,i})$ are $T \times n$ matrices with elements $b_{t,i} = \theta_i \bar{I}_t$ and $c_{t,i} = \theta_i(I_{i,t} - \bar{I}_t)$, the Weyl inequality (12) implies $\mu_2(A)^{1/2} \leq \mu_2\left(\frac{1}{nT}BB'\right)^{1/2} + \mu_1\left(\frac{1}{nT}CC'\right)^{1/2} = \mu_1\left(\frac{1}{nT}CC'\right)^{1/2}$, since matrix BB' has rank 1.

Now $\frac{1}{nT}CC' = \frac{1}{nT}\tilde{C}D\tilde{C}'$, where the elements of the $T \times n$ matrix $\tilde{C}$ are $\tilde{c}_{t,i} = I_{i,t} - \bar{I}_t$ and D is a $n \times n$ diagonal matrix with elements θ_i^2. From Assumption A.5b), we have $\mu_1\left(\frac{1}{nT}CC'\right) \leq M^2\mu_1(W)$, where the elements of matrix $W = \frac{1}{nT}\tilde{C}\tilde{C}'$ are $w_{t,s} = \frac{1}{nT}\sum_i \left(I_{i,t} - \bar{I}_t\right)\left(I_{i,s} - \bar{I}_s\right)$. Thus, from Assumption A.11, we get $\mu_2(A) = O_p(C_{n,T}^{-2})$. From bound (25), 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 $\tilde{a}_{t,s} = \frac{1}{nT}\sum_i I_{i,t}I_{i,s}\theta_i'h_t\theta_i'h_s = \sum_{m,l}\left(\frac{1}{nT}\sum_i I_{i,t}I_{i,s}\theta_{i,m}\theta_{i,l}\right)h_{t,m}h_{s,l} =: \sum_{m,l} a_{t,s}^{(m,l)}h_{t,m}h_{s,l}$, where summation w.r.t. m, l is from 1 to k. Then, we have $\tilde{A} = \sum_{m,l} D^{(m)}A^{(m,l)}D^{(l)} = DAD'$, 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 the $(Tk) \times (Tk)$ 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 $\tilde{A} \leq 2^{k-1}\sum_m D^{(m)}A^{(m,m)}D^{(m)}$. Since two symmetric matrices are ranked if, and only if, their corresponding eigenvalues are ranked, we get:

$$\mu_{k+1}(\tilde{A}) \leq 2^{k-1}\mu_{k+1}\left(\sum_m D^{(m)}A^{(m,m)}D^{(m)}\right). \tag{28}$$

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 (28) 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 $\mu_2(D^{(m)}A^{(m,m)}D^{(m)}) \leq M^2\mu_2(A^{(m,m)})$. 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 inequality and Assumptions A.5b) and A.11 imply $\mu_2(A^{(m,m)}) \leq M^2\mu_1(W) = O_p(C_{n,T}^{-2})$. Thus $\mu_{k+1}(\tilde{A}) = O_p(C_{n,T}^{-2})$.

A.2.9 Proof of Lemma 6

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

$$\mu_{k+1}\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right) = \max_{\mathcal{G}:\dim(\mathcal{G})=k+1}\ \min_{x\in\mathcal{G}:\|x\|=1} x'\left(\frac{1}{nT}\sum_i \tilde{H}_i\theta_i\theta_i'\tilde{H}_i'\right)x, \qquad (29)$$

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 $\mathcal{G}_{\mathbb{A}} := \left\{x \in \mathbb{R}^T : \ x = \frac{1}{\sqrt{T}}Ha,\ a \in \mathbb{A}\right\}$ is a linear subspace of $\mathbb{R}^T$ of dimension $k+1$. We deduce from (29):

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

By similar arguments as in the proof of Lemma 4, under Assumptions A.8 and A.9, 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 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 4, the conclusion follows.

Appendix 3 Check of Assumptions A.1 and A.3 under block dependence

In this appendix, we verify that the high-level Assumptions A.1 and A.3 on serial and cross-sectional dependences of error terms are satisfied under a block-dependence structure in a serially i.i.d. framework.

Assumption BD.1 *The error terms $u_t(\gamma)$ are i.i.d. over time with $E[u_t(\gamma)] = 0$, for all $\gamma \in [0, 1]$. For any n, there exists a partition of the interval $[0, 1]$ into $b_n \leq n$ subintervals of approximate length $B_n = O(1/b_n)$, such that $u_t(\gamma)$ and $u_t(\gamma')$ are independent if γ and γ' belong to different subintervals, and $b_n^{-1} = O(n^{-\alpha})$ as $n \to \infty$, where $\alpha \in (0, 1]$.*

Assumption BD.2 *The error terms $(u_t(\gamma))$, the factors (f_t), and the instruments (Z_t), $(Z_t(\gamma))$, $\gamma \in [0, 1]$, are mutually independent.*

The block-dependence structure as in Assumption BD.1 is satisfied for instance when there are unobserved industry-specific factors independent among industries and over time, as in Ang et al. (2010). In empirical applications, blocks can match industrial sectors. Then, the number b_n of blocks amounts to a couple of dozens, and the number of assets n amounts to a couple of thousands. There are approximately nB_n assets in each block, when n is large. In the asymptotic analysis, Assumption BD.1 requires that the number of independent blocks grows with n fast enough. Within blocks, covariances do not need to vanish asymptotically.

