Optimal Maximin GMM Tests for Sphericity in Latent Factor Analysis of Short Panels

Alain-Philippe Fortin[1], Patrick Gagliardini[2,3], Olivier Scaillet[4,3*]

August 27, 2025

Abstract

We derive optimal maximin tests for parametric hypotheses in short panels with latent common factors. We rely on a Generalized Method of Moments setting with optimal weighting under a large cross-sectional dimension n and a fixed time series dimension T. We outline the asymptotic distributions of the estimators as well as the asymptotic maximin optimality of the Wald, Lagrange Multiplier, and Likelihood Ratio-type tests. The characterisation of optimality relies on finding the limit Gaussian experiment in strongly identified GMM models under a block-dependence structure and unobserved heterogeneity. We reject sphericity of idiosyncratic errors in an empirical application to a large cross-section of U.S. stocks, which casts doubt on the validity of routinely applying Principal Component Analysis to short panels of monthly financial returns.

Keywords: Latent factor analysis, Generalized Method of Moments, maximin test, Gaussian experiment, fixed effects, panel data, sphericity, large n and fixed T asymptotics, equity returns.

JEL codes: C12, C23, C38, C58, G12.

---

*[1]Université de Montréal, [2]Università della Svizzera Italiana, [3]Swiss Finance Institute, [4]University of Geneva.
Acknowledgements: We are grateful to A. Onatski for his very insightful discussion (Onatski (2025)) of our paper Fortin, Gagliardini, Scaillet (2025) at the 14th Annual SoFiE Conference in Cambridge, which prompted us to explore sphericity in short panels. We thank M. Arellano, A. Horenstein, F. Kleibergen, and participants at FinEML 2024, QFFE 2025, SFI research days, SoFiE 2025, ANR 2025, and ESWC 2025. The authors also acknowledge financial support by the Swiss National Science Foundation (grants UN11140, 100018_215573 and 204106).

1 Introduction

Principal Component Analysis (PCA) and Factor Analysis (FA) aim at summarising the common latent linear structure of multivariate data; see Anderson (2003) Chapters 11 and 14. When we impose the sphericity restriction on the $T \times T$ error covariance matrix V_ε in the estimation procedure, where T is the time series dimension, the FA estimator $\hat{F}$ of the latent factors F boils down to the PCA estimator; see Anderson and Rubin (1956). Sphericity corresponds to $V_\varepsilon = \bar{\sigma}^2 I_T$ being a multiple of the identity matrix I_T with unknown parameter $\bar{\sigma}^2 > 0$. It is a necessary and sufficient condition for consistency of PCA estimates $\hat{F}$, when the cross-sectional dimension n is large and T is fixed (Theorem 4 of Bai (2003); see discussion in Fortin et al. (2025) and Onatski (2025)).[1] Fortin et al. (FGS, 2023b) develop inferential tools for FA in short panels. Their Pseudo Maximum Likelihood (PML) setting (White (1982), Gouriéroux et al. (1984)) under large n and fixed T relies on a diagonal $T \times T$ covariance matrix of the errors without imposing sphericity, Gaussianity, or cross-sectional independence.[2] They derive 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)). This paper derives Generalized Method of Moments (GMM) testing procedures of sphericity in short panels through comparison of constrained and unconstrained versions of the FA-GMM estimators under optimal weighting (Hansen (1982)).[3] Such an approach differs

[1]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 (2025) considers inference for latent factors in conditional linear asset pricing models under sphericity based on PCA, including estimation of the number of factors.

[2]Alvarez and Arellano (2022) develop PML approaches for dynamic panel models with non-sphericity, also achieving consistency in a fixed T setting (see Chamberlain and Moreira (2009) and Bai (2013, 2024) for other desirable properties such as minimax optimality and efficiency from a fixed-effect perspective).

[3]Omitted latent factors are also called interactive fixed effects in the panel literature (Pesaran (2006), Bai (2009), Moon and Weidner (2015), Freyberger (2018)). We find them in asset embeddings (Gabaix et al. (2023)). Ahn et al. (2001,2013) use the terminology time-varying individual effects. They use optimal GMM criterion to estimate their panel data model with random interactive effects and i.i.d. errors under fixed T, while Hayakawa et al. (2023) rely on

from other tests of sphericity available in the literature; see e.g. Mauchly (1940), Bartlett (1951), John (1972), Ledoit and Wolf (2002), Anderson (2003), Schott (2006), Onatski et al. (2013). Unlike those papers, we target a test of sphericity for a covariance matrix with a low rank structure recovered from FA estimates. Besides, it allows us to characterise the asymptotic optimal maximin properties of the usual trinity of tests, namely the Wald, Lagrange Multiplier, and Likelihood Ratio-type tests, in a GMM framework (Newey and McFadden (1994)). These properties are novel to the literature, even in the i.i.d. case, and have broad applicability for GMM tests of panel models as shown in Section 3 below for the model of Chamberlain (1992) (see Arellano and Bonhomme (2012) for identification of distributional characteristics when coefficients are random in that model). Testing for sphericity with GMM tests in FA is simply a particular case. We establish an upper bound on the power in the maximin sense, i.e., the optimal power against the least favorable directions among the local alternative hypotheses (Lehmann and Romano (LR, 2005), Chapters 8 and 13.5.3). Maximin optimality does not impose restricting tests through concepts like unbiasedness, conditioning, monotonicity, and invariance (see Romano et al. (2010) for a survey of optimality approaches in testing problems). It has a wider applicability to obtain tests with asymptotically guaranteed power. The derivation of optimal maximin tests relies on finding the limit Gaussian experiment in strongly identified GMM models under a block-dependence structure and unobserved heterogeneity,[4] before applying those maximin results to our FA model. In FA, heterogeneity is driven by means affine in fixed effects. Andrews and Mikusheva (2022) exploit a limit Gaussian experiment to design Bayesian decision rules and characterize optimal similar tests[5] in the sense of maximizing weighted average power (WAP)[6] in weakly identified GMM models. We rely on the same strategy but to derive optimal maximin results in strongly identified GMM models in a non-i.i.d. setting. Chen and Santos (2018) investigate maximin results for specification testing, in particular the J-test of Hansen (1982) and the incremental J-test of Eichenbaum et al.

a transformed Gaussian PML for a dynamic panel data model.

[4]We refer to Bonhomme and Denis (2024a,b) for two recent surveys on accounting for heterogeneity in panel data.

[5]Optimal similar tests are also developed in e.g. Moreira (2003) and Andrews et al. (2006).

[6]WAP optimality is also used by e.g. Sowell (1996), Andrews (1998), and Mueller (2011).

(1988), in an i.i.d. setting (see Newey (1985) for early work on optimality within the class of GMM tests of overidentifying restrictions and Chen et al. (2024) for use in finance). Our maximin results developed under a block-dependence structure and unobserved heterogeneity suit the particular case of our FA-GMM sphericity tests and are new to the literature on the power optimality of the trinity of GMM tests (see Engle (1984) for Asymptotically Uniformly Most Powerful Invariant tests in a maximum likelihood framework). The characterisation of the Gaussian experiment in a non-i.i.d. context is new. It is not a direct application of the available results for the i.i.d. setting, and of independent interest. It can be exploited for other applications such as designing Bayesian priors and optimal similar tests. The idea of using a Gaussian experiment based on the concept of Local Asymptotic Normality (LAN) to study asymptotic optimal tests is rooted in Le Cam's statistical work (see e.g. Le Cam (1986), van der Vaart (1998, 2002)). Choi et al. (1996) investigate asymptotically uniformly most powerful (AUMP) tests in parametric and semiparametric models through LAN technology.

The outline of the paper is as follows. In Section 2, we consider a linear latent factor model and introduce GMM estimation and testing procedures for sphericity based on FA. We work under a block-dependence structure to allow for weak dependence in the cross-section and to get consistency of asymptotic variance estimators without imposing independence. Section 3 provides a general theory for optimal maximin tests in strongly identified GMM models. The theory builds on characterizing a limit Gaussian experiment under a block-dependence structure and unobserved heterogeneity. Section 4 is dedicated to local asymptotic power, asymptotic distributions, and maximin properties of the trinity of tests for sphericity in the FA model. The maximin properties are a by-product of the broad optimality results of Section 3. We run Monte Carlo (MC) experiments in Section 5 to gauge the empirical size and power of our tests in small samples. We provide our empirical application in Section 6. We reject sphericity on a large cross-section of U.S. stocks in all subperiods between 1966 and 2023, which casts doubt on the validity of routinely applying PCA to short panels of monthly financial returns. The presence of a common component driving the variance of the error terms (Barigozzi and Hallin (2016), Renault et al. (2023)) might explain such a rejection. We collect our concluding remarks in Section 7. Appendices A and B gather the regularity assumptions and proofs of the main theoretical results. Appendix C gives a spectral characterisation of spherical models useful to distinguish constrained and unconstrained FA models. Appendix D gives the characterisation of the unconstrained and constrained FA-GMM estimators, and their feasible asymptotic distributions as well as the ones of the trinity of GMM test statistics for sphericity under local alternative hypotheses. Appendix E in the Online Appendix (OA) provides the detailed proofs of technical Lemmas 5-8 supporting the computations of Appendix D. Appendix F discusses practical implementation and provides a numerical study of the performance of the new FA-GMM estimators. We put additional MC experiments in Appendix G. Appendix H makes the link with the panel model of Chamberlain (1992) Section 4 and discusses how we can incorporate second-order moment information in sets of orthogonality restrictions for that model as in Arellano and Bonhomme (2012) Section 3.4 and our FA setting. Appendix I gives the detailed proof of Proposition 3.

2 GMM Testing for Sphericity in Latent Factor Analysis

2.1 Latent Factor Model

We consider the linear 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 - error unconditional variance-covariance matrix. The diagonality condition in Assumption 1 is standard in FA (in the more restrictive formulation involving i.i.d. data). Assumption 1 allows for serial dependence in idiosyncratic errors in the form of martingale difference sequences, like individual GARCH and stochastic volatility 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. The diagonality condition in Assumption 1 corresponds to the unconstrained model versus the constrained model with $V_\varepsilon = \bar{\sigma}^2 I_T$, for an unknown scalar $\bar{\sigma}^2 > 0$, i.e., sphericity.

This paper focuses on the trinity of GMM tests for sphericity when T is fixed and $n \to \infty$. However, we can embed Model (1) as a particular case of the panel model of Chamberlain (1992); see the beginning of Section 3. The theoretical results in Section 3 cover the optimal maximin properties of the trinity of GMM tests in such a general framework. Hence our optimal theory has broader applicability than only testing for sphericity with GMM tests in FA. In this section, we do not outline explicitly the estimators and tests for the panel model of Chamberlain (1992) since we do not use it in our empirics. They can been developed under the same lines as the ones below.

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. 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.[7] 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. Working with fixed effects avoids specific assumptions on the randomness of heterogeneity. Moreover, in Assumption 1, we neither assume Gaussianity

[7] The use of FA in this paper shares similarities with the applications of FA in psychometrics, in which the components of vector y_i are mental tests scores and the observations units $i = 1, ..., n$ are individuals (Anderson (2003)).

nor cross-sectional independence. Hence, the FA-GMM estimators defined below correspond to maximizers of a GMM criterion in a more general setting than in the standard literature on FA (Anderson and Amemiya (1988)).

Let us introduce the usual 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 $\mu_\beta = \lim_{n\to\infty} \bar{\beta}$ and $V_\beta = \lim_{n\to\infty} \tilde{V}_\beta$ with $\bar{\beta} := \frac{1}{n}\sum_{i=1}^n \beta_i$ and $\tilde{V}_\beta := \frac{1}{n}\sum_{i=1}^n \beta_i\beta_i'$. Then, under our assumptions, we have $V_y := \operatorname{plim}_{n\to\infty}\hat{V}_y = FF' + V_\varepsilon$, 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) and $V_y V_\varepsilon^{-1} F_j = (1+\gamma_j)F_j$, i.e., the F_j are eigenvectors of matrix $V_y V_\varepsilon^{-1}$ associated with eigenvalues $1+\gamma_j$, $j = 1, ..., k$. In Assumption A.1, we also normalize the mean and variance of the betas in sample, namely $\bar{\beta} = 0$ and $\tilde{V}_\beta = I_k$. Those standardizations of the factor loadings wash out the incidental parameter problem (Neyman and Scott (1948); see Lancaster (2000) for a review) since the individual loadings do not appear in $\mu_y := \operatorname{plim}_{n\to\infty} \bar{y}$ nor in V_y, and we do not need to estimate them. It explains why we are able to get consistent FA-GMM estimators for large n and fixed T in the below.[8]

The parameter set Θ is a compact subset of $\{\theta = (\mu', vec(F)', diag(V_\varepsilon)')' \in \mathbb{R}^p \ : \ V_\varepsilon$ is diagonal and positive definite, $F'V_\varepsilon^{-1}F$ is diagonal, with diagonal elements ranked in decreasing order$\}$ with $p = T(k+2)$. Model (1) under Assumption 1 corresponds to the hypothesis $H(k)$, while the k-factor spherical model corresponds to the restriction $V_\varepsilon = \bar{\sigma}^2 I_T$, for an unknown constant $\bar{\sigma}^2 > 0$, and yields the hypothesis $H_s(k) \subset H(k)$. The complement of $H_s(k)$ in $H(k)$, denoted by $\bar{H}_s(k)$, agrees with a non-spherical k-factor model with $V_{\varepsilon,tt} \neq V_{\varepsilon,ss}$, for at least one pair $t \neq s$.

