Latent Factor Analysis in Short Panels

Fortin, Alain-Philippe; Gagliardini, Patrick; Scaillet, Olivier

Latent Factor Analysis in Short Panels

Alain-Philippe Fortin[1,2], Patrick Gagliardini[3,2], Olivier Scaillet[1,2*]

September 13, 2023

Abstract

We develop inferential tools for latent factor analysis in short panels. The pseudo maximum likelihood setting under a large cross-sectional dimension n and a fixed time series dimension T relies on a diagonal $T \times T$ covariance matrix of the errors without imposing sphericity or Gaussianity. We outline the asymptotic distributions of the latent factor and error covariance estimates as well as of an asymptotically uniformly most powerful invariant (AUMPI) test based on the likelihood ratio statistic for tests of the number of factors. We derive the AUMPI characterization from inequalities ensuring the monotone likelihood ratio property for positive definite quadratic forms in normal variables. An empirical application to a large panel of monthly U.S. stock returns separates date after date systematic and idiosyncratic risks in short subperiods of bear vs. bull market based on the selected number of factors. We observe an uptrend in idiosyncratic volatility while the systematic risk explains a large part of the cross-sectional total variance in bear markets but is not driven by a single factor. Rank tests reveal that observed factors struggle spanning latent factors with a discrepancy between the dimensions of the two factor spaces decreasing over time.

Keywords: Latent factor analysis, uniformly most powerful invariant test, panel data, large n and fixed T asymptotics, equity returns. **JEL codes:** C12, C23, C38, C58, G12.

[*1]University of Geneva, [2]Swiss Finance Institute, [3]Università della Svizzera italiana. Acknowledgements: We are grateful to A. Onatski for his very insightful discussion of our paper Fortin, Gagliardini, Scaillet (2022) at the 14th Annual SoFiE Conference in Cambridge, which prompted us to exploit factor analysis for estimation in short panels. We thank L. Barras, S. Bonhomme, M. Caner, F. Carlini, I. Chaieb, F. Ghezzi, A. Horenstein, S. Kim, F. Kleibergen, H. Langlois, T. Magnac, S. Ng, E. Ossola, Y. Potiron, F. Trojani, participants at (EC)ˆ2 2022, QFFE 2023, SoFiE 2023, NASM 2023, IPDC 2023, and participants at seminars at UNIMIB, UNIBE, UNIGE, Warwick University.

1 Introduction

Latent variable models have been used for a long time in econometrics (Aigner et al. (1984)). Here, we study large cross-sectional latent factor models with small time dimension. Two common methods for estimation of latent factor spaces are principal component analysis (PCA) and factor analysis (FA), see Anderson (2003) Chapters 11 and 14. They cover multiple applications in finance and economics as well as in social sciences in general. They are often used in exploratory analysis of data. In recent work, Fortin, Gagliardini and Scaillet (FGS, 2022) show how we can use PCA to conduct inference on the number of factors in such models without making Gaussian assumptions. Their methodology relies on sphericity of the idiosyncratic variances since this restriction is both necessary and sufficient for consistency of latent factor estimates with small T (Theorem 4 of Bai (2003)). In PCA, sphericity allows to identify the number k of factors from the k first eigenvalue spacings being larger than zero, and being zero the subsequent ones. On the contrary, the FA strategy does not exploit eigenvalue spacings and does not require sphericity. However, inference with small T up to now mostly relies on (often restrictive) assumptions such as Gaussian variables (with a notable exception by Anderson and Amemiya (1988)) and error homoschedasticity across sample units. Those are untenable assumptions in our application with stock returns.

A central and practical issue in applied work with latent factors is to determine the number of factors. 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 balanced panels with static factors 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. Ando and Bai (2015) extend that approach when explanatory variables are present in the linear specification (see Bai (2009) for homogeneous regression coefficients). Onatski (2010) looks at the behavior of differences in adjacent eigenvalues to determine the number of factors when n and T are both large and comparable. Ahn and Horenstein (2013) opt for a similar strategy based on eigenvalue ratios. Caner and Han (2014) propose an estimator with a group bridge penalization to determine the number of unobservable factors. Based on the framework of Gagliardini, Ossola and Scaillet (2016), Gagliardini, Ossola and Scaillet (2019) build a simple diagnostic criterion for approximate factor structure in large panel datasets. Given observable factors, the criterion checks whether the errors are weakly cross-sectionally correlated, or share one or more unobservable common factors (interactive effects), and selects their number; see Gagliardini, Ossola and Scaillet (2020) for a survey of estimation of large dimensional conditional factor models in finance. 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.

This paper puts forward methodological and empirical contributions that complement the above literature. (i) On the methodological side, we extend the inferential tools of FA to non-Gaussian and non-i.i.d. settings. First, we characterize the asymptotic distribution of FA estimators obtained under a pseudo maximum likelihood approach where the time-series dimension is held fixed while the cross-sectional dimension diverges. Hence, the asymptotic analysis targets short panels, and allows for cross-sectionally heteroschedastic and weakly dependent errors. Cochrane (2005, p. 226) argues in favour of the development of appropriate large-n small-T tools for evaluating asset pricing models, a problem only partially addressed in finance. In a short panel setting, Zaffaroni (2019) considers inference for latent factors in conditional linear asset pricing models under sphericity based on PCA, including estimation of the number of factors. [1] The small T setting

[1] Raponi, Robotti and Zaffaroni (2020) develop tests of beta-pricing models and a two-pass methodology to estimate

mitigates concerns for panel unbalancedness and corresponds to a locally time-invariant factor structure accommodating globally time-dependent features of general forms. It is also appealing to macroeconomic data observed quarterly. For the sake of space, we put part of the theory, namely inference for FA estimates, in the Online Appendix (OA). We refer to Bai and Li (2016) for inference when n and T are both large (see Bai and Li (2012) for the cross-sectional independent case). Second, we use our new theoretical results for FA to develop testing procedures for the number of latent factors in a short panel which rely neither on sphericity nor Gaussianity, thereby extending tests based on eigenvalues, as in Onatski (2009), to small T, and as in FGS, to non-spherical errors, thanks to an FA device. We derive the Asymptotically Uniformly Most Powerful Invariant (AUMPI) property of the FA likelihood ratio (LR) test statistic in the non-Gaussian case under inequality restrictions on the DGP parameters, and cover inference with weak factors. The AUMPI property is rare and sought-after in testing procedures (see Engle (1984) for a discussion), and often holds only under restrictive assumptions such as Gaussianity. (ii) On the empirical side, we apply our FA methodology to panels of monthly U.S. stock returns with large cross-sectional and small time-series dimensions, and investigate how the number of driving factors changes over time and particular periods. Furthermore, date after date, we provide a novel separation of the risk coming from the systematic part and the risk coming from the idiosyncratic part of returns in short subperiods of bear vs. bull market based on the selected number of factors. We observe an uptrend in idiosyncratic volatility (see also Campbell et al. (2023)) while the systematic risk explains a large part of the cross-sectional total variance in bear markets but is not driven by a single factor. We also investigate whether standard observed factors span the estimated latent factors using rank tests in order to suit our fixed T setting. Observed factors struggle spanning latent factors with a discrepancy between the dimensions of the two factor spaces decreasing over time.

the ex-post risk premia (Shanken (1992)) associated to observable factors (see Kleibergen and Zhan (2023) for robust-identification inference based a continuous updating generalized method of moments). Kim and Skoulakis (2018) deals with the error-in-variable problem of the two-pass methodology with small T by regression-calibration under sphericity and a block-dependence structure.

The outline of the paper is as follows. In Section 2, we consider a linear latent factor model and introduce test statistics on the number of latent factors based on FA. Section 3 presents the asymptotic distributional theory for inference in short panels under a block-dependence structure to allow for weak dependence in the cross-section. Section 4 discusses three special cases, i.e. Gaussian errors, settings where the asymptotic distribution under Gaussian errors still holds for the test statistics, and spherical errors. Section 5 is dedicated to local asymptotic power and AUMPI tests. We provide our empirical application in Section 6 and our concluding remarks in Section 7. Appendices A and B gather the regularity assumptions and proofs of the main theoretical results. We place all omitted proofs and additional analyses in Appendices C-E in Online Appendix (OA). Besides, we gather all explicit formulas not listed in the core text but useful for coding in an online "Supplementary Materials for Coding" (SMC) attached to the replication files. We also put there other numerical checks and a Monte Carlo assessment of size and power for the LR test statistic.

2 Test statistics based on Factor Analysis

We consider the linear Factor Analysis (FA) model (e.g. Anderson (2003)):

$$y_i = \mu + F\beta_i + \varepsilon_i, \qquad i = 1, ..., n, \tag{1}$$

where $y_i = (y_{i,1}, ..., y_{i,T})'$ and $\varepsilon_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$ are T-dimensional vectors of observed data and unobserved error terms for individual i. The k-dimensional vectors $\beta_i = (\beta_{i,1}, ..., \beta_{i,k})'$ are latent individual effects, while μ and F are a $T \times 1$ vector and a $T \times k$ matrix of unknown parameters. The number of latent factors k is an unknown integer smaller than T. In matrix notation, model (1) reads $Y = \mu 1_n' + F\beta' + \varepsilon$, where Y and ε are $T \times n$ matrices, β is the $n \times k$ matrix with rows β_i', and 1_n is a n-dimensional vector of ones.

Assumption 1 *The $T \times T$ matrix $V_\varepsilon = \lim_{n\to\infty} E[\frac{1}{n}\varepsilon\varepsilon']$ is diagonal.*

Matrix V_ε is the limit cross-sectional average of the - possibly heterogeneous - errors' unconditional variance-covariance matrices. The diagonality condition in Assumption 1 is standard in FA (in the more restrictive formulation involving i.i.d. data).

In our empirics with a large cross-sectional panel of returns for n assets over a short time span with T periods, vectors y_i and ε_i stack the monthly returns and the idiosyncratic errors of stock i. Any row vector $f_t' := (f_{t,1}, ..., f_{t,k})$ of matrix F yields the latent factor values in a given month t, and vector β_i collects the factor loadings of stock i. In our finance application, we assume the No-Arbitrage (NA) principle to hold, so that the entries μ_t in the intercept vector in Equation (1) account for the (possibly time-varying) risk-free rate and (possibly non-zero) cross-sectional mean of stock betas. [2] Thus, the linear FA model (1) yields $y_{i,t} = \mu_t + f_t'\beta_i + \varepsilon_{i,t}$, that is the standard formulation in asset pricing. We cover the Capital Asset Pricing Model (CAPM) when the single latent factor is the excess return of the market portfolio. Assumption 1 allows for serial dependence in idiosyncratic errors in the form of martingale difference sequences, like individual GARCH and Stochastic Volatility (SV) processes, as well as weak cross-sectional dependence (see Assumption 2 below). It also accommodates common time-varying components in idiosyncratic volatilities by allowing different entries along the diagonal of V_ε; see Renault, Van Der Heijden and Werker (2022) for arbitrage pricing in such settings. [3]

This paper focuses mainly on testing hypotheses on the number of latent factors k when T is

[2] Under NA, the intercept term in the asset return model $y_i = \mu_i + \tilde{F}\tilde{\beta}_i + \varepsilon_i$ is $\mu_i = r_f + 1_T\nu'\tilde{\beta}_i$, where r_f is the T-dimensional vector whose entries collect the (possibly time-varying) risk-free rates, $\nu = (\nu_1, ..., \nu_k)'$ is a k-dimensional vector of parameters, and 1_T is a T-dimensional vector of ones (see e.g. Gagliardini, Ossola and Scaillet (2016)). We can absorb term $1_T\nu'\tilde{\beta}_i$ into the systematic part to get $y_i = r_f + F\tilde{\beta}_i + \varepsilon_i$ with $F = \tilde{F} + 1_T\nu'$. It holds irrespective of the latent factors being tradable or not. If the factors are tradable, we further have $\nu = 0$ from the NA restriction. Akin to standard formulation of FA, we recenter the latent effects by subtracting their mean $\tilde{\mu}_{\tilde{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\beta}_i$, to get model (1) with $\beta_i = \tilde{\beta}_i - \tilde{\mu}_{\tilde{\beta}}$ and $\mu = r_f + F\tilde{\mu}_{\tilde{\beta}}$.

[3] When there is a common random component in idiosyncratic volatilities, we have $V_\varepsilon = \underset{n\to\infty}{plim}\frac{1}{n}\varepsilon\varepsilon'$ by a suitable version of the Law of Large Number (LLN) conditional on the sigma-field generated by this common component. With fixed T, we treat the sample realizations of the common component in idiosyncratic volatilities as unknown time fixed effects (the diagonal elements of matrix V_ε), which yields time heterogeneous distributions for the errors. It is how the unconditional expectation in Assumption 1 has to be understood.

fixed and $n \to \infty$. The fixed T perspective makes FA especially well-suited for applications with short panels. Indeed, we work conditionally on the realizations of the latent factors F and treat their values as parameters to estimate. In comparison with the standard small n and large T framework in traditional asset pricing (see e.g. Shanken (1992) with observable factors), here factors and loadings are interchanged in the sense that the β_i and F play the roles of the "factors" and the "factor loadings" in FA. We depart from classical FA since the β_i are not considered as random effects (e.g. with a Gaussian distribution) but rather as fixed effects, namely incidental parameters. [4] Moreover, in Assumption 1, we neither assume Gaussianity nor we impose sphericity of the covariance matrix of the error terms. Besides we accommodate weak cross-sectional dependence and ARCH effects in idiosyncratic errors (see Section 3). Hence, the FA estimators defined below correspond to maximizers of a Gaussian pseudo likelihood. By-products of our analysis are the feasible asymptotic distributions of FA estimators of F and V_ε in more general settings than in the available literature (e.g. Anderson and Amemiya (1988)), which we present in Appendix D.

The test statistics we consider for conducting inference on the number of latent factors k are functions of the elements of the symmetric matrix

$$\hat{S} = \hat{V}_\varepsilon^{-1/2} M_{\hat{F},\hat{V}_\varepsilon} (\hat{V}_y - \hat{V}_\varepsilon) M'_{\hat{F},\hat{V}_\varepsilon} \hat{V}_\varepsilon^{-1/2}, \tag{2}$$

where $\hat{V}_y = \frac{1}{n}\tilde{Y}\tilde{Y}'$ is the sample (cross-sectional) variance matrix (the n columns of $\tilde{Y}$ are $y_i - \bar{y}$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ is the vector of cross-sectional means), $M_{F,V} := I_T - F(F'V^{-1}F)^{-1}F'V^{-1}$ is the Generalized Least Squares (GLS) projection matrix orthogonal to F for variance V, and $\hat{F}$ and $\hat{V}_\varepsilon$ are the FA estimators computed under the assumption that there are k latent factors. In the following, we use the same notation for the matrix-to-vector diag operator and the vector-to-matrix diag operator. Hence, $diag(A)$ for a matrix A denotes the vector in which we stack the diagonal elements of matrix A, and $diag(a)$ for a vector a denotes a diagonal matrix with the elements of a on the diagonal. From Anderson (2003) Chapter 14, the FA estimators $\hat{F}$, $\hat{V}_\varepsilon$ maximize a Gaussian

[4] Chamberlain (1992) studies semiparametrically efficient estimation in panel models with fixed effects and short T using moment restrictions from instrumental variables. Our approach does not rely on availability of valid instruments.

pseudo likelihood (Appendix D.1) and meet the first order conditions: [5]

(FA1) $diag(\hat{V}_y) = diag(\hat{F}\hat{F}' + \hat{V}_\varepsilon)$, and

(FA2) $\hat{F}$ is the $T \times k$ matrix of eigenvectors of $\hat{V}_y\hat{V}_\varepsilon^{-1}$ associated to the k largest eigenvalues $1 + \hat{\gamma}_j$, $j = 1, ..., k$, normalized such that $\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F} = diag(\hat{\gamma}_1, ..., \hat{\gamma}_k)$.

The number of degrees of freedom is $df = \frac{1}{2}((T-k)^2 - T - k)$ and it is required that $df \geq 0$. [6]

Statistic $\hat{S}$ in Equation (2) checks if the difference between the sample variance-covariance $\hat{V}_y$ and diagonal matrix $\hat{V}_\varepsilon$ is a symmetric matrix of reduced rank k, with range spanned by the range of $\hat{F}$. The probability limit of $\hat{S}$ is nil under the null hypothesis of k latent factors. We get further insights from the next result.

Proposition 1 *Under Assumption 1, (a) the eigenvalues of matrix $\hat{S}$ are: $\hat{\gamma}_j$, for $j = k+1, ..., T$, and 0, with multiplicity k, where $1 + \hat{\gamma}_j$ for $j = k+1, ..., T$ are the $T - k$ smallest eigenvalues of $\hat{V}_y\hat{V}_\varepsilon^{-1}$, (b) the squared Frobenius norm is $\|\hat{S}\|^2 = \sum_{j=k+1}^{T} \hat{\gamma}_j^2$, (c) $diag(\hat{S}) = 0$, and (d) we get*

$$\hat{S} = \hat{V}_\varepsilon^{-1/2}\left(\frac{1}{n}\hat{\varepsilon}\hat{\varepsilon}'\right)\hat{V}_\varepsilon^{-1/2} - \hat{V}_\varepsilon^{-1/2}M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon^{1/2}, \tag{3}$$

where $\hat{\varepsilon} = M_{\hat{F},\hat{V}_\varepsilon}\tilde{Y}$ is the $T \times n$ matrix of GLS residuals.

From Proposition 1 (a)-(c), the squared Frobenius norm of matrix $\hat{S}$ multiplied by $n/2$ coincides at second order with the classical LR statistic in FA, i.e., $LR(k) = -n\sum_{j=k+1}^{T}\log(1 + \hat{\gamma}_j)$. [7] Moreover, from (d) we can interpret matrix $\hat{S}$ in terms of scaled cross-sectional averages of

[5] The normalization in (FA2) applies for $\hat{\gamma}_j \geq 0$, which holds with probability approaching 1. Otherwise, the first-order conditions of the FA estimators hold with $\hat{\gamma}_j$ replaced by its positive part; see Anderson (2003) for similar positivity constraints.

[6] Integer df is the number of different elements in data matrix $\hat{V}_y$, i.e., $\frac{1}{2}T(T+1)$, plus the number of normalization constraints $\frac{1}{2}k(k-1)$ in equations $F'V_\varepsilon^{-1}F = diag$, minus the number of unknown parameters $(k+1)T$ (Anderson (2003)). In SMC Table 3, we list the largest admissible number k of latent factors as a function of T such that $df \geq 0$.

[7] It comes from second-order expansion of the $\log$ function using $\sum_{j=k+1}^{T}\hat{\gamma}_j = 0$, which is a consequence of Proposition 1 (a) and (c) (see also Anderson (2003)), and $\sqrt{n}\hat{\gamma}_j = O_p(1)$ for $j = k+1, ..., T$, which follows from Propositions 1 and 4.

squared and cross-products of GLS residuals. In (3), we subtract $\hat{V}_\varepsilon^{-1/2} M_{\hat{F},\hat{V}_\varepsilon} \hat{V}_\varepsilon^{1/2}$ and not the identity because the residuals are orthogonal to $\hat{F}$ by construction. From Proposition 1 (c), the diagonal elements of matrix $\hat{S}$ vanish. Those elements are not informative for inference on the number of factors, and can be ignored when constructing the test statistics. This finding is natural because we expect that only the out-of-diagonal elements of $\frac{1}{n}\hat{\varepsilon}\hat{\varepsilon}'$, i.e., the cross-sectional averages of cross-products of residuals for two different dates, are useful to check for omitted factors. [8]

We summarize the statistics to test null hypotheses on the number of latent factors next.

Definition 1 *The statistics to test the null hypothesis $H_0(k)$ of k latent factors are: (a) the squared norm statistic $\mathscr{T}(k) := n \sum_{j=k+1}^{T} \hat{\gamma}_j^2 = n\|\hat{S}\|^2$, and (b) the LR statistic $LR(k) := -n \sum_{j=k+1}^{T} \log(1+\hat{\gamma}_j) = \frac{n}{2}\|\hat{S}\|^2 + o_p(1)$, where $\hat{\gamma}_{k+j} = \delta_{k+j}(\hat{V}_y \hat{V}_\varepsilon^{-1}) - 1 = \delta_j(\hat{S})$, for $j = 1, ..., T-k$, and we denote by $\delta_j(\cdot)$ the jth largest eigenvalue of a symmetric matrix.*

The statistics in Definition 1 only use the information contained in the eigenvalues of matrix $\hat{S}$. Next we establish the asymptotic distributions of those test statistics with $n \to \infty$ and T fixed.

3 Asymptotic distributional theory

We start by defining the normalization for the latent factor matrix $F = [F_1 : \cdots : F_k]$ in population. Following classical FA, we set $\mu_\beta = 0$, $V_\beta = I_k$, and $F' V_\varepsilon^{-1} F = diag(\gamma_1, ..., \gamma_k)$, where $V_\beta = \lim_{n\to\infty} \frac{1}{n}\beta'\beta$ and $\mu_\beta = \lim_{n\to\infty} \bar{\beta}$ with $\bar{\beta} = \frac{1}{n}\sum_{i=1}^{n} \beta_i$. Then, under our assumptions, we have $V_y := \operatorname{plim}_{n\to\infty} \hat{V}_y = FF' + V_\varepsilon$ and $V_y V_\varepsilon^{-1} F_j = (1+\gamma_j) F_j$, i.e., the F_j are eigenvectors of matrix $V_y V_\varepsilon^{-1}$

[8] The test in Connor and Korajczyk (1993) is built on cross-sectional averages of squared residuals, akin to diagonal terms of $\hat{S}$, but obtained by PCA instead of FA. However, their test statistic involves the difference of such cross-sectional averages for two consecutive dates, and relies on error sphericity. Furthermore, from Proposition 1 (a), we note that test statistics based on the spacings between the small eigenvalues of $\hat{V}_y \hat{V}_\varepsilon^{-1}$ use the non-zero eigenvalues of $\hat{S}$. Such tests rely on the possibility to identify the number of latent factors k from the fact that the k smallest eigenvalues of $V_y V_\varepsilon^{-1}$ are all equal to 1.

associated with eigenvalues $1+\gamma_j$, $j=1,...,k$. [9] For any given n, we define $\tilde{V}_\varepsilon = \frac{1}{n}E[\varepsilon\varepsilon']$ and $\tilde{V}_\beta = \frac{1}{n}\beta'\beta$, and use a factor normalization in sample that is analogue to the one in population, i.e., $\bar{\beta}=0$, $\tilde{V}_\beta = I_k$ (see Assumption A.1) and $F'\tilde{V}_\varepsilon^{-1}F$ is diagonal. Thus, the normalization of the factor values $F=F_{(n)}$ is sample dependent; we skip index n for the purpose of easing notation.

We use a block-dependence structure to allow for weak cross-sectional dependence in errors.

Assumption 2 *(a) The errors are such that* $\varepsilon = V_\varepsilon^{1/2}W\Sigma^{1/2}$*, where* $W=[w_1:\cdots:w_n]$ *is a* $T\times n$ *random matrix of standardized errors terms* $w_{i,t}$ *that are independent across* i *and uncorrelated across* t*, and* $\Sigma=(\sigma_{i,j})$ *is a positive-definite symmetric* $n\times n$ *matrix, such that* $\lim\limits_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sigma_{ii} = 1$*. (b) Matrix* Σ *is block diagonal with* J_n *blocks of size* $b_{m,n}=B_{m,n}n$*, for* $m=1,...,J_n$*, where* $J_n\to\infty$ *as* $n\to\infty$*, and* I_m *denotes the set of indices in block* m*. (c) There exist constants* $\delta\in[0,1]$ *and* $C>0$ *such that* $\max\limits_{i\in I_m}\sum_{j\in I_m}|\sigma_{i,j}| \leq Cb_{m,n}^\delta$*. (d) The block sizes* $b_{m,n}$ *and block number* J_n *are such that* $n^{2\delta}\sum_{m=1}^{J_n} B_{m,n}^{2(1+\delta)} = o(1)$.

As already remarked, the diagonal elements of V_ε are the sample realizations of the common component driving the variance of the error terms at times $t=1,...,T$; see e.g. Barigozzi and Hallin (2016), Renault, Van Der Heijden and Werker (2022) for theory and empirical evidence pointing to variance factors. A sphericity assumption cannot accommodate such a common time-varying component. In empirical applications on individual stocks, blocks in Σ can match industrial sectors (Gagliardini, Ossola, and Scaillet (2016)). Assumption 2 (a) is coherent with Assumption 1. Indeed, $\frac{1}{n}E[\varepsilon\varepsilon'] = V_\varepsilon^{1/2}\frac{1}{n}\sum_{i,j=1}^n \sigma_{i,j}E[w_iw_j']V_\varepsilon^{1/2} = \frac{1}{n}\sum_{i=1}^n \sigma_{ii}V_\varepsilon$ is diagonal, and converges to matrix V_ε in the limit $n\to\infty$ under the normalization $\lim\limits_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\sigma_{ii}=1$. That normalization is without loss of generality by rescaling of the parameters. Assumption 2 (c) builds on Bickel and Levina (2008), and $\delta<1$ holds under sparsity, vanishing correlations or mixing dependence within blocks. With blocks of equal size, Assumption 2 (d) holds for $J_n = n^{\bar{\alpha}}$ and $\bar{\alpha} > \frac{2\delta}{2\delta+1}$. Hav-

[9] The remaining eigenvalue is equal to 1 with multiplicity $T-k$. We have $F_j = \sqrt{\gamma_j}V_\varepsilon^{1/2}U_j$, where the U_j are the orthonormal eigenvectors of $V_\varepsilon^{-1/2}V_yV_\varepsilon^{-1/2}$ for the k largest eigenvalues $1+\gamma_j$.

ing $\delta < 1$ helps relaxing this condition on block granularity, however it is not strictly necessary because we allow value $\delta = 1$.

3.1 Asymptotic expansions of estimators $\hat{V}_\varepsilon$ and $\hat{F}$

Using Equation (1) and the factor normalization in sample, we have $\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}\Psi_y + o_p(\frac{1}{\sqrt{n}})$, where $\tilde{V}_y = FF' + \tilde{V}_\varepsilon$ and $\Psi_y = \frac{1}{\sqrt{n}}(\varepsilon\beta F' + F\beta'\varepsilon') + \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right)$ (see Appendix D.2). The FA estimators $\hat{V}_\varepsilon$ and $\hat{F}$ are consistent M-estimators under nonlinear constraints, and admit expansions at first order for fixed T and $n \to \infty$, namely $\hat{V}_\varepsilon = \tilde{V}_\varepsilon + \frac{1}{\sqrt{n}}\Psi_\varepsilon + o_p(\frac{1}{\sqrt{n}})$ and $\hat{F}_j = F_j + \frac{1}{\sqrt{n}}\Psi_{F_j} + o_p(\frac{1}{\sqrt{n}})$ (see Appendix D.4.1). The next proposition characterizes the diagonal random matrix Ψ_ε and the random vectors Ψ_{F_j} by using conditions (FA1) and (FA2) above.

Assumption 3 *The non-zero eigenvalues of $V_y V_\varepsilon^{-1} - I_T$ are distinct, i.e., $\gamma_1 > ... > \gamma_k > 0$.*

Proposition 2 *Under Assumptions 1-3 and A.1-A.4, we have (a) for $j = 1, ..., k$*

$$\Psi_{F_j} \quad = \quad R_j(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \Lambda_j \Psi_\varepsilon V_\varepsilon^{-1}F_j, \tag{4}$$

where $R_j := \frac{1}{2\gamma_j}P_{F_j,V_\varepsilon} + \frac{1}{\gamma_j}M_{F,V_\varepsilon} + \sum_{\ell=1,\ell\neq j}^k \frac{1}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}$ and $\Lambda_j := -\sum_{\ell=1,\ell\neq j}^k \frac{\gamma_l}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}$ and $P_{F_j,V_\varepsilon} = F_j(F_j'V_\varepsilon^{-1}F_j)^{-1}F_j'V_\varepsilon^{-1} = \frac{1}{\gamma_j}F_jF_j'V_\varepsilon^{-1}$ is the GLS orthogonal projection onto F_j. Further, (b) the diagonal matrix Ψ_ε is such that:

$$diag\left(M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M_{F,V_\varepsilon}'\right) = 0. \tag{5}$$

Equation (4) yields the asymptotic expansion of the eigenvectors by accounting for estimation errors of matrix $\hat{V}_y\hat{V}_\varepsilon^{-1}$ (first term) and of the normalization constraint (second term). To interpret Equation (5), we can observe that the matrix $M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M_{F,V_\varepsilon}'$ yields the first-order term in the asymptotic expansion of the test statistic $\sqrt{n}\hat{S}$ (up to the left- and right-multiplication by diagonal matrix $V_\varepsilon^{-1/2}$); see Equation (8). Thus, Equation (5) is implied by the property that the diagonal terms of matrix $\hat{S}$ are equal to zero as stated in Proposition 1 (c).

Let us now give the explicit expression of Ψ_ε. By using $M_{F,V_\varepsilon}\Psi_y M'_{F,V_\varepsilon} = M_{F,V_\varepsilon} Z_n M'_{F,V_\varepsilon}$, where $Z_n := \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \frac{1}{n}E[\varepsilon\varepsilon']\right)$ is the standardized, centered sample mean of cross-moments of errors, we can rewrite Equation (5) as $diag\left(M_{F,V_\varepsilon}(Z_n - \Psi_\varepsilon)M'_{F,V_\varepsilon}\right) = 0$. Now, because Ψ_ε is diagonal, we have $diag\left(M_{F,V_\varepsilon}\Psi_\varepsilon M'_{F,V_\varepsilon}\right) = M^{\odot 2}_{F,V_\varepsilon} diag(\Psi_\varepsilon)$, where $M^{\odot 2}_{F,V_\varepsilon} = M_{F,V_\varepsilon} \odot M_{F,V_\varepsilon}$ and $\odot$ denotes the Hadamard product (i.e., element-wise matrix product). Thus, we get the equation:

$$M^{\odot 2}_{F,V_\varepsilon} diag(\Psi_\varepsilon) = diag(M_{F,V_\varepsilon} Z_n M'_{F,V_\varepsilon}). \tag{6}$$

To have a unique solution for vector $diag(\Psi_\varepsilon)$, we need the non-singularity of the $T \times T$ matrix on the LHS of this linear equation system. It is the local identification condition in the FA model (see Lemma 7 in Appendix D.3 i), where we show equivalence with invertibility of the bordered Hessian, i.e., the Hessian of the Lagrangian function in a constrained M-estimation).

Assumption 4 *Matrix $M^{\odot 2}_{F,V_\varepsilon}$ is non-singular.*

Under Assumption 4, we get from Equation (6):

$$\Psi_\varepsilon = \mathcal{T}_{F,V_\varepsilon}(Z_n), \tag{7}$$

where $\mathcal{T}_{F,V_\varepsilon}(V) := diag\left([M^{\odot 2}_{F,V_\varepsilon}]^{-1} diag(M_{F,V_\varepsilon} V M'_{F,V_\varepsilon})\right)$, for any $T \times T$ matrix V. Mapping $\mathcal{T}_{F,V_\varepsilon}(\cdot)$ is linear and such that $\mathcal{T}_{F,V_\varepsilon}(V) = V$, for a diagonal matrix V.