Lemma 11 *Under Assumptions A.2 and BD.1: (i) Assumption A.1 holds. (ii) Assumption A.3 holds if $n \geq T^{\bar{\gamma}}$ and:*

$$\alpha > 4/q, \quad \bar{\gamma} > \frac{1}{\alpha - 4/q}. \tag{30}$$

The conditions in (30) provide 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 (via α), and (ii) the existence of higher-order moments of the error terms (via q). 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. These conditions are compatible with $T/n = o(1)$.

SUPPLEMENTARY MATERIALS

A diagnostic criterion for approximate factor structure

Patrick Gagliardini, Elisa Ossola and Olivier Scaillet

These supplementary materials provide the proofs of technical Lemmas 7-11 (Appendix 4), the verification that conditional independence implies Assumption 2 (Appendix 5), the link with Stock and Watson (2002) (Appendix 6), and the results of Monte-Carlo experiments (Appendix 7).

Appendix 4 Proofs of technical Lemmas

A.4.1 Proof of Lemma 7

We follow the arguments in the proof of Lemma 2.2 in Yin et al. (1988). From the conditions $\delta \geq n^{\beta}$ and $T \leq C_2 n$, we have:

$$\begin{aligned} P(\mathcal{E} \neq \Xi \; i.o.) &\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} P\left(\bigcup_{2^{m-1}\leq n<2^m} \bigcup_{i=1}^{n} \bigcup_{t=1}^{T} \{|\varepsilon_{i,t}| > \delta\} \right) \\ &\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} P\left(\bigcup_{i=1}^{2^m} \bigcup_{t=1}^{C_2 2^m} \{|\varepsilon_{i,t}| > 2^{(m-1)\beta}\} \right) \\ &\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} C_2 2^{2m} P\left(|\varepsilon_{i,t}| > 2^{(m-1)\beta} \right). \end{aligned}$$

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

$$\sum_{m=1}^{\infty} 2^{2m} P\left(|\varepsilon_{i,t}| > 2^{(m-1)\beta} \right) < \infty. \tag{31}$$

To prove the summability condition (31), 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 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,

$$tr(A'B) \leq \sum_{k=1}^{T} \mu_k(A'A)^{1/2}\mu_k(B'B)^{1/2}. \tag{32}$$

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 (32):

$$\begin{aligned}\left[\mu_1^{1/2}\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) - \mu_1^{1/2}\left(\frac{1}{n}\tilde{E}\tilde{E}'\right)\right]^2 &\leq tr\left(\frac{1}{n}\tilde{\Xi}\tilde{\Xi}'\right) + tr\left(\frac{1}{n}\tilde{E}\tilde{E}'\right) - 2\frac{1}{n}tr\left(\tilde{\Xi}\tilde{E}'\right) \\ &= \frac{1}{n}tr\left[(\tilde{\Xi} - \tilde{E})(\tilde{\Xi} - \tilde{E})'\right]. \end{aligned} \tag{33}$$

The elements of matrix $\tilde{\Xi} - \tilde{E}$ are $I_{i,t}E[\varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}|\gamma_i]$. 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:

$$\frac{1}{n}tr\left[(\tilde{\Xi} - \tilde{E})(\tilde{\Xi} - \tilde{E})'\right] = \frac{1}{n}\sum_i\sum_t I_{i,t}E[\varepsilon_{i,t}1\{|\varepsilon_{i,t}| \leq \delta\}]^2 \leq T\frac{M^2}{\delta^{2(q-1)}}. \tag{34}$$

From inequalities (33) and (34), 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 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 10

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

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

A.4.5 Proof of Lemma 11

A.4.5.1 Proof of Part (i)

By the serial independence of the error terms, we have:

$$\begin{aligned} & \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4} E\left[u_{i,t_1}u_{i,t_2}u_{j,t_3}u_{j,t_4}\left|x_{i,\underline{T}},x_{j,\underline{T}},\gamma_i,\gamma_j\right.\right] \\ = & \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1,t_2,t_3,t_4} E\left[u_{i,t_1}u_{i,t_2}u_{j,t_3}u_{j,t_4}\left|\gamma_i,\gamma_j\right.\right] \\ = & \frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1} E\left[u_{i,t_1}u_{i,t_1}u_{j,t_1}u_{j,t_1}\left|\gamma_i,\gamma_j\right.\right] \\ & +\frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1\neq t_2} E\left[u_{i,t_1}u_{i,t_1}\left|\gamma_i\right.\right]E\left[u_{j,t_2}u_{j,t_2}\left|\gamma_j\right.\right] \\ & +\frac{1}{n^2T^2}\sum_{i,j}\sum_{t_1\neq t_2} E\left[u_{i,t_1}u_{j,t_1}\left|\gamma_i,\gamma_j\right.\right]E\left[u_{i,t_2}u_{j,t_2}\left|\gamma_i,\gamma_j\right.\right]. \end{aligned}$$

The conclusion follows by taking absolute values and expectation, and using the triangular inequality, the Cauchy-Schwarz inequality and Assumption A.2.