[8] The sample normalization of the fixed effects can be simply obtained by linear transformation of parameters μ and F that are drifting with n. Indeed, suppose the DGP is $y_i = \mu_0 + F_0\tilde{\beta}_i + \varepsilon_i$, where $\tilde{\mu}_{\tilde{\beta}} := \frac{1}{n}\sum_{i=1}^n \tilde{\beta}_i \to 0$ and $\tilde{V}_{\tilde{\beta}} := \frac{1}{n}\sum_{i=1}^n (\tilde{\beta}_i - \tilde{\mu}_{\tilde{\beta}})(\tilde{\beta}_i - \tilde{\mu}_{\tilde{\beta}})' \to I_k$ as $n \to \infty$. Then, we have $y_i = \tilde{\mu} + \tilde{F}\beta_i + \varepsilon_i$ with $\beta_i = C'\tilde{V}_{\tilde{\beta}}^{-1/2}(\tilde{\beta}_i - \tilde{\mu}_{\tilde{\beta}})$, $\tilde{\mu} = \mu_0 + F_0\tilde{\mu}_{\tilde{\beta}}$ and $\tilde{F} = F_0\tilde{V}_{\tilde{\beta}}^{1/2}C$, where C is the orthogonal matrix of the eigenvectors of $\tilde{V}_{\tilde{\beta}}^{1/2}F'V_\varepsilon^{-1}F\tilde{V}_{\tilde{\beta}}^{1/2}$. The fixed effects β_i meet the desired sample normalization. The drifting parameters $\tilde{\mu}$ and $\tilde{F}$ converge to μ_0 and F_0 as $n \to \infty$. We omit dependence of parameters on n to ease notation.

The model under $H(k)$ is a strict subset of the unconstrained model with general variance V_y for any k up to $k_{\max}$, where $k_{\max} = k_{\max}(T)$ is the largest integer such that the number of degrees of freedom $df = \frac{1}{2}((T-k)^2 - T - k)$ is strictly positive. Instead, $H_s(k)$ gives a strict subset of the unconstrained model for k up to $T-2$ (see Lemma 4 and Corollary 2 in Appendix C). We show in Appendix C that some Data Generating Processes (DGP) may admit FA representations with different numbers of latent factors. We consider the representation with minimal k when defining models under $H(k)$ and $H_s(k)$. In the design of the GMM tests in Section 2.3 below, we work under the maintained assumption that the DGP admits an FA representation, i.e., $k \leq k_{\max}$, which is the leading case for empirical applications in finance. Then, we test the null hypothesis $H_s(k)$ against the alternative hypothesis $\bar{H}_s(k)$. We determine the number of latent factors k by using the consistent estimator $\hat{k}$ in FGS.[9] We implement the test with $\hat{k}$ and not $k_{\max}$, since there is no guarantee that we can write V_y as $FF' + V_\varepsilon$ with a $T \times k_{\max}$ matrix F having full column rank and a $T \times T$ positive definite diagonal matrix V_ε.

2.2 GMM Estimators in FA

A GMM approach to FA relies on the $q \times 1$ orthogonality vector $\lim\limits_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E[g(y_i, \theta_0)] = 0$, with $g(y_i, \theta) := [(y_i - \mu)', vech(y_i y_i' - \Sigma(\vartheta) - \mu\mu')']'$ and $q = (T(T+3)/2) \times 1$, where $\Sigma(\vartheta) := FF' + V_\varepsilon$ and $\vartheta := (vec(F)', diag(V_\varepsilon)')'$. Here, for a $T \times T$ symmetric matrix $Z = (z_{i,j})$, we define the $\frac{1}{2}T(T+1) \times 1$ vector $vech(Z) = \left(\frac{1}{\sqrt{2}}z_{11}, ..., \frac{1}{\sqrt{2}}z_{T,T}, \{z_{i,j}\}_{i<j}\right)'$, where the out-of-diagonal elements with indices $i < j$ are ranked as $(1,2)$, $(1,3)$, ... , $(2,3)$, ... $(T-1,T)$.[10] The sample average moment vector is $\hat{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^n g(y_i, \theta) = \left[(\bar{y} - \mu)', vech(\hat{V}_y + \bar{y}\bar{y}' - \Sigma(\vartheta) - \mu\mu')'\right]'$.

[9] We can develop distributional results for testing $H_s(k)$ against the alternative hypothesis of a non-spherical model for $k = k_{\max}+1, ..., T-2$ as well, i.e., when the DGP does not admit an FA representation. For the sake of conciseness, we do not cover this case explicitly.

[10] 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 lines of proof. For instance, it holds $\frac{1}{2}\|A\|^2 = vech(A)'vech(A)$, for a symmetric matrix A.

Then, the FA-GMM estimator with optimal weighting matrix is:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{g}_n(\theta)'(\hat{V}_g)^{-1}\hat{g}_n(\theta), \tag{2}$$

where matrix $\hat{V}_g$ is a consistent estimator of the asymptotic variance in $\sqrt{n}\hat{g}_n(\tilde{\theta}_0) \Rightarrow \mathcal{N}(0, V_g)$ defined in Section D.3 accounting for cross-sectional dependence. Vector $\tilde{\theta} = (\mu', vec(F)', diag(\tilde{V}_\varepsilon)')'$ involves $\tilde{V}_\varepsilon := \frac{1}{n}E[\varepsilon\varepsilon']$ instead of its large sample limit V_ε in order to center the moment vector and guarantee the CLT after rescaling by $\sqrt{n}$. The degree of overidentification is $q - (p - \frac{1}{2}k(k-1)) = df$. In Section F of the OA, we provide an asymptotically equivalent FA-GMM estimator which is easier to compute numerically as it yields the estimate of μ in closed form as a function of ϑ. We also explain how we can compute numerically the FA-GMM estimators of ϑ based either on a Newton-Raphson (NR) algorithm or a zigzag algorithm (Magnus and Neudecker (2007), Hautsch et al. (2023)). In the latter, the step to update parameter matrix F for given V_ε corresponds to a weighted low-rank approximation problem. Due to the general form of the weighting, the solution is not obtained through Singular Value Decomposition as in the zigzag algorithm for PML estimation. We use the ideas in Manton et al. (2003), namely we solve an inner minimization problem for $diag(F'V_\varepsilon^{-1}F)$ in closed form, and then apply the NR method to the outer minimization after concentration to obtain the normalized columns of $V_\varepsilon^{-1/2}F$. In Section F.3 of OA, we provide a comparison of the numerical performance of these algorithmic choices. We find a strong advantage of the NR algorithm in terms of computational speed, and thus we run our MC experiments (Section 5) and our empirics (Section 6) with that one.

To establish the asymptotic normality of vector $\sqrt{n}\hat{g}_n(\tilde{\theta}_0)$ beyond an i.i.d. setting, we use a block-dependence structure for the error terms as in FGS. It allows for weak cross-sectional dependence and heteroschedasticity in idiosyncratic errors as in approximate factor models (Chamberlain and Rothschild (1983)).

Assumption 2 *(a) The errors are such that* $\varepsilon = V_\varepsilon^{1/2} W \check{\Sigma}^{1/2}$*, 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, with $E(|w_{i,t}|^{4r}) \leq C$ uniformly in i, t, for constants $C > 0$ and $r > 1$, and $\check{\Sigma} = (\check{\sigma}_{i,j})$ is a positive-definite symmetric $n \times n$ matrix, such that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \check{\sigma}_{ii} = 1$. (b) Matrix $\check{\Sigma}$ is block diagonal with J_n blocks of size $b_{m,n}$, for $m = 1, ..., J_n$, where $J_n \to \infty$ as $n \to \infty$, and B_m denotes the set of indices in block m. (c) There exist constants $\check{\delta} \in [0, 1]$ and $\check{C} > 0$ such that $\max_{i \in B_m} \sum_{j \in B_m} |\check{\sigma}_{i,j}| \leq \check{C} b_{m,n}^{\check{\delta}}$. (d) The block sizes $b_{m,n}$ and block number J_n are such that $n^{-r} \sum_{m=1}^{J_n} b_{m,n}^{r(1+\check{\delta})} = o(1)$ and $n^{-3/2} \sum_{m=1}^{J_n} b_{m,n}^2 = o(1)$.*

The block-dependence structure as in Assumption 2(a)-(b) is satisfied, for instance, when there are unobserved industry-specific factors independent among industries and over time, as in Ang et al. (2020). In empirical applications, blocks can match industrial sectors (Fan et al. (2016)). Assumption 2(c) covers within-blocks sparsity when $\check{\delta} < 1$, but it does not impose it since we allow $\check{\delta} = 1$. Assumption 2(d) generalizes the condition used in FGS, which applies for $r = 2$, to any $r > 1$, i.e., we require existence of error moments at an order slightly above the fourth one (finite kurtosis). With blocks of homogeneous size $b_{m,n} = Cn^{\bar{\alpha}}$, $\bar{\alpha} > 0$, Assumption 2(d) holds if $\bar{\alpha} < \min\{\frac{1}{2}, \frac{r-1}{r(\check{\delta}+1)-1}\}$. Assumption 2 covers both the null hypothesis of sphericity of V_ε and the alternative hypothesis of deviations from it.

2.3 GMM Tests in FA

We consider the usual trinity of test statistics in the GMM framework outlined in the previous subsection, but in an FA context. The sphericity of V_ε under the null hypothesis corresponds to $T - 1$ linear constraints $a(\theta) := L_{1_T}' diag(V_\varepsilon) = 0$, where L_{1_T} is a $T \times (T - 1)$ full-rank matrix such that $L_{1_T} L_{1_T}' = I_T - \frac{1}{T} 1_T 1_T'$ and $L_{1_T}' L_{1_T} = I_{T-1}$. The Wald (W), Lagrange Multiplier (LM) and Likelihood Ratio-type (LR) statistics are defined by (see e.g. Newey and McFadden (1994) for the general case and Satorra (1989) for covariance structure analysis):

$$\xi_n^W = na(\hat{\theta})'(\hat{\Omega}_W)^{-1} a(\hat{\theta}), \qquad \xi_n^{LM} = n\hat{\lambda}'(\hat{\Omega}_{LM})^{-1}\hat{\lambda}, \qquad \xi_n^{LR} = n\left(Q_n(\hat{\theta}^c) - Q_n(\hat{\theta})\right), \tag{3}$$

where vector $\hat{\lambda}$ stacks the $T-1$ Lagrange multipliers for the minimization of the GMM criterion $Q_n(\theta) = \hat{g}_n(\theta)'(\hat{V}_g)^{-1}\hat{g}_n(\theta)$ for $\theta \in \Theta$ under the constraint $a(\theta) = 0$, and vector $\hat{\theta}^c$ is the vector of constrained FA-GMM estimates. Matrix $\hat{\Omega}_W = L'_{1_T}\hat{\Sigma}_{V_\varepsilon}L_{1_T}$ is a consistent estimator of the asymptotic variance in $\sqrt{n}a(\hat{\theta}) \Rightarrow \mathcal{N}(0, \Omega_W)$, based on the unconstrained FA-GMM estimator $\hat{\theta}$, where Σ_{V_ε} is the asymptotic variance of $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon)$ characterised in Appendix D.1. Matrix $\hat{\Omega}_{LM} = \left(L'_{1_T}\hat{\Sigma}^c_{V_\varepsilon}L_{1_T}\right)^{-1}$ is a consistent estimator of the asymptotic variance in $\sqrt{n}\hat{\lambda} \Rightarrow \mathcal{N}(0, \Omega_{LM})$, based on the constrained estimator $\hat{\theta}^c$.[11] The trinity of FA-GMM test statistics is made of (3).