Anderson and Rubin (1956), Theorem 12.1, show that the FA estimator is asymptotically normal if $\sqrt{n}(\hat{V}_y - V_y)$ is asymptotically normal. They use a linearization of the first-order conditions similar as the one of Proposition 2. Their Equation (12.16) corresponds to our Equation (5). However, they only provide an implicit characterization of the Ψ_{F_j} and not an explicit expression for Ψ_ε and Ψ_{F_j} in terms of asymptotically Gaussian random matrices like Z_n as we do. These key developments pave the way to establishing the asymptotic distributions of estimators $\hat{F}$ and $\hat{V}_\varepsilon$ in general settings, that we cover in OA Section D.4, and of the test statistics for the number of factors, that we address next.

3.2 Asymptotic expansions of the test statistics

By expanding the terms in the definition of $\hat{S}$ in Equation (2), and using Equation (7) and the $\sqrt{n}$-consistency of FA estimators (see Appendix D.4.1), we have:

$$\begin{aligned}
\sqrt{n}\hat{S} &= V_\varepsilon^{-1/2} M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon) M'_{F,V_\varepsilon} V_\varepsilon^{-1/2} + o_p(1) \\
&= V_\varepsilon^{-1/2} M_{F,V_\varepsilon}(Z_n - \mathcal{T}_{F,V_\varepsilon}(Z_n)) M'_{F,V_\varepsilon} V_\varepsilon^{-1/2} + o_p(1). \qquad (8)
\end{aligned}$$

Let us now rework the RHS. First, we use that $Z_n = \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right)$, where $\tilde{V}_\varepsilon = \frac{1}{n}E[\varepsilon\varepsilon']$ is diagonal and such that $\mathcal{T}_{F,V_\varepsilon}(\tilde{V}_\varepsilon) = \tilde{V}_\varepsilon$. Thus, we have $Z_n - \mathcal{T}_{F,V_\varepsilon}(Z_n) = \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \mathcal{T}_{F,V_\varepsilon}(\frac{1}{n}\varepsilon\varepsilon')\right) =: \bar{Z}_n$. We get that $E[\bar{Z}_n] = 0$ because diagonal matrices are invariant under mapping $\mathcal{T}_{F,V_\varepsilon}(\cdot)$. Second, we write the orthogonal projection as $M_{F,V_\varepsilon} = GG'V_\varepsilon^{-1}$, where G be a $T \times (T-k)$ matrix such that $F'V_\varepsilon^{-1}G = 0$ and $G'V_\varepsilon^{-1}G = I_{T-k}$. Matrix G is unique up to post-multiplication by an orthogonal matrix. Then, from Equation (8), we get the asymptotic expansion of matrix $\hat{S}$ as

$$\sqrt{n}\hat{S} = V_\varepsilon^{-1/2} G \bar{Z}_n^* G' V_\varepsilon^{-1/2} + o_p(1), \qquad (9)$$

where $\bar{Z}_n^* := G'V_\varepsilon^{-1}\bar{Z}_n V_\varepsilon^{-1}G$. The asymptotic distribution of $\sqrt{n}\hat{S}$ is driven by the symmetric $(T-k)\times(T-k)$ zero-mean matrix $\bar{Z}_n^*$. We have $diag(V_\varepsilon^{-1/2}G\bar{Z}_n^*G'V_\varepsilon^{-1/2}) = 0$ as a consequence of Equation (5) and results above (see also Proposition 1 (c)). We can rewrite the number of degrees of freedom as $df = \frac{1}{2}(T-k)(T-k+1) - T$, i.e., the number of different elements in $\bar{Z}_n^*$ minus the number of linear constraints in $diag(V_\varepsilon^{-1/2}G\bar{Z}_n^*G'V_\varepsilon^{-1/2}) = 0$. Matrices $V_\varepsilon^{-1/2}G\bar{Z}_n^*G'V_\varepsilon^{-1/2}$ and $\bar{Z}_n^*$ have the same Frobenius norm because the columns of $V_\varepsilon^{-1/2}G$ are orthonormal. From Equation (9), the asymptotic expansions for the test statistics in Definition 1 are:

$$\mathscr{T}(k) = \|\bar{Z}_n^*\|^2 + o_p(1), \qquad LR(k) = \frac{1}{2}\|\bar{Z}_n^*\|^2 + o_p(1), \qquad (10)$$

under the null hypothesis of k latent factors. [10]

[10] We can extend results like (10) to test statistics that are generic functions of the eigenvalues of matrix $\hat{S}$ by using the Weyl's inequalities (see e.g. Bernstein (2009)), and develop test statistics along the lines of FGS.

We can get further insights in the above results by using the next proposition. Let the T-dimensional vectors g_j for $j = 1, ..., T-k$ be the columns of matrix G, and let us define the $p \times T$ matrix $\boldsymbol{X} = \left[g_1 \odot g_1 \, : \, \cdots \, : \, g_{T-k} \odot g_{T-k} \, : \, \{\sqrt{2}(g_i \odot g_j)\}_{i<j} \right]'$, where the pairs of indices (i, j) with $i < j$ are ranked as $(1, 2), (1, 3), ..., (1, T-k), (2, 3), ..., (T-k-1, T-k)$, and $p = \frac{1}{2}(T-k)(T-k+1)$. Moreover, for a $(T-k) \times (T-k)$ symmetric matrix $Z = (z_{i,j})$, let us define the p-dimensional vector $vech(Z) = \left(\frac{1}{\sqrt{2}} z_{11}, ..., \frac{1}{\sqrt{2}} z_{T-k,T-k}, \{z_{i,j}\}_{i<j} \right)'$, where the out-of-diagonal elements with indices $i < j$ are ranked as above. [11]

Proposition 3 *Under Assumptions 1-4, we have (a)* $M^{\odot 2}_{F,V_\varepsilon} = \boldsymbol{X}'\boldsymbol{X} V_\varepsilon^{-2}$, *(b)* $diag(\Psi_\varepsilon) = \sqrt{2} V_\varepsilon^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' vech(Z_n^*)$, *where* $Z_n^* = G' V_\varepsilon^{-1} Z_n V_\varepsilon^{-1} G$, *and (c)* $vech(\bar{Z}_n^*) = (I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\, vech(Z_n^*)$.

From Proposition 3 (a), we can state the local identification condition in Assumption 4 as a full-rank condition for matrix $\boldsymbol{X}$ analogously as in linear regression. In part (b), we write the diagonal of Ψ_ε via the coefficients of a OLS regression of the half-vectorization of Z_n^* onto $\boldsymbol{X}$. Part (c) shows that, after half-vectorization, we can represent the elements of matrix variate $\bar{Z}_n^*$ as the residual of the orthogonal projection of $vech(Z_n^*)$ onto the columns of $\boldsymbol{X}$. Matrix $I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is idempotent of rank $p - T = df$. Using $\frac{1}{2}\|\bar{Z}_n^*\|^2 = vech(\bar{Z}_n^*)' vech(\bar{Z}_n^*)$ and Proposition 3 (c), the leading term in the asymptotic expansions (10) is $\frac{1}{2}\|\bar{Z}_n^*\|^2 = vech(Z_n^*)' (I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\, vech(Z_n^*)$.

3.3 Feasible Central Limit Theorem

We now establish the distributional convergence $\bar{Z}_n^* \Rightarrow \bar{Z}^*$ as $n \to \infty$ and T is fixed, where $\bar{Z}^*$ is a Gaussian symmetric matrix variate. We have that $\bar{Z}^* = G' V_\varepsilon^{-1} (Z - \mathcal{T}_{F,V_\varepsilon}(Z))\, V_\varepsilon^{-1} G$, where Z is the distributional limit of Z_n. Establishing a feasible Central Limit Theorem (CLT) via a nonparametric estimator of the asymptotic variance is easier for $\bar{Z}_n^*$ than for Z_n, and is sufficient for

[11] This definition of the half-vectorization operator for symmetric matrices differs from the usual one by the ordering of the elements, and the rescaling of the diagonal elements. It is more convenient for our purposes (see proof of Lemma 11). For instance, it holds $\frac{1}{2}\|A\|^2 = vech(A)' vech(A)$, for a symmetric matrix A.

testing purposes. [12] By the block structure in Assumption 2 (b), we can write $\bar{Z}_n^*$ as a sum of independent zero-mean terms: $\bar{Z}_n^* = \frac{1}{\sqrt{n}} G' V_\varepsilon^{-1} \left(\varepsilon\varepsilon' - \mathcal{T}_{F,V_\varepsilon}(\varepsilon\varepsilon')\right) V_\varepsilon^{-1} G = \frac{1}{\sqrt{n}} \sum_{m=1}^{J_n} z_{m,n}$, where the variables in the triangular array $z_{m,n} = \sum_{i,j \in I_m} \sigma_{i,j} G' V_\varepsilon^{-1/2} \left[w_i w_j' - \mathcal{T}_{F,V_\varepsilon}(w_i w_j')\right] V_\varepsilon^{-1/2} G = \sum_{i \in I_m} G' V_\varepsilon^{-1} [\varepsilon_i \varepsilon_i' - \mathcal{T}_{F,V_\varepsilon}(\varepsilon_i \varepsilon_i')] V_\varepsilon^{-1} G$ are independent across m and such that $E[z_{m,n}] = 0$. In Appendix B, we invoque the CLT for independent heterogeneous variables to $vech(\bar{Z}_n^*) = \frac{1}{\sqrt{n}} \sum_{m=1}^{J_n} vech(z_{m,n})$ and use Assumptions 2 (c) and (d) to check the Liapunov condition. Then, we get $\bar{Z}_n^* \Rightarrow \bar{Z}^*$, where $vech(\bar{Z}^*) \sim N(0, \Omega_{\bar{Z}^*})$ and $\Omega_{\bar{Z}^*} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{J_n} V[vech(z_{m,n})]$. Our assumptions imply that $\Omega_{\bar{Z}^*}$ is finite. From Proposition 3 (c), we have $\Omega_{\bar{Z}^*} = (I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$ $\Omega_{Z^*}(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$, where $vech(Z^*) \sim N(0, \Omega_{Z^*})$, for $Z^* = G' V_\varepsilon^{-1} Z V_\varepsilon^{-1} G$. We characterize the variance $\Omega_Z = V[vech(Z)]$ of the distributional limit of Z_n in Lemma 1 in Appendix B. In particular, matrix $\Omega_{\bar{Z}^*}$ is singular with rank df. Then, the asymptotic expansions in (10) yield the asymptotic distributions for the test statistics.

Proposition 4 *Let Assumptions 1-4 and A.1-A.5 hold. As $n \to \infty$ and T is fixed, under the null hypothesis $H_0(k)$ of k latent factors, (a) $\mathscr{T}(k) \Rightarrow \|\bar{Z}^*\|^2$, $LR(k) \Rightarrow \frac{1}{2}\|\bar{Z}^*\|^2$, where $vech(\bar{Z}^*)$ $\sim N(0, \Omega_{\bar{Z}^*})$, and (b) $\hat{\mathscr{R}} \hat{\Omega}_{\bar{Z}^*} \hat{\mathscr{R}}^{-1} \xrightarrow{p} \Omega_{\bar{Z}^*}$, where $\hat{\Omega}_{\bar{Z}^*} = \frac{1}{n} \sum_{m=1}^{J_n} vech(\hat{z}_{m,n}) vech(\hat{z}_{m,n})'$ and $\hat{z}_{m,n} = \sum_{i \in I_m} \hat{G}' \hat{V}_\varepsilon^{-1} \left(\hat{\varepsilon}_i \hat{\varepsilon}_i' - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\hat{\varepsilon}_i \hat{\varepsilon}_i')\right) \hat{V}_\varepsilon^{-1} \hat{G}$, with $\hat{\varepsilon}_i = M_{\hat{F},\hat{V}_\varepsilon}(y_i - \bar{y})$, for an orthogonal matrix $\hat{\mathscr{R}}$. Under the alternative hypothesis $H_1(k)$ of more than k latent factors, (c) $\mathscr{T}(k) \geq Cn$ and $LR(k) \geq Cn$, w.p.a. 1 for a constant $C > 0$, and $\hat{\Omega}_{\bar{Z}^*} = O_p(n \sum_{m=1}^{J_n} B_{m,n}^2) = o_p(n)$.*

From Proposition 4 (a), using $\frac{1}{2}\|\bar{Z}^*\|^2 = vech(\bar{Z}^*)' vech(\bar{Z}^*) \sim \sum_{j=1}^{df} \mu_j \chi_j^2(1)$, the asymptotic distribution of the LR statistic is a weighted average of df mutually independent chi-square variates with weights μ_j that are the non-zero eigenvalues of matrix $\Omega_{\bar{Z}^*}$. [13] In Proposition 4 (b), we use $\hat{G} = \hat{V}_\varepsilon^{1/2} \hat{Q}$, where $\hat{Q}$ is a $T \times (T-k)$ matrix with orthonormal columns that span the range

[12] In Appendix D.4.3, we present a parametric estimator for the asymptotic variance of $vech(Z_n)$ under additional conditions on the error terms.

[13] In Proposition 10 in Appendix D.5, we study how $\boldsymbol{X}$ and $\Omega_{\bar{Z}^*}$ are transformed under different choices for the rotation of G. The eigenvalues μ_j are invariant to such rotation as expected.

of $I_T - \hat{V}_\varepsilon^{-1/2}\hat{F}(\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}\hat{F}'\hat{V}_\varepsilon^{-1/2}$. [14] The orthogonal matrix $\hat{\mathscr{R}}$ accounts for an arbitrary choice of that orthonormal basis. With fixed T, the GLS residuals $\hat{\varepsilon}_i$ are asymptotically close to $M_{F,V_\varepsilon}\varepsilon_i$ and not to the true errors ε_i. This fact does not impede the consistency of $\hat{\Omega}_{\bar{Z}^*}$, because $G'V_\varepsilon^{-1}M_{F,V_\varepsilon} = G'V_\varepsilon^{-1}$ and $\mathcal{T}_{F,V_\varepsilon}(M_{F,V_\varepsilon}VM'_{F,V_\varepsilon}) = \mathcal{T}_{F,V_\varepsilon}(V)$, for any V. We can consistently estimate the critical values of the asymptotic statistics $\|\bar{Z}^*\|^2$ by simulating a large number of draws from a Gaussian symmetric matrix variate with vectorized variance $\hat{\Omega}_{\bar{Z}^*}$, whose norms are unaffected by the orthogonal matrix $\hat{\mathscr{R}}$. Finally, Proposition 4 (c) gives test consistency against global alternative hypotheses.

4 Discussion of three special cases

In this section, we particularize the general distributional results of Proposition 4 to three important cases, namely Gaussian errors, settings where the asymptotic distribution under Gaussian errors still holds for the test statistics (up to scaling), and spherical errors.

4.1 Gaussian errors

Let us consider the case where the errors $\varepsilon_i \overset{ind}{\sim} N(0, \sigma_{ii}V_\varepsilon)$ are independent Gaussian vectors. From classical FA theory, we expect that the statistic $LR(k)$ admits asymptotically a chi-square distribution with df degrees of freedom in the cross-sectionally homoschedastic case, i.e., $\sigma_{ii} = 1$ for all assets i. We cannot expect that this distributional result applies to the Gaussian framework in full generality, since - even in such a case - our setting corresponds to a pseudo model (because the σ_{ii} may be heterogeneous across i, and the β_i are treated as fixed effects, namely incidental parameters, instead of Gaussian random effects). In order to establish the distribution of $\frac{1}{2}\|\bar{Z}^*\|^2 = vech(\bar{Z}^*)'vech(\bar{Z}^*)$, we use Proposition 3 (c) written for the distributional limits to get $\frac{1}{2}\|\bar{Z}^*\|^2 =$

[14] For instance, we can set $\hat{Q} = \tilde{Q}(\tilde{Q}'\tilde{Q})^{-1/2}$, where matrix $\tilde{Q}$ consists of the first $T-k$ columns of $I_T - \hat{V}_\varepsilon^{-1/2}\hat{F}(\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}\hat{F}'\hat{V}_\varepsilon^{-1/2}$, if those columns are linearly independent.

$vech(Z^*)' \left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right) vech(Z^*)$, where $I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is idempotent of rank df. Under the normality assumption for the error terms, we have $\varepsilon_i^* := G'V_\varepsilon^{-1}\varepsilon_i \overset{ind}{\sim} N(0, \sigma_{ii} I_{T-k})$. Thus, by the Liapunov CLT, the distributional limit of $\frac{1}{\sqrt{q}} Z_n^* = \sqrt{n/q}\left(\frac{1}{n}\varepsilon^*(\varepsilon^*)' - E\left[\frac{1}{n}\varepsilon^*(\varepsilon^*)'\right]\right)$ is in the Gaussian Orthogonal Ensemble (GOE) for dimension $T-k$ (see e.g. Tao (2012)), i.e., $\frac{1}{\sqrt{q}} vech(Z^*) \sim N(0, I_p)$, where $q := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sigma_{ii}^2$. Then, we get $LR(k) \Rightarrow \frac{1}{2}\|\bar{Z}^*\|^2 \sim q\chi^2(df)$, i.e., a scaled chi-square distribution with df degrees of freedom. In the cross-sectionally homoschedastic case, we have $q = 1$ yielding the classical $\chi^2(df)$ result. On the contrary, cross-sectional heterogeneity in the unconditional idiosyncratic variances yields $q > 1$ and a deviation from classical FA theory even in the Gaussian case. Hence, unobserved heterogeneity across asset idiosyncratic variances would lead to an oversized LR test if we use critical values from the chi-square table without proper scaling.

4.2 Validity of the scaled asymptotic chi-square test

In this subsection, we investigate sufficient conditions for the validity of the scaled asymptotic $\chi^2(df)$ distribution of the LR statistic in special cases beyond Gaussianity of errors. For this purpose, let us first notice that $\bar{Z} = Z - \mathcal{T}_{F,V_\varepsilon}(Z)$ only involves the out-of-diagonal elements of Z. Under independent Gaussian errors (Section 4.1), by the Liapunov CLT, we have $Z_{t,s} \sim N(0, qV_{\varepsilon,tt}V_{\varepsilon,ss})$, for $t > s$, mutually independent, where the $V_{\varepsilon,tt}$ are the diagonal elements of matrix V_ε. We deduce that any setting featuring the same joint asymptotic distribution for the out-of-diagonal elements of random matrix Z_n leads to the same asymptotic distribution of the LR statistic as in the Gaussian case, namely the scaled $\chi^2(df)$ distribution.

Proposition 5 *Let Assumptions 1-4, A.1-A.4 hold with (a)* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,r}\varepsilon_{i,p}] = qV_{\varepsilon,tt}V_{\varepsilon,ss}$, *when* $t = r > s = p$, *for a constant* $q > 0$, *and* $= 0$ *in all other cases with* $t > s$ *and* $r > p$, *and (b) let* $\kappa = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{J_n}\sum_{i\neq j\in I_m} \sigma_{ij}^2$ *as in Assumption A.4 (b). Then,* $LR(k) \Rightarrow \bar{q}\chi^2(df)$ *under* $H_0(k)$ *for* $\bar{q} := q + \kappa$.

Conditions (a) and (b) in Proposition 5 generalize the correctness of the scaled chi-square test beyond Gaussianity and error independence across time and assets. Under Assumption 2, Condition (a) is satisfied if the standardized error terms $w_{i,t}$ are conditionally homoschedastic martingale difference sequences. However, Condition (a) excludes empirically relevant cases such as ARCH processes for $w_{i,t}$, because, in that case, $\frac{1}{V_{\varepsilon,tt}V_{\varepsilon,ss}}E[\varepsilon_{i,t}^2\varepsilon_{i,s}^2]$ depends on lag $t-s$. Hence, serial correlation in squared idiosyncratic errors is responsible for the deviation of the LR test from the scaled chi-square asymptotic distribution. This setting is covered by the general results in Proposition 4.

Anderson and Amemiya (1988) establish the asymptotic distribution of FA estimates assuming that the error terms are i.i.d. across sample units and deploy an assumption that is analogue to Condition (a) above in their Corollary 2. The i.i.d. assumption in our case implies $\sigma_{ii} = 1$ for all i, which results in a cross-sectionally homoschedastic setting. [15] That setting is irrealistic in our application, as it would imply that the idiosyncratic variance is the same for all assets. Our results show that establishing the asymptotic distribution of the test statistics, especially the AUMPI property of LR test (see Section 5), in a general setting with non-Gaussian errors, heterogeneous idiosyncratic variances and ARCH effects, is challenging, but still possible.

4.3 Spherical errors

When errors are spherical, i.e., matrix $V_\varepsilon = \bar{\sigma}^2 I_T$ is a multiple of the identity with unknown parameter $\bar{\sigma}^2 > 0$, and this restriction on V_ε is imposed in the estimation procedure, the FA estimator $\hat{F}$ boils down to the Principal Component Analysis (PCA) estimator; see Anderson and Rubin (1956) Section 7.3. Then, $\hat{F}$ is the matrix of eigenvectors of matrix $\hat{V}_y$ standardized such that $\hat{F}'\hat{F} = diag(\hat{\delta}_1 - \hat{\sigma}^2, ..., \hat{\delta}_k - \hat{\sigma}^2)$, and $\hat{\sigma}^2 = \frac{1}{T-k}\sum_{j=k+1}^T \hat{\delta}_j$, where $\hat{\delta}_j = \delta_j(\hat{V}_y)$. The statistic $\hat{S}$ becomes

[15] If the σ_{ii} were treated as i.i.d. random effects independent of errors, and we exclude cross-sectional correlation of errors to simplify, we would recover the i.i.d. condition of the data. However, the random σ_{ii} would yield a stochastic common factor across time that breaks the condition in Corollary 2 of Anderson and Amemiya (1988).

$\hat{S} = \frac{1}{\hat{\sigma}^2} M_{\hat{F}}(\hat{V}_y - \hat{\sigma}^2 I_T) M_{\hat{F}} = \frac{1}{\hat{\sigma}^2}\left(\frac{1}{n}\hat{\varepsilon}\hat{\varepsilon}'\right) - M_{\hat{F}}$, where $M_{\hat{F}} = I_T - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$ and $\hat{\varepsilon} = M_{\hat{F}}\tilde{Y}$ is the matrix of OLS residuals. Then, the statistic $LR(k)$ boils down to the LR statistic invoqued by Onatski (2022) in his discussion of FGS. By repeating the arguments of Sections 2 and 3 in the constrained setting of spherical errors, we get $Tr\left(M_F(\Psi_y - \Psi_\varepsilon)M_F\right) = 0$ instead of Equation (5). It yields the asymptotic expansion $\hat{V}_\varepsilon = \tilde{\sigma}^2 I_T + \frac{1}{\sqrt{n}}\Psi_\varepsilon + O_p(1/n)$, where $\tilde{\sigma}^2 = \bar{\sigma}^2 \frac{1}{n}\sum_{i=1}^n \sigma_{ii}$ and $\Psi_\varepsilon = \frac{1}{T-k} Tr(M_F Z_n) I_T$. We get the asymptotic distribution $\sqrt{n}\hat{S} \Rightarrow \frac{1}{\bar{\sigma}^2} G\bar{Z}^* G'$ as $n \to \infty$ and T is fixed, where we have $\bar{Z}^* = \frac{1}{\bar{\sigma}^4} G'\left(Z - \frac{1}{T-k} Tr(M_F Z) I_T\right) G = Z^* - \frac{1}{T-k} Tr(Z^*) I_{T-k}$, and G is a $T \times (T-k)$ matrix such that $F'G = 0$ and $G'G = \bar{\sigma}^2 I_{T-k}$. It yields the asymptotic distribution of test statistic $LR(k) \Rightarrow \frac{1}{2}\left(Tr[(Z^*)^2] - \frac{1}{T-k}[Tr(Z^*)]^2\right)$ obtained by Onatski (2022). [16]

5 Local asymptotic power

In this section, we study the asymptotic power of the test statistics against local alternative hypotheses in which we have k (strong) factors plus a weak factor. Specifically, under $H_{1,loc}(k)$, we have $\sqrt{n}\gamma_{k+1} \to c_{k+1}$ as $n \to \infty$, with $c_{k+1} > 0$. The (drifting) DGP is $Y = \mu 1_n' + F\beta' + F_{k+1}\beta_{loc}' + \varepsilon$, where β_{loc} is the loading vector for the $(k+1)$th factor, and the factor vector is normalized such that $F_{k+1} = \sqrt{\gamma_{k+1}}\rho_{k+1}$ with $\rho_{k+1}' V_\varepsilon^{-1} \rho_{k+1} = 1$ and $F' V_\varepsilon^{-1} \rho_{k+1} = 0$. Thus, we can write $\rho_{k+1} = G\xi_{k+1}$ for a $T-k$ dimensional vector ξ_{k+1} with unit norm. Scalar c_{k+1} and vector ξ_{k+1} yield the (normalized) strength and the direction of the local alternative.

5.1 Asymptotic distributions under local alternative hypotheses

The derivation of the asymptotic distribution of $\hat{S}$ under $H_{1,loc}(k)$ uses the asymptotic expansion $\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}\Psi_{y,loc} + o_p(\frac{1}{\sqrt{n}})$, where $\Psi_{y,loc} = c_{k+1}\rho_{k+1}\rho_{k+1}' + \frac{1}{\sqrt{n}}(\varepsilon\beta F' + F\beta'\varepsilon') + \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right)$

[16] We have $\hat{\gamma}_j = \hat{\delta}_j/\hat{\sigma}^2 - 1$, and $LR(k) = \frac{n}{2\hat{\sigma}^4}\sum_{j=k+1}^T (\hat{\delta}_j^2 - \hat{\sigma}^2)^2 + o_p(1)$, i.e., the LR statistic is asymptotically equivalent to the sum of squared deviations of the $T-k$ smallest eigenvalues from their mean. Besides, by similar results, we have that the eigenvalue spacing statistic $\sqrt{n}\mathscr{S}(k) := \hat{\gamma}_{k+1} - \hat{\gamma}_T$ corresponds to the statistic considered in FGS divided by $\hat{\sigma}^2$, and its asymptotic distribution coincides with that obtained by FGS.

and $\tilde{V}_y = FF' + \tilde{V}_\varepsilon$ (see the proof of Proposition 6 in Appendix B for the derivation). Thus, the arguments deployed in Section 3 now apply with $\Psi_{y,loc}$ instead of Ψ_y and lead to the next result.

Proposition 6 *Let Assumptions 1-4, A.1-A.4 hold. Under the local alternative hypothesis $H_{1,loc}(k)$, we have as $n \to \infty$ and T is fixed (a) $\sqrt{n}\hat{S} \Rightarrow V_\varepsilon^{-1/2} G \bar{Z}^*_{loc} G' V_\varepsilon^{-1/2}$, and (b) $\mathscr{T}(k) \Rightarrow \|\bar{Z}^*_{loc}\|^2$ and $LR(k) \Rightarrow \frac{1}{2}\|\bar{Z}^*_{loc}\|^2$, with Gaussian matrix $\bar{Z}^*_{loc} := \bar{Z}^* + \Delta$ and vector $vech(\Delta) := c_{k+1}\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right) vech(\xi_{k+1}\xi'_{k+1})$.*

Matrix variate $\bar{Z}^*_{loc}$ is a non-central symmetric Gaussian matrix. The non-zero mean depends in general on both c_{k+1} and ξ_{k+1}, while the variances and covariances of the elements of $\bar{Z}^*_{loc}$ are the same as those of $\bar{Z}^*$. The non-centrality term $vech(\Delta)$ is in charge of the asymptotic local power of the statistics. When this vector is null, the asymptotic local power is zero. Indeed, for some local "alternatives" the $(k+1)$th weak factor can be absorbed in the diagonal variance matrix V_ε of the error terms. More precisely, in Appendix D.3 ii), we show that $V_y + \frac{c_{k+1}}{\sqrt{n}}\rho_{k+1}\rho'_{k+1} = F^*(F^*)' + V^*_\varepsilon + \frac{1}{\sqrt{n}} G\Delta G' + o(1/\sqrt{n})$ for some $T \times k$ matrix F^* and diagonal matrix V^*_ε, which yields asymptotically a k-factor model when $\Delta = 0$.

Using the expression $\frac{1}{2}\|\bar{Z}^*_{loc}\|^2 = [vech(Z^*) + c_{k+1} vech(\xi_{k+1}\xi'_{k+1})]'\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)$ $[vech(Z^*) + c_{k+1} vech(\xi_{k+1}\xi'_{k+1})]$ and $vech(Z^*) \sim N(0, \Omega_{Z^*})$, from Proposition 6 we deduce that the asymptotic distribution of the $LR(k)$ statistic under the local alternative is a weighted average of df mutually independent non-central chi-square distributions:

$$LR(k) \Rightarrow \sum_{j=1}^{df} \mu_j \chi^2(1, \lambda_j^2), \tag{11}$$

for $\lambda_j^2 = c_{k+1}^2 \mu_j^{-1}\left[v'_j\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right) vech(\xi_{k+1}\xi'_{k+1})\right]^2$, where the μ_j and v_j are the non-zero eigenvalues and the associated standardized eigenvectors of matrix $\Omega_{\bar{Z}^*}$. We have $\lambda^2 := \sum_{j=1}^{df} \mu_j \lambda_j^2 = c_{k+1}^2 vech(\xi_{k+1}\xi'_{k+1})'\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right) vech(\xi_{k+1}\xi'_{k+1}) = \frac{1}{2}\|\Delta\|^2$, i.e., the half squared Frobenius norm of the matrix measuring local distance from the k-factor specification. It follows that the asymptotic local power of the LR statistic is non null as long as $\lambda^2 > 0$, i.e., it has non-trivial asymptotic power against any proper local alternative hypothesis. In our Monte Carlo experiments reported in the SMC, we find that the LR statistic has size close to the nominal value, and power against global as well as local alternatives with time dimension as small as $T = 6$.

Under the normality of errors, or more generally the conditions of Proposition 5, using that matrix $\frac{1}{\sqrt{\bar{q}}} Z^*$ is in the GOE for dimension $T - k$, i.e. $vech(Z^*) \sim N(0, \bar{q} I_p)$, we have $LR(k) \Rightarrow \bar{q}\chi^2(df, \lambda^2/\bar{q})$ from (11). The local power is a function solely of the squared Euclidean norm of the vector $vech(\Delta)$ measuring local distance from the k-factor specification, divided by $\bar{q}$.

5.2 AUMPI tests

In this subsection, we investigate asymptotic local optimality of the LR statistic for testing hypotheses on the number of latent factors. In our framework with composite null and alternative hypotheses and multi-dimensional parameter, we cannot expect in general to establish Uniformly Most Powerful (UMP) tests. Instead, we can establish an optimality property by restricting the class of tests to invariant tests (e.g. Lehmann and Romano (2005)). We focus on statistics with test functions ϕ written on the elements of matrix $\hat{S}$. To eliminate the asymptotic redundancy in the elements of $\hat{S}$, we actually consider the test class $\mathscr{C} = \left\{ \phi \, : \, \phi = \phi(\hat{W}) \right\}$ with $\hat{W} := \sqrt{n} \boldsymbol{D}' vech(\hat{S}^*)$, where $\hat{S}^* = \hat{G}' \hat{V}_\varepsilon^{-1/2} \hat{S} \hat{V}_\varepsilon^{-1/2} \hat{G}$ and $\boldsymbol{D}$ is a $p \times df$ full-rank matrix, such that $I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{D}\boldsymbol{D}'$ and $\boldsymbol{D}'\boldsymbol{D} = I_{df}$. Symmetric matrix $\hat{S}^*$ contains the information in $\hat{S} = \hat{V}_\varepsilon^{-1/2} \hat{G} \hat{S}^* \hat{G}' \hat{V}_\varepsilon^{-1/2}$ beyond orthogonality to $\hat{V}_\varepsilon^{-1/2} \hat{F}$. Vector $\hat{W}$ contains the information in $\sqrt{n} vech(\hat{S}^*)$ beyond asymptotic orthogonality to $\boldsymbol{X}$. From Proposition 6 we have $\hat{W} \Rightarrow N(0, \boldsymbol{D}'\Omega_{Z^*}\boldsymbol{D})$ under the null hypothesis $H_0(k)$ and $\hat{W} \Rightarrow N(c_{k+1} \boldsymbol{D}' vech(\xi_{k+1}\xi_{k+1}'), \boldsymbol{D}'\Omega_{Z^*}\boldsymbol{D})$ under the local alternative $H_{1,loc}(k)$, i.e., an asymptotic Gaussian testing problem.