A.4.5.2 Proof of Part (ii)

Here, we treat $\varepsilon_{i,t}$ as a random variable but not through the random draw γ_i. This avoids the notational burden coming from conditional expectations. We show directly the inequality $\frac{1}{n^k}\sum_{i_1,\ldots,i_k}\sum_{t_1,\ldots,t_k}|E[e_{i_1,t_k}e_{i_1,t_1}e_{i_2,t_1}e_{i_2,t_2}e_{i_3,t_2}\cdots e_{i_{k-1},t_{k-1}}e_{i_k,t_{k-1}}e_{i_k,t_k}]| \leq M^k$, which implies Assumption A.3. Under Assumption BD.1, 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:

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

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, $N_1 = O(bd^k T^k)$. 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 $O(b^k d^k)$. 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, $N_k = O(b^k d^k T)$ 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 $N_k = O(b^k d^k T)$ trivially holds also in this case. However, this case will not occur with our choice of sequence k, since (41) 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

$$N_m = O(b^m) \times \left\{ \begin{matrix} k \\ m \end{matrix} \right\} \times O(d^k) \times O(T^{k-m+1}), \tag{37}$$

$$Q_m = O(\delta^{2(k-m)}), \tag{38}$$

  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:

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

4) We exploit the following upper bound for the Stirling numbers of the second kind (see Rennie and Dobson (1969), Theorem 3) $\left\{ \begin{matrix} k \\ m \end{matrix} \right\} \leq \frac{1}{2} \binom{k}{m} m^{k-m}$. Then, we get: $\sum_{m=1}^{k} \left\{ \begin{matrix} k \\ m \end{matrix} \right\} (\delta^2 T/b)^{k-m} \leq$

$\frac{1}{2}\sum_{m=1}^{k} m^{k-m}\binom{k}{m}(\delta^2 T/b)^{k-m} \leq \frac{1}{2}\sum_{m=0}^{k}\binom{k}{m}(k\delta^2 T/b)^{k-m} = \frac{1}{2}(1+k\delta^2 T/b)^k$, from the binomial theorem. Thus, from (39), we get:

$$I_k \leq constT(1+k\delta^2 T/b)^k. \tag{40}$$

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

$$k\delta^2 T/b = o(1), \quad T = O(e^k). \tag{41}$$

From (40) and (41), we get $I_k \leq (2e)^k$. Then, from (35):

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

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

6) Let us now verify the 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 $\sqrt{T}/\delta^{q-1} = o(1)$, $\beta > 2/q$, and they match conditions (41). 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 $k\delta^2 T/b = o(1)$ implies:

$$\beta < \frac{1}{2}(\alpha - 1/\bar{\gamma}). \tag{42}$$

Condition $\sqrt{T}/\delta^{q-1} = o(1)$ implies $\beta > \frac{1}{2\bar{\gamma}(q-1)}$. The latter inequality is implied by

$$\beta > \frac{2}{q}, \tag{43}$$

since $\bar{\gamma} > 1$ and $q \geq 8$ in Assumption A.2. Then, there exists a power $\beta > 0$ satisfying conditions (42) and (43) if, and only if, $\frac{1}{2}(\alpha - 1/\bar{\gamma}) > \frac{2}{q}$, which corresponds to Condition (30). This condition 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.

Appendix 5 Verification that conditional independence implies Assumption 2

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

We have:

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

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) applied to residuals $\hat{\varepsilon}_{i,t}$:

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

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

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

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

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

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

We have

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

Thus, $Q_1\left(\zeta,\tilde{\zeta}\right)=\frac{1}{nT}\sum_i\sum_t\left(\tilde{\tilde{\varepsilon}}_{i,t}-\theta_i h_t\right)^2+\frac{1}{nT}\sum_i\sum_t\left(1-I_{i,t}\right)\sigma_{i,t}^2$, 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 $\tilde{\tilde{\varepsilon}}$: $\min_\zeta\frac{1}{nT}\sum_i\sum_t\left(\tilde{\tilde{\varepsilon}}_{i,t}-\theta_i h_t\right)^2=\frac{1}{nT}\sum_i\sum_t\tilde{\tilde{\varepsilon}}_{i,t}^2-\mu_1\left(\frac{\tilde{\tilde{\varepsilon}}\tilde{\tilde{\varepsilon}}'}{nT}\right)$. Therefore, at convergence with $\hat{\zeta}=\tilde{\zeta}$, we have

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

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 criterion ξ^{SW} after penalization.