3 Optimal Maximin GMM Tests

In this section, we provide a general theory for optimal maximin tests in strongly identified GMM models. The theory exploits a limit Gaussian experiment given in the next subsection for a block-dependence structure (like Assumption 2) and unobserved heterogeneity instead of i.i.d. observations. The characterisation of the Gaussian experiment in a non-i.i.d. context is not a direct application of the available results for the i.i.d. setting. The moment specification covers our FA model where heterogeneity is driven by means affine in fixed effects (see Section 4). It also covers the panel data model of Chamberlain (1992) Section 4. To recall that framework, let us consider the i.i.d. random vector $y_i = (\mathcal{Y}_i, z_i')'$ with $\mathcal{Y}_i$ being vector-valued, so that

$$\mathcal{Y}_i = d(z_i, \zeta) + R(z_i, \zeta)\beta_i + \varepsilon_i, \tag{4}$$

where z_i is a vector of observed regressors, ζ is the vector of parameters for known functions d and R, and β_i are treated as random coefficients. His approach consists in using the assumpion $E[\varepsilon_i|z_i, \beta_i] = 0$, to get a system of orthogonality restrictions involving the finite-dimensional pa-

[11] The link between the Lagrange multipliers vector and the score, i.e. $\hat{\lambda} = -L'_{1_T}\frac{\partial Q_n(\hat{\theta}^c)}{\partial diag(V_\varepsilon)}$, implies the equivalence of the LM and score statistics in the GMM setting with optimal weighting matrix (see Newey and McFadden (1994) for the general result).

rameters ζ and $\phi = E[\beta_i]$, such that the associated GMM estimator is semiparametrically efficient. Specifically, he shows how to obtain a matrix of instruments $B(z_i, \zeta)$ such that $B(z_i, \zeta)R(z_i, \zeta) =: F(\zeta)$ is independent of z_i. Then, from (4), we have $B(z_i, \zeta)[\mathcal{Y}_i - d(z_i, \zeta)] = F(\zeta)\beta_i + u_i$, where $u_i = B(z_i, \zeta)\varepsilon_i$ and $E[u_i] = 0$. Chamberlain (1992) gets the orthogonality restrictions:

$$E[B(z_i, \zeta)(\mathcal{Y}_i - d(z_i, \zeta)) - F(\zeta)\phi] = 0. \tag{5}$$

In our fixed effect setting, we get $\phi = \mu_\beta = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \beta_i$. Equation (5) takes the form: $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n E[g(y_i, \theta)] = 0$, where θ gathers ζ and ϕ. In the next subsection, we consider similar sets of orthogonality restrictions but under heterogeneous distributions $P_{0,i}$ for y_i and allowing for cross-sectional dependence. Hence, our theory on maximin GMM tests also applies to panel model (4). The particularity of the FA model presented in Section 2 is that we can simply take $B(z_i, \zeta) = I_T$ since matrix R does not depend on z_i. Moreover, by the latent factor structure, we can take the normalisation $\phi = 0$, so that those parameters do not appear in the orthogonality restrictions. In Appendix H in the OA, building on Arellano and Bonhomme (2012) Section 3.4, we further show that the stacked vector $(\mathcal{Y}_i', (\mathcal{Y}_i \otimes \mathcal{Y}_i)')'$ yields conditional moment restrictions of the type studied by Chamberlain (1992) written for an augmented vector of individual effects and parameters. It paves the way to incorporate second-order moment information in sets of orthogonality restrictions for panel model (4) as in our FA setting.

3.1 Gaussian Experiment for Strongly Identified GMM

Let us consider the GMM orthogonality restriction $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n E_{P_{0,i}}[g(y_i, \theta_0)] = 0$, where the $P_{0,i}$ are unknown possibly heterogeneous distributions for random vectors y_i, $g(\cdot, \theta)$ is a q-vector of known orthogonality functions and $\theta_0 \in \Theta \subset \mathbb{R}^p$ is an unknown parameter vector, with $q \geq p$. In the i.i.d. case, $P_{0,i} = P_0$, $i = 1, ..., n$. We test the null hypothesis $H_0 : a(\theta_0) = 0$, or equivalently $\theta_0 \in \Theta_0 := \{\theta \in \Theta : a(\theta) = 0\}$, where $a : \Theta \to \mathbb{R}^r$ is a differentiable function, with $r \leq p$ and $A := \frac{\partial a(\theta_0)}{\partial \theta'}$ has full row-rank. To study the local power of test statistics, the sequence of local alternative hypotheses $H_{1,n}$ is such that $\frac{1}{n} \sum_{i=1}^n E_{P_{n,i}}[g(y_i, \theta_n)] = 0$ for all $n \in \mathbb{N}$, where

$\theta_n = \theta_0 + \frac{1}{\sqrt{n}}h$, with $h \in \mathbb{R}^p$, and $\{P_{n,i} : i = 1, ..., n\}$ is a triangular array of probability distributions. We assume that:

$$[y_1 : \cdots : y_n] = [x_1 : \cdots : x_n]S, \tag{6}$$

where the x_i are independent random vectors with possibly heterogeneous distributions $Q_{n,i}$ with pdf $q_{n,i}$, for $i = 1, ..., n$, and S is a nonsingular block-diagonal $n \times n$ matrix, with diagonal blocks S_m of size $b_{m,n}$, for $m = 1, ..., J_n$. The block structure is known, while matrices S_m are unknown nuisance parameters, $|S_m| = 1$ without loss of generality, and subject to conditions introduced below to control the degree of cross-sectional dependence when the number J_n of blocks diverges as $n \to \infty$. Setting $S = I_n$ gives the independence case.

Assumption 3 *The family $q_{n,i}$ is quadratic mean differentiable (q.m.d.) such that $\sqrt{q_{n,i}(x)} - \sqrt{q_{0,i}(x)} = \frac{1}{2\sqrt{n}} f_i^q(x)\sqrt{q_{0,i}(x)} + R_{n,i}(x)$, for any $i = 1, ..., n$ and $x \in \mathbb{R}^T$, where $f_i^q \in T(Q_{0,i}) := \{f \in L_2(Q_{0,i}) : \int f(x)q_{0,i}(x)dx = 0\}$ and $\int [R_{n,i}^q(x)]^2 dx = O(\frac{1}{n^\alpha})$ uniformly in i, with $\alpha > 1$.*

The linear space $T(Q_{0,i})$ is the tangent space for distribution $Q_{0,i}$ with pdf $q_{0,i}$ prevailing under the null hypothesis, and a measurable function $f \in T(Q_{0,i})$ is called a score (van der Vaart (1998) Chapter 25). For $h = 0$ and $f_i^q = 0$ for all i, we get the null hypothesis, while for $Ah \neq 0$, we get a sequence of local alternative hypotheses converging to the null hypothesis at rate $n^{1/2}$. In the definition of the q.m.d. property in Assumption 3, we use a stronger decay $O(\frac{1}{n^\alpha})$, with $\alpha > 1$, for the squared L^2 norm of the remainder term, instead of the usual decay $o(\frac{1}{n})$ of the i.i.d. case (LR, Chapter 12, Chen and Santos (2018), Section 3.1.2, Andrews and Mikusheva (2022), Eq. (2)), since we want to accommodate a block-dependence structure (Assumption 2). Here, we face a trade-off between the rate α and the granularity of blocks.

Assumption 4 *In block structure (6), the block sizes $b_{m,n}$ are such that $n^{-\rho/2}\left(\frac{\log n}{n}\sum_{m=1}^{J_n} b_{m,n}^3 + \frac{\log n}{n^{\alpha/2}}\sum_{m=1}^{J_n} b_{m,n}^2\right) = o(1)$, where $\rho := \min\{\frac{3}{2}(1 - \frac{1}{2r-1}), 1\} > 0$ for $r > 1$. Constant $r > 1$ is related to higher-order moments, i.e., we have $E_{P_{0,i}}[\|g(y_i, \theta_0)\|^{2r}] \leq C$ and $E_{Q_{0,j}}[(f_j^q(x_j))^{2r}] \leq C$, for all i, j and a constant $C > 0$, and $E_{P_{n,i}}[\|g(y_i, \theta_0)\|^{2r}] \leq C$ for all i, n.*

Assumption 4 is stronger than Assumption 2(d).[12] It entails a trade-off between the existence of higher-order moments and the strength of cross-sectional dependence. Indeed, for $r \leq 2$, smaller values of moment order $2r$ imply smaller values of coefficient ρ, which in turns requires more granular blocks. With cross-sectional independent errors (and $\alpha \geq 2$), Assumption 4 is met as soon as $r > 1$, i.e. the orthogonality vector and the score admit moments at order slightly above two. With blocks of homogeneous size $b_{m,n} = Cn^{\bar{\alpha}}$, $\bar{\alpha} > 0$, Assumption 4 holds if $\bar{\alpha} < \min\{\rho/4, (\rho+\alpha)/2 - 1\}$. It requires $\rho + \alpha > 2$, which holds if, and only if, $r > \frac{\alpha+1}{2\alpha-1}$. To simplify the proofs, in Assumption 4 we state the bounds on higher-order moments uniformly across i. We can relax uniformity at the cost of more cumbersome conditions on the block sizes $b_{m,n}$.

The blocks $Y_m = [y_i : i \in B_m]$, for $m = 1, ..., J_n$, are independent with joint densities $p_n^m(Y_m) = q_n^m(Y_m S_m^{-1})$, where $q_n^m(X_m) := \prod_{i \in B_m} q_{n,i}(x_i)$ and B_m denotes the set of indices in block m. We obtain the individual densities $p_{n,i}$, for $i \in B_m$, from p_n^m by marginalization. We define similarly the densities p_0^m, q_0^m and $p_{0,i}$ by replacing $q_{n,i}$ with $q_{0,i}$ in the definition of q_n^m. The next lemma shows that the densities q_n^m and $p_{n,i}$ inherit the q.m.d. behavior of $q_{n,i}$ through an adequate block-aggregation and marginalization of the initial $f_i^q \in T(Q_{0,i})$.

Lemma 1 *Let Assumptions 3 and 4 hold. We have (a)* $\sqrt{q_n^m(X_m)} - \sqrt{q_0^m(X_m)} = \frac{1}{2\sqrt{n}} f_m(X_m)\sqrt{q_0^m(X_m)} + R_n^m(X_m)$, *for* $X_m \in \mathbb{R}^{T \times b_{m,n}}$, *where* $f_m(X_m) = \sum_{i \in B_m} f_i^q(x_i)$ *is such that* $f_m \in T(Q_0^m)$, *and* $\int [R_n^m(X_m)]^2 dX_m = O(b_{m,n}^2(\frac{b_{m,n}^2}{n^2} + \frac{1}{n^\alpha}))$. *(b)* $p_{n,i}(y) - p_{0,i}(y) = \frac{1}{\sqrt{n}} f_i^p(y) p_{0,i}(y) + R_{n,i}^p(y)$, *for* $y \in \mathbb{R}^T$ *and* $i \in B_m$, *where* $f_i^p(y) = E_{P_0^m}[f_m(Y_m S_m^{-1}) | y_i = y]$ *is such that* $f_i^p \in T(P_{0,i})$ *with the conditional expectation* $E_{P_0^m}[\cdot | y_i = y]$ *being w.r.t. the r.v.* Y_m *in block* m, *and the remainder* $R_{n,i}^p$ *is such that* $\frac{1}{n}\sum_{i=1}^n \int g(y;\theta_0) 1^\tau(y) R_{n,i}^p(y) dy = o(\frac{1}{\sqrt{n}})$ *where* $1^\tau(y) := 1\{\|g(y,\theta_0)\| \leq \tau_n\}$ *with* $\tau_n = n^{\frac{1}{2(2r-1)}} \log n$.

[12] Indeed, Assumption 4 implies $b_{m,n} = O(\sqrt{n})$ as a necessary condition. Then, $n^{-r}\sum_{m=1}^{J_n} b_{m,n}^{r(1+\check{\delta})} \leq n^{-r}\sum_{m=1}^{J_n} b_{m,n}^{2r} = O(\frac{1}{n^{3/2}}\sum_{m=1}^{J_n} b_{m,n}^3) = o(1)$ for $r \geq 3/2$. Moreover, for $r < 3/2$, by the Hölder inequality we have $n^{-r}\sum_{m=1}^{J_n} b_{m,n}^{2r} \leq \frac{J_n^{1-2r/3}}{n^r}\left(\sum_{m=1}^{J_n} b_{m,n}^3\right)^{2r/3} \leq \left(\frac{1}{n^{5/2-3/(2r)}}\sum_{m=1}^{J_n} b_{m,n}^3\right)^{2r/3} = o(1)$ because $5/2 - 3/(2r) > 1 + \rho/2$. Further, because $\alpha \leq 2$ and $\rho \leq 1$, the condition $\frac{\log n}{n^{(\rho+\alpha)/2}}\sum_{m=1}^{J_n} b_{m,n}^2 = o(1)$ implies $\frac{1}{n^{3/2}}\sum_{m=1}^{J_n} b_{m,n}^2 = o(1)$.