Matrices G and $\boldsymbol{D}$ are both defined up to post-multiplication by an orthogonal matrix. This point yields a group of orthogonal transformations under which we require the test statistics to be invariant. [17] In Appendix D.5, we show that the maximal invariant under this group is provided

[17]Here, we do not deal with invariance to data transformations but rather with invariance to parameterization of G

by $\hat{W}'\hat{W} = n \cdot vech(\hat{S}^*)'(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')vech(\hat{S}^*)$. Because $\sqrt{n} \cdot vech(\hat{S}^*)$ belongs to the range of matrix $I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ up to $o_p(1)$ terms under both $H_0(k)$ and $H_{1,loc}(k)$, we have $\hat{W}'\hat{W} = \frac{n}{2}\|\hat{S}^*\|^2 + o_p(1) = \frac{n}{2}\|\hat{S}\|^2 + o_p(1) = \frac{1}{2}\mathscr{T}(k) + o_p(1)$. Therefore, the invariant tests are functions of the squared norm statistic $\mathscr{T}(k)$, which is asymptotically equivalent to the LR statistic (up to the factor $1/2$).

In the Gaussian case, or more generally under the conditions of Proposition 5, the LR statistic follows asymptotically a scaled non-central chi-square distribution with df degrees of freedom and non-centrality parameter $\lambda^2 = \sum_{j=1}^{df} \lambda_j^2$ as shown in the previous subsection. Thus, we can simplify the null and alternative hypotheses of our testing problem asymptotically and locally to a one-sided test with null hypothesis $H_0(k) : \lambda^2 = 0$ vs. alternative hypothesis $H_{1,loc}(k) : \lambda^2 > 0$. The scaling constant $q > 0$ plays no role in the power analysis. It means that the LR test is an AUMPI test (Lehmann and Romano (2005) Chapters 3 and 13). Indeed, the density $g(z; df, \lambda^2)$ of the $\chi^2(df, \lambda^2)$ distribution is Totally Positive of order 2 ($TP2$) in z and λ^2 (Eaton (1987) Example A.1 p. 468); see Miravete (2011) for a review of applications of TP2 in economics. A density, which is $TP2$ in z and λ^2, has the Monotone Likelihood Ratio (MLR) property (Eaton (1987) p. 467). Since $g(z; df, \lambda^2)/g(z; df, 0)$ is an increasing function in z, it gives the AUMPI property.

In the general case with $df > 1$, when neither Gaussianity nor the conditions of Proposition 5 apply, we cannot use the same reasoning, since the density $f(z; \lambda_1, ..., \lambda_{df})$ of $\sum_{j=1}^{df} \mu_j \chi^2(1, \lambda_j^2)$, with $\mu_j > 0$, $j = 1, ..., df$, is not a function of $\lambda^2 = \sum_{j=1}^{df} \mu_j \lambda_j^2$ only, and thus cannot be $TP2$ in z and λ^2. Instead, we use a power series representation of the density of $\sum_{j=1}^{df} \mu_j \chi^2(1, \lambda_j^2)$ in terms of central chi-square densities from Kotz, Johnson, and Boyd (1967). Under the sufficient condition (12) in Proposition 7, the density ratio $\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)}$ is monotone increasing in z.

Proposition 7 *Let Assumptions 1-4, A.1-A.4 hold. (a) Let us assume that, for any DGP in the*

and $\boldsymbol{D}$. However, if we consider tests based on the elements of vector $\hat{W}$, this difference is immaterial.

subset $\bar{H}_{1,loc}(k) \subset H_{1,loc}(k)$ of the local alternative hypothesis, we have for any integer $m \geq 3$:

$$\sum_{j>l\geq 0, j+l=m} \frac{(j-l)\Gamma(\frac{df}{2})^2}{\Gamma(\frac{df}{2}+j)\Gamma(\frac{df}{2}+l)}[c_j(\lambda_1,...,\lambda_{df})c_l(0,...,0) - c_l(\lambda_1,...,\lambda_{df})c_j(0,...,0)] \geq 0, \qquad (12)$$

where $\Gamma(\cdot)$ is the Gamma function, $c_j(\lambda_1,...,\lambda_{df}) := E[Q(\lambda_1,...,\lambda_{df})^j]/j!$ for $Q(\lambda_1,...,\lambda_{df}) = \frac{1}{2}\sum_{j=1}^{df}(\sqrt{\nu_j}X_j + \sqrt{1-\nu_j}\lambda_j)^2$, $\nu_j = 1 - \frac{1}{\mu_j}\mu_1$ with the μ_j ranked in increasing order, and $X_j \sim N(0,1)$ are mutually independent. Then, the statistics $\mathscr{T}(k)$ and $LR(k)$ yield AUMPI tests against $\bar{H}_{1,loc}(k)$. (b) Suppose that either $\lambda_1^2 + (1-\nu_2)\lambda_2^2 \geq \nu_2$ and $(1-\nu_2)\lambda_2^2 \geq \frac{1}{2}\nu_2$ when $df = 2$, or

$$1\{i=0\}\lambda_1^2 + \sum_{j=2}^{df-1}\rho_j^i(1-\nu_j)\lambda_j^2 + (1-\nu_{df})\lambda_{df}^2 \geq \frac{\nu_{df}}{i+1}\left(df - 2 - \sum_{j=2}^{df-1}\rho_j^{i+1}\right), \qquad (13)$$

for all $i \geq 0$, where $\rho_j := \frac{\nu_j}{\nu_{df}}$, when $df \geq 3$. Then, Inequalities (12) hold for any $m \geq 3$.

Conditions (12) involve polynomial inequalities in the parameters λ_j of the alternative hypothesis, and parameters ν_j of the weights of the non-central chi-square distributions, $j = 1,...,df$. It is challenging to establish an explicit characterization of the λ_j and ν_j equivalent to Inequalities (12), unless $df = 1$. [18] By deploying a novel characterization of the $c_j(\lambda_1,...,\lambda_{df})$ in terms of a recurrence relation (Lemma 3), we establish explicit sufficient conditions in part b) of Proposition 7. Inequalities (13) are linear in the λ_j^2, and define a non-empty convex domain in the $(\lambda_1^2,..,\lambda_{df}^2)$ space, that does not contain the origin $\lambda_1 = ... = \lambda_{df} = 0$ (unless the DGP is such that $\nu_2 = ... = \nu_{df}$, in which case the RHS of (13) is nil for all i and thus any λ_j^2 meet the inequalities). Proposition 7 b) implies that, for a given set of values of df, the MLR property holds if $\lambda_j \geq \underline{\lambda}$ for all j, uniformly for $\nu_j \leq \bar{\nu}$, where $\underline{\lambda} > 0$ is a constant that depends on $\bar{\nu} < 1$. Vanishing values of the ν_j correspond to homogenous weights μ_j, i.e., the scaled non-central chi-square distribution with df degrees of freedom. Hence, the AUMPI property in Proposition 7 holds in neighborhoods of DGPs that match the conditions of Proposition 5 (e.g. Gaussian errors) for alternative hypotheses that are sufficiently separated from the null hypothesis. Besides, Proposition 7 shows that the

[18] Inequalities (12) with $df = 1$ are easily proved to hold. In such a case, we can use the asymptotic distribution of a scaled chi-square variable and its MLR property.

Gaussian case is not the only design delivering an AUMPI test. Further, in the SMC, we establish an analytical representation of the coefficients $c_k(\lambda_1, ..., \lambda_{df})$ in terms of matrix product iterations. That analytical representation allows us to check numerically the validity of Inequalities (12) for given df, λ_j, ν_j, and $m = 1, ..., M$, for a large bound M (see Appendix E). In Appendix E, when Inequalities (13) are met, we always conclude to the MLR property in the numerical checks as predicted by the theory of Proposition 7. There, we also provide numerical evidence that the domain of validity of the MLR property is relevant for our empirical application. The sufficient conditions (12) and (13) in Proposition 7 yielding the monotone property of density ratios have potentially broad application outside the current setting to show AUMPI properties of other tests based on an asymptotic distribution characterized by a positive definite quadratic form in normal vectors.

6 Empirical application

In this section, we test hypotheses about the number of latent factors driving stock returns in short subperiods of the Center for Research in Securities Prices (CRSP) panel. Then, we decompose the cross-sectional variance into systematic and idiosyncratic components. We also check whether there is spanning between the estimated latent factors and standard observed factors.

6.1 Testing for the number of latent factors

We consider monthly returns of U.S. common stocks trading on the NYSE, AMEX or NASDAQ between January 1963 and December 2021, and having a non-missing Standard Industrial Classification (SIC) code. We partition subperiods into bull and bear market phases according to the classification methodology of Lunde and Timmermann (2004). [19] We implement the tests using a rolling window of $T = 20$ months, moving forward 12 months each time (adjacent windows over-

[19] We fix their parameter values $\lambda_1 = \lambda_2 = 0.2$ for the classification based on the nominal S&P500 index. Bear periods are close to NBER recessions.

lap by 8 months), thereby ensuring that we can test up to 14 latent factors in each subperiod. The size of the cross-section n ranges from 1768 to 6142, and the median is 3680. We only consider stocks with available returns over the whole subperiod, so that our panels are balanced. In each subperiod, we sequentially test $H_0(k)$ v.s. $H_1(k)$, for $k = 0, \ldots, k_{max}$, where $k_{max} = 14$ is the largest nonnegative integer such that $df > 0$ (see Table 3 in the SMC). We compute the variance-covariance estimator $\hat{\Omega}_{\bar{Z}^*}$ using a block structure implied by the partitioning of stocks by the first two digits of their SIC code. The number of blocks ranges from 61 to 87 over the sample, and the number of stocks per block ranges from 1 to 641. The median number of blocks is 76 and the median number of stocks per block is 21. We display the p-values of the statistic $LR(k)$ for each subperiod in the upper panel of Figure 1, stopping at the smallest k such that $H_0(k)$ is not rejected at level $\alpha_n = 10/n_{max}$, where n_{max} is the largest cross-sectional sample size over all subperiods, so that $\alpha_n = 0.16\%$ in our data. If no such k is found then p-values are displayed up to k_{max}. The n-dependent size adjustment controls for the over-rejection problem induced by sequential testing (see Section 6.2 below). Overall, the results point to a higher number of latent factors during bear market phases compared to bull market phases and a decrease of the number of factors over time. [20] It remains true for the three-month recession periods 1987/09-1987/11 and 2020/01-2020/03, which represent only a fraction of their respective subperiods, although there are "bull" market periods finding a similar number of latent factors. In particular, our results based on a fixed T and large n approach contradict the common wisdom of a single factor model during market downturns due to estimated correlations between equities approaching 1. It is consistent with the presence of risk factors, such as tail risk or liquidity risk, only showing in stress periods. A rise in the estimated k often happens towards the end of the recession periods. It is consistent with the methodology of Lunde and Timmermann (2004) being early in detecting bear periods (early warning system). The

[20] We also investigate stability of the factor structure by dividing each window of 20 months into two overlapping subperiods of 16 months (overlap of 12 months) and by estimating canonical correlations between the betas in each subperiod (see SMC). We find that the fraction of common factors is 1 in 70% of the windows. The fraction is between 0.8 and 1 in 25%. It is between 0.5 and 0.8 in the remaining periods.

results with statistic $\mathscr{T}(k)$ are similar and not reported. The average estimated number of factors is around 7, close to the 4 to 6 factors found by PCA in Bai and Ng (2006) on large time spans of individual stocks. [21]

6.2 Decomposing the cross-sectional variance

Building on the results in Pötscher (1983), we can obtain a consistent estimator of the number of latent factors in each subperiod by allowing the asymptotic size α go to zero as $n \to \infty$ in the sequential testing procedure. We let $\hat{k}$ be defined as the smallest nonnegative integer k satisfying $\text{pval}(k) > \alpha_n$, where $\text{pval}(k)$ is the p-value from testing $H_0(k)$, and α_n is a sequence in $[0,1]$ with $\alpha_n \to 0$. In practice, we take $\alpha_n = 10/n_{max}$. [22] If no such k is found after sequentially testing $H_0(k)$, for $k = 0, \dots, k_{max}$ at level α_n, then we take $\hat{k} = k_{max} + 1$. We use the estimate $\hat{k}$ at each subperiod to decompose the path of the cross-sectional variance of stock returns into its systematic and idiosyncratic parts: $\hat{V}_{y,tt} = \hat{F}_t'\hat{F}_t + \hat{V}_{\varepsilon,tt}$, where $\hat{F}$ and $\hat{V}_\varepsilon$ are the FA estimates obtained by extracting $\hat{k}$ latent factors. The condition (FA1) ensures that the decomposition holds for any t. Such a decomposition is invariant to the choice of normalization for the latent factors. If we look at time averages on a subperiod, we get the decomposition $\overline{\hat{V}}_y = \overline{\hat{F}'\hat{F}} + \overline{\hat{V}}_\varepsilon$, where the overline indicates averaging $\hat{V}_{y,tt} = \hat{F}_t'\hat{F}_t + \hat{V}_{\varepsilon,tt}$ on t. In the lower panels of Figure 1, the blue dots correspond to the square root of those quantities for the volatilities, while the ratios $\hat{R}^2 = \overline{\hat{F}'\hat{F}}/\overline{\hat{V}}_y$ and $\hat{R}^2$ under a single-factor model in the two last panels give measures of goodness-of-fit. [23] We

[21] With fixed T, the selection procedure of Zaffaroni (2019), being by construction more conservative than a (multiple) testing procedure (see the discussion on p. 508 of Gagliardini, Scaillet, and Ossola (2019)), yields a smaller number of factors. Imposing cross-sectional independence (resp., Gaussianity and cross-sectional independence) for the LR test gives most of the time an increase by 1 or 2 (1 or 3). We have an average increase of 3 under sphericity.

[22] This choice satisfies the theoretical rule $\log \alpha_n / n \to 0$ given in Pötscher (1983).

[23] We do not plot the whole paths date t by date t, but only averages, for readability. If we sum over time instead of averaging the estimated variances, we get a quantity similar to an integrated volatility (see e.g. Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev, Diebold, and Labys (2003), and references in Aït-Sahalia and Jacod (2014)), and $\hat{R}^2$ is the ratio of such quantities.

can observe an uptrend in total and idiosyncratic volatilities, while the systematic volatility appears to remain stable over time even if the number of factors has overall decreased over time. [24] As a result, $\hat{R}^2$ is lower on average after the year 2000, indicating a more noisy environment. During the 2007-2008 financial crisis, we can observe a rise in systematic volatility, causing $\hat{R}^2$ to reach 59% during that period. In bear markets, $\hat{R}^2$ is often higher. It means that over a bear subperiod, the systematic risk explains a large part of the cross-sectional total variance even if it is not driven by a single factor as reported in Section 6.1. The lowest panel in Figure 1 also signals that $\hat{R}^2$ under the constraint of a single-factor model can be way below the one given by the multifactor model. It also means that the idiosyncratic volatility is overestimated if we use a single latent factor only. The plots of the equal-weighted market and firm volatilities used as measures of total and idiosyncratic volatility from a CAPM decomposition in Campbell et al. (2023) show similar patterns as our panels in Figure 1. [25] Section 4 of Campbell et al. (2023) discusses economic forces (firm fundamentals and investor sentiments) driving the observed time-series variation in average idiosyncratic volatility.

6.3 Spanning with observed factors

As discussed in Bai and Ng (2006), we get economic interpretation of latent factors with observed factors when we have spanning between the latent factors and the observed factors to be used as proxies in asset pricing (Shanken (1992)). When n and T are large, Bai and Ng (2006) exploit the asymptotic normality of the empirical canonical correlations between the two sets of factors to investigate spanning under a symmetric role of the two sets. When T is fixed, we suggest the following strategy based on testing for the rank of a matrix. Let us consider $k^O \geq k$ empirical

[24] Cross-sectional independence (resp. cross-sectional independence and Gaussianity / sphericity) increases estimated systematic risk in average by 0.6% (resp. 0.6%/ 1%) and decreases estimated idiosyncratic volatility in average by 0.5% (resp. 0.5% / 1.3%), so that estimated R^2 is inflated in average by 4% (resp. 4% / 12.8%).

[25] As in Campbell et al. (2023), we have also made the estimation on value-weighted returns and we confirm that the results are qualitatively similar.

factors that are excess returns of portfolios [26], and let $\hat{F}^O$ denote the $T \times k^O$ matrix of their values with row t given by the transpose of $\hat{f}_t^O = \frac{1}{n}\sum_{i=1}^n (y_{i,t} - r_{f,t}) z_{i,t}$, where $\frac{1}{n} z_{i,t}$ is a $k^O \times 1$ vector of time-varying portfolio weights (long or short positions) based on stocks characteristics. Let matrix F^O with rows $f_t^O = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[(y_{i,t} - r_{f,t}) z_{i,t}]$ be the corresponding large-n population limit. The notation $\hat{F}^O$ makes clear that the sample average of weighted excess returns is an estimate of the population values F^O. We need to take this into account in the asymptotic analysis of the rank test statistics when $n \to \infty$. From the factor model under NA, $y_{i,t} = r_{f,t} + f_t' \tilde{\beta}_i + \varepsilon_{i,t}$ (see footnote 2), and assuming cross-sectional non-correlation of idiosyncratic errors and portfolio weights, we get $F^O = F\Phi'$, where $\Phi = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[z_{i,t}]\tilde{\beta}_i'$ is assumed independent of t, $t = 1, ..., T$. Hence, the range of F^O is a subset of the range of F, namely the latent factors span the observed factors (in the population limit sense) by construction. Moreover, Rank$(F^O) \leq k$. We can test the null hypothesis that F and F^O span the same linear spaces, namely matrices F and F^0 have the same range. Such a null hypothesis is equivalent to the rank condition: Rank$(F^O) = k$.

We build on the rank testing literature; see e.g. Cragg and Donald (1996), Robin and Smith (RS, 2000), Kleibergen and Paap (KP, 2006), Al-Sadoon (2017). [27] We use in particular the RS and KP statistics. For those tests, the null hypothesis is that a given matrix has a reduced rank r against the alternative that the rank is greater than r. Hence, to test for spanning by the empirical factors, we consider the null hypothesis $H_{0,sp}(r)$: Rank$(F^O) = r$ against the alternative $H_{1,sp}(r)$: Rank$(F^O) > r$, for any integer $r < k$. [28] We use the asymptotic expansion $\hat{F}^O = \tilde{F}^O + \frac{1}{\sqrt{n}}\Psi_{F^O,n}$, where $\tilde{F}^O = F\Phi_n'$ with $\Phi_n = \frac{1}{n}\sum_{i=1}^n E[z_{i,t}]\tilde{\beta}_i'$, and the rows of matrix $\Psi_{F^0,n}$ are given by $\Psi_{F,n,t} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\eta_{i,t} f_t + \varepsilon_{i,t} z_{i,t})$ with $\eta_{i,t} := (z_{i,t} - E[z_{i,t}])\tilde{\beta}_i'$. Under $H_{0,sp}(r)$, we assume that Φ_n has

[26] If $k^O < k$, empirical factors cannot span the latent space by construction. The condition $k^O \geq k$ eases discussion but is not needed for the rank tests.

[27] Ahn, Horenstein and Wang (2018) use that technology in a fixed-n large-T setting, and find that ranks of beta matrices estimated from either portfolios, or individual stocks, excess returns are often substantially smaller than the (potentially large) number k^O of observed factors. The explanation in large economies is that the portfolio beta matrices coincide with Φ, and thus they cannot have a rank above the (potentially small) number k of latent factors.

[28] Spanning holds if we can reject $H_{0,sp}(r)$ for any $r < k$.

the same null space as Φ, in particular Φ_n has rank r, for n large enough. [29] We assume the CLT $vec(\Psi_{F^0,n}) \Rightarrow N(0, \Omega_\Psi)$. Further, we use the Singular Value Decomposition (SVD) of matrix $\hat{F}^O = \hat{U}\hat{\mathcal{S}}\hat{V}'$. Then, the RS and KP statistics are the quadratic forms $\mathscr{S}_{RS} = nvec(\hat{\mathcal{S}}_{22})'vec(\hat{\mathcal{S}}_{22})$ and $\mathscr{S}_{KP} = nvec(\hat{\mathcal{S}}_{22})'\hat{\Omega}_{\mathcal{S}}^{-1}vec(\hat{\mathcal{S}}_{22})$, where $\hat{\mathcal{S}}_{22}$ is the lower-right $(T-r)\times(k^O-r)$ block of matrix $\hat{\mathcal{S}}$. Here, $\hat{\Omega}_{\mathcal{S}} = (\hat{V}_{k^O-r}\otimes\hat{U}_{T-r})'\hat{\Omega}_\Psi(\hat{V}_{k^O-r}\otimes\hat{U}_{T-r})$, where $\hat{U}_{T-r}$ and $\hat{V}_{k^O-r}$ are the $T\times(T-r)$ and $k^O\times(k^O-r)$ matrices in the block forms $\hat{U} = [\hat{U}_r : \hat{U}_{T-r}]$ and $\hat{V} = [\hat{V}_r : \hat{V}_{k^O-r}]$. In the SMC, we design a consistent estimator $\hat{\Omega}_\Psi$ of Ω_Ψ building on a block structure for the characteristics akin to Assumption 2 and a stationarity condition. The definitions of the test statistics $\mathscr{S}_{RS}$ and $\mathscr{S}_{KP}$ are equivalent to those in the original RS and KP papers. The asymptotic distributions under $H_{0,sp}(r)$ are $\mathscr{S}_{RS} \Rightarrow \sum_{j=1}^{(T-r)(k^O-r)}\delta_j(\Omega_{\mathcal{S}})\chi_j^2(1)$ and $\mathscr{S}_{KP} \Rightarrow \chi^2[(T-r)(k^O-r)]$, where $\Omega_{\mathcal{S}} = (V_{k^O-r}\otimes U_{T-r})'\Omega_\Psi(V_{k^O-r}\otimes U_{T-r})$ is assumed non-singular.

We build the empirical matrix $\hat{F}^O$ with the time-varying portfolio weights of the Fama-French five-factor model (Fama and French (2015)) plus the momentum factor (Carhart (1997)), i.e., $k^O = 6$. In the two panels of Figure 2, we can observe that the rank tests point most of the time at a low reduced rank r either 1 or 2, with only occasionally 3 or 4, for the matrix $\hat{F}^O$. Observed factors struggle spanning latent factors since their associated linear space is of a dimension smaller than the one of the latent factor space. The discrepancy between the dimensions of the two factor spaces has decreased over time. According to the KP statistic, the rank deficiency of $\hat{F}^O$ is often less pronounced in bear markets indicating less redundancy between the observed factors.

7 Concluding remarks

In this paper, we develop a new theory of Factor Analysis in short panels beyond the Gaussian and i.i.d. cases. We establish the AUMPI property of the LR statistic for testing hypotheses on the number of latent factors. Our results for short subperiods of the CRSP panel of US stock returns contradict the common wisdom of a single factor during market downturns. In bear markets,

[29] Under $\text{Rank}(F) = k$, we have $\text{Rank}(F^O) = \text{Rank}(\Phi)$. Hence, under $H_{0,sp}(r)$, matrix Φ has reduced rank r.

systematic risk explains a large part of the cross-sectional variance, and is not spanned by traditional empirical factors. Our new methodology can be used to address relevant empirical questions in applications beyond asset pricing. For example, in analysis of education when a panel consists of students repeatedly tested along different cognitive domains in mathematics and science (Freyberger (2018)) or interviewed in successive waves (Sarzosa and Urzúa (2021)), in analysis of particular time spans in long panel data of wages (Gobillon, Magnac, and Roux (2022)) or in analysis of unemployment for a panel of counties followed on a couple of years (Hagedorn, Manovskii, and Mitman (2015)).

References

Ahn, S., and Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81 (3), 1203-1227.

Ahn, S., Horenstein, A., and Wang, N., 2018. Beta matrix and common factors in stock returns. Journal of Financial and Quantitative Analysis 53 (3), 1417-1440.

Aigner, D., Hsiao, C., Kapteyn A., and Wansbeek, T., 1984. Latent variable models in econometrics, in Handbook of Econometrics, Volume II, Z. Griliches and M.D. Intriligator Eds., 1321-1393.

Aït-Sahalia, Y., and Jacod, J., 2014. High-frequency financial econometrics. Princeton University Press.

Al-Sadoon, M., 2017. A unifying theory of tests of rank. Journal of Econometrics 199 (1), 49-62.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71(2), 579-625.

Anderson, T. W., 2003. An introduction to multivariate statistical analysis. Wiley.

Anderson, T. W., and Rubin, H., 1956. Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability 5, 11-150.

Anderson, T. W. and Amemiya, Y., 1988. The asymptotic normal distribution of estimators in factor analysis under general conditions. Annals of Statistics 16 (2), 759-771.

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

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

Bai, J., 2009. Panel data models with interactive effects. Econometrica, 77 (4), 1229-1279.

Bai, J., and Li, K., 2012. Statistical analysis of factor models of high dimension. Annals of Statistics 40 (1), 436-465.

Bai, J., and Li, K., 2016. Maximum likelihood estimation and inference for approximate factor models of high dimension. Review of Economics and Statistics 98 (2), 298-309.

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

Bai, J., and Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131 (1), 507-537.

Barigozzi, M., and Hallin, M., 2016. Generalized dynamic factor models and volatilities: recovering the market volatility shocks. Econometrics Journal 19 (1), 33-60.

Barndorff-Nielsen, O., and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B, 64 (2), 253-280.

Bell, E. T., 1934. Exponential polynomials. Annals of Mathematics 35 (2), 258-277.

Bernstein, D., 2009. Matrix mathematics: Theory, facts and formulas. Princeton University Press.

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

Campbell, J., Lettau, M., Malkiel, B., and Xu, Y., 2023. Idiosyncratic equity risk two decades later. Critical Finance Review, 12, 203-223.

Caner, M., and Han, X., 2014. Selecting the correct number of factors in approximate factor models: the large panel case with group bridge estimator. Journal of Business and Economic Statistics 32 (3), 359-374.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance. 52 (1), 57-82.

Chamberlain, G., 1992. Efficiency bounds for semi-parametric regression. Econometrica 60 (3), 567-596.

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

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

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

Cragg, J., and Donald, S., 1996. On the asymptotic properties of LDU-based tests of the rank of a matrix. Journal of the American Statistical Association 91 (435), 1301-1309.

Eaton, M., 1987. Multivariate statistics. A vector space approach. Institute of Mathematical Statistics Lecture notes-Monograph series, Vol. 53.

Engle, R., 1984. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in Handbook of Econometrics, Volume II, Z. Griliches and M.D. Intriligator Eds., 775-826.

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

Fortin, A.-P., Gagliardini, P., and Scaillet, O., 2022. Eigenvalue tests for the number of latent factors in short panels. Journal of Financial Econometrics, forthcoming.

Freyberger, J., 2018. Non-parametric panel data models with interactive fixed effects. Review of Economic Studies 85 (3), 1824-1851.

Gagliardini, P., Ossola, E., and Scaillet, O., 2016. Time-varying risk premium in large cross-sectional equity datasets. Econometrica 84 (3), 985-1046.

Gagliardini, P., Ossola, E., and Scaillet, O., 2019. A diagnostic criterion for approximate factor structure. Journal of Econometrics 21 (2), 503-521.

Gagliardini, P., Ossola, E., and Scaillet, O., 2020. Estimation of large dimensional conditional factor models in finance, in Handbook of Econometrics, Volume 7A, S. Durlauf, L. Hansen, J. Heckman, and R. Matzkin Eds., 219-282.

Gobillon, L., Magnac, T., and Roux, S., 2022. Lifecycle wages and human capital investments: Selection and missing data. TSE working paper.

Hagedorn, M., Manovskii, I., and Mitman, K., 2015. The impact of unemployment benefit extensions on employment: The 2014 employment miracle. NBER working paper.

Kapetanios, G., 2010. 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.

Kim, S., and Skoulakis, G., 2018. Ex-post risk premia estimation and asset pricing tests using large cross-sections: the regression-calibration approach. Journal of Econometrics 204 (2), 159-188.

Kleibergen, F., and Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics 133 (1), 97-126.

Kleibergen, F., and Zhan, Z., 2023. Identification-robust inference for risk premia in short panels. Working paper, University of Amsterdam.

Kotz, S., Johnson, N., and Boyd, D., 1967. Series representations of distributions of quadratic forms in Normal variables II. Non-central case. Annals of Mathematical Statistics 38 (3), 838-848.

Lehmann, E., and Romano, D., 2005. Testing statistical hypotheses, Springer Texts in Statistics.

Lunde, A., and Timmermann, A., 2004. Duration dependence in stock prices: An analysis of bull and bear markets. Journal of Business and Economic Statistics 22 (3), 253-273.

Magnus, J., and Neudecker, H., 2007. Matrix differential calculus, with applications in statistics and econometrics. Wiley.

Miravete, E., 2011. Convolution and composition of totally positive random variables in economics. Journal of Mathematical Economics 47 (4), 479-490.

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

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

Onatski, A. 2022. Comment on “Eigenvalue tests for the number of latent factors in short panels" by A.-P. Fortin, P. Gagliardini and O. Scaillet, Journal of Financial Econometrics, forthcoming.

Pötscher, B., 1983. Order estimation in ARMA-models by Lagrangian multiplier tests, Annals of Statistics 11 (3), 872-885.

Raponi, V., Robotti, C., and Zaffaroni, P., 2020. Testing beta-pricing models using large cross-sections. Review of Financial Studies 33 (6), 2796-2842.

Renault, E., Van Der Heijden, T., and Werker, B., 2022. Arbitrage pricing theory for idiosyncratic variance factors. Journal of Financial Econometrics, forthcoming.

Robin, J.-M., and Smith, R., 2000. Tests of rank. Econometric Theory 16 (2), 151-175.

Sarzosa, M., and Urzúa, S., 2021. Bullying among adolescents: The role of skills. Quantitative Economics 12 (3), 945-980.

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

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

Tao, T. 2012. Topics in random matrix theory. Graduate Studies in Mathematics, Volume 132, American Mathematical Society.

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

Zaffaroni, P., 2019. Factor models for conditional asset pricing. Working Paper.

Appendix

A Regularity assumptions

In this appendix, we list and comment the additional assumptions used to derive the large sample properties of the estimators and test statistics. We often denote by $C > 0$ a generic constant. Set Θ is a compact subset of $\{\theta = (vec(F)', diag(V_\varepsilon)')' \in \mathbb{R}^r \; : \; V_\varepsilon$ is diagonal and positive definite, $F'V_\varepsilon^{-1}F$ is diagonal, with diagonal elements ranked in decreasing order$\}$ with $r = (T +$

$1)k$, and function $L_0(\theta) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(V_y\Sigma(\theta)^{-1}\right)$ is the population FA criterium, where $\Sigma(\theta) = FF' + V_\varepsilon$ and $V_y = \operatorname*{plim}_{n\to\infty} \hat{V}_y$. Further, $\theta_0 = (vec(F_0)', diag(V_\varepsilon^0)')'$ denotes the vector of true parameter values under $H_0(k)$ and is an interior point of set Θ.