Appendix 7 Monte-Carlo experiments

From the core text, we know that our selection procedure is equivalent to the penalized least-squares strategy of Bai and Ng (2002) when we use the same penalty term. The first part of our Monte Carlo experiments in Section ... aims to show, as expected, that the penalisation introduced in Section 5 delivers a performance similar to the one observed in the literature with other penalisations in the presence of latent factors only. We investigate settings with n and T of comparable sizes as well as n much larger than T, as covered by our theory. Then we investigate how our selection procedure performs with unbalanced panels. The main result is that the performance is similar when the effective size of an unbalanced panel, i..e, the cross-sectional dimension n^{χ} obtained after trimming, is close to the size n in a balanced panel. The second part of our Monte Carlo experiments in Section ... aims to extend the performance study to settings where we face observable factors, and apply the selection procedure to residuals. We show that the performance is close to the one obtained without observable factors.

TO UPDATE!!!!

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. In the balanced case, we consider two simulation designs: (i) the simulation process in Ahn and Horenstein (2013), (ii) the simulation design that 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.

Under $\mathcal{M}_1$, we generate $S = 1,000$ dataset of $n \times T$ dimension from the simulation process in Ahn and Horenstein (2013), with $\theta = 1$ and $r = 1$. In particular, at each simulation $s = 1, ..., S$, we generate the one-factor model:

$$y_{i,t}^s = b_i^s f_t^s + \varepsilon_{i,t}^s, \text{ with } \varepsilon_{i,t}^s = \sqrt{\frac{1-\rho^2}{1+2J\beta^2}} e_{i,t}^s, \text{ for } i = 1, .., n, \text{ and } t = 1, ..., T, \tag{44}$$

where $e_{i,t}^s = \rho e_{i,t-1}^s + v_{i,t}^s + \sum_{h=\max(i-J,1)}^{i=1} \beta v_{h,t}^s + \sum_{h=i+1}^{\min(i+J,n)} \beta v_{h,t}^s$, and the error term $v_{i,t}^s$, the factor loading b_i^s, and the factor f_t^s are drawn from normal standardized distribution. Bai and Ng (2002) and Onatski (2010) use a similar data generating process (DGP) in their simulation exercises. The DGP in Equation (44) depends on three parameters: (i) ρ measures the magnitude of time-series correlation in the idiosyncratic errors $e_{i,t}^s$, (ii) β measures the magnitude of cross-sectional correlation between the errors $e_{i,t}^s$, (iii) J defines the number of units i that are cross-correlated. At each simulation, we estimate the one-factor model and we compute our diagnostic criterion on the fitted residuls $\hat{\varepsilon}_{i,t}^s$. The trimming levels does not affect the number of assets n in the simulations since the panel is balanced and all individual regressions are well estimated. Under $\mathcal{M}_2$, we generate $y_{i,t}^s$ from two different factor models: a one factor model and a three-factor model, and we estimate a one-factor model to get the residuals. The correct number of omitted factors is zero and two, respectively. In order to define the number of unobservable factors, we compare our procedure with the criteria proposed in Bai and Ng (2002), Onatski (2010), and Ahn and Horenstein (2013). At each simulation, we compute our diagnostic criteria $\xi(k)$ (see Section 5). We also maximize the estimators PC, IC, AIC and BIC proposed by Bai and Ng (2002), the edge distribution (ED) estimator introduced by Onatski (2010), and the eigenvalue ratio (ER) and growth ratio (GR) estimators described in Ahn and Horenstein (2013). We fix the maximum possible number of factors ($kmax$) equal to eight.

In our second design for balanced panel, we explore the properties of diagnostic criterion using a simulation design that mimics the empirical feature of our data. 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 $n \leq 6,775$ assets from the sample of our empirical analysis that comprises $6,775$ 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 $6,775 \times 6,775$ thresholded variance-covariance matrix of estimated residuals (see GOS for the thresholding tecnique). 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 $R_{i,t}^s$ the simulated excess returns of asset i at time t as follows

$$R_{i,t}^s = b_i^s f_t + \varepsilon_{i,t}^s, \text{ for } i = 1, ..., n, \text{ and } t = 1, ..., T, \tag{45}$$

where f_t is the market excess returns and $\varepsilon_{i,t}^s$ is the error term. In Equation (45), we impose the intercepts to be zero to satisfy the no-arbitrage restrictions for tradable factors. The $n \times 1$ error vectors $\boldsymbol{\varepsilon}_t^s$ 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 one and three-factor alternative hypothesis, i.e., zero and 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 the unbalanced case, we provide simulation exercises that mimics the empirical features of our data. To account for the unbalanced property, we introduce 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. The excess returns $R_{i,t}^s$ of asset i at time t under $\mathcal{M}_1$ is:

$$R_{i,t}^s = b_i^s f_t + \varepsilon_{i,t}^s, \text{ if } I_{i,t}^s = 1, \text{ for } i = 1, ..., n, \text{ and } t = 1, ..., T, \tag{46}$$

where $I_{i,t}^s$ is the observability indicator of asset i at time t in simulation s. Under $\mathcal{M}_2$, we again consider a CAPM and the Fama-French model.