In Lemma 1(b), we consider the difference of two pdf, instead of the difference of square roots thereof as in the usual definition of q.m.d., and we control the order of the remainder term $R^p_{n,i}$ through the average truncated L^2 scalar product with the orthogonality function, since this bound is needed for the rest of our analysis (see proof of Lemma 2).

We assume that $\frac{1}{n}\sum_{i=1}^n E_{P_{0,i}}[g(y_i,\theta_0)] = o(\frac{1}{\sqrt{n}})$. Otherwise, the null hypothesis H_0 for the parametric test implies a local alternative hypothesis for the validity of the moment restrictions, which instead is a maintained hypothesis under both the null and local alternative hypotheses for the parametric test. Then, we have the next result linking the population covariances of $f_i^p(y_i)$ and $g(y_i,\theta_0)$, $i = 1,...,n$, with vector h. Indeed, the required validity of the orthogonality restriction $\frac{1}{n}\sum_{i=1}^n E_{P_{n,i}}[g(y_i,\theta_n)] = 0$ for any n for the sequence of local alternative hypotheses defined by θ_n and $P_{n,i}$, implies that vector $h \in \mathbb{R}^p$ and functions $f_i^p \in T(P_{0,i})$, $i = 1,...,n$, are linked. In an heterogeneous context driven by $P_{n,i}$, $i = 1,...,n$, that link is instrumental for the local power analysis of a semiparametric model based on a limit Gaussian experiment that we establish below. Without ensuring $\frac{1}{n}\sum_{i=1}^n E_{P_{n,i}}[g(y_i,\theta_n)] = 0$, we cannot control the 'directions' f_i^p obtained from the initial scores f_i^q, from which the sequence of the data generating process $P_{n,i}$ approaches $P_{0,i}$, $i = 1,...,n$, at rate $1/\sqrt{n}$ in the sequence of local alternative hypotheses. It exemplifies a key difference w.r.t. an i.i.d. setting where it holds trivially.

Lemma 2 *Under Assumptions 3-4,* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \left(E_{P_{0,i}}[g(y_i,\theta_0)f_i^p(y_i)] + E_{P_{0,i}}[\frac{\partial g(y_i,\theta_0)}{\partial \theta'}]h \right) = 0.$

Let $P^n_{n,f}$ and P^n_0 be the joint probability distributions for n-tuple samples $[y_1 : \cdots : y_n] = [x_1 : \cdots : x_n]S$ with the x_i independent draws from $Q_{n,i}$, and from $Q_{0,i}$. Matrix S is the same under $P^n_{n,f}$ and P^n_0 and meets Assumption 4. The index f subsumes dependence of $P^n_{n,f}$ on score functions f_i^q and vector h. The likelihood ratio $L_{n,f} := \frac{dP^n_{n,f}}{dP^n_0}$ is

$$L_{n,f} = \prod_{m=1}^{J_n} \frac{p_n^m(Y_m)}{p_0^m(Y_m)} = \prod_{m=1}^{J_n} \frac{q_n^m(X_m)}{q_0^m(X_m)} = \prod_{m=1}^{J_n} \frac{\prod_{i\in B_m} q_{n,i}(x_i)}{\prod_{i\in B_m} q_{0,i}(x_i)} = \prod_{i=1}^{n} \frac{q_{n,i}(x_i)}{q_{0,i}(x_i)}, \qquad (7)$$

by the independence of the Y_m among blocks, and the independence of the x_i. This product representation of the likelihood ratio does not depend from the characteristics of the block structure but only on the assumption that data is a one-to-one right-transformation of random matrix $[x_1 : \cdots : x_n]$ with independent columns.[13]

Let $J_0 := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E_{P_{0,i}}[\frac{\partial g(y_i,\theta_0)}{\partial \theta'}]$, $V_g := \lim_{n\to\infty} E_{P_0^n}\left[(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(y_i,\theta_0))(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(y_i,\theta_0))'\right]$, and $\Sigma_0 := (J_0' V_g^{-1} J_0)^{-1}$, and assume validity of a CLT for the sample moment vector under the null hypothesis (see Lemma 7 for the statement in the FA-GMM setting).

Assumption 5 *We have $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(y_i,\theta_0) \Rightarrow \mathcal{N}(0, V_g)$, under P_0^n.*

Proposition 1 *Let Assumptions 3-5 hold. (a) We have $\log L_{n,f} = Z_{n,f} - \frac{1}{2}\sigma_f^2 + o_{P_0^n}(1)$, where $Z_{n,f} = \frac{1}{\sqrt{n}}\sum_{i=1}^n f_i^q(x_i) = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n} f_m(Y_m S_m^{-1})$ and $\sigma_f^2 := \lim_{n\to\infty} E_{P_0^n}[(Z_{n,f})^2] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \int [f_i^q(x)]^2 q_{0,i}(x)dx$. Moreover, (b) under P_0^n, we have $Z_{n,f} \Rightarrow \mathcal{N}(0,\sigma_f^2)$ and*

$$\log L_{n,f} \Rightarrow \mathcal{N}(-\frac{1}{2}\sigma_f^2, \sigma_f^2). \tag{8}$$

(c) The sequences $P_{n,f}^n$ and P_0^n are contiguous. (d) We have :

$$\log L_{n,f} = h' Z_n^* - \frac{1}{2}h'\Sigma_0^{-1}h + Z_n^{\perp} - \frac{1}{2}\sigma_{\perp}^2 + o_{P_0^n}(1), \tag{9}$$

where $Z_n^* = -J_0' V_g^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g(y_i,\theta_0)$ and $Z_n^{\perp}$ are asymptotically mutually independent, and distributed as $\mathcal{N}(0,\Sigma_0^{-1})$ and $\mathcal{N}(0,\sigma_{\perp}^2)$ under P_0^n, with $\sigma_f^2 = h'\Sigma_0^{-1}h + \sigma_{\perp}^2$.

When we multiply σ_f^2 by 4, we get the limit cross-sectional average of the so-called Fisher Information for the q.m.d. families associated to the scores $f_i^q \in T(Q_{0,i})$. The latter are mapped into the random variable $Z_{n,f} \in T(P_0^n)$, where $T(P_0^n)$ denotes the linear space of square integrable, zero-mean random variables under P_0^n. The moment restriction $\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^n E_{P_{0,i}}[g(y_i,\theta_0)] = 0$ implies that the components of vector $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(y_i,\theta_0)$ belongs to space $T(P_0^n)$ up to a term $o(1)$.

[13] We cover model (4) by assuming $\mathcal{Y} := [\mathcal{Y}_1 : \cdots : \mathcal{Y}_n] = [x_1 : \cdots : x_n]S$, with the x_i independent draws from $Q_{n,i}$, and from $Q_{0,i}$, but conditionally on $Z = [z_1 : \cdots : z_n]$. We further assume that regressors Z are exogenous for the testing problem at hand, i.e., the distribution of Z is the same under $P_{n,f}^n$ and P_0^n. Then, we get decomposition (7) as follows: $L_{n,f} = \frac{p_n(Y)}{p_0(Y)} = \frac{p_n(\mathcal{Y}|Z)}{p_0(\mathcal{Y}|Z)} = \prod_{m=1}^{J_n}\frac{q_n^m(X_m)}{q_0^m(X_m)} = \prod_{i=1}^n \frac{q_{n,i}(x_i)}{q_{0,i}(x_i)}$.

Hence, we have the decomposition $T(P_0^n) = \mathscr{H}_n^* \oplus \mathscr{H}_n^\perp$, where $\mathscr{H}_n^*$ is the q-dimensional linear subspace of $T(P_0^n)$ spanned by $\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(g_j(y_i, \theta_0) - E_{P_{0,i}}[g_j(y_i, \theta_0)] \right)$, for $j = 1, ..., q$, and $\mathscr{H}_n^\perp$ is the orthogonal complement of $\mathscr{H}_n^*$ in $T(P_0^n)$ w.r.t. the standard $L_2(P_0^n)$ scalar product. In the proof of Proposition 1(d), we use Lemma 2 to show that the projection of $Z_{n,f}$ onto $\mathscr{H}_n^*$ is given by $h' Z_n^* + o_{P_0^n}(1)$ while $Z_n^\perp$ is the projection of $Z_{n,f}$ onto the orthogonal complement $\mathscr{H}_n^\perp$. In particular, the projection onto $\mathscr{H}_n^*$ depends asymptotically on the sequence of local alternative hypotheses via vector $h \in \mathbb{R}^p$ only. In (9) the log likelihood ratio is the sum of a first part that characterizes a LAN parametric model (see Definition 13.4.1 in LR), with 'sufficient statistic' Z_n^* and 'parameter' h, and a second part that corresponds to the nonparametric 'nuisance' contribution induced by $Z_n^\perp$. From Proposition 1 and joint asymptotic normality of $\frac{1}{\sqrt{n}} \sum_{i=1}^n g(y_i, \theta_0)$ and $\log L_{n,f}$ with asymptotic covariance $-J_0 h$ under P_0^n, Le Cam's Third Lemma (LR, Corollary 12.3.2) implies under $P_{n,f}^n$:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g(y_i, \theta_0) \Rightarrow \mathcal{N}(-J_0 h, V_g). \tag{10}$$

The Gaussian random vector $\mathcal{N}(-J_0 h, V_g)$ yields the Gaussian experiment for our problem. We can estimate matrices J_0 and V_g consistently (see Section D.3 for the FA-GMM setting). If we consider matrices J_0 and V_g given, the Gaussian experiment corresponds to a generalized linear regression model under Gaussian errors with a single q-dimensional observation, a p-dimensional unknown parameter vector h, and design matrix J_0. The null hypothesis is the linear restriction $Ah = 0$.

3.2 Maximin GMM tests

Let us now establish a characterisation of asymptotic power of tests in terms of Gaussian experiments building on the arguments used in the proof of Theorem 13.4.1 in LR. Consider a sequence of tests ϕ_n, i.e., a sequence of functions on sample space $\mathcal{Y}^n$, where $\mathcal{Y}$ is the sample space of one observation, with values in $[0, 1]$. Let $\beta_{n,f}(\phi_n) := E_{P_{n,f}^n}[\phi_n]$ denote its power under the sequence of local alternative hypotheses. From Proposition 1(d), and by using a tightness argument and Prohorov Theorem, for any subsequence n_j, there exists a further subsequence n_{j_m} such that $(\phi_{n_{j_m}}, Z^{*\prime}_{n_{j_m}}, Z^{\perp}_{n_{j_m}})' \Rightarrow (\bar{\phi}, Z^{*\prime}, Z^{\perp})'$, under $P_0^{n_{j_m}}$, where $\bar{\phi}$ is a random variable admitting values in $[0,1]$ living on the same probability space as Z^* and $Z^{\perp}$, i.e., the weak limits of Z_n^* and $Z_n^{\perp}$. Then, from (9), we get $(\phi_{n_{j_m}}, L_{n_{j_m},f}) \Rightarrow (\bar{\phi}, \exp\{h'Z^* - \frac{1}{2}h'\Sigma_0^{-1}h + Z^{\perp} - \frac{1}{2}\sigma_{\perp}^2\})$, under $P_0^{n_{j_m}}$, and thus

$$\beta_{n_{j_m},f}(\phi_{n_{j_m}}) = E_{P_0^{n_{j_m}}}[\phi_{n_{j_m}} L_{n_{j_m},f}] \to E[\bar{\phi} \exp\{h'Z^* - \frac{1}{2}h'\Sigma_0^{-1}h + Z^{\perp} - \frac{1}{2}\sigma_{\perp}^2\}], \tag{11}$$