Assumption A.1 *The loadings are normalized such that $\bar{\beta} = \frac{1}{n}\sum_{i=1}^n \beta_i = 0$ and $\tilde{V}_\beta := \frac{1}{n}\sum_{i=1}^n \beta_i\beta_i'$ $= I_k$, for any n. Moreover, $|\beta_i| \leq C$, for all i.*

Assumption A.2 *We have $E[w_{i,t}^8] \leq C$ and $|\sigma_{i,j}| \leq C$, for all i, j, t.*

Assumption A.3 *Under the null hypothesis $H_0(k)$, we have: $\Sigma(\theta) = \Sigma(\theta_0)$, $\theta \in \Theta \Rightarrow \theta = \theta_0$, up to sign changes in the columns of F.*

Assumption A.4 *(a) The $\frac{T(T+1)}{2} \times \frac{T(T+1)}{2}$ symmetric matrix $D = \lim_{n\to\infty} D_n$ exists, where $D_n = \frac{1}{n}\sum_{i=1}^n \sigma_{ii}^2 V[vech(w_i w_i')]$. (b) We have $\delta_{T(T+1)/2}\left(V[vech(w_i w_i')]\right) \geq \underline{c}$, for all $i \in \bar{S}$, where $\bar{S} \subset \{1, ..., n\}$ with $\frac{1}{n}\sum_{i=1}^n 1_{i\in\bar{S}} \geq 1 - \frac{1}{2\bar{C}}$, for constants $\bar{C}, \bar{c} > 0$, such that $\sigma_{ii} \leq \bar{C}$. (c) We have $\lim_{n\to\infty} \kappa_n = \kappa$ for a constant $\kappa \geq 0$, where $\kappa_n := \frac{1}{n}\sum_{m=1}^{J_n}\left(\sum_{i\neq j\in I_m} \sigma_{ij}^2\right)$.*

Assumption A.5 *Under the alternative hypothesis $H_1(k)$, (a) function $L_0(\theta)$ has a unique maximizer $\theta^* = (vec(F^*)', diag(V_\varepsilon^*)')'$ over Θ, and (b) we have $V_y \neq FF' + V_\varepsilon$, for any $T \times k$ matrix F and any diagonal positive definite matrix V_ε.*

Assumption A.6 *Matrix $Q_\beta := \lim_{n\to\infty} \frac{1}{n}\beta'\Sigma\beta$ is positive definite.*

Assumptions A.1 and A.2 require uniform bounds on factor loadings as well as on covariances and higher-order moments of the idiosyncratic errors. Assumption A.3 implies global identification in the FA model (see Lemma 5). Assumptions A.1-A.3 yield consistency of FA estimators (see proof of Lemma 6). We use Assumption A.4 together with Assumption A.2 to invoke a CLT based on a multivariate Lyapunov condition (see proof of Lemma 1) to establish the asymptotic distribution of the test statistics. To ease the verification of the Lyapunov condition, we bound a fourth-order moment of squared errors, which explains why we require finite eight-order moments in Assumption A.2. We could relax this condition at the expense of a more sophisticated proof of Lemma 1. The mild Assumption A.4 (b) requires that the smallest eigenvalue of $V[vech(w_i w_i')]$ is bounded away from 0 for all assets i up to a small fraction. In Assumption A.4 (c), in order to have κ_n bounded, we need either mixing dependence in idiosyncratic errors within blocks, i.e., $|\sigma_{i,j}| \leq C\rho^{|i-j|}$ for $i, j \in I_m$ and $0 \leq \rho < 1$, or vanishing correlations, i.e., $|\sigma_{i,j}| \leq Cb_{m,n}^{-\bar{s}}$ for all $i \neq j \in I_m$ and a constant $\bar{s} \geq 1/2$, with blocks of equal size. In Assumption A.5, part (a) defines the pseudo-true parameter value (White (1982)) under the alternative hypothesis, and part (b) is used to establish the consistency of the LR test under global alternatives (see proof of Proposition 4). Finally, Assumption A.6 is used to apply a Lyapunov CLT (see proof of Lemma 8) when deriving the asymptotic normality of the FA estimators.

B Proofs of Propositions 1-7

Proof of Proposition 1: Let $\hat{U}$ be the $T \times k$ matrix whose orthonormal columns are the eigenvectors for the k largest eigenvalues of matrix $\hat{V}_\varepsilon^{-1/2}\hat{V}_y\hat{V}_\varepsilon^{-1/2}$. Those eigenvalues are $1 + \hat{\gamma}_j$, $j = 1, ..., k$, while it holds $\hat{F} = \hat{V}_\varepsilon^{1/2}\hat{U}\hat{\Gamma}^{1/2}$, where $\hat{\Gamma} = diag(\hat{\gamma}_1, ..., \hat{\gamma}_k)$. We have $I_T - \hat{U}\hat{U}' = I_T - \hat{V}_\varepsilon^{-1/2}\hat{F}\hat{\Gamma}^{-1}\hat{F}'\hat{V}_\varepsilon^{-1/2} = I_T - \hat{V}_\varepsilon^{-1/2}\hat{F}(\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}\hat{F}'\hat{V}_\varepsilon^{-1/2} = \hat{V}_\varepsilon^{-1/2}M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon^{1/2} = \hat{V}_\varepsilon^{1/2}M'_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon^{-1/2}$. Thus, $\hat{S} = (I_T - \hat{U}\hat{U}')\left(\hat{V}_\varepsilon^{-1/2}\hat{V}_y\hat{V}_\varepsilon^{-1/2} - I_T\right)(I_T - \hat{U}\hat{U}')$. By the spectral decomposition of $\hat{V}_\varepsilon^{-1/2}\hat{V}_y\hat{V}_\varepsilon^{-1/2}$, we get $(I_T - \hat{U}\hat{U}')\left(\hat{V}_\varepsilon^{-1/2}\hat{V}_y\hat{V}_\varepsilon^{-1/2} - I_T\right)(I_T - \hat{U}\hat{U}') = \sum_{j=k+1}^{T}\hat{\gamma}_j\hat{P}_j$, where the $\hat{P}_j$ are the orthogonal projection matrices onto the eigenspaces for the $T - k$ smallest eigenvalues. Then, Part (a) follows. Part (b) is a consequence of the squared Frobenius norm of a symmetric matrix being equal to the sum of its squared eigenvalues. For Part (c), let $P_{\hat{F},\hat{V}_\varepsilon} = I_T - M_{\hat{F},\hat{V}_\varepsilon}$ and note that $\hat{F}\hat{F}' = P_{\hat{F},\hat{V}_\varepsilon}(\hat{V}_y - \hat{V}_\varepsilon) + (\hat{V}_y - \hat{V}_\varepsilon)P'_{\hat{F},\hat{V}_\varepsilon} - P_{\hat{F},\hat{V}_\varepsilon}(\hat{V}_y - \hat{V}_\varepsilon)P'_{\hat{F},\hat{V}_\varepsilon}$ $= \hat{V}_y - \hat{V}_\varepsilon - M_{\hat{F},\hat{V}_\varepsilon}(\hat{V}_y - \hat{V}_\varepsilon)M'_{\hat{F},\hat{V}_\varepsilon}$, where the first equality is because the three terms on the RHS are all equal to $\hat{F}\hat{F}'$ by (FA2). The conclusion follows from (FA1) and $\hat{V}_\varepsilon$ being diagonal. Finally, Part (d) follows because $\frac{1}{n}\hat{\varepsilon}\hat{\varepsilon}' = M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_yM'_{\hat{F},\hat{V}_\varepsilon}$, $\hat{V}_\varepsilon M'_{\hat{F},\hat{V}_\varepsilon} = M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon$, and $M_{\hat{F},\hat{V}_\varepsilon}$ is idempotent, which implies $\hat{V}_\varepsilon^{-1/2}M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon M'_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon^{-1/2} = \hat{V}_\varepsilon^{-1/2}M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon^{1/2}$.

Proof of Proposition 2: Let us substitute $\hat{V}_y = FF' + \tilde{V}_\varepsilon + \frac{1}{\sqrt{n}}\Psi_y + o_p(\frac{1}{\sqrt{n}})$ into (FA2) and rearrange to obtain $\hat{F}\hat{\Gamma} - FF'\hat{V}_\varepsilon^{-1}\hat{F} = \frac{1}{\sqrt{n}}\Psi_y\hat{V}_\varepsilon^{-1}\hat{F} + (\tilde{V}_\varepsilon\hat{V}_\varepsilon^{-1} - I_T)\hat{F} + o_p(\frac{1}{\sqrt{n}})$. From $\hat{V}_\varepsilon = \tilde{V}_\varepsilon + \frac{1}{\sqrt{n}}\Psi_\varepsilon + o_p(\frac{1}{\sqrt{n}})$, we have $\tilde{V}_\varepsilon\hat{V}_\varepsilon^{-1} - I_T = -\frac{1}{\sqrt{n}}\Psi_\varepsilon\hat{V}_\varepsilon^{-1} + o_p(\frac{1}{\sqrt{n}})$. Substituting into the above equation and right multiplying both sides by $(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}$ gives $\hat{F}\hat{\mathcal{D}} - F = \frac{1}{\sqrt{n}}(\Psi_y - \Psi_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{F}(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1} + o_p(\frac{1}{\sqrt{n}})$, where $\hat{\mathcal{D}} := \hat{\Gamma}(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}$. By the root-$n$ convergence of the FA estimates (see Section D.4.1), we get

$$\hat{F}\hat{\mathcal{D}} - F = \frac{1}{\sqrt{n}}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}}), \tag{B.1}$$

and $\hat{\mathcal{D}} = I_k + O_p(\frac{1}{\sqrt{n}})$, where $\Gamma = diag(\gamma_1, ..., \gamma_k)$. We can push the expansion by plugging into (B.1) the expansion of $\hat{\mathcal{D}}$. We have $F'\hat{V}_\varepsilon^{-1}\hat{F} = [I_k - (\hat{F} - F)'\hat{V}_\varepsilon^{-1}\hat{F}\hat{\Gamma}^{-1}]\hat{\Gamma}$, so that $\hat{\mathcal{D}} = [I_k - (\hat{F} - F)'\hat{V}_\varepsilon^{-1}\hat{F}\hat{\Gamma}^{-1}]^{-1} = I_k + (\hat{F} - F)'V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}})$. By plugging into (B.1), we get:

$$\hat{F} - F + F[(\hat{F} - F)'V_\varepsilon^{-1}F\Gamma^{-1}] = \frac{1}{\sqrt{n}}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}}). \tag{B.2}$$

By multiplying both sides with M_{F,V_ε}, we get $M_{F,V_\varepsilon}(\hat{F} - F) = \frac{1}{\sqrt{n}}M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(\frac{1}{\sqrt{n}})$. Then, $\hat{F} - F = \frac{1}{\sqrt{n}}M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + \frac{1}{\sqrt{n}}FA + o_p(\frac{1}{\sqrt{n}})$, where A is a random $k \times k$ matrix to be determined next. By plugging into (B.2), we get $F(A + A') = P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + o_p(1)$. By multiplying both sides by $\frac{1}{2}\Gamma^{-1}F'V_\varepsilon^{-1}$ and using $F'V_\varepsilon^{-1}P_{F,V_\varepsilon} = F'V_\varepsilon^{-1}$, we get the symmetric part of matrix A, i.e., $\frac{1}{2}(A + A') = \frac{1}{2}\Gamma^{-1}F'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1}$ (we include higher-order terms in the remainder $o_p(\frac{1}{\sqrt{n}})$). Thus, $\hat{F} - F = \frac{1}{\sqrt{n}}\Psi_F + o_p(\frac{1}{\sqrt{n}})$, where

$$\Psi_F = M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + \frac{1}{2}P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + F\tilde{A}, \tag{B.3}$$

and $\tilde{A} = \frac{1}{2}(A - A')$ is an antisymmetric $k \times k$ random matrix. To find the antisymmetric matrix $\tilde{A} = (\tilde{a}_{\ell,j})$, we use that $\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F}$ is diagonal. Plugging the expansions of the FA estimates, for the term at order $1/\sqrt{n}$ we get that the out-of-diagonal elements of matrix $\Psi_F'V_\varepsilon^{-1}F + F'V_\varepsilon^{-1}\Psi_F - F'V_\varepsilon^{-1}\Psi_\varepsilon V_\varepsilon^{-1}F = \frac{1}{2}\Gamma^{-1}F'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F + \frac{1}{2}F'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F\Gamma^{-1} + \Gamma\tilde{A} - \tilde{A}\Gamma - F'V_\varepsilon^{-1}\Psi_\varepsilon V_\varepsilon^{-1}F$ are nil. Setting the (ℓ, j) element of this matrix equal to 0, we get $\tilde{a}_{\ell,j} = -\tilde{a}_{j,\ell} = \frac{1}{\gamma_j - \gamma_\ell}\left[\frac{1}{2}(\frac{1}{\gamma_j} + \frac{1}{\gamma_\ell})F_\ell'V_\varepsilon^{-1}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j - F_\ell'V_\varepsilon^{-1}\Psi_\varepsilon V_\varepsilon^{-1}F_j\right]$, for $j \neq \ell$. Then, from Equation (B.3), the jth column of Ψ_F is $\Psi_{F_j} = \frac{1}{\gamma_j}M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \frac{1}{2\gamma_j}P_{F_j,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \sum_{\ell=1:\ell\neq j}^k \frac{1}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j - \sum_{\ell=1:\ell\neq j}^k \frac{\gamma_\ell}{\gamma_j - \gamma_\ell}P_{F_\ell,V_\varepsilon}\Psi_\varepsilon V_\varepsilon^{-1}F_j$, where we use $P_{F,V_\varepsilon} = \sum_{\ell=1}^k P_{F_\ell,V_\varepsilon}$. Part (a) follows.

Let us now prove part (b). The asymptotic expansion of condition (FA1) yields:

$$diag(\Psi_y) = diag\left(\sum_{j=1}^{k}(F_j\Psi'_{F_j} + \Psi_{F_j}F'_j) + \Psi_\varepsilon\right). \tag{B.4}$$

From part (a) and the definition of P_{F_j,V_ε} we have $\sum_{j=1}^k \Psi_{F_j}F'_j = \frac{1}{2}\sum_{j=1}^k P_{F_j,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} + M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F,V_\varepsilon} + \sum_{\ell\neq j}\frac{\gamma_j}{\gamma_j-\gamma_\ell}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} - \sum_{\ell\neq j}^k \frac{\gamma_\ell\gamma_j}{\gamma_j-\gamma_\ell}P_{F_\ell,V_\varepsilon}\Psi_\varepsilon P'_{F_j,V_\varepsilon} =: N_1 + N_2 + N_3 + N_4$, where $P_{F,V_\varepsilon} = \sum_{j=1}^k P_{F_j,V_\varepsilon} = I_T - M_{F,V_\varepsilon}$ and $\sum_{\ell\neq j}$ denotes the double sum over $j,\ell = 1,...,k$ such that $\ell \neq j$. Matrix N_1 is symmetric and it contributes $2N_1$ to the RHS of (B.4). Instead, matrix N_4 is antisymmetric (it can be seen by interchanging indices j and ℓ in the summation) and it does not contribute to the RHS of (B.4). For matrix N_3 we have $N_3 + N'_3 = \sum_{\ell\neq j}\frac{\gamma_j}{\gamma_j-\gamma_\ell}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} + \sum_{\ell\neq j}\frac{\gamma_\ell}{\gamma_\ell-\gamma_j}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} = \sum_{\ell\neq j}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} = \sum_{\ell,j}P_{F_\ell,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} - \sum_j P_{F_j,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F_j,V_\varepsilon} = P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F,V_\varepsilon} - 2N_1$, where we have interchanged ℓ and j in the first equality when writing N'_3. Thus, we get:

$$\begin{aligned}\sum_{j=1}^{k}(F_j\Psi'_{F_j} + \Psi_{F_j}F'_j) &= M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F,V_\varepsilon} + P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M'_{F,V_\varepsilon} + P_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)P'_{F,V_\varepsilon} \\ &= (\Psi_y - \Psi_\varepsilon) - M_{F,V_\varepsilon}(\Psi_y - \Psi_\varepsilon)M'_{F,V_\varepsilon}. \end{aligned}\tag{B.5}$$

Then, Equation (B.4) with (B.5) yields Equation (5).

Proof of Proposition 3: We use the following properties of the Hadamard product: $(ab')\odot(cd') = (a\odot c)(b\odot d)'$, $diag(ab') = a\odot b$, $a'\Delta b = (a\odot b)'diag(\Delta)$, and $(\Delta a)\odot b = a\odot(\Delta b) = \Delta(a\odot b)$ for conformable vectors a,b,c,d and diagonal matrix Δ. Moreover, we deploy the following facts about the $vech$ operator: $diag(GAG') = \sqrt{2}\boldsymbol{X}'vech(A)$ and $vech(G'\Delta G) = \frac{1}{\sqrt{2}}\boldsymbol{X}\,diag(\Delta)$ for $(T-k)\times(T-k)$ symmetric matrix A and diagonal $T\times T$ matrix Δ.

(a) With $G = [g_1 \;:\; \cdots \;:\; g_{T-k}]$, we have $M_{F,V_\varepsilon} = GG'V_\varepsilon^{-1} = \sum_{j=1}^{T-k} g_j(V_\varepsilon^{-1}g_j)'$. Then, we get the Hadamard product $M_{F,V_\varepsilon}^{\odot 2} = \sum_{i,j=1}^{T-k}[g_i(V_\varepsilon^{-1}g_i)']\odot[g_j(V_\varepsilon^{-1}g_j)'] = \left[\sum_{i,j=1}^{T-k}(g_i\odot g_j)(g_i\odot g_j)'\right] V_\varepsilon^{-2} = \left[\sum_{i=1}^{T-k}(g_i\odot g_i)(g_i\odot g_i)' + 2\sum_{i<j}(g_i\odot g_j)(g_i\odot g_j)'\right] V_\varepsilon^{-2} = (\boldsymbol{X}'\boldsymbol{X})\,V_\varepsilon^{-2}$.

(b) From part (a) and Equation (7), $\Psi_\varepsilon = diag\left(V_\varepsilon^2\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,diag(M_{F,V_\varepsilon}Z_nM'_{F,V_\varepsilon})\right)$, with $diag(M_{F,V_\varepsilon}Z_nM'_{F,V_\varepsilon}) = diag\,(GG'V_\varepsilon^{-1}Z_nV_\varepsilon^{-1}GG') = \sqrt{2}\boldsymbol{X}'vech\,(G'V_\varepsilon^{-1}Z_nV_\varepsilon^{-1}G)\,.$

(c) We use $\bar{Z}_n^{\;*} = Z_n^* - G'V_\varepsilon^{-1}\mathcal{T}_{F,V_\varepsilon}(Z_n)V_\varepsilon^{-1}G$ in vectorized form. From part (b), we have $diag(\mathcal{T}_{F,V_\varepsilon}(Z_n)) = \sqrt{2}V_\varepsilon^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'vech\left(Z_n^*\right)$. Moreover, $vech\left(G'V_\varepsilon^{-1}\mathcal{T}_{F,V_\varepsilon}(Z_n)V_\varepsilon^{-1}G\right) = \frac{1}{\sqrt{2}}\boldsymbol{X}diag(V_\varepsilon^{-1}\mathcal{T}_{F,V_\varepsilon}(Z_n)V_\varepsilon^{-1}) = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'vech\left(Z_n^*\right)$. The conclusion follows.

Proof of Proposition 4: (a) We first establish asymptotic normality of $\mathscr{Z}_n := V_\varepsilon^{-1/2}Z_nV_\varepsilon^{-1/2}$.

Lemma 1 *(a) Under Assumptions 1-2, A.2, A.4 (a)-(b), we have* $\Omega_n^{-1/2}vech(\mathscr{Z}_n) \Rightarrow N(0, I_{\frac{T(T+1)}{2}})$ *as* $n \to \infty$ *and* T *is fixed, where* $\Omega_n = D_n + \kappa_n I_{\frac{T(T+1)}{2}}$, *and* $\kappa_n = \frac{1}{n}\sum_{m=1}^{J_n}\left(\sum_{i\neq j\in I_m}\sigma_{ij}^2\right)$. *If additionally Assumption A.4 (c) holds, then* $vech(\mathscr{Z}_n) \Rightarrow N(0,\Omega)$, *with* $\Omega := D + \kappa I_{\frac{T(T+1)}{2}}$.

The proof of the next Lemma on $vech(Z_n^*)$ being a linear transformation of $vech(\mathscr{Z}_n)$ uses $vec(S) = A_m vech(S)$ for any symmetric $m \times m$ matrix S, where A_m is the $m^2 \times \frac{1}{2}m(m+1)$ duplication matrix (Magnus, Neudecker (2007)) suited to our definition of the half-vectorization operator $vech$ and given by $A_m = \left[\sqrt{2}(e_1 \otimes e_1) : \cdots : \sqrt{2}(e_m \otimes e_m) : \{e_i \otimes e_j + e_j \otimes e_i\}_{i<j}\right]$, with e_i being the ith unit vector in dimension m.

Lemma 2 *Under Assumption 1, we have* $vech(Z_n^*) = vech(Q'\mathscr{Z}_nQ) = \boldsymbol{R}'vech(\mathscr{Z}_n)$, *where* $\boldsymbol{R} := \frac{1}{2}A_T'(Q\otimes Q)A_{T-k}$ *is a* $\frac{1}{2}T(T+1)\times p$ *matrix with orthonormal columns, and* $Q := V_\varepsilon^{-1/2}G$.

From Proposition 3 (c) and Lemma 2, we get $vech(\bar{Z}_n^*) = \left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)\boldsymbol{R}'vech(\mathscr{Z}_n)$. Then, Lemma 1 yields part (a) with $\Omega_{\bar{Z}^*} = \left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)\boldsymbol{R}'\Omega\boldsymbol{R}\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)$.

(b) We have $\hat{z}_{m,n} = \sum_{i\in I_m}\hat{G}'\hat{V}_\varepsilon^{-1}\left[\tilde{y}_i\tilde{y}_i' - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\tilde{y}_i\tilde{y}_i')\right]\hat{V}_\varepsilon^{-1}\hat{G}$ with $\tilde{y}_i = y_i - \bar{y}$, because $\hat{\varepsilon}_i = M_{\hat{F},\hat{V}_\varepsilon}\tilde{y}_i$ and $\hat{G}'\hat{V}_\varepsilon^{-1}M_{\hat{F},\hat{V}_\varepsilon} = \hat{G}'\hat{V}_\varepsilon^{-1}$. Using $\tilde{y}_i\tilde{y}_i' = \tilde{\varepsilon}_i\tilde{\varepsilon}_i' + F\beta_i\beta_i'F' + F\beta_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\beta_i'F'$, we get $\hat{z}_{m,n} = \sum_{i\in I_m}\hat{G}'\hat{V}_\varepsilon^{-1}\left[\tilde{\varepsilon}_i\tilde{\varepsilon}_i' - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\tilde{\varepsilon}_i\tilde{\varepsilon}_i')\right]\hat{V}_\varepsilon^{-1}\hat{G} + \sum_{i\in I_m}\hat{G}'\hat{V}_\varepsilon^{-1}\left[F\beta_i\beta_i'F' - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(F\beta_i\beta_i'F')\right]\hat{V}_\varepsilon^{-1}\hat{G} + \sum_{i\in I_m}\hat{G}'\hat{V}_\varepsilon^{-1}\left[F\beta_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\beta_i'F' - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(F\beta_i\tilde{\varepsilon}_i' + \tilde{\varepsilon}_i\beta_i'F')\right]\hat{V}_\varepsilon^{-1}\hat{G} =: \tilde{z}_{m,n} + z_{m,n,1} + z_{m,n,2}$, where $\tilde{\varepsilon}_i = \varepsilon_i - \bar{\varepsilon}$. Then, we can decompose $\hat{\Omega}_{\bar{Z}^*}$ into a sum of a leading term and other terms, which are asymptotically negligible, so that $\hat{\Omega}_{\bar{Z}^*} = \tilde{\Omega}_{\bar{Z}^*} + o_p(1)$, with $\tilde{\Omega}_{\bar{Z}^*} = \frac{1}{n}\sum_{m=1}^{J_n}vech(\bar{z}_{m,n})vech(\bar{z}_{m,n})'$, with $\bar{z}_{m,n}$ defined as $\tilde{z}_{m,n}$ after replacing $\tilde{\varepsilon}_i$ with ε_i. Let us now show that $\tilde{\Omega}_{\bar{Z}^*} = \Omega_{\bar{Z}^*} + o_p(1)$ up to pre- and post-multiplication by a rotation matrix and its inverse. First, we note that

$vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}\left[\varepsilon_i\varepsilon_i' - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\varepsilon_i\varepsilon_i')\right]\hat{V}_\varepsilon^{-1}\hat{G}\right) = vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}\left[\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon)\right]\right.$ $\left.\hat{V}_\varepsilon^{-1}\hat{G}\right)$ because $\mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\cdot)$ is the identity transformation for diagonal matrices. Moreover, we have:

$$
\begin{aligned}
vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}\left[\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon - \mathcal{T}_{\hat{F},\hat{V}_\varepsilon}(\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon)\right]\hat{V}_\varepsilon^{-1}\hat{G}\right) &= M_{\hat{\boldsymbol{X}}} vech\left(\hat{G}'\hat{V}_\varepsilon^{-1}(\varepsilon_i\varepsilon_i' - \sigma_{ii}V_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}\right) \\
= \; M_{\hat{\boldsymbol{X}}} vech\left(\tilde{Q}'(e_ie_i' - \sigma_{ii}I_T)\tilde{Q}\right) &= M_{\hat{\boldsymbol{X}}}\hat{\boldsymbol{R}}' vech(e_ie_i' - \sigma_{ii}I_T), \qquad \text{(B.6)}
\end{aligned}
$$

where $M_{\hat{\boldsymbol{X}}} := I_p - \hat{\boldsymbol{X}}(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'$ with $\hat{\boldsymbol{X}}$ defined as $\boldsymbol{X}$ by replacing G with $\hat{G}$, we define $e_i = V_\varepsilon^{-1/2}\varepsilon_i$, and $\hat{\boldsymbol{R}} := \frac{1}{2}A_T'(\tilde{Q}\otimes\tilde{Q})A_{T-k}$ with $\tilde{Q} = V_\varepsilon^{1/2}\hat{V}_\varepsilon^{-1}\hat{G}$. The first equality in (B.6) uses an argument similar to Proposition 3 (c), and the third equality is similar to Lemma 2. We get $vech(\bar{z}_{m,n}) = M_{\hat{\boldsymbol{X}}}\hat{\boldsymbol{R}}' vech(\zeta_{m,n})$, where $\zeta_{m,n} := \sum_{i\in I_m}(e_ie_i' - \sigma_{ii}I_T)$. Besides, $vech(\mathscr{Z}_n) = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n} vech(\zeta_{m,n})$. Then, $\tilde{\Omega}_{\bar{Z}^*} = M_{\hat{\boldsymbol{X}}}\hat{\boldsymbol{R}}'\tilde{\Omega}_n\hat{\boldsymbol{R}}M_{\hat{\boldsymbol{X}}}$ for $\tilde{\Omega}_n := \frac{1}{n}\sum_{m=1}^{J_n} vech(\zeta_{m,n})vech(\zeta_{m,n})'$. Further, we have $E[\tilde{\Omega}_n] = V[vech(\mathscr{Z}_n)] = \Omega_n$. Moreover, $\tilde{\Omega}_n - E[\tilde{\Omega}_n] = o_p(1)$, by using $vec(\tilde{\Omega}_n) = \frac{1}{n}\sum_{m=1}^{J_n} vech(\zeta_{m,n})\otimes vech(\zeta_{m,n})$ and $\|V[vec(\tilde{\Omega}_n)]\| \le C\frac{1}{n^2}\sum_{m=1}^{J_n} E\left[\|vech(\zeta_{m,n})\|^4\right] = o(1)$, where the latter bound is shown in the proof of Lemma 1 using Assumption 2 (d). Additionally, by Assumption A.4, we have $\Omega_n = \Omega + o(1)$. Thus, $\tilde{\Omega}_n = \Omega + o_p(1)$. Now, we use consistency of the FA estimates and $\hat{G}\hat{O} = G + o_p(1)$ for a (possibly data-dependent) $T-k$ dimensional orthogonal matrix $\hat{O}$. Then, by Proposition 10 (e) in Appendix D.5, we have $\hat{\boldsymbol{R}}M_{\hat{\boldsymbol{X}}}\hat{\mathscr{R}}^{-1} = \hat{\boldsymbol{R}}\left(I_p - \hat{\boldsymbol{X}}(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'\right)\mathscr{R}(\hat{O})^{-1} = \boldsymbol{R}(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') + o_p(1)$, for a p dimensional orthogonal matrix $\hat{\mathscr{R}} \equiv \mathscr{R}(\hat{O})$. We conclude that $\hat{\mathscr{R}}\tilde{\Omega}_{\bar{Z}^*}\hat{\mathscr{R}}^{-1}$ is a consistent estimator of $\Omega_{\bar{Z}^*}$ as $n\to\infty$ and T is fixed, which yields part (b).

(c) Under $H_1(k)$ and Assumption A.5 (a), we have $\hat{F} \xrightarrow{p} F^*$ and $\hat{V}_\varepsilon \xrightarrow{p} V_\varepsilon^*$. Then, $\hat{S} \xrightarrow{p} S^*$ with $S^* = (V_\varepsilon^*)^{-1/2}M_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*)M_{F^*,V_\varepsilon^*}'(V_\varepsilon^*)^{-1/2} \neq 0$. Indeed, if S^* were the null matrix, then we would have $M_{F^*,V_\varepsilon^*}(V_y - V_\varepsilon^*)M_{F^*,V_\varepsilon^*}' = 0$, which implies $V_y - V_\varepsilon^* = F^*A(F^*)'$ for a symmetric matrix A, in contradiction with Assumption A.5 (b). Thus, $n\|\hat{S}\|^2 \ge Cn$, w.p.a. 1, for a constant $C > 0$. Moreover, using $vech(\hat{z}_{m,n}) = \left(I_p - \hat{\boldsymbol{X}}(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'\right)vech(\hat{G}'\hat{V}_\varepsilon^{-1}(\sum_{i\in I_m}\tilde{y}_i\tilde{y}_i')\hat{V}_\varepsilon^{-1}\hat{G})$ and the conditions on Θ, we get $\|vech(\hat{z}_{m,n})\| \le C\sum_{i\in I_m}\|\tilde{y}_i\|^2$. Then, from Assumptions A.1 and A.2, $E[\|\hat{\Omega}_{\bar{Z}^*}\|] \le C\frac{1}{n}\sum_{m=1}^{J_n} b_{m,n}^2 = O(n\sum_{m=1}^{J_n} B_{m,n}^2)$. Moreover, $\sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$. Indeed, Assumption 2 (d) implies $B_{m,n} \le cn^{-\frac{\delta}{\delta+1}}$ uniformly in m, for any $c > 0$ and n large enough, and hence $\sum_{m=1}^{J_n} B_{m,n}^2 = cn^{-\frac{\delta}{\delta+1}} \sum_{m=1}^{J_n} B_{m,n} \leq c$, for any $c > 0$ and n large. Part (c) follows.

Proof of Proposition 5: We have $LR(k) = \frac{n}{2}\|\hat{S}\|^2 + o_p(1) = nvech(\hat{S})'vech(\hat{S}) + o_p(1)$. Moreover, from the asymptotic expansion (9), we can write $\sqrt{n}vech(\hat{S}) = A(F, V_\varepsilon)z_n^{AD} + o_p(1)$, where vector z_n^{AD} stacks the $T(T-1)/2$ above-diagonal elements of matrix Z_n and $A(F, V_\varepsilon)$ is a deterministic matrix whose elements only depend on F, V_ε. From Conditions (a) and (b) of Proposition 5, and Lemma 1, we have $z_n^{AD} \Rightarrow N(0, \Omega_z)$, where the diagonal matrix Ω_z is the same as if the errors were independent normally distributed - up to replacing q with $q + \kappa$.