A.7.1 Simulation under $\mathcal{M}_1$

In this section, we provide simulation experiments based on two different data generating processes. In our first experiment,

In order to understand how the criteria work, we consider several covariances structures, i.e., several combinations of parameters (ρ, β, J), and several combinations of cross-sectional dimension n and time dimension T. Table (6) 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. The selection probabilities are close to 1 for most combinations of cross-sectional sample size n and time dimension T. Table 6 compares the selection probabilities when the magnitude of time-series correlation changes in the error structure. When the time-series observations are highly correlated ($\rho = 0.7$, see Panel D) and the time-serie dimension is small

($T = 150$), the selection probabilities is approximately zero. Table 6, Panels E-G contains the results when the magnitude of cross-sectionally correlation increases throught parameters β and J. The increase of the correlation in the cross-section affects the selection probabilities when the ratio T/n is too far from zero. The magnitude of cross-sectional correlation in the errors has a large effect on the selection probabilities than the presence of time-series correlation in the errors. Tables 7-11 report the percentage of the replications that result in overestimation w.r.t. the percentage of the replications that result in underestimation of the number of factors. The tables reveal that the criteria $\xi(k)$ introduced in Equation (9), estimate the correct number of unobservable factors for the most combinations of n and T, in the different designs for the error structure. In particular, the criterion $\xi(k)$ performs equally to the maximization criteria proposed by Ahn and Horenstein (2013), and better than the maximization criteria in Bai and Ng (2002), and Onatski (2010). The results for the PC, IC, AIC and BIC and ED estimators are essentially the same as the results reported in Bai and Ng (2002), and Onatski (2010).

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 12 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. The diagnostic criterion selects the correct model in all the several combinations of n and T. Table 13 show the percentage of the replications that result in overestimation w.r.t. the percentage of the replications that result in underestimation for several criteria that define the number of unobservable factors. We provide results for several combinations of n and T. For a small number of time-series observations T, the criterion $\xi_l(k)$ and the maximization criteria tend to underestimate the correct number of factors. However, when the cross-sectional and time-series dimensions are large, the criterion $\xi(k)$ performs better to the maximization criteria.

A.7.2 Simulation under $\mathcal{M}_2$

Let us repeat the second experiment described in the previous section, but with unbalanced characteristics for the simulated datasets.

In Tables 14 and 15, 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

14, 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 $\chi_{2,T} = T/12$, we keep all assets in all simulations, while for the level of trimming $\chi_{2,T} = T/60$ we keep about two thirds of the assets on average. In Table 15, 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 16 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 14, 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 (17) reports the overestimation and underestimation percentages of the number of factors for criteria $\xi\,(k)$ in Equation (9).

Table 12: Selection probabilities in Monte Carlo simulations of DGP (45), balanced case

T	150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500
$Pr(\xi < 0\|\mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi > 0\|\mathcal{M}_2)$	0.0000	0.0160	0.9130	1.0000	0.0370	1.0000	1.0000	1.0000

The table contains the selection probabilities of the correct model, $Pr\left(\xi_l < 0|\mathcal{M}_1\right)$ and $Pr\left(\check{\xi}_l > 0|\mathcal{M}_2\right)$ estimated from the simulated balanced dataset based on the DGP in Equation (45).

Table 6: Selection probabilities in Monte Carlo replications of DGP (44)

T	150				500				150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500
	Panel A: i.i.d. errors ($\rho = 0, \beta = 0, J = 0$)								Panel B: serially correlated errors ($\rho = 0.3, \beta = 0, J = 0$)							
$Pr(\xi < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Panel C: serially correlated errors ($\rho = 0.5, \beta = 0, J = 0$)								Panel D: serially correlated errors ($\rho = 0.7, \beta = 0, J = 0$)							
$Pr(\xi < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0260	1.0000	1.0000	1.0000	0.0000	0.9920	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Panel E: cross-sectionally correlated errors ($\rho = 0, \beta = 0.2, J = 5$)								Panel F: serially and cross-sectionally correlated errors ($\rho = 0, \beta = 0.2, J = 10$)							
$Pr(\xi < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Panel G: cross-sectionally correlated errors ($\rho = 0, \beta = 0.5, J = 5$)								Panel H: serially and cross-sectionally correlated errors ($\rho = 0.2, \beta = 0.2, J = 5$)							
$Pr(\xi < 0 \vert \mathcal{M}_1)$	0.8500	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Panel I: serially and cross-sectionally correlated errors ($\rho = 0.5, \beta = 0.2, J = 5$)								Panel L: serially and cross-sectionally correlated errors ($\rho = 0.2, \beta = 0.5, J = 5$)							
$Pr(\xi < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7480	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The table contains the selection probabilities of the correct model, $Pr\left(\xi_l < 0|\mathcal{M}_1\right)$ and $Pr\left(\check{\xi}_l > 0|\mathcal{M}_2\right)$ estimated from the simulated dataset based on the DGP in Equation (44). Panels A-L contain the results assuming several combinations of parameters ρ, β and J.