where $E[\cdot]$ denotes the expectation w.r.t. $(\bar{\phi}, Z^{*\prime}, Z^{\perp})'$. Suppose that the test sequence ϕ_n is such that its weak limit $\bar{\phi}$ is independent of $Z^{\perp}$. It occurs e.g. for tests that are based on LAN GMM estimators such that $\hat{\theta}_n = \theta_0 + \Sigma_0 \frac{1}{\sqrt{n}} Z_n^* + o_{P_0^n}(\frac{1}{\sqrt{n}})$, or more generally tests that are based asymptotically on transformations of $\frac{1}{\sqrt{n}} \sum_{i=1}^n g(y_i, \theta_0)$. This class includes the W, LM and LR statistics. Then, since $E[\exp\{Z^{\perp} - \frac{1}{2}\sigma_{\perp}^2\}] = 1$, the expectation in the RHS of (11) is equal to $E[\bar{\phi} \exp\{h'Z^* - \frac{1}{2}h'\Sigma_0^{-1}h\}] = \frac{|\Sigma_0|^{1/2}}{(2\pi)^{p/2}} \int m_{\bar{\phi}}(z^*) \exp\{h'z^* - \frac{1}{2}h'\Sigma_0^{-1}h - \frac{1}{2}z^{*\prime}\Sigma_0 z^*\} dz^* = E_h[m_{\bar{\phi}}(Z)]$, where $m_{\bar{\phi}}(z^*) = E[\bar{\phi} | Z^* = z^*]$ is the conditional expectation under the joint distribution of $(\bar{\phi}, Z^{*\prime})'$, and $E_h[\cdot]$ denotes expectation w.r.t. random variable $Z \sim \mathcal{N}(\Sigma_0^{-1}h, \Sigma_0^{-1})$ with parameter h. It follows:

$$\beta_{n_{j_m},f}(\phi_{n_{j_m}}) \to E_h[m_{\bar{\phi}}(Z)], \tag{12}$$

for a further subsequence n_{j_m} of any subsequence n_j, and a test function $m_{\bar{\phi}}$ in the Gaussian experiment $Z \sim \mathcal{N}(\Sigma_0^{-1}h, \Sigma_0^{-1})$. Since $\tilde{Z} := \Sigma_0 Z \sim \mathcal{N}(h, \Sigma_0)$, we can write the limit in (12) equivalently as $E_h[\tilde{\phi}(\tilde{Z})]$ for a test $\tilde{\phi}$ in the Gaussian experiment $\tilde{Z} \sim \mathcal{N}(h, \Sigma_0)$. The matrix Σ_0 corresponds to the inverse of the Fisher Information matrix in usual likelihood theory for the Gaussian experiment with unknown h and known Σ_0. Moreover, $\tilde{Z}$ corresponds to the GLS estimator, i.e., the sufficient statistic in the Gaussian linear regression model implied by (10).

We now use the subsequence convergence in (12) to get an upper bound in the maximin sense, i.e., the optimal power against the least favorable directions among the local alternative hypotheses (LR, Chapters 8 and 13.5.3). For linear hypotheses on vector h in the Gaussian experiment $\tilde{Z} \sim$

$\mathcal{N}(h, \Sigma_0)$, the chi-square test is maximin optimal (see Lemma 3 and Problem 8.29 in LR) and we can build on the arguments in the proof of Theorem 13.5.4 in LR to get the next result.

Proposition 2 *Let ϕ_n be a sequence of tests such that $E_{P_0^n}[\phi_n] \to \alpha$, with $\alpha \in (0,1)$, and the subsequence weak limit is independent of $Z^\perp$. Under Assumptions 3-5, we have for any $\lambda_{nc}^2 > 0$:*

$$\limsup_{n\to\infty} \inf\{\beta_{n,f}(\phi_n) \ : \ h'A'(A\Sigma_0 A')^{-1}Ah \geq \lambda_{nc}^2\} \leq 1 - F_{\chi^2(r,\lambda_{nc}^2)}(c_{r,1-\alpha}), \tag{13}$$

where $F_{\chi^2(r,\lambda_{nc}^2)}$ is the cdf of the chi-square distribution with r degrees of freedom and non-centrality parameter λ_{nc}^2, and $c_{r,1-\alpha}$ is the $(1-\alpha)$-quantile of the central chi-square distribution with r degrees of freedom.

The quadratic form $h'A'(A\Sigma_0 A')^{-1}Ah$ is the non-centrality parameter of the non-central chi-square distribution of $\zeta = \tilde{Z}'A'(A\Sigma_0 A')^{-1}A\tilde{Z}$ under the Gaussian experiment $\tilde{Z} \sim \mathcal{N}(h, \Sigma_0)$. The test that rejects for large values of ζ is maximin optimal for the linear hypothesis $Ah = 0$.

The W, LM and LR GMM tests are $\phi_n = \{\xi_n^U \geq c_{r,1-\alpha}\}$, for $U = W, LM, LR$. Under standard regularity conditions, which we establish for the sphericity test in the FA model in the next section, they satisfy the next assumption.

Assumption 6 *For the W, LM and LR GMM tests, ϕ_n meets the conditions of Proposition 2 and the asymptotic local power is such that $\beta_{n,f}(\phi_n) \to 1 - F_{\chi^2(r,\lambda_{nc}^2)}(c_{r,1-\alpha})$, with $\lambda_{nc}^2 = h'A'(A\Sigma_0 A')^{-1}Ah$.*

In particular, independence of the weak limit $\bar{\phi}$ from $Z^\perp$ holds, because the GMM tests are asymptotically quadratic forms of affine transformations of the sample orthogonality vector $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(y_i, \theta_0)$. The asymptotic power is increasing with λ_{nc}^2. Then, Inequality (13) holds as an equality, and we deduce the maximin optimality for the trinity of GMM tests.

Corollary 1 *Under Assumption 3-6, the W, LM and LR GMM tests of the null hypothesis $H_0 : \ a(\theta_0) = 0$ are asymptotically maximin optimal.*

4 Maximin Sphericity Tests

In the framework of Section 2 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 optimality with the broad maximin results of the previous section. To see this, let us study the asymptotic power of the test statistics against local alternative hypotheses in which we have a local deviation from sphericity. Specifically, under $H_{1,loc}(k)$, we use the matrix $V_{\varepsilon,loc} := \bar{\sigma}^2 I_T + \frac{1}{\sqrt{n}}\Delta_\varepsilon$, where $\bar{\sigma}^2 > 0$ and $\Delta_\varepsilon \neq 0$ is a diagonal matrix normalized such that $tr(\Delta_\varepsilon) = 0$. Variance $V_{\varepsilon,loc}$ is positive-definite for n large enough. The normalization $tr(\Delta_\varepsilon) = 0$ is feasible by subtracting a multiple of the identity and including the latter in the spherical component. By setting $tr(\Delta_\varepsilon) = 0$, we use $r = T - 1$ parameters to describe the local alternative, that is the number of restrictions of the test. Under the local alternative $H_{1,loc}$, the parameter θ_n is such that $a(\theta_n) = \frac{1}{\sqrt{n}}\delta$, with $\delta := L'_{1_T} diag(\Delta_\varepsilon)$.

To link this section with the previous one, let us characterize the heterogeneity and cross-sectional dependence in the sphericity test for the FA model. We have $Y = [x_1 : \cdots : x_n]S$, with $x_i = \mathfrak{m}_i(\theta) + V_{\varepsilon,n}^{1/2} w_i$, and $S := \check{\Sigma}^{1/2}$, where the mean $\mathfrak{m}_i(\theta) = \mu s_i + F\gamma_i$ involves scalar s_i, i.e., the ith element of vector $(S^{-1})'1_n$, vector γ_i, i.e. the ith column of $\beta' S^{-1}$, and the variance is $V_{\varepsilon,n} = \bar{\sigma}^2 I_T + \frac{1}{\sqrt{n}}\Delta_\varepsilon$. Hence, we have $q_{n,i}(x_i) = |V_{\varepsilon,n}|^{-1/2}\varphi_i\left(V_{\varepsilon,n}^{-1/2}(x_i - \mathfrak{m}_i(\theta_0))\right)$, where the pdf φ_i of vector w_i is normalized with mean zero and variance the identity matrix for all i. The score function for the q.m.d. condition in Assumption 3 is $f_i^q(x_i) = -\frac{1}{2\bar{\sigma}^2}[\nabla \log \varphi_i\left(\bar{\sigma}^{-1}(x_i - \mathfrak{m}_i(\theta_0))\right)]'\Delta_\varepsilon$ $(\bar{\sigma}^{-1}(x_i - \mathfrak{m}_i(\theta_0)))$, where ∇ denotes the gradient operator. In Proposition 3 below we show that Assumptions 3 and 6 are met in the FA model under the regularity conditions in our Assumptions A.1-A.8. Moreover, the mean $\mathfrak{m}_i(\theta)$ is affine in the fixed effects. As already mentioned, in our FA-GMM setting, we avoid the incidental parameter problem through the standardizations $\bar{\beta} = 0$ and $\tilde{V}_\beta = I_k$. On the contrary, if the mean function $\mathfrak{m}_i(\theta)$ is nonlinear in the fixed effects, then we get inconsistent estimates for θ_0, when T is fixed. As discussed in Hahn and Newey (2004) for example, this inconsistency occurs because only a finite number of observations are available to estimate each individual effect. Hence, the estimation error for the individual effects does not vanish as the sample size n grows, and this error contaminates the estimates of parameters of interest. Arellano and Bonhomme (2012) and Bonhomme (2012) investigate the incidental parameter problem in random coefficient models for fixed T. The next proposition states the asymptotic distributional equivalence of the trinity of GMM test statistics for sphericity under local alternative hypotheses. Appendix I of OA gives its detailed proof.

Proposition 3 *Under Assumptions 1, 2 with condition (d) replaced by condition* $\frac{\log n}{n^{1+\rho/2}} \sum_{m=1}^{J_n} b_{m,n}^3 + \frac{\log}{n^{(\rho+\alpha)/2}} \sum_{m=1}^{J_n} b_{m,n}^2 = o(1)$*, A.1-A.8, and the local alternative hypothesis* $H_{1,loc}(k)$*, as* $n \to \infty$ *and* T *is fixed, we have* $\xi_n^U = \xi_n^W + o_p(1)$*, for* $U = LM, LR$*, and* $\xi_n^U \Rightarrow \chi^2(T-1, \lambda_{loc}^2)$*, for* $U = W, LM, LR$*, where the noncentrality parameter of the chi-square random variable* $\chi^2(T-1, \lambda_{loc}^2)$ *with* $T-1$ *degrees of freedom is*

$$\lambda_{loc}^2 := \delta' \Omega_W^{-1} \delta = diag(\Delta_\varepsilon)' L_{1_T} (L_{1_T}' \Sigma_{V_\varepsilon} L_{1_T})^{-1} L_{1_T}' diag(\Delta_\varepsilon). \tag{14}$$

Moreover, Assumptions 3-6 for establishing validity of the Gaussian experiment hold.

The maximin properties of the GMM tests in FA is thus a direct consequence of Corollary 1 by taking $r = T - 1$ and $\lambda_{nc}^2 = \lambda_{loc}^2$. The asymptotic distribution under the null hypothesis of sphericity obtains by setting $\delta = 0$. The proof of Proposition 3 in Appendix D.3 relies on adapting the standard distributional results of the GMM literature (see e.g. Newey and McFadden (1994)). In particular, for FA-GMM inference, we need to explicitly address the normalization of the parameters corresponding to the latent factors. The asymptotic equivalence stated in Proposition 3 breaks down if we do not use an optimal weighting matrix in the GMM criterion, and we end up with weighted sums of noncentral chi-square random variables. For example, it happens when we compare directly a constrained and unconstrained Gaussian pseudo likelihood (see FGS), which do not exploit optimal weighting in their construction. We can show that the second-order expansion of the Gaussian pseudo likelihood criterion underlying the FA estimator of FGS yields (minus) a GMM criterion with a non-optimal weighting matrix $(V_y^0)^{-1} \otimes (V_y^0)^{-1}$. Besides, we also have asymptotic distributional equivalence with a test statistic based on the Hausman principle (Hausman (1978)), i.e., considering a test statistic based on a weighted quadratic form in the difference between constrained estimator $\hat{\theta}^c$ (efficient but not robust to deviation from sphericity) and unconstrained estimator $\hat{\theta}$ (robust but not as efficient).