Proof of Proposition 6: Let us first get the asymptotic expansion of $\hat{V}_y = \frac{1}{n}\tilde{Y}\tilde{Y}'$. With the drifting DGP $Y = \mu 1_n' + F\beta' + F_{k+1}\beta_{loc}' + \varepsilon$, and using $\bar{\beta} = 0$, $\bar{\beta}_{loc} = 0$, $\frac{1}{n}[\beta : \beta_{loc}]'[\beta : \beta_{loc}] = I_{k+1}$ and Lemma 6 (a) in Appendix D, we get $\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}\Psi_{y,loc} + R_y$, where $\tilde{V}_y = FF' + \tilde{V}_\varepsilon$,

$$\Psi_{y,loc} = c_{k+1}\rho_{k+1}\rho_{k+1}' + \frac{1}{\sqrt{n}}(\varepsilon\beta F' + F\beta'\varepsilon') + \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right), \tag{B.7}$$

and $R_y = \frac{1}{n}(\varepsilon\beta_{loc}F_{k+1}' + F_{k+1}\beta_{loc}'\varepsilon') + [F_{k+1}F_{k+1}' - n^{-1/2}c_{k+1}\rho_{k+1}\rho_{k+1}'] + o_p(\frac{1}{\sqrt{n}})$. Using $F_{k+1} = \sqrt{\gamma_{k+1}}\rho_{k+1}$ and $\sqrt{n}\gamma_{k+1} = c_{k+1} + o(1)$, we get $R_y = o_p(1/\sqrt{n})$. We use Equation (5) with $\Psi_{y,loc}$ given in (B.7) instead of Ψ_y, and get $\Psi_\varepsilon = \mathcal{T}_{F,V_\varepsilon}(Z_{n,loc})$, where $Z_{n,loc} := \sqrt{n}\left(\frac{1}{n}\varepsilon\varepsilon' - \tilde{V}_\varepsilon\right) + c_{k+1}\rho_{k+1}\rho_{k+1}'$ and $diag(\Psi_\varepsilon) = \sqrt{2}V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z_{n,loc}^*)$. Then, as in Equation (8), $\sqrt{n}\hat{S} = V_\varepsilon^{-1/2}M_{F,V_\varepsilon}\left(Z_{n,loc} - \mathcal{T}_{F,V_\varepsilon}(Z_{n,loc})\right)M_{F,V_\varepsilon}'V_\varepsilon^{-1/2} + o_p(1) = V_\varepsilon^{-1/2}G(\bar{Z}_n^* + \Delta)G'V_\varepsilon^{-1/2} + o_p(1)$, where $\Delta = c_{k+1}G'V_\varepsilon^{-1}\left(\rho_{k+1}\rho_{k+1}' - \mathcal{T}_{F,V_\varepsilon}(\rho_{k+1}\rho_{k+1}')\right)V_\varepsilon^{-1}G$. As in Proposition 3 (c), we have $vech(\Delta) = c_{k+1}\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)vech(G'V_\varepsilon^{-1}\rho_{k+1}\rho_{k+1}'V_\varepsilon^{-1}G) = c_{k+1}\left(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)$ $vech(\xi_{k+1}\xi_{k+1}')$ since $\rho_{k+1} = G\xi_{k+1}$. From the proof of Proposition 4 (a), $\bar{Z}_n^* \Rightarrow \bar{Z}^*$ as $n \to \infty$, and Part (a) follows. Part (b) is a consequence of Part (a) and the Continuous Mapping Theorem.

Proof of Proposition 7: The proof of part (a) is in three steps. (i) The testing problem asymptotically simplifies to the null hypothesis $H_0 : \lambda_1 = ... = \lambda_{df} = 0$ vs. the alternative hypothesis $H_1 : \exists\lambda_j > 0, j = 1, ..., df$. Let us define $\boldsymbol{\lambda}_0 = (0, ..., 0)'$ for the null hypothesis and pick a given vector $\boldsymbol{\lambda}_1 = (\lambda_1, ..., \lambda_{df})'$ in the alternative hypothesis, and consider the test of $\boldsymbol{\lambda}_0$ versus $\boldsymbol{\lambda}_1$ (simple hypothesis). By Neyman-Pearson Lemma, the most powerful test for $\boldsymbol{\lambda}_0$ versus $\boldsymbol{\lambda}_1$ rejects the null hypothesis when $f(z; \lambda_1, ..., \lambda_{df})/f(z; 0, ..., 0)$ is large, i.e., the test function is

$\phi(z) = \mathbf{1}\left\{\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)} \geq C\right\}$ for a constant $C > 0$ set to ensure the correct asymptotic size.

(ii) Let us now show that the density ratio $\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)}$ is an increasing function of z. To show this, we can rely on an expansion of the density of $\sum_{j=1}^{df} \mu_j \chi^2(1, \lambda_j^2)$ in terms of central chi-square densities (Kotz, Johnson, and Boyd (1967) Equations (144) and (151)):

$$f(z; \lambda_1, ..., \lambda_{df}) = \sum_{k=0}^{\infty} \bar{c}_k(\lambda_1, ..., \lambda_{df}) g(z; df + 2k, 0), \tag{B.8}$$

where the coefficients $\bar{c}_k(\lambda_1, ..., \lambda_{df}) = Ae^{-\sum_{j=1}^{df} \lambda_j^2/2} E[Q(\lambda_1, ..., \lambda_{df})^k]/k!$ involve moments of the quadratic form $Q(\lambda_1, ..., \lambda_{df}) = (1/2) \sum_{j=1}^{df} \left(\nu_j^{1/2} X_j + \lambda_j (1 - \nu_j)^{1/2}\right)^2$ of the mutually independent variables $X_j \sim N(0, 1)$, $A = \prod_{j=1}^{df} \mu_j^{-1/2}$, and $\nu_j = 1 - \frac{1}{\mu_j} \min_\ell \mu_\ell$. Without loss of generality for checking the monotonicity, we have rescaled the density so that $\min_j \mu_j = 1$. Then, from (B.8), we get the ratio: $\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)} = \frac{\sum_{k=0}^{\infty} \bar{c}_k(\lambda_1,...,\lambda_{df}) g(z;df+2k,0)}{\sum_{k=0}^{\infty} \bar{c}_k(0,...,0) g(z;df+2k,0)}$. By dividing both the numerator and the denominator by the central chi-square density $g(z; df, 0)$, we get $\frac{f(z;\lambda_1,...,\lambda_{df})}{f(z;0,...,0)} = e^{-\sum_{j=1}^{df} \lambda_j^2/2} \frac{\sum_{k=0}^{\infty} c_k(\lambda_1,...,\lambda_{df}) \psi_k(z)}{\sum_{k=0}^{\infty} c_k(0,...,0) \psi_k(z)} =: e^{-\sum_{j=1}^{df} \lambda_j^2/2} \Psi(z; \lambda_1, ..., \lambda_{df})$, where $\psi_k(z) := g(z; df + 2k, 0)/g(z; df, 0) = \frac{\Gamma(\frac{df}{2})}{2^k \Gamma(\frac{df}{2}+k)} z^k$ is the ratio of central chi-square distributions with $df + 2k$ and df degrees of freedom, and $c_k(\lambda_1, ..., \lambda_{df}) = E[Q(\lambda_1, ..., \lambda_{df})^k]/k!$. We use the short notation $c_k(\lambda) := c_k(\lambda_1, ..., \lambda_{df})$ and $c_k(0) := c_k(0, ..., 0)$. The factor $e^{-\sum_{j=1}^{df} \lambda_j^2/2}$ does not impact on the monotonicity of the density ratio. We take the derivative of $\Psi(z; \lambda_1, ..., \lambda_{df})$ with respect to argument z and get $\partial_z \Psi(z; \lambda_1, ..., \lambda_{df}) = \frac{\left(\sum_{k=1}^{\infty} c_k(\lambda) \psi_k'(z)\right)\left(1+\sum_{k=1}^{\infty} c_k(0) \psi_k(z)\right)}{\left(\sum_{k=0}^{\infty} c_k(0) \psi_k(z)\right)^2} - \frac{\left(1+\sum_{k=1}^{\infty} c_k(\lambda) \psi_k(z)\right)\left(\sum_{k=1}^{\infty} c_k(0) \psi_k'(z)\right)}{\left(\sum_{k=0}^{\infty} c_k(0) \psi_k(z)\right)^2}$. The sign is given by the difference of the numerators, which is $\sum_{k=1}^{\infty} [c_k(\lambda) - c_k(0)] \psi_k'(z) + \sum_{k,l=1,k \neq l}^{\infty} c_k(\lambda) c_l(0) [\psi_k'(z) \psi_l(z) - \psi_k(z) \psi_l'(z)] = \sum_{k=1}^{\infty} [c_k(\lambda) - c_k(0)] \psi_k'(z) + \sum_{k,l=1,k>l}^{\infty} [c_k(\lambda) c_l(0) - c_l(\lambda) c_k(0)] [\psi_k'(z) \psi_l(z) - \psi_k(z) \psi_l'(z)]$. We use $\psi_k'(z) = \frac{\Gamma(\frac{d}{2}) k}{2^k \Gamma(\frac{d}{2}+k)} z^{k-1}$ and $\psi_k'(z) \psi_l(z) - \psi_k(z) \psi_l'(z) = (k - l) \frac{\Gamma(\frac{d}{2})^2}{2^{k+l} \Gamma(\frac{d}{2}+k) \Gamma(\frac{d}{2}+l)} z^{k+l-1}$ for $k > l$ and $z \geq 0$. The difference of the numerators in the derivative of the density ratio becomes: $\frac{1}{2} \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+1)} [c_1(\lambda) - c_1(0)] + \frac{1}{2^2} \frac{2\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+2)} [c_2(\lambda) - c_2(0)] z + \sum_{m=3}^{\infty} \frac{1}{2^m} \left(m \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+m)} [c_m(\lambda) - c_m(0)] + \sum_{k>l \geq 1, k+l=m} \frac{(k-l)\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d}{2}+k)\Gamma(\frac{d}{2}+l)} [c_k(\lambda) c_l(0) - c_l(\lambda) c_k(0)] \right) z^{m-1} = \sum_{m=1}^{\infty} \frac{1}{2^m} \kappa_m z^{m-1}$, with $\kappa_m :=$

$\sum_{k>l\geq 0,k+l=m}(k-l)\frac{\Gamma(\frac{d}{2})^2}{\Gamma(\frac{d}{2}+k)\Gamma(\frac{d}{2}+l)}\left[c_k(\lambda)c_l(0)-c_l(\lambda)c_k(0)\right]$. A direct calculation shows that $\kappa_1,\kappa_2\geq 0$. Hence, a sufficient condition for monotonicity of the density ratio is $\kappa_m\geq 0$, for all $m\geq 3$, i.e., Inequalities (12). Thus, the test rejects for large values of the argument, i.e., $\phi(z)=\mathbf{1}\{z\geq\bar{C}\}$, where the constant $\bar{C}$ is determined by fixing the asymptotic size under the null hypothesis.

(iii) Since the test function ϕ does not depend on $\boldsymbol{\lambda}_1$, it is AUMPI in the class of hypothesis tests based on the LR statistic (or the squared norm statistic). It yields part (a).

Let us now turn to the proof of part (b). From the definition of the κ_m coefficients written as $\kappa_m=\sum_{j>l\geq 0,j+l=m}\frac{(j-l)\Gamma(\frac{df}{2})^2}{\Gamma(\frac{df}{2}+j)\Gamma(\frac{df}{2}+l)}c_j(0)c_l(0)[\frac{c_j(\lambda)}{c_j(0)}-\frac{c_l(\lambda)}{c_l(0)}]$, it is sufficient to get $\kappa_m\geq 0$, for all m, that sequence $\frac{c_j(\lambda)}{c_j(0)}$, for $j=0,1,...$, is increasing. To prove that, we link the coefficients $c_j(\lambda)$ to the complete exponential Bell's polynomials (Bell (1934)) and establish the following recurrence.

Lemma 3 *We have* $c_{l+1}(\lambda)=\frac{1}{l+1}\sum_{i=0}^{l}\left(\frac{1}{2}\sum_{j=1}^{df}\nu_j^i\left[\nu_j+(i+1)(1-\nu_j)\lambda_j^2\right]\right)c_{l-i}(\lambda)$, *for* $l\geq 0$.

We use $\frac{c_l(\lambda)}{c_l(0)}=\frac{\tilde{c}_l(\lambda)}{\gamma_l}$, where we obtain the sequences $\gamma_l:=c_l(0)\nu_{df}^{-l}$ and $\tilde{c}_l(\lambda):=c_l(\lambda)\nu_{df}^{-l}$ by standardization with ν_{df}^{-l}. From Lemma 3, we have $\gamma_{l+1}=\frac{1}{l+1}\sum_{i=0}^{l}\frac{1}{2}\left(1+\sum_{j=2}^{df-1}\rho_j^{i+1}\right)\gamma_{l-i}$ with $\gamma_0=1$, and $\tilde{c}_{l+1}(\lambda)=\frac{1}{l+1}\sum_{i=0}^{l}\left(\frac{1}{2}\sum_{j=1}^{df}\rho_j^i\left[\rho_j+\frac{i+1}{\nu_{df}}(1-\nu_j)\lambda_j^2\right]\right)\tilde{c}_{l-i}(\lambda)$ with $\tilde{c}_0(\lambda)=1$ (note that $\rho_1=0$ and $\rho_{df}=1$). To prove that sequence $\frac{\tilde{c}_l(\lambda)}{\gamma_l}$ is increasing, the next lemma provides a sufficient condition from "separation" of the coefficients that define the recursive relations.

Lemma 4 (Separation Lemma) *Let* (a_i) *be a real sequence, and let* $b_i=\frac{1}{2}\left(1+\sum_{j=2}^{df-1}\rho_j^i\right)$, *for* $i\geq 1$, *where* $0\leq\rho_j\leq 1$. *Let sequences* (g_l) *and* (c_l) *be defined recursively by* $g_{l+1}=\frac{1}{l}(b_1g_l+b_2g_{l-1}+...+b_l)$ *and* $c_{l+1}=\frac{1}{l}(a_1c_l+a_2c_{l-1}+...+a_l)$, *with* $g_1=c_1=1$. *Suppose that* $a_i\geq\max\{\frac{df-1}{2},1\}$, *for all* i *(separation condition). Then, sequence* $(\frac{c_l}{g_l})$ *is increasing.*

We apply Lemma 4 to sequences $\tilde{c}_l(\lambda)$ and γ_l. We detail the case $df\geq 3$ (for $df=2$ the analysis is simpler). The separation condition $\frac{1}{2}\sum_{j=1}^{df}\rho_j^i\left[\rho_j+\frac{i+1}{\nu_{df}}(1-\nu_j)\lambda_j^2\right]\geq\frac{df-1}{2}$, for $i=0$, yields $\lambda_1^2+\sum_{j=2}^{df}(1-\nu_j)\lambda_j^2\geq\nu_{df}\left(df-2-\sum_{j=2}^{df-1}\rho_j\right)$, and, for $i\geq 1$, it yields $\sum_{j=2}^{df-1}\rho_j^i(1-\nu_j)\lambda_j^2+(1-\nu_{df})\lambda_{df}^2\geq\frac{\nu_{df}}{i+1}\left(df-2-\sum_{j=2}^{df-1}\rho_j^{i+1}\right)$. Inequalities (13) follow.

Figure 1: The upper panel displays the p-values for the statistic $LR(k)$ for the subperiods from January 1963 to December 2021, stopping at the smallest k such that $H_0(k)$ is not rejected at level $\alpha_n = 10/n_{max}$. If no such k is found then p-values are displayed up to k_{max}. We use rolling windows of $T = 20$ months moving forward by 12 months each time. The first bar of p-values covers the whole 20 months. Other bars cover the last 12 months of the 20 months subperiod. We flag bear market phases with grey shaded vertical bars. The five lower panels display $\left(\widehat{\bar{V}}_y\right)^{1/2}$ for total cross-sectional volatility, $\left(\overline{\hat{F}'\hat{F}}\right)^{1/2}$ for systematic volatility, $\left(\widehat{\bar{V}}_\varepsilon\right)^{1/2}$ for idiosyncratic volatility, as well as $\hat{R}^2$ and $\hat{R}^2$ under a single-factor model.

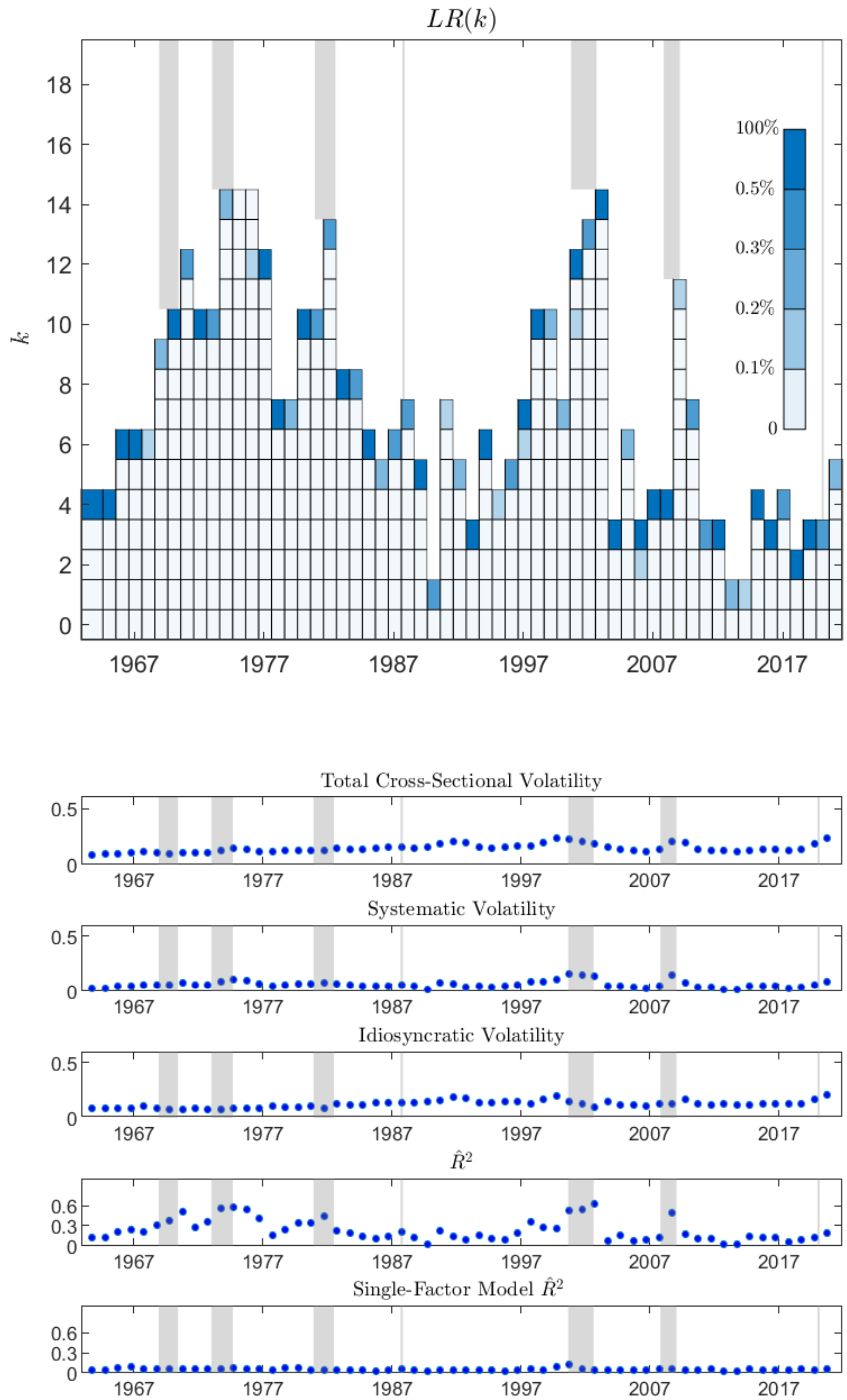
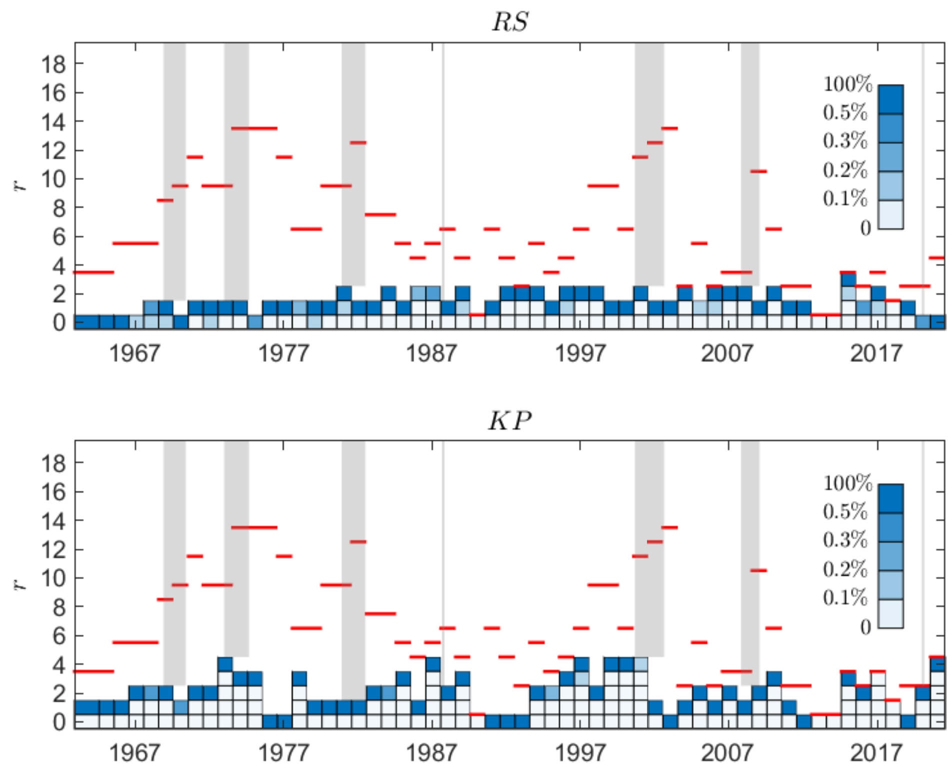


Figure 2: The upper and lower panels display the p-values for the RS and KP statistics for the subperiods from January 1963 to December 2021, for the rank test of the null hypothesis $H_{0,sp}(r)$ that F^O has rank r against the alternative hypothesis of rank larger than r, for any integer $r \leq k-1$. The empirical matrix $\hat{F}^O$ is computed with the time-varying portfolio weights of the Fama-French five-factor model plus momentum. We stop at the smallest r such that $H_{0,sp}(r)$ is not rejected at level $\alpha_n = 10/n$. If no such r is found then p-values are displayed up to $k-1$. The red horizontal segments give $\hat{k}-1$, i.e., the estimated number of latent factors obtained from Figure 1 minus 1. We flag bear market phases with grey shaded vertical bars, and use the same rolling windows as in Figure 1.



ONLINE APPENDIX

Latent Factor Analysis in Short Panels

Alain-Philippe Fortin, Patrick Gagliardini, and Olivier Scaillet

We prove Lemmas 1-4 of the paper in Section C. We provide additional theory in Appendix D, namely the characterization of the pseudo likelihood and the PML estimator (Subsection D.1), the conditions for global identification and consistency (D.2), the local analysis of the first-order conditions of FA estimators (D.3), the asymptotic normality of FA estimators (D.4), the definition of invariant tests (D.5), and proofs of additional lemmas (D.6). Finally, we give numerical checks of Inequalities (12) of Proposition 7 in Appendix E.

C Proofs of Lemmas 1-4

Proof of Lemma 1: We have $\mathscr{Z}_n = \frac{1}{\sqrt{n}}(W\Sigma W' - Tr(\Sigma)I_T)$. Hence, $(\mathscr{Z}_n)_{tt} = \frac{1}{\sqrt{n}}\sum_{i,j}(w_{i,t}w_{j,t} - 1_{\{i=j\}})\sigma_{ij} = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n}\zeta_{m,n}^{tt}$, with $\zeta_{m,n}^{tt} = \sum_{i\in I_m}[w_{i,t}^2 - 1]\sigma_{ii} + 2\sum_{\substack{i,j\in I_m\\ i<j}} w_{i,t}w_{j,t}\sigma_{ij}$, together with $(\mathscr{Z}_n)_{ts} = \frac{1}{\sqrt{n}}\sum_{i,j} w_{i,t}w_{j,s}\sigma_{ij} = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n}\zeta_{m,n}^{ts}, t \neq s$, with $\zeta_{m,n}^{ts} = \sum_{i\in I_m} w_{i,t}w_{i,s}\sigma_{ii} + \sum_{\substack{i,j\in I_m\\ i<j}} w_{i,t}w_{j,s}\sigma_{ij} + \sum_{\substack{i,j\in I_m\\ i>j}} w_{i,t}w_{j,s}\sigma_{ij}, t\neq s$, so that $vech(\mathscr{Z}_n) = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n} vech(\zeta_{m,n})$, where $\zeta_{m,n}$ is the $T\times T$ matrix having element $\zeta_{m,n}^{ts}$ in position (t,s). Hence, $vech(\mathscr{Z}_n)$ is the row sum of a triangular array $\{vech(\zeta_{m,n})\}_{1\leq m\leq n}$ of independent centered random vectors. Let $\Omega_{m,n} := V[vech(\zeta_{m,n})]$. Using Assumption 2 (a), we compute (i) $E[(\zeta_{m,n}^{tt})^2] = \sum_{i\in I_m}(E[w_{i,t}^4] - 1)\sigma_{ii}^2 + 2\sum_{\substack{i,j\in I_m\\ i\neq j}}\sigma_{ij}^2$; (ii) $E[(\zeta_{m,n}^{ts})^2] = \sum_{i\in I_m} E[w_{i,t}^2 w_{i,s}^2]\sigma_{ii}^2 + \sum_{\substack{i,j\in I_m\\ i\neq j}}\sigma_{ij}^2, t\neq s$; (iii) $E[\zeta_{m,n}^{tt}\zeta_{m,n}^{ss}] = \sum_{i\in I_m} E[w_{i,t}^2 w_{i,s}^2 - 1]\sigma_{ii}^2, t\neq s$; (iv) $E[\zeta_{m,n}^{tt}\zeta_{m,n}^{rp}] = \sum_{i\in I_m} E[w_{i,t}^2 w_{i,r}w_{ip}]\sigma_{ii}^2, r\neq p$; (v) $E[\zeta_{m,n}^{ts}\zeta_{m,n}^{rp}] = \sum_{i\in I_m} E[w_{i,t}w_{i,s}w_{i,r}w_{i,p}]\sigma_{ii}^2, t\neq s, r\neq p$. It follows that $V[vech(\mathscr{Z}_n)] = \frac{1}{n}\sum_{m=1}^{J_n}\Omega_{m,n} = D_n + \kappa_n I_{\frac{T(T+1)}{2}} = \Omega_n$. The eigenvalues of D_n are bounded away from 0 under Assumption A.4 (b), because for any unit vector $\xi\in\mathbb{R}^{T(T+1)/2}$ we have $\xi' D_n\xi \geq \frac{1}{n}\sum_{i=1}^n 1_{i\in\bar{S}}\sigma_{ii}^2\xi' V[vech(w_i w_i')]\xi \geq \underline{c}\frac{1}{n}\sum_{i=1}^n 1_{i\in\bar{S}}\sigma_{ii}^2 \geq \underline{c}\left(1 - \frac{1}{n}\sum_{i=1}^n (1 - 1_{i\in\bar{S}})\sigma_{ii}\right)^2 \geq \underline{c}\left(1 - \bar{C}\frac{1}{n}\sum_{i=1}^n (1 - 1_{i\in\bar{S}})\right)^2 \geq \frac{\underline{c}}{4}$, for all n.

We use the multivariate Lyapunov condition $\|\Omega_n^{-1/2}\|^4 \frac{1}{n^2}\sum_{m=1}^{J_n} E[\|vech(\zeta_{m,n})\|^4] \to 0$ to invoke a CLT. Since $\|A^{-1/2}\|^4 \leq \frac{k^2}{\delta_k^2(A)}$ and $\|x\|^4 \leq k\sum_{j=1}^k x_j^4$, for any $k \times k$ positive semi-definite matrix A and $k \times 1$ vector x, it suffices to check that $\frac{1}{n^2}\sum_{m=1}^{J_n} E[(\zeta_{m,n}^{ts})^4] \to 0$, for all t, s. Besides, we can show that there exists a constant $M > 0$, such that $E[(\zeta_{m,n}^{ts})^4] \leq M b_{m,n}^{2(1+\delta)}$, for all m, n, t, s. We get $\frac{1}{n^2}\sum_{m=1}^{J_n} E[(\zeta_{m,n}^{ts})^4] \leq M\frac{1}{n^2}\sum_{m=1}^{J_n} b_{m,n}^{2(1+\delta)} = Mn^{2\delta}\sum_{m=1}^{J_n} B_{m,n}^{2(1+\delta)} = o(1)$, under Assumption 2 (d). Then, $\Omega_n^{-1/2} vech(\mathscr{Z}_n) \Rightarrow N(0, I_{\frac{T(T+1)}{2}})$ by the multivariate Lyapunov CLT. Under Assumptions A.4 (a)-(c), $\Omega_n \to \Omega$ follows from the Slutsky theorem, and Ω is positive definite.

Proof of Lemma 2: We use that matrix A_m is such that $A_m' A_m = 2I_{\frac{1}{2}m(m+1)}$, $A_m A_m' = I_{m^2} + K_{m,m}$, and $K_{m,m}A_m = A_m$, where $K_{m,m}$ is the commutation matrix (see also Magnus, Neudecker (2007) Theorem 12 in Chapter 2.8). Then, we have: $vech(Q'\mathscr{Z}_n Q) = \frac{1}{2}A_{T-k}' vec(Q'\mathscr{Z}_n Q) = \frac{1}{2}A_{T-k}'(Q' \otimes Q')vec(\mathscr{Z}_n) = \boldsymbol{R}' vech(\mathscr{Z}_n)$. The columns of matrix $\boldsymbol{R}$ are orthonormal: $\boldsymbol{R}'\boldsymbol{R} = \frac{1}{4}A_{T-k}'(Q' \otimes Q')A_T A_T'(Q \otimes Q)A_{T-k} = \frac{1}{4}A_{T-k}'(Q' \otimes Q')(I_{T^2} + K_{T,T})(Q \otimes Q)A_{T-k} = \frac{1}{4}A_{T-k}'(I_{(T-k)^2} + K_{T-k,T-k})A_{T-k} = \frac{1}{2}A_{T-k}'A_{T-k} = I_p$, since $Q'Q = I_{T-k}$.