Table 7: (% of overestimation)/(% of underestimation) in Monte Carlo replications of DGP (44)

	Panel A: i.i.d. errors ($\rho=0,\beta=0,J=0$)								Panel B: serially correlated errors ($\rho=0.3,\beta=0,J=0$)							
T	150				500				150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500
$\xi(k)$	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_3	99.7/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
IC_1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
IC_2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
IC_3	1.6/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	98.5/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
IC_4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
AIC_1	100/0	100/0	100/0	52.7/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_3	100/0	100/0	42/0	0/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_1	99.7/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0
BIC_2	99.7/0	100/0	100/0	100/0	0/0	0/0	88/0	100/0	100/0	100/0	100/0	100/0	0/0	0/0	100/0	100/0
BIC_3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
ED	0.8/0	0.7/0	0.6/0	0.7/0	0.5/0	0.4/0	0.7/0	0.8/0	1.1/0	0.5/0	0.4/0	0/0	0.6/0	0.3/0	0.5/0	0.5/0
ER	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
GR	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100-x-y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the DGP in Equation (44). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). Panel A contains the results assuming that error are i.i.d.. Panel B reports the results assuming that errors are serially correlated with $\rho = 0.3$.

Table 8: (% of overestimation)/(% of underestimation) in Monte Carlo replications of DGP (44)

	Panel A: serially correlated errors ($\rho = 0.5, \beta = 0, J = 0$)								Panel B: serially correlated errors ($\rho = 0.7, \beta = 0, J = 0$)							
T	150				500				150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500
$\xi(k)$	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_1	31.7/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	100/0	100/0	100/0	100/0	7.6/0	48/0	83.7/0	96.9/0
PC_2	0.2/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	100/0	100/0	100/0	100/0	1/0	0/0	1.9/0	19.7/0
PC_3	100/0	96.9/0	0.1/0	0/0	0/0	100/0	0/0	0/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
PC_4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	6.7/0	14.9/0	68.3/0	97.9/0	0/0	0/0	0/0	0/0
IC_1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	99.1/0	100/0	100/0	100/0	0/0	0/0	0.3/0	1.2/0
IC_2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	21.1/0	100/0	100/0	100/0	0/0	0/0	0/0	0/0
IC_3	100/0	0.1/0	0/0	0/0	0/0	100/0	0/0	0/0	100/0	100/0	100/0	100/0	29.7/0	100/0	100/0	100/0
IC_4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
AIC_1	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_3	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_1	100/0	96.9/0	0.1/0	0/0	100/0	100/0	0/0	0/0	100/0	100/0	100/0	100/0	100/0	100/0	0/0	0/0
BIC_2	100/0	100/0	100/0	100/0	0/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
ED	1/0	0.5/0	0.8/0	0.4/0	0.8/0	0.6/0	0.7/0	0.5/0	0.3/0	0.5/0	0/0	0/0	1.5/0	0.7/0	0.3/0	0.7/0
ER	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
GR	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100 - x - y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the DGP in Equation (44). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). Panel A and B contain the results assuming that errors are serially correlated with $\rho = 0.5$ and $\rho = 0.7$.

Table 9: (% of overestimation)/(% of underestimation) in Monte Carlo replications of DGP (44)