5 Monte Carlo Experiments

This section gives a Monte Carlo assessment of size and power for our trinity of sphericity tests under non-Gaussian errors. Let us start with a description of the DGP similar to the one underlying the MC experiments in FGS. The good numerical performance reported in Section F.5 of OA for FA-GMM estimates motivates us to choose a Newton-Raphson algorithm for the criterion minimization (see description in Section F.2.1 of OA). In the DGP, the betas are $\beta_i \overset{i.i.d.}{\sim} N(0, I_k)$, with $k = 2$, and the matrix of factor values is $F = V_\varepsilon^{1/2} U \Gamma^{1/2}$, where $U = \tilde{F}(\tilde{F}'\tilde{F})^{-1/2}$ and $vec(\tilde{F}) \sim N(0, I_{Tk})$. The diagonal matrix $\Gamma = T diag(3, 2)$ yields $\frac{1}{T} F' V_\varepsilon^{-1} F = diag(3, 2)$, i.e., the "signal-to-noise" ratios equal 3 and 2 for the two factors. Under the null hypothesis of sphericity (DGP1), we set $V_\varepsilon = I_T$, and we generate the idiosyncratic errors by $\varepsilon_{i,t} = h_{i,t}^{1/2} z_{i,t}$, where $h_{i,t} = c_i + \alpha_i h_{i,t-1} z_{i,t-1}^2$, with $z_{i,t} \sim IIN(0, 1)$. We use the constraint $c_i = \sigma_{ii}(1 - \alpha_i)$ with uniform draws for the idiosyncratic variances $V[\varepsilon_{i,t}] = \sigma_{ii} \sim U[1, 4]$, so that $V[\varepsilon_{i,t}] = \frac{c_i}{1-\alpha_i} = \sigma_{ii}$. Such a setting allows for cross-sectional heterogeneity in the variances of $\varepsilon_{i,t}$ under sphericity with $\bar{\sigma}^2 = \lim_{n\to\infty} \frac{1}{n} \sum_i \sigma_{ii}$. The ARCH parameters are uniform draws $\alpha_i \overset{i.i.d.}{\sim} U[0.2, 0.5]$ with an upper boundary of the interval ensuring existence of fourth-order moments. Under the alternative hypothesis (DGP2), we generate the diagonal elements of $V_\varepsilon = diag(h_1, ..., h_T)$ through a common time-varying component in idiosyncratic volatilities (Barigozzi and Hallin (2016), Renault et al. (2023)) via the ARCH $h_t = 0.6 + 0.5 h_{t-1} z_{t-1}^2$, with $z_t \sim IIN(0, 1)$. This common component with unconditional variance $V[h_t^{1/2} z_t] = 0.6/(1 - 0.5) = 1.2$ induces a deviation from spherical errors. We generate the idiosyncratic errors by $\varepsilon_{i,t} = h_t^{1/2} h_{i,t}^{1/2} z_{i,t}$, where $h_{i,t} = c_i + \alpha_i h_{i,t-1} z_{i,t-1}^2$, with $z_{i,t} \sim IIN(0, 1)$ mutually independent of z_t. We use the constraint

$c_i = \sigma_{ii}(1-\alpha_i)$ with uniform draws for the idiosyncratic variances $V[\varepsilon_{i,t}] = \sigma_{ii} \overset{i.i.d.}{\sim} U[1,4]$, so that $V[\varepsilon_{i,t}/h_t^{1/2}] = \frac{c_i}{1-\alpha_i} = \sigma_{ii}$. Such a setting allows for cross-sectional heterogeneity in the variances of the scaled $\varepsilon_{i,t}/h_t^{1/2}$. To study local power (DGP3), we set the diagonal elements of $V_\varepsilon = diag(\check{h}_1, ..., \check{h}_T)$ with $\check{h}_t = 1.2 + 1/\sqrt{n}$, $t = 1, ..., T-1$, and $\check{h}_T = 1.2 - (T-1)/\sqrt{n}$, so that $\varepsilon_{i,t} = \check{h}_t^{1/2} h_{i,t}^{1/2} z_{i,t}$. We generate 5000 panels of returns of size $n \times T$ for each of the 100 draws of the $T \times k$ factor matrix F and common ARCH process h_t, $t = 1, ..., T$, in order to keep the factor values constant within repetitions, but also to study the potential heterogeneity of size and power results across different factor paths. The factor betas β_i, idiosyncratic variances σ_{ii}, and individual ARCH parameters α_i are the same across all repetitions in all designs of the section. We opt for three different cross-sectional sizes $n = 500, 1000, 5000$, and three values of time-series dimension $T = 6, 12, 24$. The p-values are computed over $5,000$ draws.

We provide the empirical size and power results in % in Table 1 for the W test. The results for the LM and LR tests gathered in Section G of the OA are similar. The number of latent factors is set equal to two.[14] Size of all tests is close to its nominal level 5%, with size distortions smaller than 1%, except for $n = 500$. Power computation is not size adjusted. Global power is close to 100%. Local power ranges is above 22% and reaches 100% for $T = 24$. The asymptotic local power can be computed from the noncentrality parameter λ^2_{loc} given in (14) using the DGP parameters, and its value for $T = 6$, 12 and 24 is 19%, 83%, and 100%. The MC outputs for $n = 5000$ are close to those numbers as expected from asymptotic theory. The approximate constancy of local power w.r.t. n, for large n, is coherent with theory implying convergence to asymptotic local power. We can conclude that our testing procedure for sphericity with FA-GMM estimates works well in our simulations, and should be relevant for applied work.

As a final remark, when sphericity fails, the huge Bias, Standard Deviation, and Root Mean Square Error exhibited by the PCA estimates in the numerical study of Section F.5 in OA play against relying on them in short panels. It is not the case for the unconstrained FA-GMM and PML

[14]The numbers are similar with an estimated k.

estimators.

ξ_n^W	Size (%)			Global Power (%)			Local Power (%)		
T	6	12	24	6	12	24	6	12	24
$n = 500$	6.5	6.2	6.2	93	99	100	29	98	97
	(0.4)	(0.3)	(0.3)	(17.2)	(0.0)	(0.0)	(11.4)	(4.6)	(1.2)
$n = 1000$	5.8	5.7	5.6	97	100	100	26	95	100
	(0.3)	(0.3)	(0.3)	(12.9)	(0.0)	(0.0)	(10.3)	(8.0)	(0.0)
$n = 5000$	5.4	5.3	5.2	100	100	100	22	90	100
	(0.3)	(0.3)	(0.3)	(0.0)	(0.0)	(0.0)	(9.0)	(12.9)	(0.0)

Table 1: For each sample size combination (n, T), we provide the average size, power, and local power in % for the test statistic ξ_n^W under DGP1-3. Nominal size is 5%. In parentheses, we report the standard deviations for size, power, and local power across 100 different draws of the factor path. The number of latent factors is set equal to two.

6 Empirical Application

In this section, we test the sphericity hypothesis in short subperiods of the Center for Research in Securities Prices (CRSP) panel of stock returns. We consider monthly returns of U.S. common stocks trading on the NYSE, AMEX, or NASDAQ between January 1963 and December 2023, 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).[15] We implement the sphericity tests using nonoverlapping windows of $T = 20$ months, thereby ensuring that we can allow for up to 14 latent factors in each subperiod

[15]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.

(see FGS).[16] 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 for the number of factors to get a consistent estimate $\hat{k}$ of the number of latent factors based on the likelikood ratio test of FGS.[17] We compute the variance-covariance estimator 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 values for the W test statistic over time for each subperiod in Figure 1 on the right vertical axis, while the red horizontal segments give $\hat{k}$, i.e., the estimated number of latent factors on the right vertical axis. The red dotted line corresponds to the critical value of a $\chi^2(19)$ distribution at significance level $5\%/35$, where 35 is the number of windows, i.e., a Bonferroni correction. We reject strongly the null hypothesis of sphericity,[18] and this for all subperiods with values ranging from 64.6 to 466.8. Evidence against the null hypothesis is not necessarily larger around market downturns with large $\hat{k}$. We prefer to plot the values of the Wald test statistic instead of their associated p-values since the p-values are tiny. For example, the asymptotic distribution, which is here the one of a $\chi^2(19)$ r.v., already gives a p-value of 5.3557×10^{-13} when $\xi_n^W = 100$. In line with our Monte Carlo results in Table 1, since we have large cross-sections of stocks in each subperiod, we benefit from the good power

[16] Our empirical results also hold with $T = 16$ and $T = 24$ or overlapping windows. For longer sample sizes such as $T = 36$ and $T = 60$, sphericity is also strongly rejected.

[17] Alternatively, we can use the specification test (J-test) of Hansen (1982) based on the estimated value of the unconstrained optimal FA-GMM criterion (see Ahn et al. (2013) for use in a panel data model with random interactive effects and i.i.d. errors) instead of the classical FA likelihood ratio statistic to determine the number of factors. Its asymptotic distribution is given by a chi-square distribution with df degrees of freedom under the null hypothesis of k factors. In our empirics and unreported simulations, we find that the use of optimal versus non-optimal weighting does not affect the estimated value $\hat{k}$.

[18] Unreported results show that we also reject most of the time (60%) sphericity on monthly U.S. macroeconomic indicators (FRED-MD database developed by McCracken and Ng (2016)) with nonoverlapping windows of $T = 20$.

properties of our testing procedure. The pictures are very similar for the LM and LR tests, and thus omitted. The strong rejection is corroborated by running sphericity tests on windows of 6 months within subperiods (15 tests inside each subperiod of length $T = 20$). It thus points to a time-varying behavior of $V_{\varepsilon,tt}^{1/2}$, especially in the second part of the sample. The rejection of sphericity might be explained by the presence of a common component driving the variance of the error terms; see e.g. Barigozzi and Hallin (2016), Renault et al. (2023) for theory and empirical evidence in favour of variance factors.

7 Concluding Remarks

Sphericity is key to achieve consistency of factor estimates with PCA in a large-n and fixed-T setting. This paper provides optimal maximin GMM tests to check whether it holds on the data or not. If not, empirical researchers should refrain from running PCA in short panels. Our empirics show that the assumption of sphericity is doubtful in our financial data. The optimal maximin properties in our FA-GMM framework are a by-product of obtaining the limit Gaussian experiment in strongly identified GMM models under a block-dependence structure and unobserved heterogeneity. They have a broader pertinence in panel models than only in our FA setting. The characterisation of the Gaussian experiment in a non-i.i.d. context is new and is of independent interest. It might be useful for other applications such as designing Bayesian priors and optimal similar tests or AUMP tests.

References

Ahn, S., Lee, H., and Schmidt, P., 2001. GMM estimation of linear panel data models with time-varying individual effects. Journal of Econometrics 101(2), 219-255.

Ahn, S., Lee, H., and Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. Journal of Econometrics 174(1), 1-14.

Alvarez, J., and Arellano M., 2022. Robust likelihood estimation of dynamic panel data models. Journal of Econometrics 226(1), 21-61.

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.

Andrews, D., 1998. Hypothesis testing with a restricted parameter space. Journal of Econometrics 84(1), 155-199.

Andrews, I., and Mikusheva, A., 2022. Optimal decision rules for weak GMM. Econometrica 90(2), 715-748.

Andrews, D., Moreira, M, and Stock, J., 2006. Optimal two-sided invariant tests for instrumental variables regression. Econometrica 74(3), 715-752.

Ang, A., Liu, J., and Schwarz, K., 2020. Using stocks or portfolios in tests of factor models. Journal of Financial and Quantitative Analysis 55(3), 709-750.

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

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

Bai, J., 2013. Fixe-effects dynamic panel models, a factor analytical approach. Econometrica 81(1), 285-314.

Bai, J., 2024. Likelihood approach to dynamic panel models with interactive effects. Journal of Econometrics 240(1), 105636.

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

Bartlett, M., 1951. The effect of standardization on a χ^2 approximation in factor analysis. Biometrika 38(3/4), 337-344

Bonhomme, S., 2012. Functional differencing. Econometrica 80(4), 1337-1385.

Bonhomme, S., and Denis, A., 2024a. Fixed effects and beyond bias reduction, groups, shrinkage, and factors in panel data. Working paper University of Chicago.

Bonhomme, S., and Denis, A., 2024b. Estimating heterogeneous effects: applications to labor economics. Labour Economics 91, 102638.

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

Chamberlain, G., and Moreira, M., 2009. Decision theory applied to a linear panel data model. Econometrica 77(1), 107-133.

Chamberlain, G., and Rothschild, M., 1983. Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. Econometrica 51(5), 1281-1304.

Chen, H., Dou, W., and Kogan, L., 2024. Measuring “dark matter" in asset pricing models. Journal of Finance 79(2), 843-902.

Chen, X., and Santos, A., 2018. Overidentification in regular models. Econometrica 86(5), 1771-1817.

Choi, S., Hall, W., and Schick, A., 1996. Asymptotically uniformly most powerful tests in parametric and semiparametric models. Annals of Statistics 24(2), 841-861.

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

de la Peña, V., 1992. Decoupling and Khintchine’s inequalities for U-statistics. The Annals of Probability 20(4), 1877-1892.

Eichenbaum, M., Hansen, L., and Singleton, K., 1988. A time series analysis of representative agent models of consumption and leisure choice under uncertainty. Quarterly Journal of Economics 103(1), 51-78.

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.