Proof of Lemma 3: We have $c_j(\lambda) = \frac{1}{j!}E[Q^j] = \frac{1}{j!}\frac{d^j\Psi(0)}{du^j}$ where $\Psi(u) := E[\exp(uQ)] = \exp[\psi(u)]$ is the Moment Generating Function (MGF) of $Q = \frac{1}{2}\sum_{j=1}^{df}(\sqrt{\nu}_j X_j + \sqrt{1-\nu_j}\lambda_j)^2$ with $X_j \sim i.i.d.N(0,1)$. By the independence of variables X_j, we get $\Psi(u) = \prod_{j=1}^{df} E[\exp(\frac{u}{2}(\sqrt{\nu}_j X_j + \sqrt{1-\nu_j}\lambda_j)^2]$ where $E[\exp(\frac{u}{2}(\sqrt{\nu}_j X_j + \sqrt{1-\nu_j}\lambda_j)^2] = (1-\nu_j u)^{-1/2} e^{\frac{1}{2}\frac{(1-\nu_j)u}{1-\nu_j u}\lambda_j^2}$, for $u < 1/\nu_j$. Thus we get the log MGF $\psi(u) = \frac{1}{2}\sum_{j=1}^{df}\left[-\log(1-\nu_j u) + \frac{(1-\nu_j)u}{1-\nu_j u}\lambda_j^2\right]$, for $u < 1/\nu_{df}$. Its lth order derivative evaluated at $u = 0$ is

$$\psi^{(l)}(0) = \frac{(l-1)!}{2}\sum_{j=1}^{df}\nu_j^{l-1}\left[\nu_j + l(1-\nu_j)\lambda_j^2\right], \quad l \geq 0. \tag{C.1}$$

By using the Faa di Bruno formula for the derivatives of a composite function, we have $\frac{d^l}{du^l}e^{\psi(u)} = e^{\psi(u)}B_l(\psi'(u), \psi''(u), ..., \psi^{(l)}(u))$, where B_l is the lth complete exponential Bell's polynomial (Bell (1934)). Hence, $\Psi^{(l)}(0) = B_l(\psi'(0), \psi''(0), ..., \psi^{(l)}(0))$. The complete Bell's polynomials satisfy the recurrence relation $B_{l+1}(x_1, x_2, ..., x_{l+1}) = \sum_{i=0}^{l}\binom{l}{i}B_{l-i}(x_1, ..., x_{l-i})x_{i+1}$. Thus, $\Psi^{(l+1)}(0) = \sum_{i=0}^{l}\binom{l}{i}\Psi^{(l-i)}(0)\psi^{(i+1)}(0)$. After standardization with the factorial term, and using equation (C.1), the conclusion follows.

Proof of Lemma 4: The proof is in four steps. (i) We first show that (c_i) is increasing, i.e., $G_i^c := c_{i+1} - c_i \geq 0$ for all i. For this purpose, from the recursive relation defining c_{i+1} we have:

$$\begin{aligned} c_{i+1} &= \frac{1}{i}\left(a_1(c_{i-1} + G_{i-1}^c) + a_2(c_{i-2} + G_{i-2}^c) + \cdots + a_{i-1}(c_1 + G_1^c) + a_i\right) \\ &= \frac{1}{i}\left((a_1 - 1)G_{i-1}^c + (a_2 - 1)G_{i-2}^c + \cdots + (a_{i-1} - 1)G_1^c + (a_i - 1)\right) \\ &\quad + \frac{1}{i}\left(G_{i-1}^c + G_{i-2}^c + \cdots + G_1^c + 1\right) + \frac{1}{i}\left(a_1 c_{i-1} + a_2 c_{i-2} + \cdots + a_{i-1}\right). \end{aligned}$$

The second term in the RHS is equal to $\frac{1}{i}c_i$. Using $a_1 c_{i-1} + a_2 c_{i-2} + \cdots + a_{i-1} = (i-1)c_i$, the third term in the RHS is equal to $\frac{i-1}{i}c_i$. Thus, by bringing these two terms in the LHS, we get $G_i^c = \frac{1}{i}\left((a_1 - 1)G_{i-1}^c + (a_2 - 1)G_{i-2}^c + \cdots + (a_{i-1} - 1)G_1^c + (a_i - 1)\right)$, for all $i \geq 2$, with $G_1^c = a_1 - 1$. Because $a_i \geq 1$ for all i, by an induction argument we get $G_i^c \geq 0$ for all $i \geq 1$.

(ii) We now strengthen the result in step (i) and show that $H_i^c := c_{i+1} - c_i \frac{\zeta + i - 1}{i} \geq 0$ for all i, with $\zeta = \max\{\frac{df-1}{2}, 1\}$. Similarly as in step (i), we have

$$\begin{aligned} c_{i+1} &= \frac{1}{i}\left((a_1 - \zeta)G_{i-1}^c + (a_2 - \zeta)G_{i-2}^c + \cdots + (a_{i-1} - \zeta)G_1^c + (a_i - \zeta)\right) \\ &\quad + \frac{\zeta}{i}\left(G_{i-1}^c + G_{i-2}^c + \cdots + G_1^c + 1\right) + \frac{1}{i}\left(a_1 c_{i-1} + a_2 c_{i-2} + \cdots + a_{i-1}\right), \end{aligned}$$

where the second term in the RHS equals $\frac{\zeta}{i}c_i$, and the third term equals $\frac{i-1}{i}c_i$. Thus, we get $H_i^c = \frac{1}{i}\left((a_1 - \zeta)G_{i-1}^c + (a_2 - \zeta)G_{i-2}^c + \cdots + (a_{i-1} - \zeta)G_1^c + (a_i - \zeta)\right)$, for all i. By step (i), we have $G_i^c \geq 0$ for $i \geq 1$. Using the separation condition $a_i \geq \zeta$ for all i, we get $H_i^c \geq 0$ for all i.

(iii) We show that $H_i^g := g_{i+1} - g_i \frac{\zeta + i - 1}{i} \leq 0$ for all $i \geq 1$. For $df = 2$ this statement follows with $\zeta = 1$ because $g_{i+1} = \frac{1}{2i}(g_i + g_{i-1} + ... + 1) = \frac{2i-1}{2i} g_i$ and hence (g_i) is decreasing. Let us now consider the case $df \geq 3$ with $\zeta = \frac{df-1}{2}$. As above we have $H_i^g = \frac{1}{i}\sum_{l=1}^{i}(b_l - \zeta)G_{i-l}^g$, where $G_i^g := g_{i+1} - g_i$. We plug in $b_l - \zeta = \frac{1}{2}\sum_{j=2}^{df-1}(\rho_j^l - 1) = \frac{1}{2}\sum_{j=2}^{df-1}(\rho_j - 1)(1 + \rho_j + ... + \rho_j^{l-1}) = \frac{1}{2}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{k=1}^{l}\rho_j^{k-1}$. Thus, we get:

$$\begin{aligned} H_i^g &= \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{l=1}^{i}\sum_{k=1}^{l}\rho_j^{k-1}G_{i-l}^g = \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{k=1}^{i}\rho_j^{k-1}\sum_{l=k}^{i}G_{i-l}^g \\ &= \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\sum_{k=1}^{i}\rho_j^{k-1}g_{i-k+1} = \frac{1}{2i}\sum_{j=2}^{df-1}(\rho_j - 1)\left(g_i + \rho_j g_{i-1} + ... + \rho_j^{i-1}\right) \leq 0. \end{aligned}$$

(iv) The inequalities established in steps (ii) and (iii) imply $\frac{c_{i+1}}{c_i} \geq \frac{\zeta+i-1}{i}$ and $\frac{g_{i+1}}{g_i} \leq \frac{\zeta+i-1}{i}$ for all i. Then, we get $\frac{c_{i+1}}{c_i} \geq \frac{g_{i+1}}{g_i}$, that is equivalent to $\frac{c_{i+1}}{g_{i+1}} \geq \frac{c_i}{g_i}$, for all i, because the sequences c_i and g_i are strictly positive. The conclusion follows.

D Additional theory

D.1 Pseudo likelihood and PML estimator

The FA estimator is the PML estimator based on the Gaussian likelihood function obtained from the pseudo model $y_i = \mu + F\beta_i + \varepsilon_i$ with $\beta_i \sim N(0, I_k)$ and $\varepsilon_i \sim N(0, V_\varepsilon)$ mutually independent and i.i.d. across $i = 1, ..., n$. Then, $y_i \sim N(\mu, \Sigma(\theta))$ under this pseudo model, where $\Sigma(\theta) := FF' + V_\varepsilon$ and $\theta := (vec(F)', diag(V_\varepsilon)')' \in \mathbb{R}^r$ with $r = (k+1)T$. It yields the pseudo log-likelihood function $\hat{L}(\theta, \mu) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2n}\sum_{i=1}^n (y_i - \mu)'\Sigma(\theta)^{-1}(y_i - \mu) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(\hat{V}_y \Sigma(\theta)^{-1}\right) - \frac{1}{2}(\bar{y} - \mu)'\Sigma(\theta)^{-1}(\bar{y} - \mu)$, up to constants, where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ and $\hat{V}_y = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})'$. We concentrate out parameter μ to get its estimator $\hat{\mu} = \bar{y}$. Then, estimator $\hat{\theta} = (vec(\hat{F})', diag(\hat{V}_\varepsilon)')'$ is defined by the maximization of

$$\hat{L}(\theta) := -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(\hat{V}_y \Sigma(\theta)^{-1}\right), \tag{D.1}$$

subject to the normalization restriction that $F'V_\varepsilon^{-1}F$ is a diagonal matrix, with diagonal elements ranked in decreasing order. [30]

D.2 Global identification and consistency

The population criterium $L_0(\theta)$ is defined in Appendix A, with $V_y = V_y^0 = \Sigma(\theta_0) = F_0F_0' + V_\varepsilon^0$.

[30] If the risk-free rate vector is considered observable, we can rewrite the model as $\tilde{y}_i = F\tilde{\beta}_i + \varepsilon_i = \mu + F\beta_i + \varepsilon_i$, where $\tilde{y}_i = y_i - r_f$ is the vector of excess returns and $\mu = F\mu_{\tilde{\beta}}$. It corresponds to a constrained model with parameters θ and $\mu_{\tilde{\beta}}$. The maximization of the corresponding Gaussian pseudo likelihood function leads to a constrained FA estimator, that we do not consider in this paper since it does not match a standard FA formulation.

Lemma 5 *The following conditions are equivalent: a) the true value θ_0 is the unique maximizer of $L_0(\theta)$ for $\theta \in \Theta$; b) $\Sigma(\theta) = \Sigma(\theta_0)$, $\theta \in \Theta \Rightarrow \theta = \theta_0$, up to sign changes in the columns of F. They yield the global identification in the FA model.*

In Lemma 5, condition a) is the standard identification condition for a M-estimator with population criterion $L_0(\theta)$. Condition (b) is the global identification condition based on the variance matrix as in Anderson and Rubin (1956). Condition (b) corresponds to our Assumption A.3.

Let us now establish the consistency of the FA estimators in our setting. Write $\hat{V}_y = \frac{1}{n}\sum_{i=1}^n(\varepsilon_i - \bar{\varepsilon})(\varepsilon_i - \bar{\varepsilon})' + F[\frac{1}{n}\sum_{i=1}^n(\beta_i - \bar{\beta})(\beta_i - \bar{\beta})']F' + F[\frac{1}{n}\sum_{i=1}^n(\beta_i - \bar{\beta})(\varepsilon_i - \bar{\varepsilon})'] + [\frac{1}{n}\sum_{i=1}^n(\varepsilon_i - \bar{\varepsilon})(\beta_i - \bar{\beta})']F'$, where $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^n \varepsilon_i$ and $\bar{\beta} = \frac{1}{n}\sum_{i=1}^n \beta_i$. Under the normalization in Assumption A.1 we have:

$$\hat{V}_y = \frac{1}{n}\varepsilon\varepsilon' - \bar{\varepsilon}\bar{\varepsilon}' + FF' + F\left(\frac{1}{n}\varepsilon\beta\right)' + \left(\frac{1}{n}\varepsilon\beta\right)F'. \tag{D.2}$$

Lemma 6 *Under Assumptions 1, 2, and A.1, A.2, as $n \to \infty$, we have: (a) $\bar{\varepsilon} = o_p(\frac{1}{n^{1/4}})$, (b) $\frac{1}{n}\varepsilon\varepsilon' \xrightarrow{p} V_\varepsilon^0$, and (c) $\frac{1}{n}\varepsilon\beta \xrightarrow{p} 0$.*

From Equation (D.2) and Lemma 6, we have $\hat{V}_y \xrightarrow{p} V_y^0$. Thus, $\hat{L}(\theta)$ converges in probability to $L_0(\theta)$ as $n \to \infty$, uniformly over Θ compact. From standard results on M-estimators, we get consistency of $\hat{\theta}$. Moreover, from $\bar{y} = \mu + \bar{\varepsilon}$, we get the consistency of $\hat{\mu}$.

Proposition 8 *Under Assumptions 1, 2, and A.1-A.3, the FA estimators $\hat{F}$, $\hat{V}_\varepsilon$ and $\hat{\mu}$ are consistent as $n \to \infty$ and T is fixed.*

Anderson and Rubin (1956) establish consistency in Theorem 12.1 (see beginning of the proof, page 145) within a Gaussian ML framework. Anderson and Amemiya (1988) provide a version of this result in their Theorem 1 for generic distribution of the data, dispensing for compacity of the parameter set but using a more restrictive identification condition.

D.3 Local analysis of the first-order conditions of FA estimators

Consider the criterion $L(\theta) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Tr\left(V_y\Sigma(\theta)\right)$, where V_y is a p.d. matrix in a neighbourhood of V_y^0. In our Assumptions, θ_0 is an interior point of Θ. Let $\theta^* = (vec(F^*)', diag(V_\varepsilon^*)')'$ denote the maximizer of $L(\theta)$ subject to $\theta \in \Theta$. According to Anderson (2003), the first-order conditions (FOC) for the maximization of $L(\theta)$ are: (a) $diag(V_y) = diag(F^*(F^*)' + V_\varepsilon^*)$ and (b) F^* is the matrix of eigenvectors of $V_y(V_\varepsilon^*)^{-1}$ associated to the k largest eigenvalues $1+\gamma_j^*$ for $j = 1, ..., k$, normalized such that $(F^*)'(V_\varepsilon^*)^{-1}F^* = diag(\gamma_1^*, ..., \gamma_k^*)$.

i) Local identification

Let $V_y = V_y^0$. The true values F_0 and V_ε^0 solve the FOC. Let $F = F_0 + \epsilon\Psi_F^\epsilon$ and $V_\varepsilon = V_\varepsilon^0 + \epsilon\Psi_{V_\varepsilon}^\epsilon$, where ϵ is a small scalar and $\Psi_F^\epsilon, \Psi_{V_\varepsilon}^\epsilon$ are deterministic conformable matrices, be in a neighbourhood of F_0 and V_ε^0 and solve the FOC up to terms $O(\epsilon^2)$. The model is locally identified if, and only if, it implies $\Psi_{V_\varepsilon}^\epsilon = 0$ and $\Psi_F^\epsilon = 0$.

Lemma 7 *Under Assumption 1, the following four conditions are equivalent: (a) Matrix $M_{F_0,V_\varepsilon^0}^{\odot 2}$ is non-singular, (b) Matrix $\boldsymbol{X}$ is full-rank, (c) Matrix $\Phi^{\odot 2}$ is non-singular, where $\Phi := V_\varepsilon^0 - F_0(F_0'(V_\varepsilon^0)^{-1}F_0)^{-1}F_0'$, (d) Matrix $B_0'J_0B_0$ is non-singular, where $J_0 := -\frac{\partial^2 L_0(\theta_0)}{\partial\theta\partial\theta'}$ and B_0 is any full-rank $r \times (r - \frac{1}{2}k(k-1))$ matrix such that $\frac{\partial g(\theta_0)}{\partial\theta'}B_0 = 0$, for $g(\theta) = \{[F'V_\varepsilon^{-1}F]_{i,j}\}_{i<j}$ the $\frac{1}{2}k(k-1)$ dimensional vector of the constraints. They yield the local identification of our model.*

In Lemma 7, condition (a) corresponds to Assumption 4. Condition (c) is used in Theorem 5.9 of Anderson and Rubin (1956) to show local identification. Condition (d) involves the second-order partial derivatives of the population criterion function. While the Hessian matrix J_0 itself is singular because of the rotational invariance of the model to latent factors, the second-order partial derivatives matrix along parameter directions, which are in the tangent plan to the contraint set, is non-singular. Condition (d) is equivalent to invertibility of the bordered Hessian.

ii) Local misspecification

Now let $V_y = V_y^0 + \epsilon\Psi_y^\epsilon$ be in a neighbourhood of V_y^0. Let $F^* = F_0 + \epsilon\Psi_F^\epsilon + O(\epsilon^2)$ and $V_\varepsilon^* = V_\varepsilon^0 + \epsilon\Psi_{V_\varepsilon}^\epsilon + O(\epsilon^2)$ be the solutions of the FOC. Consider $V_y - \Sigma^*$, where $\Sigma^* = F^*(F^*)' + V_\varepsilon^*$, i.e., the difference between variance V_y and its k-factor approximation with population FA. We want to find the first-order development of $V_y - \Sigma^*$ for small ϵ. From the FOC, we have that the diagonal of such symmetric matrix is null, but not necessarily the out-of-diagonal elements.

From the arguments in the proof of Proposition 2, Equations (B.4) and (B.5), we get:

$$\Psi_F^\epsilon F_0' + F_0(\Psi_F^\epsilon)' = \Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon - M_{F_0,V_\varepsilon^0}(\Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon)M_{F_0,V_\varepsilon^0}', \tag{D.3}$$

$$diag(M_{F_0,V_\varepsilon^0}(\Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon)M_{F_0,V_\varepsilon^0}') = 0. \tag{D.4}$$

Similarly as above, Equation (D.4) yields $diag(\Psi_{V_\varepsilon}^\epsilon) = (V_\varepsilon^0)^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} diag(M_{F_0,V_\varepsilon^0}\Psi_y^\epsilon M_{F_0,V_\varepsilon^0}')$. Moreover, $diag(M_{F_0,V_\varepsilon^0}\Psi_y^\epsilon M_{F_0,V_\varepsilon^0}') = \sqrt{2}\boldsymbol{X}'vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$. Thus, we have:

$$diag(\Psi_{V_\varepsilon}^\epsilon) = \sqrt{2}(V_\varepsilon^0)^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right). \tag{D.5}$$

Now, using Equation (D.3), we get $V_y - \Sigma^* = \epsilon\left(\Psi_y^\epsilon - F_0(\Psi_F^\epsilon)' - \Psi_F^\epsilon F_0' - \Psi_{V_\varepsilon}^\epsilon\right) + O(\epsilon^2)$ $= \epsilon M_{F_0,V_\varepsilon^0}(\Psi_y^\epsilon - \Psi_{V_\varepsilon}^\epsilon)M_{F_0,V_\varepsilon^0}' + O(\epsilon^2) = \epsilon G_0\Delta^*G_0' + O(\epsilon^2)$, where $\Delta^* := G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0 - G_0'(V_\varepsilon^0)^{-1}\Psi_{V_\varepsilon}^\epsilon(V_\varepsilon^0)^{-1}G_0$. Using that $vech(G_0'diag(a)G_0) = \frac{1}{\sqrt{2}}\boldsymbol{X}a$, and Equation (D.5), the vectorized form of matrix Δ^* is: $vech(\Delta^*) = vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right) - \frac{1}{\sqrt{2}}\boldsymbol{X}(V_\varepsilon^0)^{-2}diag(\Psi_{V_\varepsilon}^\epsilon) = \left(I_p - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$. Thus, we have shown that, at first order in ϵ, the difference between $V_y = V_y^0 + \epsilon\Psi_y^\epsilon$ and the FA k-factor approximation Σ^* is $\epsilon G_0\Delta^*G_0'$, with $vech(\Delta^*) = \left(I_p - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$. It shows that the small perturbation $\epsilon\Psi_y^\epsilon$ around V_y^0 keeps the DGP within the k-factor specification (at first order) if, and only if, vector $vech\left(G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0\right)$ is spanned by the columns of $\boldsymbol{X}$.

Consider $\Psi_y^\epsilon = H\xi\xi'H'$, where $H := [F_0 : G_0]$ and vector $\xi = (\xi_F', \xi_G')'$ are partitioned in k and $T-k$ dimensional components, which corresponds to a local alternative with $(k+1)$th factor $H\xi$ and small loading ϵ in the perturbation $\epsilon\Psi_y^\epsilon$. Then, we have $G_0'(V_\varepsilon^0)^{-1}\Psi_y^\epsilon(V_\varepsilon^0)^{-1}G_0 = \xi_G\xi_G'$

since $F_0'(V_\varepsilon^0)^{-1}G_0 = 0$ and $G_0'(V_\varepsilon^0)^{-1}G_0 = I_{T-k}$. Thus, $vech(\Delta^*) = \Big(I_p - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Big)$ $vech\left(\xi_G\xi_G'\right)$. Hence, it is only the component of $vech\left(\xi_G\xi_G'\right)$ that is orthogonal to the range of $\boldsymbol{X}$, which generates a local deviation from a k-factor specification through the multiplication by the projection matrix $I_p - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$. It clarifies the role of the projector in the local power. On the contrary, the component spanned by the columns of $\boldsymbol{X}$ can be "absorbed" in the k-factor specification by a redefinition of the factor F and the variance V_ε through F^* and V_ε^*.

D.4 Feasible asymptotic normality of the FA estimators

D.4.1 Asymptotic expansions

We first establish the asymptotic expansion of $\hat{\theta}$ along the lines of pseudo maximum likelihood estimators (White (1982)). The sample criterion is $\hat{L}(\theta)$ given in Equation (D.1), where $\theta = (vec(F)', diag(V_\varepsilon)')'$ is subject to the nonlinear vector constraint $g(\theta) := \{[F'V_\varepsilon^{-1}F]_{i,j}\}_{i<j} = 0$, i.e., matrix $F'V_\varepsilon^{-1}F$ is diagonal. By standard methods for constrained M-estimators, we consider the FOC of the Lagrangian function: $\frac{\partial \hat{L}(\hat{\theta})}{\partial\theta} - \frac{\partial g(\hat{\theta})'}{\partial\theta}\hat{\lambda}_L = 0$ and $g(\hat{\theta}) = 0$, where $\hat{\lambda}_L$ is the $\frac{1}{2}k(k-1)$ dimensional vector of estimated Lagrange multipliers. Define vector $\tilde{\theta} := \Big(vec(F_0)', diag(\tilde{V}_\varepsilon)'\Big)'$, which also satisfies the constraint $g(\tilde{\theta}) = 0$ by the in-sample factor normalization. We apply the mean value theorem to the FOC around $\tilde{\theta}$ and get:

$$\hat{J}(\bar{\theta})\sqrt{n}(\hat{\theta} - \tilde{\theta}) + A(\hat{\theta})\sqrt{n}\hat{\lambda}_L = \sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial\theta}, \tag{D.6}$$

$$A(\bar{\theta})'\sqrt{n}(\hat{\theta} - \tilde{\theta}) = 0, \tag{D.7}$$

where $\hat{J}(\theta) := -\frac{\partial^2 \hat{L}(\theta)}{\partial\theta\partial\theta'}$ is the $r \times r$ Hessian matrix, $A(\theta) := \frac{\partial g(\theta)'}{\partial\theta}$ is the $r \times \frac{1}{2}k(k-1)$ dimensional gradient matrix of the constraint function, and $\bar{\theta}$ is a mean value vector between $\hat{\theta}$ and $\tilde{\theta}$ componentwise. Matrix $A(\theta)$ is full rank for θ in a neighbourhood of θ_0. For any θ define the $r \times (r - \frac{1}{2}k(k-1))$ matrix $B(\theta)$ with orthonormal columns that span the orthogonal complement of the range of $A(\theta)$. Matrix function $B(\theta)$ is continuous in θ in a neighbourhood of θ_0.

[31] Then, by multiplying Equation (D.6) times $B(\hat{\theta})'$ to get rid of the Lagrange multiplier vector, using the identity $I_r = A(\theta)(A(\theta)'A(\theta))^{-1}A(\theta)' + B(\theta)B(\theta)'$ for $\theta = \bar{\theta}$ and Equation (D.7), we get $[B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})]B(\bar{\theta})'\sqrt{n}(\hat{\theta} - \tilde{\theta}) = B(\hat{\theta})'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta}$. By the uniform convergence of $\hat{J}(\theta)$ to $J(\theta) := -\frac{\partial^2 L_0(\theta)}{\partial\theta\partial\theta'}$, and the consistency of the FA estimator $\hat{\theta}$ (Section D.2), matrix $B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})$ converges to $B_0'J_0B_0$, where $J_0 := J(\theta_0)$ and $B_0 := B(\theta_0)$. Matrix $B_0'J_0B_0$ is invertible under the local identification Assumption 4 (see Lemma 7 condition d)). Then, $B(\bar{\theta})'\sqrt{n}(\hat{\theta} - \tilde{\theta}) = [B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})]^{-1}B(\hat{\theta})'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta}$ w.p.a. 1. By using again $I_r = A(\bar{\theta})(A(\bar{\theta})'A(\bar{\theta}))^{-1}A(\bar{\theta})' + B(\bar{\theta})B(\bar{\theta})'$ and Equation (D.7), we get $\sqrt{n}(\hat{\theta} - \tilde{\theta}) = B(\bar{\theta})[B(\hat{\theta})'\hat{J}(\bar{\theta})B(\bar{\theta})]^{-1}B(\hat{\theta})'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta}$. The distributional results established below imply $\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} = O_p(1)$. Thus, we get $\sqrt{n}$-consistency:

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) = B_0(B_0'J_0B_0)^{-1}B_0'\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} + o_p(1). \tag{D.8}$$

Let us now find the score $\frac{\partial \hat{L}(\theta)}{\partial \theta}$. We have $\frac{\partial \hat{L}(\theta)}{\partial \theta} = \left(\frac{\partial vec(\Sigma(\theta))}{\partial \theta'}\right)' vec\left(\frac{\partial \hat{L}(\theta)}{\partial \Sigma}\right)$, where $vec\left(\frac{\partial \hat{L}(\theta)}{\partial \Sigma}\right) = \frac{1}{2}\left(\Sigma(\theta)^{-1} \otimes \Sigma(\theta)^{-1}\right) vec\left(\hat{V}_y - \Sigma(\theta)\right)$. Moreover, by using $vec(\Sigma(\theta)) = \sum_{j=1}^k F_j \otimes F_j + [e_1 \otimes e_1 \ : \ \cdots \ : \ e_T \otimes e_T]diag(V_\varepsilon)$, where e_t is the t-th column of I_T, we get: $\frac{\partial vec(\Sigma(\theta))}{\partial \theta'} = [(I_T \otimes F_1) + (F_1 \otimes I_T) \ : \ \cdots \ : \ (I_T \otimes F_k) + (F_k \otimes I_T) \ : \ e_1 \otimes e_1 \ : \ \cdots \ : \ e_T \otimes e_T]$. Thus, we get: $\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} = \frac{1}{2}\left(\frac{\partial vec(\Sigma(\tilde{\theta}))}{\partial \theta'}\right)'\left(\tilde{V}_y^{-1} \otimes \tilde{V}_y^{-1}\right)\sqrt{n}vec\left(\hat{V}_y - \tilde{V}_y\right)$. From Equation (D.2) and Lemma 6 we have $\hat{V}_y = \tilde{V}_y + \frac{1}{\sqrt{n}}(Z_n + W_nF' + FW_n') + o_p(\frac{1}{\sqrt{n}})$, where $W_n := \frac{1}{\sqrt{n}}\varepsilon\beta$. Thus, $\sqrt{n}\frac{\partial \hat{L}(\tilde{\theta})}{\partial \theta} = \frac{1}{2}\left(\frac{\partial vec(\Sigma(\theta_0))}{\partial \theta'}\right)'\left(V_y^{-1} \otimes V_y^{-1}\right) vec\left(W_nF' + FW_n' + Z_n\right) + o_p(1)$ and, from Equation (D.8), we get:

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) = \frac{1}{2}B_0\left(B_0'J_0B_0\right)^{-1}B_0'\left(\frac{\partial vec(\Sigma(\theta_0))}{\partial \theta'}\right)'\left(V_y^{-1} \otimes V_y^{-1}\right) vec\left(W_nF' + FW_n' + Z_n\right) + o_p(1). \tag{D.9}$$

D.4.2 Asymptotic normality

In this subsection, we establish the asymptotic normality of estimators $\hat{F}$ and $\hat{V}_\varepsilon$. From Lemma 1, as $n \to \infty$ and T is fixed, we have the Gaussian distributional limit $Z_n \Rightarrow Z$ with $vech(Z) \sim$

[31] Matrix $B(\theta)$ is uniquely defined up to rotation and sign changes in their columns. We can pick a unique representer such that matrix $B(\theta)$ is locally continuous, e.g., by taking $B(\theta) = \tilde{B}(\theta)[\tilde{B}(\theta)'\tilde{B}(\theta)]^{-1/2}$, where matrix $\tilde{B}(\theta)$ consists of the first $r - \frac{1}{2}k(k-1)$ columns of $I_r - A(\theta)[A(\theta)'A(\theta)]^{-1}A(\theta)'$, if those columns are linearly independent.

$N(0, \Omega_Z)$, where the asymptotic variance Ω_Z is related to the asymptotic variance Ω of $\mathscr{Z}$ such that $Cov(Z_{ts}, Z_{rp}) = \sqrt{V_{\varepsilon,tt}V_{\varepsilon,ss}V_{\varepsilon,rr}V_{\varepsilon,pp}}Cov(\mathscr{Z}_{ts}, \mathscr{Z}_{rp})$. Moreover, $Z_n^* \Rightarrow Z^* = G'V_\varepsilon^{-1}ZV_\varepsilon^{-1}G$ and $\bar{Z}_n \Rightarrow \bar{Z}$, where $\bar{Z} = Z - \mathcal{T}_{F,V_\varepsilon}(Z) = Z - V_\varepsilon^2 diag\left((\boldsymbol{X}'\boldsymbol{X})^{-1}diag(GZ^*G')\right) = Z - \sqrt{2}V_\varepsilon^2 diag\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z^*)\right)$. The distributional limit of W_n is given next.

Lemma 8 *Under Assumptions 1, 2 and A.1, A.2, A.6, as $n \to \infty$, (a) we have $W_n \Rightarrow \bar{W}$, where $vec(\bar{W}) \sim N(0, \Omega_W)$ with $\Omega_W = Q_\beta \otimes V_\varepsilon$, and (b) if additionally $E[w_{i,t}w_{i,r}w_{i,s}] = 0$, for all t, r, s and i, then Z and $\bar{W}$ are independent.*

We get the following proposition from Lemmas 1 and 8 (see proof at the end of the section).

Proposition 9 *Under Assumptions 1-4 and A.1-A.4, A.6, as $n \to \infty$ and T is fixed, for $j = 1, .., k$:*

$$\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) \Rightarrow \sqrt{2}V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z^*), \tag{D.10}$$

$$\sqrt{n}(\hat{F}_j - F_j) \Rightarrow R_j(\bar{W}F' + F\bar{W}' + \bar{Z})V_\varepsilon^{-1}F_j + \sqrt{2}\Lambda_j\{[V_\varepsilon(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z^*)] \odot F_j\}, \tag{D.11}$$

$$\sqrt{n}(\hat{F}_j\hat{\mathcal{D}} - F_j) \Rightarrow \frac{1}{\gamma_j}(\bar{W}F' + F\bar{W}' + \bar{Z})V_\varepsilon^{-1}F_j, \tag{D.12}$$

where deterministic matrices R_j and Λ_j are defined in Proposition 2, and $\hat{\mathcal{D}} := \hat{\Gamma}(F'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}$ and $\hat{\Gamma} := diag(\hat{\gamma}_1, ..., \hat{\gamma}_k)$.