	Panel A: cross-sectionally correlated errors ($\rho=0,\beta=0.2,J=5$)								Panel B: cross-sectionally correlated errors ($\rho=0,\beta=0.2,J=10$)							
T	150				500				150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500
$\xi(k)$	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_1	100/0	0.5/0	0/0	0/0	100/0	100/0	0/0	0/0	100/0	100/0	62.1/0	0/0	100/0	100/0	100/0	100/0
PC_2	100/0	0/0	0/0	0/0	100/0	2.8/0	0/0	0/0	100/0	100/0	35.7/0	0/0	100/0	100/0	100/0	99.9/0
PC_3	100/0	99.8/0	0/0	0/0	100/0	100/0	100/0	0/0	100/0	100/0	100/0	0/0	100/0	100/0	100/0	100/0
PC_4	39.7/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	20.2/0	0/0	0/0	100/0	100/0	0/0	0/0
IC_1	100/0	0/0	0/0	0/0	100/0	4.9/0	0/0	0/0	100/0	100/0	0/0	0/0	100/0	100/0	100/0	96.8/0
IC_2	56.9/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	100/0	0/0	0/0	100/0	100/0	100/0	34/0
IC_3	100/0	4.7/0	0/0	0/0	100/0	100/0	100/0	0/0	100/0	100/0	14.8/0	0/0	100/0	100/0	100/0	100/0
IC_4	0/0	0/0	0/0	0/0	0.1/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	67.6/0	0/0	0/0
AIC_1	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_3	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_1	100/0	96.9/0	0/0	0/0	100/0	100/0	100/0	0/0	100/0	100/0	100/0	0/0	100/0	100/0	100/0	100/0
BIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	22.6/0	0/0	0/0
ED	0.9/0	0.5/0	0.5/0	0.6/0	0/0	0.7/0	0.3/0	0.1/0	0.7/0	0.7/0	0.9/0	0.9/0	0/0	0.6/0	0.4/0	0.8/0
ER	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
GR	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100-x-y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the DGP in Equation (44). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). Panel A and B contain the results assuming that errors are cross-sectionally correlated with $\beta=0.2$, $J=5$ and $J=10$.

Table 10: (% of overestimation)/(% of underestimation) in Monte Carlo replications of DGP (44)

	Panel A: cross-sectionally correlated errors ($\rho = 0, \beta = 0.5, J = 5$)								Panel B: serially and cross-sectionally correlated errors ($\rho = 0.2, \beta = 0.2, J = 5$)							
T	150				500				150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500
$\xi(k)$	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_1	100/0	100/0	0.4/0	0/0	100/0	100/0	100/0	17.6/0	100/0	9.2/0	0/0	0/0	100/0	100/0	0/0	0/0
PC_2	100/0	100/0	0/0	0/0	100/0	100/0	100/0	0.1/0	100/0	0.4/0	0/0	0/0	100/0	44.9/0	0/0	0/0
PC_3	100/0	100/0	72.2/0	0/0	100/0	100/0	100/0	100/0	100/0	100/0	0/0	0/0	100/0	100/0	100/0	0/0
PC_4	100/0	0/0	0/0	0/0	100/0	50/0	0/0	0/0	66.9/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0
IC_1	100/0	97.2/0	0/0	0/0	100/0	100/0	100/0	0/0	100/0	0/0	0/0	0/0	100/0	49.2/0	0/0	0/0
IC_2	100/0	59/0	0/0	0/0	100/0	100/0	76/0	0/0	79.5/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0
IC_3	100/0	100/0	0/0	0/0	100/0	100/0	100/0	100/0	100/0	27.6/0	0/0	0/0	100/0	100/0	100/0	0/0
IC_4	98.9/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
AIC_1	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_3	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_1	100/0	100/0	72.2/0	0/0	100/0	100/0	100/0	0/0	100/0	100/0	0/0	0/0	100/0	100/0	100/0	0/0
BIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_3	100/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
ED	0.4/0	0.4/0	0.8/0	0.8/0	0/0	1.7/0	0.6/0	0.6/0	1.8/0	0.8/0	0.4/0	0.2/0	0/0	1.1/0	0.5/0	0.3/0
ER	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
GR	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100 - x - y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the DGP in Equation (44). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). Panel A contains the results assuming that errors are cross-sectionally correlated with $\beta = 0.5$ and $J = 5$. Panel B reports the results assuming that errors are serially and cross-sectionally correlated with $\rho = 0.2$, $\beta = 0.2$ and $J = 5$.

Table 11: (% of overestimation)/(% of underestimation) in Monte Carlo replications of DGP (44)