Fan, J., Furger, A., and Xiu, D., 2016. Incorporating Global Industrial Classification Standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. Journal of Business and Economic Statistics 34(4), 489-503.

Fortin, A.-P., Gagliardini, P., and Scaillet, O., 2025. Eigenvalue tests for the number of latent factors in short panels. Journal of Financial Econometrics, 23(1), nbad024.

Fortin, A.-P., Gagliardini, P., and Scaillet, O., 2023b. Latent factor analysis in short panels. Swiss Finance Institute DP 2023.44.

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

Gabaix, X., Koijen, R., Richmond, R., and Yogo, M., 2023. Asset embeddings. Working paper University of Chicago.

Giné, E., Latala, R., and Zinn, J., 2000. Exponential and moment inequalities for U-Statistics. In *High Dimensional Probability II*, ed. E. Giné, D. Mason, and J. Wellner.

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

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

Hansen, L., 1982. Large sample properties of Generalized Method of Moments estimators. Econometrica 50(4), 1029-1054.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46(6), 1251-1271.

Hautsch, N., Okhrin, O., and Ristig, A., 2023. Maximum-likelihood estimation using the zig-zag algorithm. Journal of Financial Econometrics 21(4), 1346-1375.

Hayakawa, K., Pesaran, H., and Smith, V., 2023. Short T dynamic panel data models with individual, time and interactive effects. Journal of Applied Econometrics 38(6), 940-967.

John, S., 1972. The distribution of a statistic used for testing sphericity of normal distributions. Biometrika 59(1), 169-173.

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.

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

Le Cam, L., 1986. Asymptotic theory of statistical inference. John Wiley & Sons.

Ledoit, O., and Wolf, M., 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Annals of Statistics 30(4), 1081-1102.

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.

Manton, J., Mahony, R., and Hua, Y., 2003. The geometry of weighted low-rank approximation. IEEE Transactions on Signal Processing 51(2), 500-514.

Marcinkiewicz, J., and Zygmund, A., 1937. Sur les fonctions indépendantes. Fundamenta Mathematicae 28, 60-90.

Mauchly, J., 1940. Significance test for sphericity of a Normal n-Variate Distribution. The Annals of Mathematical Statistics 11(2), 204-209.

McConnell, T., and Taqqu, M., 1986. Decoupling inequalities for multilinear forms in independent symmetric Banach-valued random variables. Probability Theory and Related Fields 75, 499-507.

McCracken, M., and Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. Journal of Business and Economic Statistics 34(4), 574-589.

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

Moreira, M., 2003. A conditional likelihood ratio test for structural models. Econometrica 71(4), 1027-1048.

Mueller, U., 2011. Efficient tests under a weak convergence assumption. Econometrica 79(2), 395-435.

Newey, W., 1985. Generalized method of moments specification testing. Journal of Econometrics 29(3), 229-256.

Newey, W., and McFadden, D.,1984. Large sample estimation and hypothesis testing, in Handbook of Econometrics, Volume IV, R. Engle and D. McFadden Eds., 2111-2245.

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

Onatski, A., 2025. Comment on: Eigenvalue tests for the number of latent factors in short panels. Journal of Financial Econometrics, 23(1), nbad028.

Onatski, A., Moreira, M., and Hallin, M., 2013. Asymptotic power of sphericity tests for high-dimensional data. Annals of Statistics 41(3), 1204-1231.

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

Renault, E., Van Der Heijden, T., and Werker, B., 2023. Arbitrage pricing theory for idiosyncratic variance factors. Journal of Financial Econometrics 21(5), 1403-1442.

Romano, J., Shaikh, A., and Wolf, M., 2010. Hypothesis testing in econometrics. Annual Review of Economics 2, 75-104.

Satorra, A., 1989. Alternative test criteria in covariance structure analysis: a unified approach. Psychometrika 54(1), 131-151.

Schott, J., 2006. A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. Journal of Multivariate Analysis 97(4), 827-843.

Sowell, F., 1996. Optimal tests for parameter instability in the Generalized Method of Moments framework. Econometrica 64(5), 1085-1107.

van der Vaart, A., 1998. Asymptotic statistics. Cambridge University Press.

van der Vaart, A., 2002. The statistical work of Lucien Le Cam. Annals of Statistics 30(3), 631-682.

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

Zaffaroni, P., 2025. Factor models for conditional asset pricing. Journal of Political Economy 133 (8), 2615-2642.

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 in the FA model. We denote a generic constant by $C > 0$. Let $\theta_0 = (\mu_0', vec(F_0)', diag(V_\varepsilon^0)')' = (\mu_0', \vartheta_0')'$ denote the vector of true parameter values in the FA model with k latent factors, which is an interior point of compact set $\Theta = \mathcal{M} \times \mathcal{T}$, with $\mathcal{M} \subset \mathbb{R}^T$ and $\mathcal{T} \subset \{\vartheta \in \mathbb{R}^{(k+1)T} \ : \ V_\varepsilon \text{ is positive definite}, \ h(\vartheta) = 0\}$ and $h(\vartheta)$ is the $k(k-1)/2 \times 1$ vector of the above-diagonal elements of $F'V_\varepsilon^{-1}F$.

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* $|\check{\sigma}_{i,j}| \leq C$*, for all* i, j*.*

Assumption A.3 *In the FA model with* k *latent factors, we have:* $\Sigma(\vartheta) = \Sigma(\vartheta_0)$, $\vartheta \in \mathcal{T} \Rightarrow \vartheta = \vartheta_0$*, up to sign changes in the columns of* F*.*

Assumption A.4 *Matrix* $M_{F_0,V_\varepsilon^0} \odot M_{F_0,V_\varepsilon^0}$ *is non-singular, where* $M_{F,V_\varepsilon} := I_T - F(F'V_\varepsilon^{-1}F)^{-1}F'V_\varepsilon^{-1}$ *is the GLS oblique projector and* $\odot$ *denotes the Hadamard product (i.e., element-wise matrix product).*

Assumption A.5 *(a) The* $\frac{T(T+1)}{2} \times \frac{T(T+1)}{2}$ *symmetric matrix* $D = \lim_{n\to\infty} D_n$ *exists, where* $D_n = \frac{1}{n}\sum_{i=1}^n \check{\sigma}_{ii}^2 V[vech(w_i w_i')]$*. (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* $\check{\sigma}_{ii} \leq \bar{C}$*. (c) We have* $\lim_{n\to\infty} \kappa_n = \kappa$ *for a constant* $\kappa \geq 0$*, where* $\kappa_n := \frac{1}{n}\sum_{m=1}^{J_n}\left(\sum_{i\neq j\in B_m} \check{\sigma}_{ij}^2\right)$*.*

Assumption A.6 *In the FA model with* $k+1$ *factors, (a) function* $L_0(\vartheta) = -\frac{1}{2}\log|\Sigma(\vartheta)| - \frac{1}{2}Tr\left(V_y^0\Sigma(\vartheta)^{-1}\right)$ *has a unique maximizer* $\vartheta^* = (vec(F^*)', diag(V_\varepsilon^*)')'$ *over* $\mathcal{T}$*, and (b) we have* $V_y^0 \neq F^*(F^*)' + V_\varepsilon^*$*.*

Assumption A.7 *Matrix* $Q_{\mathcal{B}} := \lim_{n\to\infty} \frac{1}{n}\mathcal{B}'\check{\Sigma}\mathcal{B}$ *is positive-definite, where* $\mathcal{B} = [1_n : \beta]$*. Moreover,* $E[w_{i,t}w_{i,s}w_{i,r}] = 0$*, for all* t, s, r*.*

Assumption A.8 *The pdf* φ_i *of random vector* w_i *is such that (a)* $E[\|\nabla \log \varphi_i(w_i)\|^{2r}\|w_i\|^{2r}] \leq C$ *and* $E[\|w_i\|^{4r}] \leq C$*, for all* i *and some constants* $C > 0$ *and* $r > 1$*, and (b) it holds* $\int \left[\sqrt{\varphi_i[(I_T + \frac{1}{\bar{\sigma}^2\sqrt{n}}\Delta_\varepsilon)x]} - \sqrt{\varphi_i(x)} - \frac{1}{2\bar{\sigma}^2\sqrt{n}}[\nabla \log \varphi_i(x)]'\Delta_\varepsilon x\sqrt{\varphi_i(x)}\right]^2 dx = O(n^{-\alpha})$*, uniformly in* i*, with* $\alpha > 1$*.*

Assumption A.1 states standard normalization restrictions and uniform bounds on factor loadings. Assumption A.2 gives uniform bounds on covariances of the idiosyncratic errors. Assumption A.3 implies global identification in the FA model (see Lemma 5 in FGS). It also implies that the non-zero eigenvalues of matrix $V_y^0 V_\varepsilon^{-1}$ are distinct. Otherwise, the normalization condition of a diagonal $F'V_\varepsilon^{-1}F$ would fail to fix the rotational invariance of latent factors up to sign change. Assumptions A.1-A.3, together with Assumptions 1 and 2, yield consistency of FA-GMM estimators (see Section D.1). Assumption A.4 is the local identification condition in the FA model (Lemma 7 in FGS). We use it to establish well-defined asymptotic expansions of FA-GMM estimators and test statistics (see Sections D.1, D.2 and D.4). We use Assumption A.5 together with Assumptions 2 and A.2 to invoke a CLT based on a multivariate Lyapunov condition to establish the asymptotic normality of FA-GMM estimators and asymptotic chi-square distribution of the trinity of test statistics. It extends Lemma 2 in FGS to any $r > 1$, i.e., requiring existence of error moments slightly above order 4 (finite kurtosis). In the verification of the Lyapunov condition, we establish bounds for higher-order moments of U-statistics generalizing the results in McConnell and Taqqu (1986) beyond symmetric variables using the Marcinkiewicz and Zygmund (1937) inequality and results in de la Peña (1992) and Giné, Latala and Zinn (2000). The mild Assumption A.5(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.5(c), in order to have κ_n bounded, we need either mixing dependence in idiosyncratic errors within blocks, i.e., $|\check{\sigma}_{i,j}| \leq C\rho^{|i-j|}$, for $i, j \in B_m$ and $0 \leq \rho < 1$, or vanishing correlations, i.e., $|\check{\sigma}_{i,j}| \leq C b_m^{-\bar{s}}$, for all $i \neq j \in B_m$ and a constant $\bar{s} \geq 1/2$, with blocks of equal size. In Assumption A.6, 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 for the number of factors under global alternative hypotheses (see proof of Proposition 4 of FGS). We use a sequential testing procedure based on the LR statistic to select consistently the number of factors. We use Assumption A.7 to apply a Lyapunov CLT (see proof of Lemma 7 in FGS) when deriving the asymptotic normality of the FA estimators as well as a feasible CLT for the sample orthogonality vector based on a consistent estimator of V_g (Section D.3). Finally, we use Assumption A.8 to show that the FA model meets the q.m.d. condition in Assumptions 3, and the uniform bounds in Assumption 4, needed to establish the Gaussian experiment in Section 3. Assumption A.8 holds e.g. for a Gaussian distribution.