The joint asymptotic Gaussian distribution of the FA estimators involves the Gaussian matrices Z^*, $\bar{Z}$ and $\bar{W}$, the former two being symmetric. The asymptotic distribution of $\hat{V}_\varepsilon$ involves recentering around $\tilde{V}_\varepsilon = \frac{1}{n}\sum_{i=1}^n E[\varepsilon_i\varepsilon_i']$, i.e., the finite-sample average cross-moments of errors, and not V_ε. For the asymptotic distribution of any functional that depends on F up to one-to-one transformations of its columns, we can use the Gaussian law of (D.12) involving $\bar{W}$ and $\bar{Z}$ only.

The asymptotic expansions (D.10)-(D.11) characterize explicitly the matrices $C_1(\theta)$ and $C_2(\theta)$ that appear in Theorem 2 in Anderson and Amemiya (1988). Their derivation is based on an asymptotic normality argument treating $\hat{\theta}$ as a M-estimator, see Section C.2. However, neither the asymptotic variance nor a feasible CLT are given in Anderson and Amemiya (1988).

To further compare our Proposition 9 with Theorem 2 in Anderson and Amemiya (1988), let $\bar{Z} = Z - \mathcal{T}_{F,V_\varepsilon}(Z) = \check{Z} - \mathcal{T}_{F,V_\varepsilon}(\check{Z})$, where $\check{Z} := Z - diag(Z)$ is the symmetric matrix of the off-diagonal elements of Z with zeros on the diagonal. [32] Hence, the zero-mean Gaussian matrix $\bar{Z}$ only involves the off-diagonal elements of Z. Moreover, since $\sqrt{2}V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(\Delta_n^*) = V_\varepsilon^2 diag(V_\varepsilon^{-1}\Delta_n V_\varepsilon^{-1}) = diag(\Delta_n)$ for a diagonal matrix Δ_n and $\Delta_n^* := G'V_\varepsilon^{-1}\Delta_n V_\varepsilon^{-1}G$, we can write the asymptotic expansion of $\hat{V}_\varepsilon$ as $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) = \sqrt{2}V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(\check{Z}_n^*) + diag(Z_n) + o_p(1)$, where $\check{Z}_n^* = G'V_\varepsilon^{-1}\check{Z}_n V_\varepsilon^{-1}G$ and $\check{Z}_n := Z_n - diag(Z_n)$. Thus, we get: $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) \Rightarrow \sqrt{2}V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(\check{Z}^*) + diag(Z)$, where $\check{Z}^* = G'V_\varepsilon^{-1}\check{Z}V_\varepsilon^{-1}G$. Hence, the asymptotic distribution of the FA estimators depends on the diagonal elements of Z via term $diag(Z)$ in the asymptotic distribution of $\hat{V}_\varepsilon$. In Theorem 2 in Anderson and Amemiya (1988), this term does not appear because in their results the asymptotic distribution of $\hat{V}_\varepsilon$ is centered around $diag(\frac{1}{n}\varepsilon\varepsilon')$ instead of $\tilde{V}_\varepsilon$. Our recentering around $\tilde{V}_\varepsilon$ avoids a random bias term.

Finally, by applying the CLT to (D.9), the asymptotic distribution of vector $\hat{\theta}$ is:

$$\sqrt{n}(\hat{\theta} - \tilde{\theta}) \Rightarrow \frac{1}{2}B_0\left(B_0'J_0B_0\right)^{-1}B_0'\left(\frac{\partial vec(\Sigma(\theta_0))}{\partial\theta'}\right)'\left(V_y^{-1}\otimes V_y^{-1}\right)vec\left(\bar{W}F' + F\bar{W}' + Z\right). \quad \text{(D.13)}$$

The Gaussian asymptotic distribution in (D.13) matches those in (D.10) and (D.11) written for the components, and its asymptotic variance yields the 'sandwich formula". The result in (D.13) is analogue to Theorem 2 in Anderson and Amemiya (1988), for different factor normalization and recentering of the variance estimator.

Proof of Proposition 9: From Proposition 3 (b) and Section D.4.1, we have the asymptotic expansion: $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon) = diag(\Psi_\varepsilon) + o_p(1) = \sqrt{2}V_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z_n^*) + o_p(1)$. Moreover, from Proposition 2 (a) and using $\Psi_y - \Psi_\varepsilon = W_nF' + FW_n' + \bar{Z}_n$, we have: $\sqrt{n}(\hat{F}_j - F_j) = R_j(\Psi_y - \Psi_\varepsilon)V_\varepsilon^{-1}F_j + \Lambda_j\Psi_\varepsilon V_\varepsilon^{-1}F_j + o_p(1) = R_j(W_nF' + FW_n' + \bar{Z}_n)V_\varepsilon^{-1}F_j + \Lambda_j[diag(\Psi_\varepsilon)\odot(V_\varepsilon^{-1}F_j)] + o_p(1) = R_j(W_nF' + FW_n' + \bar{Z}_n)V_\varepsilon^{-1}F_j + \sqrt{2}\Lambda_j\{[V_\varepsilon(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'vech(Z_n^*)]\odot F_j\} + o_p(1)$. Lemmas 1 and 8 yield (D.10)-(D.11), together with (D.12) from (B.1) since $\Psi_y - \Psi_\varepsilon \Rightarrow \bar{W}F' + F\bar{W}' + \bar{Z}$.

[32] Here, $diag(Z)$ is the diagonal matrix with the same diagonal elements as Z.

D.4.3 Feasible CLT for the FA estimators

i) Feasible CLT for Z_n via a parametric estimator of the asymptotic variance

We first show that, under strengthening of Assumption 2, we get a parametric structure for the variance $V[vech(Z)] = \Omega_Z(V_\varepsilon, \vartheta)$ with a vector of unknown parameters ϑ of dimension $T+1$.

Assumption 5 *The standardized errors processes $w_{i,t}$ in Assumption 2 are (a) stationary martingale difference sequences (mds), and (b) $E[w_{i,t}^2 w_{i,r} w_{i,s}] = 0$, for $t > r > s$.*

Assumption 5 holds e.g. for conditionally homoschedastic mds, and for ARCH processes (see below). Let $\mathscr{Z} := V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}$. Then, using Lemma 1, under Assumptions 2 and 5, we have $V[\mathscr{Z}_{t,t}] = \psi(0) + 2\kappa$, $V[\mathscr{Z}_{t,s}] = \psi(t-s) + q + \kappa$ and $Cov(\mathscr{Z}_{t,t}, \mathscr{Z}_{s,s}) = \psi(t-s)$, where $\psi(t-s) := \lim_{n\to\infty} \frac{1}{n} \sum_i Cov(w_{i,t}^2, w_{i,s}^2)\sigma_{ii}^2$. Quantity $\psi(t-s)$ depends on the difference $t-s$ only, by stationarity. The other covariance terms between elements of $\mathscr{Z}$ vanish. Then, we have $\Omega = [\psi(0) - 2q]D(0) + \sum_{h=1}^{T-1} \psi(h)D(h) + (q+\kappa)I_{T(T+1)/2}$, where $D(0) = \sum_{t=1}^T vech(E_{t,t})vech(E_{t,t})'$ and $D(h) = \tilde{D}(h) + \bar{D}(h)$ with $\tilde{D}(h) = \sum_{t=1}^{T-h}[vech(E_{t,t})vech(E_{t+h,t+h})' + vech(E_{t+h,t+h})vech(E_{t,t})']$ and $\bar{D}(h) = \sum_{t=1}^{T-h} vech(E_{t,t+h} + E_{t+h,t})vech(E_{t,t+h} + E_{t+h,t})'$ for $h = 1, ..., T-1$, and where $E_{t,s}$ denote the $T \times T$ matrix with entry 1 in position (t,s) and 0 elsewhere. Hence, with $Z = V_\varepsilon^{1/2} \mathscr{Z} V_\varepsilon^{1/2}$, we get a parametrization $\Omega_Z(V_\varepsilon, \vartheta)$ for $V[vech(Z)]$ with $\vartheta = (q+\kappa, \psi(0) - 2q, \psi(1), ..., \psi(T-1))'$.

Now, from Proposition 3 (c) and Lemma 2, we obtain a parametric structure for $V[vech(\bar{Z}^*)]$.

Lemma 9 *Under Assumptions 1-5 and A.1-A.4, we have:*

$$\Omega_{\bar{Z}^*} = \sum_{h=1}^{T-1} [\psi(h) + q + \kappa](I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{R}'\bar{D}(h)\boldsymbol{R}(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'). \tag{D.14}$$

Hence, the parametric structure $\Omega_{\bar{Z}^*}(V_\varepsilon, G, \tilde{\vartheta})$ depends linearly on vector $\tilde{\vartheta}$ that stacks the $T-1$ parameters $\psi(h) + q + \kappa$, for $h = 1, ..., T-1$. It does not involve parameter $\psi(0)$, i.e., the quartic moment of errors, because the asymptotic expansion of the LR statistic does not involve the diagonal terms of Z. Moreover, the unknown parameters appear through the linear combinations $\psi(h) + q + \kappa$ that are the scaled variances of the out-of-diagonal elements of Z. We can estimate the unknown parameters in $\tilde{\vartheta}$ by least squares applied on (D.14), using the nonparametric estimator $\hat{\Omega}_{\bar{Z}^*}$ defined in Proposition 4, after half-vectorization and replacing V_ε and G by their FA estimates. It yields a consistent estimator of $\Omega_{\bar{Z}^*}$ incorporating the restrictions implied by Assumption 5.

To get a feasible CLT for the FA estimates, we need to estimate the additional parameters $\psi(0) - 2q$ and $q + \kappa$. We consider matrix $\hat{\tilde{\Xi}} = \frac{1}{n}\sum_{m=1}^{J_n} vech(\hat{h}_{m,n})vech(\hat{h}_{m,n})'$, where $\hat{h}_{m,n} = \sum_{i\in I_m} \hat{G}'\hat{V}_\varepsilon^{-1}\hat{\varepsilon}_i\hat{\varepsilon}_i'\hat{V}_\varepsilon^{-1}\hat{G}$, that involves fourth-order moments of residuals. Note that $\hat{\Omega}_{\bar{Z}^*} = M_{\hat{\boldsymbol{X}}}\hat{\tilde{\Xi}}M_{\hat{\boldsymbol{X}}}$, where $M_{\hat{\boldsymbol{X}}} = I_p - \hat{\boldsymbol{X}}(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'$.

Lemma 10 *Under Assumptions 1-5 and A.1-A.4, and $\sqrt{n}\sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$, up to pre- and post-multiplication by an orthogonal matrix and its transpose, we have $\hat{\tilde{\Xi}} = \boldsymbol{R}'\tilde{\Xi}_n\boldsymbol{R} + o_p(1)$, where $\tilde{\Xi}_n = [\psi_n(0) - 2q_n]D(0) + \sum_{h=1}^{T-1}\psi_n(h)D(h) + (q_n + \kappa_n)I_{T(T+1)/2} + (q_n + \xi_n)vech(I_T)vech(I_T)'$ and $\xi_n := \dfrac{1}{n}\sum_{m=1}^{J_n}\sum_{i\neq j\in I_m}\sigma_{ii}\sigma_{jj}$.*

With blocks of equal size, the condition $\sqrt{n}\sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$ holds if $J_n = n^{\bar{\alpha}}$ and $\bar{\alpha} > 1/2$. Now, we have the relation $3D(0) + \sum_{h=1}^{T-1} D(h) - vech(I_T)vech(I_T)' = I_{T(T+1)/2}$, which implies $3\boldsymbol{R}'D(0)\boldsymbol{R} + \sum_{h=1}^{T-1}\boldsymbol{R}'D(h)\boldsymbol{R} - vech(I_{T-k})vech(I_{T-k})' = I_p$. Hence, matrix

$$
\begin{aligned}
\boldsymbol{R}'\tilde{\Xi}_n\boldsymbol{R} &= [\psi_n(0) + q_n + 3\kappa_n]\boldsymbol{R}'D(0)\boldsymbol{R} + \sum_{h=1}^{T-1}[\psi_n(h) + q_n + \kappa_n]\boldsymbol{R}'D(h)\boldsymbol{R} \\
&\quad + (\xi_n - \kappa_n)vech(I_{T-k})vech(I_{T-k})' \qquad \text{(D.15)}
\end{aligned}
$$

depends on $T + 1$ linear combinations of the elements of $\vartheta_n = (q_n + \kappa_n, \psi_n(0) - 2q_n, \psi_n(1), ..., \psi_n(T-1))'$ and $\xi_n - \kappa_n$. Thus, the linear system (D.15) is rank-deficient to identify ϑ_n. Moreover, in Assumption A.4 (b), κ_n is defined as a double sum over squared covariances scaled by n, and is assumed to converge to a constant κ. Such a convergence is difficult to assume for ξ_n since ξ_n is a double sum over products of two variances scaled by n.

We apply half-vectorization on (D.15), replace the LHS by its consistent estimate $\hat{\tilde{\Xi}}$, and plug-in the FA estimates in the RHS. From Lemma 10, least squares estimation on such a linear regression yields consistent estimates of linear combinations $\psi(0) + q + 3\kappa$ and $\psi(h) + q + \kappa$ for

$h = 1, \ldots, T-1$. Consistency of those parameters applies independently of $\xi_n - \kappa_n$ converging as $n \to \infty$, or not. [33] In order to identify the components of ϑ, we need an additional condition. We use the assumption $\psi(T-1) = 0$. That condition is implied by serial uncorrelation in the squared standardized errors after lag $T-1$, that is empirically relevant in our application with monthly returns data. Then, parameter $q + \kappa$ is estimated by $\widehat{\psi_n(T-1) + q_n} + \kappa_n$, and by difference we get the estimators of $\psi(0) - 2q$ and $\psi(h)$, for $h = 1, ..., T-2$.

Let us now discuss the case of ARCH errors. Suppose the $w_{i,t}$ follow independent ARCH(1) processes with Gaussian innovations that are independent across assets, i.e., $w_{i,t} = h_{i,t}^{1/2} z_{i,t},\ z_{i,t} \sim IIN(0,1),\ h_{i,t} = c_i + \alpha_i w_{i,t-1}^2$ with $c_i = 1 - \alpha_i$. Then $E[w_{i,t}] = 0$, $E[w_{i,t}^2] = 1$, $\eta_i := V[w_{i,t}^2] = \frac{2}{1-3\alpha_i^2}$, $Cov(w_{i,t}^2, w_{i,t-h}^2) = \eta_i \alpha_i^h$. Moreover, $E[w_{i,t} w_{i,r} w_{i,s} w_{i,p}] = 0$ if one index among t, r, s, p is different from all the others. Indeed, without loss of generality, suppose t is different from s, p, r. By the law of iterated expectation: $E[\varepsilon_{i,t}\varepsilon_{i,s}\varepsilon_{i,p}\varepsilon_{i,r}] = E[E[\varepsilon_{i,t}|\{z_{i,\tau}^2\}_{\tau=-\infty}^{\infty}, \{z_{i,\tau}\}_{\tau \neq t}]\varepsilon_{i,s}\varepsilon_{i,p}\varepsilon_{i,r}] = E[h_{i,t}^{1/2} E[z_{i,t}|z_{i,t}^2]\varepsilon_{i,s}\varepsilon_{i,p}\varepsilon_{i,r}] = 0$. Then, Assumption 5 holds. The explicit formula of Ω involves $\psi(h) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \frac{2\alpha_i^h}{1-3\alpha_i^2}\sigma_{ii}^2$, for $h = 0, 1, ..., T-1$. Hence, setting $\psi(T-1) = 0$ is a mild assumption for identification purpose since α_i^{T-1} is small. If $\alpha_i = 0$ for all i, i.e., no ARCH effects, we have $\psi(0) = 2q$ and $\psi(h) = 0$ for $h > 0$, so that $\Omega = (q+\kappa) I_{\frac{T(T+1)}{2}}$.

ii) Feasible CLT for W_n

Let us now establish a feasible CLT for W_n. In order to estimate matrix Q_β in the asymptotic variance Ω_W in Lemma 8, we use the estimated betas and residuals, and combine them with a temporal sample splitting approach to cope with the EIV problem caused by the fixed T setting. Specifically, let us split the time spell into two consecutive sub-intervals with T_1 and T_2 observations, with $T_1 + T_2 = T$ and such that $T_1 > k$ and $T_2 \geq k$. The factor model in the two sub-intervals reads $y_{1,i} = \mu_1 + F_1\beta_i + \varepsilon_{1,i}$ and $y_{2,i} = \mu_2 + F_2\beta_i + \varepsilon_{2,i}$, and let $V_{1,\varepsilon}$ and $V_{2,\varepsilon}$ denote

[33] To see this, write the half-vectorization of the RHS of (D.15) as $\chi\eta_n$, where χ is the $\frac{p(p+1)}{2} \times (T+1)$ matrix of regressors and η_n the $(T+1)\times 1$ vector of unknown parameters. Then, $vech(\hat{\Xi}) = \hat{\chi}\eta_n + o_p(1)$, by Lemma 10, the consistency of the FA estimates, and the last column of χ not depending on unknown parameters. Thus, $\hat{\eta}_n := (\hat{\chi}'\hat{\chi})^{-1}\hat{\chi}' vech(\hat{\Xi}) = \eta_n + o_p(1)$. In particular, we also have $\widehat{\xi_n - \kappa_n} = \xi_n - \kappa_n + o_p(1)$.

the corresponding diagonal matrices of error average unconditional variances. [34] The conditions $T_1 > k$ and $T_2 \geq k$ are needed because we estimate residuals and betas in the first and the second sub-intervals, namely $\hat{\varepsilon}_{1,i} = M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}(y_{1,i} - \bar{y}_1)$ and $\hat{\beta}_i = (\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}(y_{2,i} - \bar{y}_2)$. Here, $\hat{F}_j$ and $\hat{V}_{j,\varepsilon}$ for $j = 1, 2$ are deduced from the FA estimates in the full period of T observations. Define $\hat{\Psi}_\beta = \frac{1}{n}\sum_m \sum_{i,j\in I_m}(\hat{\beta}_i\hat{\beta}_j') \otimes (\hat{\varepsilon}_{1,i}\hat{\varepsilon}_{1,j}')$. By using $\hat{\varepsilon}_{1,i} = (M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}F_1)\beta_i + M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}(\varepsilon_{1,i} - \bar{\varepsilon}_1)$, $M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}F_1 = O_p(\frac{1}{\sqrt{n}})$ and $\frac{1}{n^2}\sum_m b_{m,n}^2 = \sum_m B_{m,n}^2 = o(1)$, we get $\hat{\Psi}_\beta = (I_k \otimes M_{\hat{F}_1,\hat{V}_{1,\varepsilon}})\left(\frac{1}{n}\sum_m\sum_{i,j\in I_m}(\hat{\beta}_i\hat{\beta}_j') \otimes [(\varepsilon_{1,i} - \bar{\varepsilon}_1)(\varepsilon_{1,j} - \bar{\varepsilon}_1)']\right)(I_k \otimes M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}') + o_p(1)$. Now, we use $\hat{\beta}_i = \left[(\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}F_2\right]\beta_i + (\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}(\varepsilon_{2,i} - \bar{\varepsilon}_2)$, and $\bar{\varepsilon}_1 = o_p(n^{-1/4})$, $\bar{\varepsilon}_2 = o_p(n^{-1/4})$ from Lemma 6 (a), as well as the the mds condition in Assumption 5. We get $\hat{\Psi}_\beta = \hat{\Psi}_{\beta,1} + \hat{\Psi}_{\beta,2} + o_p(1)$, where $\hat{\Psi}_{\beta,1} = (I_k \otimes M_{F_1,V_{1,\varepsilon}})\left(\frac{1}{n}\sum_m\sum_{i,j\in I_m}(\beta_i\beta_j') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}')\right)(I_k \otimes M_{F_1,V_{1,\varepsilon}}')$ and $\hat{\Psi}_{\beta,2} = \left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)\left(\frac{1}{n}\sum_m\sum_{i,j\in I_m}(\varepsilon_{2,i}\varepsilon_{2,j}') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}')\right)$ $\left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)'$. We use $\frac{1}{n}\sum_m\sum_{i,j\in I_m}(\beta_i\beta_j') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}') = Q_\beta \otimes V_{1,\varepsilon} + o_p(1)$, and $\frac{1}{n}\sum_m\sum_{i,j\in I_m}(\varepsilon_{2,i}\varepsilon_{2,j}') \otimes (\varepsilon_{1,i}\varepsilon_{1,j}') = \Omega_{21} + o_p(1)$, where Ω_{21} is the sub-block of matrix Ω_Z that is the asymptotic variance of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_{2,i} \otimes \varepsilon_{1,i} \Rightarrow N(0, \Omega_{21})$. Then, $\hat{\Psi}_\beta = Q_\beta \otimes (M_{F_1,V_{1,\varepsilon}}V_{1,\varepsilon}) + \left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)\Omega_{21}\left([(F_2'V_{2,\varepsilon}^{-1}F_2)^{-1}F_2'V_{2,\varepsilon}^{-1}] \otimes M_{F_1,V_{1,\varepsilon}}\right)' + o_p(1)$. Thus, we get a consistent estimator of $Q_\beta \otimes (V_{1,\varepsilon}^{-1/2}M_{F_1,V_{1,\varepsilon}}V_{1,\varepsilon}^{1/2})$ by subtracting to $\hat{\Psi}_\beta$ a consistent estimator of the second term on the RHS, and then by pre- and post-multiplying times $(I_k \otimes \hat{V}_{1,\varepsilon}^{-1/2})$. To get a consistent estimator of Q_β we apply a linear transformation that amounts to computing the trace of the second term of a Kronecker product, and divide by $Tr(V_{1,\varepsilon}^{-1/2}M_{F_1,V_{1,\varepsilon}}V_{1,\varepsilon}^{1/2}) = T_1 - k$. Thus: $\hat{Q}_\beta = \frac{1}{n(T_1-k)}\sum_m\sum_{i,j\in I_m}(\hat{\beta}_i\hat{\beta}_j')(\hat{\varepsilon}_{1,j}'\hat{V}_{1,\varepsilon}^{-1}\hat{\varepsilon}_{1,i}) - \frac{1}{T_1-k}\sum_{j=1}^{T_1}(I_k \otimes e_j')\left\{\left([(\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}] \otimes [\hat{V}_{1,\varepsilon}^{-1/2}M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}]\right)\hat{\Omega}_{21}\left([\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2(\hat{F}_2'\hat{V}_{2,\varepsilon}^{-1}\hat{F}_2)^{-1}] \otimes [M_{\hat{F}_1,\hat{V}_{1,\varepsilon}}'\hat{V}_{1,\varepsilon}^{-1/2}]\right)\right\}(I_k \otimes e_j)$, where the e_j are T_1-dimensional unit vectors, and $\hat{\Omega}_{21}$ is obtained from Subsection D.4.3 i). If estimate $\hat{Q}_\beta$ is not positive definite, we regularize it by deleting the negative eigenvalues.

iii) Joint feasible CLT

[34] We can take the two sub-intervals as the halves of the time span. If this choice does not meet conditions $T_1 > k$ and $T_2 \geq k$ in a subperiod, we take the second sub-interval such that $T_2 = k$, and add to the first sub-interval a sufficient number of dates from the preceeding subperiod in order to get $T_1 = k + 1$.

To get a feasible CLT for the FA estimators from (D.10)-(D.11), we need the joint distribution of the Gaussian matrix variates Z and W. Under the condition of Lemma 8 (b), the estimates of the asymptotic variances of $vech(Z)$ and $vec(W)$ are enough, since these vectors are independent. Otherwise, to estimate the covariance $Cov(vech(Z), vec(W))$, we need to extend the approaches of the previous subsections.

D.4.4 Special cases

In this subsection, we particularize the asymptotic distributions of the FA estimators for three special cases along the lines of Section 4, plus a fourth special case that allows us to further discuss the link with Anderson and Amemiya (1988).

i) Gaussian errors

When the errors admit a Gaussian distribution $\varepsilon_i \overset{ind}{\sim} N(0, \sigma_{ii} V_\varepsilon)$ with diagonal V_ε, matrix $\frac{1}{\sqrt{q}} V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}$ is in the GOE for dimension T, i.e., $\frac{1}{\sqrt{q}} vech(V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}) \sim N(0, I_{T(T+1)/2})$, where $q = \lim_{n\to\infty} \frac{1}{n} \sum_i \sigma_{ii}^2$. Moreover, $vec(W) \sim N(0, Q_\beta \otimes V_\varepsilon)$, where $Q_\beta = \lim_{n\to\infty} \frac{1}{n} \sum_i \sigma_{ii} \beta_i \beta_i'$, mutually independent of Z because of the symmetry of the Gaussian distribution.

ii) Quasi GOE errors

As an extension of the previous case, here let us suppose that the errors meet Assumption 2, the Conditions (a) and (b) in Proposition 5 plus additionally (c) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n V(\varepsilon_{i,t}^2) = \eta V_{\varepsilon,tt}^2$, for a constant $\eta > 0$, and (d) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n E[\varepsilon_{i,t}^2 \varepsilon_{i,r} \varepsilon_{i,p}] = 0$ for $r \neq p$. This setting allows e.g. for conditionally homoschedastic mds processes in the errors, but excludes ARCH effects. Then, the arguments in Lemma 1 imply $vech(V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}) \sim N(0, \Omega)$ with $\Omega = \begin{pmatrix} (\eta/2 + \kappa) I_T & 0 \\ 0 & (q + \kappa) I_{\frac{1}{2}T(T-1)} \end{pmatrix}$. The distribution of $V_\varepsilon^{-1/2} Z V_\varepsilon^{-1/2}$ is similar to (scaled) GOE holding in the Gaussian case up to the variances of diagonal and of out-of-diagonal elements being different when $\eta \neq 2q$. Hence, contrasting with test statistics, the asymptotic distributions of FA estimates differ in cases i) and ii) beyond scaling factors. It is because the asymptotic distributions of FA estimates involve diagonal elements of Z as well.

iii) Spherical errors

Let us consider the case $\varepsilon_i \overset{ind}{\sim} (0, \sigma_{ii} V_\varepsilon)$ where $V_\varepsilon = \bar{\sigma}^2 I_T$, with independent components across time and the normalization $\lim\limits_{n\to\infty} \frac{1}{n}\sum_i \sigma_{ii} = 1$. From Sections 4.3 and D.4.1, we have the asymptotic expansions of the FA estimators $\sqrt{n}(\hat{\sigma}^2 - \tilde{\sigma}^2) = \frac{1}{T-k} Tr(M_F Z_n) + o_p(1) = \frac{\bar{\sigma}^2}{T-k} Tr(Z_n^*) + o_p(1)$, and $\sqrt{n}(\hat{F}_j - F_j) = \frac{1}{\bar{\sigma}^2} R_j(\Psi_y - \Psi_\varepsilon)F_j - \frac{1}{\bar{\sigma}^2}\Lambda_j \Psi_\varepsilon F_j + o_p(1) = \frac{1}{\bar{\sigma}^2} R_j(W_n F' + F W_n' + \bar{Z}_n)F_j + o_p(1)$, where we use $\Psi_y - \Psi_\varepsilon = W_n F' + F W_n' + \bar{Z}_n$, $\Psi_\varepsilon = \frac{1}{T-k} Tr(M_F Z_n) I_T$ and $\Lambda_j F_j = 0$, and $\bar{Z}_n = Z_n - \frac{1}{T-k} Tr(M_F Z_n) I_T$. Moreover, by sphericity, we have $R_j = \frac{1}{2\gamma_j} P_{F_j} + \frac{1}{\gamma_j} M_F + \sum_{\ell=1,\ell\neq j}^k \frac{1}{\gamma_j - \gamma_\ell} P_{F_\ell}$. Thus, we get $\sqrt{n}(\hat{\sigma}^2 - \tilde{\sigma}^2) \Rightarrow \frac{\bar{\sigma}^2}{T-k} Tr(Z^*)$ and $\sqrt{n}(\hat{F}_j - F_j) \Rightarrow \frac{1}{\bar{\sigma}^2} R_j(WF' + FW' + \bar{Z})F_j$. [35] The Gaussian matrix Z is such that $Z_{tt} \sim N(0, \eta)$ and $Z_{t,s} \sim N(0, q)$ for $t \neq s$, mutually independent, where $\eta = \lim\limits_{n\to\infty} \frac{1}{n}\sum_i V[\varepsilon_{i,t}^2]$, and $vec(W) \sim N(0, Q_\beta \otimes I_T)$. Variables Z and W are independent if $E[\varepsilon_{i,t}^3] = 0$. FGS (2022), Section 4.3.1, explain how we can estimate q and η by solving a system of two linear equations based on estimated moments of $\hat{\varepsilon}_{i,t}$.

iv) Cross-sectionally homoschedastic errors and link with Anderson and Amemiya (1988)

Let us now make the link with the distributional results in Anderson and Amemiya (1988). In our setting, the analogous conditions as those in their Corollary 2 would be: (a) random effects for the loadings that are i.i.d. with $E[\beta_i] = 0$, $V[\beta_i] = I_k$, (b) error terms are i.i.d. $\varepsilon_i \sim (0, V_\varepsilon)$ with $V_\varepsilon = diag(V_{\varepsilon,11}, ..., V_{\varepsilon,TT})$ such that $E[\varepsilon_{i,t}\varepsilon_{i,r}\varepsilon_{i,s}\varepsilon_{i,p}] = V_{\varepsilon,tt}V_{\varepsilon,ss}$, for $t = r > s = p$, and $= 0$, otherwise, and (c) β_i and ε_i are mutually independent. Thus, $\sigma_{ii} = 1$ for all i, i.e., errors are cross-sectionally homoschedastic. Under the aforementioned Conditions (a)-(c), the Gaussian distributional limits Z and W are such that $V[Z_{tt}] = \eta_t V_{\varepsilon,tt}^2$, for $\eta_t := V[\varepsilon_{i,t}^2]/V_{\varepsilon,tt}^2$, $V[Z_{ts}] = V_{\varepsilon,tt}V_{\varepsilon,ss}$, for $t \neq s$, all covariances among different elements of Z vanish, and $V[vec(W)] = I_k \otimes V_\varepsilon$. Equations (D.10)-(D.11) yield the asymptotic distributions of the FA estimates. In particular, they do not depend on the distribution of the β_i. Moreover, the distribution of the out-of-diagonal elements of Z does not depend on the distribution of the errors, while, for the diagonal term, we

[35] The asymptotic distribution of estimator $\hat{\sigma}^2$ coincides with that derived in FGS (2022) with perturbation theory methods. The asymptotic distribution of the factor estimates slightly differs from that given in FGS (2022), Section 5.1, because of the different factor normalization adopted by FA compared to PCA even under sphericity.

have $\eta_t = 2$ for Gaussian errors. As remarked in Section D.4.2, if the asymptotic distribution of estimator $\hat{V}_\varepsilon$ is centered around the realized matrix $\frac{1}{n}\sum_i \varepsilon_i\varepsilon_i'$ instead of its expected value, that distribution involves the out-of-diagonal elements of Z, and the elements of W. Hence, in that case, the asymptotic distribution of the FA estimates is the same independent of the errors being Gaussian or not, and depends on F and V_ε only, as found in Anderson and Amemiya (1988).

D.5 Orthogonal transformations and maximal invariant statistic

In this subsection, we consider the transformation $\mathcal{O}$ that maps matrix $\hat{G}$ into $\hat{G}O$, where O is an orthogonal matrix in $\mathbb{R}^{(T-k)\times(T-k)}$, and the transformation $\mathcal{O}_D$ that maps matrix $\boldsymbol{D}$ into $\boldsymbol{D}O_D$, where O_D is an orthogonal matrix in $\mathbb{R}^{df\times df}$. These transformations are induced from the freedom in chosing the orthonormal bases spanning the orthogonal complements of $\hat{F}$ and $\boldsymbol{X}$. We show that they imply a group of orthogonal transformations on the vector $\hat{W} = \sqrt{n}\boldsymbol{D}'vech(\hat{S}^*)$, with $\hat{S}^* = \hat{G}'\hat{V}_\varepsilon^{-1}(\hat{V}_y - \hat{V}_\varepsilon)\hat{V}_\varepsilon^{-1}\hat{G}$, and establish the maximal invariant.