	Panel A: serially and cross-sectionally correlated errors ($\rho = 0.5, \beta = 0.2, J = 5$)								Panel B: serially and cross-sectionally correlated errors ($\rho = 0.2, \beta = 0.5, J = 5$)							
T	150				500				150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500	150	500	1,000	1,500
$\xi(k)$	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
PC_1	100/0	100/0	91.8/0	15.2/0	100/0	100/0	96.8/0	1.1/0	100/0	100/0	23.3/0	0/0	100/0	100/0	100/0	86.1/0
PC_2	100/0	100/0	74.4/0	5.9/0	100/0	100/0	25.2/0	0/0	100/0	100/0	8.1/0	0/0	100/0	100/0	100/0	15.4/0
PC_3	100/0	100/0	100/0	90.8/0	100/0	100/0	100/0	100/0	100/0	100/0	99.6/0	0/0	100/0	100/0	100/0	100/0
PC_4	100/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	0.1/0	0/0	0/0	100/0	77.2/0	0/0	0/0
IC_1	100/0	76.2/0	0.2/0	0/0	100/0	100/0	9.2/0	0/0	100/0	99.6/0	0/0	0/0	100/0	100/0	100/0	0.7/0
IC_2	100/0	29.5/0	0/0	0/0	100/0	100/0	0.1/0	0/0	100/0	85.8/0	0/0	0/0	100/0	100/0	99.1/0	0/0
IC_3	100/0	100/0	44.3/0	0.1/0	100/0	100/0	100/0	100/0	100/0	100/0	1.8/0	0/0	100/0	100/0	100/0	100/0
IC_4	0.8/0	0/0	0/0	0/0	4.4/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0
AIC_1	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_3	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_1	100/0	100/0	100/0	90.8/0	100/0	100/0	100/0	100/0	100/0	100/0	99.6/0	0/0	100/0	100/0	100/0	100/0
BIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
BIC_3	12/0	0/0	0/0	0/0	0.3/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0	100/0	0/0	0/0	0/0
ED	2.2/0	0.6/0	0.8/0	0.4/0	0.6/0	0.8/0	1.1/0	0.7/0	1.4/0	0.7/0	0.8/0	0.6/0	0.1/0	1.1/0	0.4/0	0.6/0
ER	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
GR	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100 - x - y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the DGP in Equation (44). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). Panel A and B contain the results assuming that errors are serially and cross-sectionally correlated with $\rho = 0.5$, $\beta = 0.2$ and $J = 5$, and $\rho = 0.2$, $\beta = 0.5$ and $J = 5$.

Table 13: (% of overestimation)/(% of underestimation) in Monte Carlo simulations of DGP (45), balanced case

T	150				500			
n	150	500	1,000	1,500	150	500	1,000	1,500
$\xi(k)$	0/100	0/100	0/100	0/100	0/100	0/100	0/0	0/0
PC_1	0/100	0/100	0/100	0/100	0/88.4	0/0	96.8/0	1.1/0
PC_2	0/100	0/100	0/100	0/100	0/96.7	0/0.2	25.2/0	0/0
PC_3	0/1.6	0/99.6	0/100	0/100	0/18.2	0/0	100/0	100/0
PC_4	0/100	0/100	0/100	0/100	0/100	0/100	0/0	0/0
IC_1	0/100	0/100	0/100	0/100	0/99	0/0	9.2/0	0/0
IC_2	0/100	0/100	0/100	0/100	0/100	0/2.5	0.1/0	0/0
IC_3	0/64.6	0/100	0/100	0/100	0/60.7	0/0	100/0	100/0
IC_4	0/100	0/100	0/100	0/100	0/100	0/100	0/0	0/0
AIC_1	100/0	100/0	99.9/0	0/0	100/0	100/0	100/0	100/0
AIC_2	100/0	100/0	100/0	100/0	100/0	100/0	100/0	100/0
AIC_3	100/0	80.8/0	0/0	0/0	74.9/0	100/0	100/0	100/0
BIC_1	0/1.6	0/99.6	0/100	0/100	100/0	0/0	100/0	100/0
BIC_2	0/1.6	100/0	100/0	100/0	0/18.2	100/0	100/0	100/0
BIC_3	0/100	0/100	0/100	0/100	0/100	0/0	0/0	0/0
ED	0.7/16	0.6/0	0.5/0	0.5/0	0.8/0	0.8/0	1.1/0	0.7/0
ER	0/99.7	0/100	0/99.8	0/99.7	0/29.1	0/0	0/0	0/0
GR	0/99.7	0/100	0/99.8	0/99.3	0/26.3	0/0	0/0	0/0

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100 - x - y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the balanced DGP in Equation (45). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013).

Table 14: 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 15: Operative time-series sample size

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

Table 16: Selection probabilities in Monte Carlo simulations of DGP (46), 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
$Pr(\xi < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	0.6180	0.9980	1.0000	1.0000	1.0000	0.9380	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
$Pr(\xi < 0 \vert \mathcal{M}_1)$	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0040	1.0000	1.0000	1.0000
$Pr(\xi > 0 \vert \mathcal{M}_2)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The table contains the selection probabilities of the correct model, $Pr\left(\xi_l < 0|\mathcal{M}_1\right)$ and $Pr\left(\check{\xi}_l > 0|\mathcal{M}_2\right)$ with $l = 1, 2, 3$, estimated from the simulated unbalanced dataset based on the DGP in Equation (45).

Table 17: (% of overestimation)/(% of underestimation) in Monte Carlo simulations of DGP (46), **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
$\xi_1(k)$										
$\xi_2(k)$										
$\xi_3(k)$										
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
$\xi_1(k)$										
$\xi_2(k)$										
$\xi_3(k)$										

The table reports the $x\%$ of the replications that result in overestimation respect to the $y\%$ underestimation of the number of factors. The $(100 - x - y)\,\%$ of the replications is the probability of correct estimation of the number of unobservable factors. The simulations are based on the unbalanced DGP in Equation (46). The results refer to our criterion in Equation (9), and the criteria in Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013).