Under the transformation $\mathcal{O}$, matrix $\hat{S}^*$ is mapped into $O^{-1}\hat{S}^*O$. This transformation is mirrored by a linear mapping at the level of the half-vectorized form $vech(\hat{S}^*)$. In fact, this mapping is norm-preserving, because $\|vech(S)\|^2 = \frac{1}{2}\|S\|^2$ and $\|O^{-1}SO\| = \|S\|$ for any conformable symmetric matrix S and orthogonal matrix O. This mapping is characterized in the next lemma.

Lemma 11 *For any symmetric matrix S and orthogonal matrix O in $\mathbb{R}^{m\times m}$, we have $vech(O^{-1}SO) = \mathscr{R}(O)vech(S)$, where $\mathscr{R}(O) = \frac{1}{2}A_m'(O'\otimes O')A_m$ is an orthogonal matrix, and A_m is the duplication matrix defined in Appendix B. Transformations $\mathscr{R}(O)$ with orthogonal O have the structure of a group: (a) $\mathscr{R}(I_m) = I_{\frac{1}{2}m(m+1)}$, (b) $\mathscr{R}(O_1)\mathscr{R}(O_2) = \mathscr{R}(O_2O_1)$, and (c) $[\mathscr{R}(O)]^{-1} = \mathscr{R}(O^{-1})$.*

With this lemma, we can give the transformation rules under $\mathcal{O}$ for a set of relevant statistics in the next proposition. We denote generically with $\widetilde{\cdot}$ a quantity computed with $\hat{G}O$ instead of $\hat{G}$.

Proposition 10 *Under Assumptions 1 and 4, (a) $vech(\widetilde{\hat{S}^*}) = \mathscr{R}(O)vech(\hat{S}^*)$, (b) $\widetilde{\boldsymbol{X}} = \mathscr{R}(O)\boldsymbol{X}$, (c) $I_p - \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}' = \mathscr{R}(O)[I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\mathscr{R}(O)^{-1}$, (d) $\widetilde{\boldsymbol{R}} = \boldsymbol{R}\mathscr{R}(O)^{-1}$,* *(e)* $\widetilde{\boldsymbol{R}}(I_p - \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}') = \boldsymbol{R}(I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\mathscr{R}(O)^{-1}$.

From Proposition 10 (c), under transformation $\mathcal{O}$, matrix $\boldsymbol{D}$ is mapped into $\mathscr{R}(O)\boldsymbol{D}$. Combining with transformation $\mathcal{O}_D$, we have $\widetilde{\boldsymbol{D}} = \mathscr{R}(O)\boldsymbol{D}O_D$. Thus, using Proposition 10 (a), under $\mathcal{O}$ and $\mathcal{O}_D$, vector $\hat{W}$ is mapped into $\widetilde{\hat{W}} = \sqrt{n}\widetilde{\boldsymbol{D}}'vech(\widetilde{\hat{S}^*}) = O_D'\hat{W}_D$. Thus, statistic $\hat{W}$ is invariant under $\mathcal{O}$, while $\mathcal{O}_D$ operates as the group of orthogonal transformations. The maximal invariant under this group of transformations is the squared norm $\|\hat{W}\|^2 = \hat{W}'\hat{W}$.

Proof of Proposition 10: With $\widetilde{\hat{S}^*} = O^{-1}\hat{S}^*O$, part (a) follows from Lemma 11. Let $\tilde{G} = GO$. Then, for any diagonal matrix Δ, on the one hand, we have $vech(\tilde{G}'\Delta\tilde{G}) = \frac{1}{\sqrt{2}}\tilde{\boldsymbol{X}}diag(\Delta)$, and on the other hand, we have $vech(\tilde{G}'\Delta\tilde{G}) = vech(O^{-1}G'\Delta GO) = \mathscr{R}(O)vech(G'\Delta G) = \frac{1}{\sqrt{2}}\mathscr{R}(O)\boldsymbol{X}diag(\Delta)$. By equating the two expressions for any diagonal matrix Δ, part (b) follows. Statement (c) is a consequence thereof and $\mathscr{R}(O)$ being orthogonal. Moreover, with $\tilde{Q} = QO$ and using $vech(\tilde{Q}'Z\tilde{Q}) = vech(O^{-1}Q'ZQO) = \mathscr{R}(O)\boldsymbol{R}'vech(Z)$, we deduce part (d). Statement (e) is a consequence of (c) and (d).

D.6 Proofs of Lemmas 5-11

Proof of Lemma 5: The equivalence of conditions (a) and (b) is a consequence of the fact that function $\mathscr{L}(A) = -\frac{1}{2}\log|A| - \frac{1}{2}Tr(V_y^0A^{-1})$, where A is a p.d. matrix, is uniquely maximized for $A = V_y^0$ (see Magnus and Neudecker (2007), p. 410), and $L_0(\theta) = \mathscr{L}(\Sigma(\theta))$.

Proof of Lemma 6: (a) From Assumption 2 we have $E[\bar{\varepsilon}] = 0$ and $V[\bar{\varepsilon}] = V\left[\frac{1}{n}\sum_{i,k=1}^n s_{i,k}V_\varepsilon^{1/2}w_k\right]$ $= V_\varepsilon^{1/2}\frac{1}{n^2}\sum_{i,j,k,l=1}^n s_{i,k}s_{j,l}E[w_kw_l']V_\varepsilon^{1/2} = (\frac{1}{n^2}\sum_{i,j}^n\sigma_{i,j})V_\varepsilon$ where the $s_{i,k}$ are the elements of $\Sigma^{1/2}$. Now, $\frac{1}{n^2}\sum_{i,j=1}^n\sigma_{i,j} \leq C\frac{1}{n^2}\sum_{m=1}^{J_n}b_{m,n}^{1+\delta} = O(n^{\delta-1}\sum_{m=1}^{J_n}B_{m,n}^{1+\delta}) = O(n^{\delta-1}J_n^{1/2}(\sum_{m=1}^{J_n}B_{m,n}^{2(1+\delta)})^{1/2}) = o(n^{-1}J_n^{1/2}) = o(n^{-1/2})$ from the Cauchy-Schwarz inequality and Assumptions 2 (c) and (d). Part (a) follows. To prove part (b), we use $E[\frac{1}{n}\varepsilon\varepsilon'] \to V_\varepsilon^0$ and $V[vech((V_\varepsilon^0)^{-1/2}(\frac{1}{n}\varepsilon\varepsilon')(V_\varepsilon^0)^{-1/2})] = \frac{1}{n}\Omega_n$ from the proof of Lemma 1, and $\frac{1}{n}\Omega_n = o(1)$ by Assumption A.2. Finally, to show part (c), write $\frac{1}{n}\sum_{i=1}^n\varepsilon_i\beta_i' = (V_\varepsilon^0)^{1/2}\frac{1}{n}\sum_{i,j=1}^n s_{i,j}w_j\beta_i'$. Then, $E[\frac{1}{n}\sum_{i=1}^n\varepsilon_i\beta_i'] = 0$ while the variance of $vec(\frac{1}{n}\sum_{i=1}^n\varepsilon_i\beta_i')$ vanishes asymptotically because $V[vec(\frac{1}{n}\sum_{i,j=1}^n s_{i,j}w_j\beta_i')] = \frac{1}{n^2}\sum_{i,j,m,l=1}^n s_{i,j}s_{m,l}$

$(\beta_i\beta_l') \otimes E[w_j w_m'] = \frac{1}{n^2}\sum_{i,l=1}^n \sigma_{i,l}(\beta_i\beta_l') \otimes I_T = o(1)$ under Assumptions 2 and A.1.

Proof of Lemma 7: From the arguments in the proof of Proposition 2 with $\Psi_y = 0$, the solution of the FOC is such that $\Psi_{F,j}^{\epsilon} = (\Lambda_j^0 - R_j^0)\Psi_{V_\varepsilon}^{\epsilon}(V_\varepsilon^0)^{-1}F_j$ for $j = 1, ..., k$, and $diag(M_{F_0,V_\varepsilon^0}\Psi_{V_\varepsilon}^{\epsilon}M_{F_0,V_\varepsilon^0}') = 0$. Because $\Psi_{V_\varepsilon}^{\epsilon}$ is diagonal, the latter equation yields $M_{F_0,V_\varepsilon^0}^{\odot 2} diag(\Psi_{V_\varepsilon}^{\epsilon}) = 0$. Under condition (a) of Lemma 7, we get $\Psi_{V_\varepsilon}^{\epsilon} = 0$, which in turn implies $\Psi_F^{\epsilon} = 0$. Thus, condition (a) is sufficient for local identification. It is also necessary to get uniqueness of the solution $\Psi_{V_\varepsilon}^{\epsilon} = 0$. Moreover, conditions (a) and (b) of Lemma 7 are equivalent by Proposition 3 a). Further, conditions (a) and (c) are equivalent because $\Phi^{\odot 2} = M_{F_0,V_\varepsilon^0}^{\odot 2}(V_\varepsilon^0)^2$. Finally, let us show that condition (d) of Lemma 7 is both sufficient and necessary for local identification. The FOC for the Lagrangian problem are $\frac{\partial L_0(\theta)}{\partial\theta} - \frac{\partial g(\theta)'}{\partial\theta}\lambda_L = 0$ and $g(\theta) = 0$, where λ_L is the Lagrange multiplier vector. By expanding at first-order around θ_0 and $\lambda_0 = 0$, we get $H_0\begin{pmatrix}\theta - \theta_0 \\ \lambda\end{pmatrix} = 0$, where $H_0 := \begin{pmatrix} J_0 & A_0 \\ A_0' & 0\end{pmatrix}$, with $A_0 = \frac{\partial g(\theta_0)'}{\partial\theta}$, is the bordered Hessian. The parameters are locally identified if, and only if, H_0 is invertible. The latter condition is equivalent to $B_0'J_0B_0$ being invertible. [36]

Proof of Lemma 8: By Assumption 2, $vec(W_n) = (I_k \otimes V_\varepsilon^{1/2})\frac{1}{\sqrt{n}}\sum_{m=1}^{J_n} x_{m,n}$ where the $x_{m,n} := \sum_{i,j\in I_m} s_{i,j}(\beta_i \otimes w_j)$ are independent across m. Now, we apply the Liapunov CLT to show $\frac{1}{\sqrt{n}}\sum_{m=1}^{J_n} x_{m,n} \Rightarrow N(0, Q_\beta \otimes I_T)$. We have $E[x_{m,n}] = 0$ and $E[x_{m,n}x_{m,n}'] = \left(\sum_{i,j\in I_m}\sigma_{i,j}\beta_i\beta_j'\right) \otimes I_T$ and, by Assumption A.6, $\Omega_{W,n} := \frac{1}{n}\sum_{m=1}^{J_n} E[x_{m,n}x_{m,n}']$ converges to the positive definite matrix $Q_\beta \otimes I_T$. Let us now check the multivariate Liapunov condition $\|\Omega_{W,n}^{-1/2}\|^4 \frac{1}{n^2}\sum_{m=1}^{J_n} E[\|x_{m,n}\|^4] = o(1)$. Because $\|\Omega_{W,n}^{-1/2}\| = O_p(1)$, it suffices to prove $\frac{1}{n^2}\sum_{m=1}^{J_n} E[(x_{m,n}^{p,t})^4] = o(1)$, for any $p = 1, ..., k$ and $t = 1, ..., T$, where $x_{m,n}^{p,t} := \sum_{i,j\in I_m} s_{i,j}\beta_{i,p}w_{j,t}$. For this purpose, Assumptions A.1 and A.2 yield $E[(x_{m,n}^{p,t})^4] \leq C(\sum_{i,j\in I_m}\sigma_{i,j})^2$. Then, we get $\frac{1}{n^2}\sum_{m=1}^{J_n} E[(x_{m,n}^{p,t})^4] \leq C\frac{1}{n^2}\sum_{m=1}^{J_n} b_{m,n}^{2(1+\delta)} \leq Cn^{2\delta}\sum_{m=1}^{J_n} B_{m,n}^{2(1+\delta)} = o(1)$ by Assumptions 2 (c) and (d). Part (a) of Lemma 8 follows. Moreover, $E[vech(\zeta_{m,n})x_{m,n}'] = 0$ and the proof of Lemma 1 imply part (b).

Proof of Lemma 9: We have $\Omega_{\bar{Z}^*} = M_{\boldsymbol{X}}\boldsymbol{R}'\Omega\boldsymbol{R}M_{\boldsymbol{X}}$, where $M_{\boldsymbol{X}} = I_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$,

[36] Indeed, we can show $|H_0| = (-1)^{\frac{1}{2}k(k-1)}|A_0'A_0||B_0'J_0B_0|$ by using $J_0A_0 = 0$, where the latter equality follows because the criterion is invariant to rotations of the latent factors.

$\boldsymbol{R} = \frac{1}{2}A_T'(Q \otimes Q)A_{T-k}$, and $\Omega = D + \kappa I_{T(T+1)/2} = [\psi(0) - 2q]D(0) + \sum_{h=1}^{T-1} \psi(h)[\tilde{D}(h) + \bar{D}(h)] + (q+\kappa)I_{T(T+1)/2}$. Then, because the columns of $\boldsymbol{R}$ are orthonormal, we get $\Omega_{\bar{Z}^*} = [\psi(0)-2q]M_{\boldsymbol{X}}\boldsymbol{R}'D(0)\boldsymbol{R}M_{\boldsymbol{X}} + \sum_{h=1}^{T-1}\psi(h)M_{\boldsymbol{X}}\boldsymbol{R}'\tilde{D}(h)\boldsymbol{R}M_{\boldsymbol{X}} + \sum_{h=1}^{T-1}\psi(h)M_{\boldsymbol{X}}\boldsymbol{R}'\bar{D}(h)\boldsymbol{R}M_{\boldsymbol{X}} + (q+\kappa)M_{\boldsymbol{X}}$. Now, we show that the the first two terms in this sum are nil. Indeed, recall that matrix $M_{\boldsymbol{X}}$ is idempotent of rank $p - T = df$, and its kernel coincides with the range of the $p \times T$ matrix $\boldsymbol{X}$. By the definition of the latter matrix in Section 3.2, we can write it as $\boldsymbol{X} = \frac{1}{\sqrt{2}}[vech(G'E_{1,1}G) : \cdots : vech(G'E_{T,T}G)]$. Now, we have $G'E_{t,t}G = Q'V_\varepsilon^{1/2}E_{t,t}V_\varepsilon^{1/2}Q = V_{\varepsilon,tt}Q'E_{t,t}Q$ and thus $vech(G'E_{t,t}G) = V_{\varepsilon,tt}vech(Q'E_{t,t}Q) = V_{\varepsilon,tt}\boldsymbol{R}'vech(E_{t,t})$ by Lemma 2. Hence, the kernel of matrix $M_{\boldsymbol{X}}$ is spanned by vectors $\boldsymbol{R}'vech(E_{t,t})$, for $t = 1, ..., T$. We deduce that $M_{\boldsymbol{X}}\boldsymbol{R}'D(0) = 0$ and $M_{\boldsymbol{X}}\boldsymbol{R}'\tilde{D}(h)\boldsymbol{R}M_{\boldsymbol{X}} = 0$. Furthermore, from $I_{T(T+1)/2} = 2\sum_{t=1}^{T} vech(E_{t,t})vech(E_{t,t})' + \sum_{t<s} vech(E_{t,s}+E_{s,t})vech(E_{t,s}+E_{s,t})' = 2D(0) + \sum_{h=1}^{T-1}\bar{D}(h)$, we get $M_{\boldsymbol{X}} = M_{\boldsymbol{X}}\boldsymbol{R}'I_{T(T+1)/2}\boldsymbol{R}M_{\boldsymbol{X}} = \sum_{h=1}^{T-1} M_{\boldsymbol{X}}\boldsymbol{R}'\bar{D}(h)\boldsymbol{R}M_{\boldsymbol{X}}$. The conclusion follows.

Proof of Lemma 10: By the root-n consistency of the FA estimators, $\hat{h}_{m,n} = h_{m,n} + O_p(\frac{b_{m,n}}{\sqrt{n}})$, uniformly in m, where $h_{m,n} = \sum_{i \in I_m} G'V_\varepsilon^{-1}\varepsilon_i\varepsilon_i'V_\varepsilon^{-1}G = \sum_{i\in I_m} Q'e_ie_i'Q$. Under the condition $\frac{1}{n^{3/2}}\sum_{m=1}^{J_n} b_{m,n}^2 = \sqrt{n}\sum_{m=1}^{J_n} B_{m,n}^2 = o(1)$, we have $\hat{\tilde{\Xi}} = \frac{1}{n}\sum_{m=1}^{J_n} E[vech(h_{m,n})vech(h_{m,n})'] + o_p(1)$, up to pre- and post-multiplication by an orthogonal matrix. Moreover, $vech(h_{m,n}) = \boldsymbol{R}'[\sum_{i\in I_m} vech(e_ie_i')] = \frac{1}{2}\boldsymbol{R}'A_T'[\sum_{i\in I_m}(e_i \otimes e_i)]$, and $\sum_{i \in I_m}(e_i \otimes e_i) = \sum_{a,b}\sigma_{a,b}(w_a \otimes w_b)$. Thus, we get $E\left[vech(h_{m,n})vech(h_{m,n})'\right] = \frac{1}{4}\boldsymbol{R}'A_T'\left\{\sum_{a,b,c,d\in I_m}\sigma_{a,b}\sigma_{c,d}E[(w_a \otimes w_b)(w_c \otimes w_d)']\right\}A_T\boldsymbol{R}$. The non-zero contributions to the term in the curly brackets come from the combinations with $a = b = c = d$, $a = b \neq c = d$, $a = c \neq b = d$ and $a = d \neq b = c$, yielding: $\sum_{a,b,c,d}\sigma_{a,b}\sigma_{c,d}E[(w_a \otimes w_b)(w_c \otimes w_d)'] = \sum_a \sigma_{a,a}^2 E[(w_aw_a')\otimes(w_aw_a')] + (\sum_{a\neq c}\sigma_{a,a}\sigma_{c,c})vec(I_T)vec(I_T)' + (\sum_{a\neq b}\sigma_{a,b}^2)(I_{T^2} + K_{T,T}) = \sum_a[\sigma_{a,a}^2 V(w_a \otimes w_a)] + (\sum_a \sigma_{a,a})^2 vec(I_T)vec(I_T)' + (\sum_{a\neq b}\sigma_{a,b}^2)(I_{T^2} + K_{T,T})$. Then, using $w_a \otimes w_a = A_T vech(w_aw_a')$, we get $\frac{1}{4}A_T'\left\{\sum_{a,b,c,d\in I_m}\sigma_{a,b}\sigma_{c,d}E[(w_a\otimes w_b)(w_c \otimes w_d)']\right\}A_T = \sum_a[\sigma_{a,a}^2 V(vech(w_aw_a'))] + (\sum_a \sigma_{a,a})^2 vech(I_T)vech(I_T)' + (\sum_{a\neq b}\sigma_{a,b}^2)I_{\frac{T(T+1)}{2}}$. Then, because $\frac{1}{n}\sum_{i=1}^n \sigma_{i,i}^2 V[vech(w_iw_i')] = D_n$, where matrix D_n is defined in Assumption A.4, we get $\hat{\tilde{\Xi}} = \boldsymbol{R}'\tilde{\Xi}_n\boldsymbol{R} + o_p(1)$ where $\tilde{\Xi}_n = D_n + (q_n + \xi_n)vech(I_T)vech(I_T)' + \kappa_n I_{\frac{T(T+1)}{2}}$. Moreover, under Assumption 5, and singling out parameter q_n along the diagonal, we have $D_n = [\psi_n(0) - 2q_n]D(0) +$

$\sum_{h=1}^{T-1} \psi_n(h)[\tilde{D}(h) + \bar{D}(h)] + q_n I_{T(T+1)/2}$. The conclusion follows.

Proof of Lemma 11: We use $vec(S) = A_m vech(S)$, where the $m^2 \times \frac{1}{2}m(m+1)$ matrix A_m is such that: (i) $A_m' A_m = 2I_{\frac{1}{2}m(m+1)}$, (ii) $K_{m,m}A_m = A_m$, where $K_{m,m}$ is the commutation matrix for order m, and (iii) $A_m A_m' = I_{m^2} + K_{m,m}$ (see also Theorem 12 in Magnus, Neudecker (2007) Chapter 2.8). Then, $vech(S) = \frac{1}{2}A_m' vec(S)$ by property (i), and $vech(O^{-1}SO) = \frac{1}{2}A_m' vec(O^{-1}SO) = \frac{1}{2}A_m'(O' \otimes O')vec(S) = \frac{1}{2}A_m'(O' \otimes O')A_m vech(S)$, for all symmetric matrix S. It follows $\mathscr{R}(O) = \frac{1}{2}A_m'(O' \otimes O')A_m$. Moreover, by properties (i)-(iii), we have (a) $\mathscr{R}(I_m) = I_{\frac{1}{2}m(m+1)}$, (b) $\mathscr{R}(O_1)\mathscr{R}(O_2) = \frac{1}{4}A_m'(O_1' \otimes O_1')A_m A_m'(O_2' \otimes O_2')A_m = \frac{1}{4}A_m'(O_1' \otimes O_1')(I_{m^2} + K_{m,m})(O_2' \otimes O_2')A_m = \frac{1}{4}A_m'(O_1'O_2' \otimes O_1'O_2')(I_{m^2} + K_{m,m})A_m = \frac{1}{2}A_m'[(O_2O_1)' \otimes (O_2O_1)']A_m = \mathscr{R}(O_2O_1)$, and thus (c) $[\mathscr{R}(O)]^{-1} = \mathscr{R}(O^{-1})$.

E Numerical checks of conditions (12) of Proposition 7

In this section, we check numerically the validity of Inequalities (12) for given df, λ_j, ν_j, and $m = 3, ..., M$, for a large bound M. The idea is to compute the frequency of the LHS of (12) becoming strictly negative over a large number of potential values of λ_j and ν_j, $j = 1, ..., df$, for any given $df > 1$. [37] Table 1 provides those frequencies for $m = 3, ..., 16$ (cumulatively), with λ_j uniformly drawn in $[\underline{\lambda}, \bar{\lambda}]$ for $j = 1, ..., df$, and $\nu_1 = 0$ [38] and ν_j uniformly drawn in $[0, \bar{\nu}]$, for $j = 2, ..., df$, and different combinations of bounds $\underline{\lambda}$, $\bar{\lambda}$, $\bar{\nu}$, and degrees of freedom $df = 2, ..., 12$. Each frequency is computed from 10^8 draws of λ_j and ν_j, $j = 1, ..., df$. In the SMC, we also report a table of frequencies for large grids of equally-spaced points in $[\underline{\lambda}, \bar{\lambda}]^{df} \times [0, \bar{\nu}]^{df-1}$, which corroborate the findings of this section.

[37] From Footnote 18, we know that Inequalities (12) are automatically met with $df = 1$. A given value of df may result from several different combinations of T and k, while a given T implies different values of df depending on k. For instance, $df = 2$ applies with $(T, k) = (4, 1)$, $(8, 4)$, and $(13, 8)$, among other combinations. For $T = 20$, the tests for $k = 1, 2, ..., 14$ yield $df = 170, 151, 133, 116, 100, 85, 71, 58, 46, 35, 25, 16, 8, 1$, respectively.

[38] This normalization results from ranking the eigenvalues μ_j, so that μ_1 is the smallest one.

E.1 Calibration of $\bar{\nu}$, $\underline{\lambda}$ and $\bar{\lambda}$

To calibrate the bounds $\bar{\nu}$, $\underline{\lambda}$ and $\bar{\lambda}$ with realistic values, we run the following numerical experiment. For $T = 20$ and $k = 7$, we simulate $10,000$ draws from random $T \times k$ matrix $\tilde{F}$ such that $vec(\tilde{F}) \sim N(0, I_{Tk})$ and set $F = V_{\varepsilon}^{1/2} U \Gamma^{1/2}$, $U = \tilde{F}(\tilde{F}'\tilde{F})^{-1/2}$, $G = V_{\varepsilon}^{1/2} Q$, $Q = \tilde{Q}(\tilde{Q}'\tilde{Q})^{-1/2}$, $\tilde{Q}$ are the first $T - k$ columns of $I_T - UU'$, for $V_{\varepsilon} = diag(V_{\varepsilon,11}, ..., V_{\varepsilon,TT})$, with $V_{\varepsilon,tt} = 1.5$ for $t = 1, ..., 10$, and $V_{\varepsilon,tt} = 0.5$ for $t = 11, ..., 20$, and $\Gamma = T diag(4, 3.5, 3, 2.5, 2, 1.5, 1)$, $c_{k+1} = 10T$, and $\xi_{k+1} = e_1$. With these choices, the "signal-to-noise" $\frac{1}{T} F_j' V_{\varepsilon}^{-1} F_j$ for the seven factors $j = 1, ..., 7$ are $4, 3.5, 3, 2.5, 2, 2.5, 1$, and the "signal-to-noise" for the weak factor is $\frac{1}{T} F_{k+1}' V_{\varepsilon}^{-1} F_{k+1} = 10n^{-1/2}$. Moreover, the errors follow the ARCH model of Section D.4.3 (i) with ARCH parameters either (a) $\alpha_i = 0.2$ for all i, or (b) $\alpha_i = 0.5$ for all i, and $q = 4$, and $\kappa = 0$ (cross-sectional independence). The choices $\alpha_i = 0.2$, 0.5 both meet the condition $3\alpha_i^2 < 1$ ensuring the existence of fourth-order moments. Moreover, with $q - 1 = 3$, we have a cross-sectional variance of the σ_{ii} that is three times larger than the mean (normalized to 1). For each draw, we compute the $df = 71$ non-zero eigenvalues and associated eigenvectors of $\Omega_{\bar{Z}^*}$, and the values of parameters ν_j and λ_j. In our simulations (a) with $\alpha_i = 0.2$, the draws of $\max_{j=1,...,df} \nu_j$ range between 0.21 and 0.30, with 95% quantile equal to 0.28, while the 5% and 95% quantiles of the λ_j are 0.13 and 7.65. Instead, (b) with $\alpha_i = 0.5$, the $\max_{j=1,...,df} \nu_j$ range between 0.70 and 0.79, with 95% quantile equal to 0.77, and the 5% and 95% quantiles of the λ_j are 0.12 and 6.64. To get further insights in the choice of parameters $\bar{\nu}$, $\underline{\lambda}$, $\bar{\lambda}$, we also consider the values implied by the FA estimates in our empirical analysis. Here, when testing for the last retained k in a given subperiod, the median across subperiods of $\max_{j=1,...,df} \nu_j$ is 0.76, and smaller than about 0.90 in most subperiods. Similarly, assuming $c_{k+1} = 10T$ and $\xi_{k+1} = e_1$ as above, the median values of the smallest and the largest estimated λ_j are 0.0024 and 5.84. Inspired by these findings, we set $\bar{\lambda} = 7$, and consider $\bar{\nu} = 0.2$, 0.7, 0.9, 0.99, and $\underline{\lambda} = 0$, 0.1, 0.5, 1, to get realistic settings with different degrees of dissimilarity from the case with serially uncorrelated squared errors (increasing with $\bar{\nu}$), and separation of the alternative hypothesis from the null hypothesis (increasing with $\underline{\lambda}$).

E.2 Results with Monte Carlo draws

In Table 1, the entries are nil for $\bar{\nu}$ sufficiently small and $\underline{\lambda}$ sufficiently large, suggesting that the AUMPI property holds for those cases that are closer to the setting with uncorrelated squared errors and sufficiently separated from the null hypothesis. Violations of Inequalities (12) concern $df = 2, 3, 4, 5$. [39] Let us focus on the setting with $\bar{\nu} = 0.7$ and $\underline{\lambda} = 0.1$. We find 3752 violations of Inequalities (12) out of 10^8 simulations, all occurring for $df = 2$, except 65 for $df = 3$. For those draws violating Inequalities (12) for $df = 2$, a closer inspection shows that (a) they feature values ν_2 close to upper bound $\bar{\nu} = 0.7$, and values of λ_2 close to lower bound $\underline{\lambda} = 0.1$, and (b) several of them yield non-monotone density ratios $\frac{f(z;\lambda_1,\lambda_2)}{f(z;0,0)}$, with the non-monotonicity region corresponding to large values of z. As an illustration, let us take the density ratio for $df = 2$ with $\nu_2 = 0.666$, $\lambda_1 = 1.372$, and $\lambda_2 = 0.130$. Here, the eigenvalues of the variance-covariance matrix are $\mu_1 = 1$ (by normalization) and $\mu_2 = (1 - \nu_2)^{-1} = 2.994$, and the non-centrality parameter λ_2 is small. The quantiles of the asymptotic distribution under the null hypothesis for asymptotic size $\alpha = 20\%, 10\%, 5\%, 1\%, 0.1\%$ are $9.3, 12.8, 16.2, 24.5, 36.5$. Non-monotonicity applies for $z \geq 16$. The optimal rejection regions $\{\frac{f(z;\lambda_1,\lambda_2)}{f(z;0,0)} \geq C\}$ correspond to those of the LR test $\{z \geq \tilde{C}\}$, e.g., for asymptotic levels such as $\alpha = 20\%$, but not for $\alpha = 5\%$ or smaller. Indeed, in the latter cases, because of non-monotonicity of the density ratio, the optimal rejection regions are finite intervals in argument z. With $\bar{\nu} = 0.7$, we do not find violations with $\underline{\lambda} = 0.5$ or larger.

[39] A given number of simulated draws become increasingly sparse when considering larger values of df, which makes the exploration of the parameter space more challenging in those cases. However, unreported theoretical considerations show via an asymptotic approximation that the monotone likelihood property holds for $df \to \infty$ since the limiting distribution is then Gaussian. This finding resonates with the absence of violations in Table 1 for the larger values of df.

	df	2	3	4	5	6	7	8	9	10	11	12
$\bar{\nu} = 0.2$	$\underline{\lambda} = 0$	0.002	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.000	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.7$	$\underline{\lambda} = 0$	0.051	0.000	0.000	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.037	0.000	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0	0	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.9$	$\underline{\lambda} = 0$	0.151	0.004	0.000	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.134	0.004	0.000	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0.007	0.000	0	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0	0	0	0	0	0	0	0	0	0	0
$\bar{\nu} = 0.99$	$\underline{\lambda} = 0$	0.426	0.015	0.000	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.1$	0.411	0.014	0.000	0.000	0	0	0	0	0	0	0
	$\underline{\lambda} = 0.5$	0.218	0.007	0.000	0	0	0	0	0	0	0	0
	$\underline{\lambda} = 1$	0.078	0.001	0	0	0	0	0	0	0	0	0

Table 1: Numerical check of Inequalities (12) by Monte Carlo. We display the cumulative frequency of violations in ‰ of Inequalities (12), for $m = 3, ..., 16$, over 10^8 random draws of the parameters $\lambda_j \sim Unif[\underline{\lambda}, \bar{\lambda}]$ and $\nu_j \sim Unif[0, \bar{\nu}]$, for $\bar{\lambda} = 7$, and different combinations of bounds $\underline{\lambda}$, $\bar{\nu}$, and degrees of freedom df. An entry 0.000 corresponds to less than 100 cases out of 10^8 draws.