B Proofs of Propositions 1-2 and Lemmas 1-2

Proof of Lemma 1: *Part (a).* We use $\prod_{i=1}^N b_i - \prod_{i=1}^N a_i = \sum_{i=1}^N (\prod_{j=1}^{i-1} a_j)(b_i - a_i)(\prod_{j=i+1}^N b_j)$, for real sequences a_i and b_i. Then, we have:

$$\begin{aligned}
\sqrt{q_n^m(X_m)} - \sqrt{q_0^m(X_m)} &= \prod_{i\in B_m} \sqrt{q_{n,i}(x_i)} - \prod_{i\in B_m} \sqrt{q_{0,i}(x_i)} \\
= \quad & \frac{1}{2\sqrt{n}} \sum_{i\in B_m} \Big(\prod_{j\in B_m: j\leq i} \sqrt{q_{0,j}(x_j)}\Big) f_i^q(x_i) \Big(\prod_{j\in B_m:j>i} \sqrt{q_{n,j}(x_j)}\Big) \\
& + \sum_{i\in B_m} \Big(\prod_{j\in B_m:j<i} \sqrt{q_{0,j}(x_j)}\Big) R_{n,i}^q(x_i) \Big(\prod_{j\in B_m:j>i} \sqrt{q_{n,j}(x_j)}\Big), \qquad \text{(B.1)}
\end{aligned}$$

from Assumption 3. We get $\sqrt{q_n^m(X_m)} - \sqrt{q_0^m(X_m)} = \frac{1}{2\sqrt{n}} f_m(X_m) \sqrt{q_0^m(X_m)} + R_n^m(X_m)$, where:

$$\begin{aligned}
R_n^m(X_m) \quad &:= \quad \frac{1}{2\sqrt{n}} \sum_{i\in B_m} \Big(\prod_{j\in B_m:j\leq i} \sqrt{q_{0,j}(x_j)}\Big) f_i^q(x_i) \Big(\prod_{j\in B_m:j>i} \sqrt{q_{n,j}(x_j)} - \prod_{j\in B_m:j>i} \sqrt{q_{0,j}(x_j)}\Big) \\
&+ \sum_{i\in B_m} \Big(\prod_{j\in B_m:j<i} \sqrt{q_{0,j}(x_j)}\Big) R_{n,i}^q(x_i) \Big(\prod_{j\in B_m:j>i} \sqrt{q_{n,j}(x_j)}\Big) \\
&=: \quad \frac{1}{2\sqrt{n}} \sum_{i\in B_m} R_{n,i,1}^m(X_m) + \sum_{i\in B_m} R_{n,i,2}^m(X_m). \qquad \text{(B.2)}
\end{aligned}$$

Next, we bound the L^2 norms of functions $R^m_{n,i,1}$ and $R^m_{n,i,2}$ in (B.2). We have $\int [R^m_{n,i,1}(X_m)]^2 dX_m = (\int [f^q_i(x)]^2 q_{0,i}(x)dx) \int \left[\prod_{j\in B_m:j>i} \sqrt{q_{n,j}(x_j)} - \prod_{j\in B_m:j>i} \sqrt{q_{0,j}(x_j)}\right]^2 \prod_{j\in B_m:j>i} dx_j = O(\frac{b^2_{m,n}}{n})$, uniformly in i, by an argument similar to (B.1) and Assumption 3. Moreover, $\int [R^m_{n,i,2}(X_m)]^2 dX_m = \int [R^q_{n,i}(x)]^2 dx = O(\frac{1}{n^\alpha})$, uniformly in i, from Assumption 3. By the triangular inequality, we get $\int [R^m_n(X_m)]^2 dX_m = O(\varrho^2_{m,n})$, where $\varrho^2_{m,n} := b^2_{m,n}(\frac{b^2_{m,n}}{n^2} + \frac{1}{n^\alpha})$.

Part (b). We use $p_{n,i}(y_i) = \int p^m_n(Y_m)dY_{m,-i} = \int q^m_n(Y_m S^{-1}_m)dY_{m,-i}$ if $i \in B_m$, and similarly $p_{0,i}(y_i) = \int q^m_0(Y_m S^{-1}_m)dY_{m,-i}$, where $dY_{m,-i}$ denotes integration w.r.t. the variables y_j, for $j \in B_m$ with $j \neq i$. Then, from Lemma 1(a): $p_{n,i}(y_i) - p_{0,i}(y_i) = \int [q^m_n(Y_m S^{-1}_m) - q^m_0(Y_m S^{-1}_m)]dY_{m,-i} = \int [\sqrt{q^m_n(Y_m S^{-1}_m)} - \sqrt{q^m_0(Y_m S^{-1}_m)}][\sqrt{q^m_n(Y_m S^{-1}_m)} + \sqrt{q^m_0(Y_m S^{-1}_m)}]dY_{m,-i} = \frac{1}{\sqrt{n}} \int f_m(Y_m S^{-1}_m) q^m_0(Y_m S^{-1}_m)dY_{m,-i} + R^p_{n,i}(y_i) = \frac{1}{\sqrt{n}} E_{P^m_0}[f_m(Y_m S^{-1}_m)|y_i]p_{0,i}(y_i) + R^p_{n,i}(y_i)$, where $R^p_{n,i}(y_i) := 2\int R^m_n(Y_m S^{-1}_m)\sqrt{q^m_0(Y_m S^{-1}_m)}dY_{m,-i} + \int [\sqrt{q^m_n(Y_m S^{-1}_m)} - \sqrt{q^m_0(Y_m S^{-1}_m)}]^2 dY_{m,-i} = \int R^m_n(Y_m S^{-1}_m)[\sqrt{q^m_0(Y_m S^{-1}_m)} + \sqrt{q^m_n(Y_m S^{-1}_m)}]dY_{m,-i} + \frac{1}{4n} E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i]p_{0,i}(y_i) + \frac{1}{2\sqrt{n}} \int R^m_n(Y_m S^{-1}_m) f_m(Y_m S^{-1}_m)\sqrt{q^m_0(Y_m S^{-1}_m)}dY_{m,-i}$. By the triangular and Cauchy-Schwarz inequalities, we get the upper bound:

$$|R^p_{n,i}(y_i)| \leq (\int [R^m_n(Y_m S^{-1}_m)]^2 dY_{m,-i})^{1/2} \left(\sqrt{p_{0,i}(y_i)} + \sqrt{p_{n,i}(y_i)} + \frac{1}{2\sqrt{n}} E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i]^{1/2}\sqrt{p_{0,i}(y_i)}\right) + \frac{1}{4n} E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i]p_{0,i}(y_i). \quad \text{(B.3)}$$

Then, we have:

$$\int \|g(y_i,\theta_0)\| 1^\tau(y_i) R^p_{n,i}(y_i)dy_i \leq (\int [R^m_n(Y_m S^{-1}_m)]^2 dY_m)^{1/2} \Big(E_{P_{0,i}}[\|g(y_i,\theta_0)\|^2]^{1/2} + E_{P_{n,i}}[\|g(y_i,\theta_0)\|^2]^{1/2} + \frac{1}{2\sqrt{n}} E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|^2 1^\tau(y_i) E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i]\right]^{1/2}\Big) + \frac{1}{4n} E_{P_{0,i}}\left[\|g(y_i,\theta_0)\| 1^\tau(y_i) E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i]\right]. \quad \text{(B.4)}$$

Now, $\int [R^m_n(Y_m S^{-1}_m)]^2 dY_m = \int [R^m_n(X_m)]^2 dX_m \leq C\varrho^2_{m,n}$ from Lemma 1(a). To control the expectations on the RHS of (B.4), we use the Hölder inequality. Moreover, we use $E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i] = E_{P^m_0}[f_m(X_m)^2|y_i] \leq (\sum_{j\in B_m} \eta^q_j(y_i))^2$, where $\eta^q_j(y_i) := E_{P^m_0}[(f^q_j(x_j))^2|y_i]^{1/2}$. Then, $E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|^2 1^\tau(y_i) E_{P^m_0}[f_m(Y_m S^{-1}_m)^2|y_i]\right]^{1/2} \leq E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|^{2\bar{r}} 1^\tau(y_i)\right]^{1/(2\bar{r})}$

$E_{P_{0,i}}\left[E_{P_0^m}[f_m(Y_mS_m^{-1})^2|y_i]^r\right]^{1/(2r)} = O(\tau_n^{(2-r)\vee 0}b_{m,n})$, where $\bar{r} > 1$ is such that $1/r + 1/\bar{r} = 1$, because we have $E_{P_{0,i}}\left[E_{P_0^m}[f_m(Y_mS_m^{-1})^2|y_i]^r\right]^{1/(2r)} \leq \sum_{j\in B_m} E_{P_{0,i}}[(\eta_j^q(y_i))^{2r}]^{1/(2r)} \leq \sum_{j\in B_m} E_{Q_{0,j}}[(f_j^q(x_j))^{2r}]^{1/(2r)}$ and $E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|^{2\bar{r}}1^\tau(y_i)\right] \leq E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|^{2r}\right]\tau_n^{2(\bar{r}-r)}$ (we consider without loss of generality $r \leq 2$ for which $\bar{r} \geq r$). Similarly, we have the bound $E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|1^\tau(y_i)E_{P_0^m}[f_m(Y_mS_m^{-1})^2|y_i]\right] \leq E_{P_{0,i}}\left[\|g(y_i,\theta_0)\|^{\bar{r}}1^\tau(y_i)\right]^{1/\bar{r}} E_{P_{0,i}}\left[E_{P_0^m}[f_m(Y_mS_m^{-1})^2|y_i]^r\right]^{1/r} = O(\tau_n^{(3-2r)\vee 0}b_{m,n}^2)$ (we consider $r \leq 3/2$ for which $\bar{r} \geq 2r$). Then, from (B.4), we get: $\int \|g(y_i,\theta_0)1^\tau(y_i)\|R_{n,i}^p(y_i)dy_i \leq C\varrho_{m,n}\left(1 + \frac{b_{m,n}}{n^{1/2}}\tau_n^{(2-r)\vee 0}\right) + C\frac{b_{m,n}^2}{n}\tau_n^{(3-2r)\vee 0} \leq C\varrho_{m,n}\left(1+\frac{b_{m,n}}{\sqrt{n}}\right)\tau_n^{(2-r)\vee 0}$, for any $i \in B_m$, where the latter inequality follows from $\varrho_{m,n} \geq \frac{b_{m,n}^2}{n}$, and we have $(2-r)\vee 0 \geq (3-2r)\vee 0$ for $r > 1$. Next, we sum over i: $\sum_{i=1}^n \int \|g(y_i,\theta_0)\|1^\tau(y_i)R_{n,i}^p(y_i)dy_i = O\left(\tau_n^{(2-r)\vee 0}\sum_{m=1}^{J_n}\varrho_{m,n}b_{m,n}\left(1+\frac{b_{m,n}}{\sqrt{n}}\right)\right)$. Further, we use $\varrho_{m,n} \leq Cb_{m,n}(\frac{b_{m,n}}{n} + \frac{1}{n^{\alpha/2}})$. Thus, the conclusion follows from the condition

$$\frac{\tau_n^{(2-r)\vee 0}}{n^{1/2}}\sum_{m=1}^{J_n} b_{m,n}^2\left(\frac{b_{m,n}}{n} + \frac{1}{n^{\alpha/2}}\right)\left(1 + \frac{b_{m,n}}{\sqrt{n}}\right) = o(1). \tag{B.5}$$

The LHS is bounded from below by $\frac{1}{n^{3/2}}\sum_{m=1}^{J_n} b_{m,n}^3$. If one block size $b_{m,n}$ grows as $\sqrt{n}$ or faster, we see that this lower bound does not shrink to zero. In other words, a necessary condition is $b_{m,n} = o(\sqrt{n})$, for any m. Thus, (B.5) is equivalent to $\frac{\tau_n^{(2-r)\vee 0}}{n^{1/2}}\sum_{m=1}^{J_n} b_{m,n}^2\left(\frac{b_{m,n}}{n} + \frac{1}{n^{\alpha/2}}\right) = o(1)$, which is the condition in Assumption 4 with $\tau_n = n^{\frac{1}{2(2r-1)}}\log n$. Indeed, we have $\frac{\tau_n^{(2-r)\vee 0}}{n^{1/2}} = n^{-1/2+\frac{(r-2)\vee 0}{2(2r-1)}} = n^{-\rho}$, up to log terms, where $\rho = \frac{1}{2}\min\{\frac{3}{2}(1-\frac{1}{2r-1}), 1\}$. Q.E.D.

Proof of Lemma 2: We have $0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n E_{P_{n,i}}[g(y_i,\theta_n)] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n(E_{P_{n,i}}[g(y_i,\theta_0)] + \frac{1}{\sqrt{n}}E_{P_{n,i}}[\frac{\partial g(y_i,\theta_0)}{\partial\theta'}]h) + o(\frac{1}{\sqrt{n}}) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n(\int g(y_i,\theta_0)[p_{n,i}(y_i) - p_{0,i}(y_i)]dy_i + \frac{1}{\sqrt{n}}E_{P_{0,i}}[\frac{\partial g(y_i,\theta_0)}{\partial\theta'}]h) + o(\frac{1}{\sqrt{n}})$, from $\frac{1}{n}\sum_{i=1}^n E_{P_{0,i}}[g(y_i,\theta_0)] = o(\frac{1}{\sqrt{n}})$ and $E_{P_{n,i}}[\frac{\partial g(y_i,\theta_0)}{\partial\theta}] = E_{P_{0,i}}[\frac{\partial g(y_i,\theta_0)}{\partial\theta}] + o(1)$. We rewrite the first term on the RHS by using the expansion of $p_{n,i}(y_i) - p_{0,i}(y_i)$ from Lemma 1(b), and apply the truncation indicator $1^\tau(y_i) = 1\{\|g(y_i,\theta_0)\| \leq \tau_n\}$, and we get:

$$\int g(y_i,\theta_0)[p_{n,i}(y_i) - p_{0,i}(y_i)]dy_i = \int g(y_i,\theta_0)1^\tau(y_i)[p_{n,i}(y_i) - p_{0,i}(y_i)]dy_i$$
$$+\int g(y_i,\theta_0)(1 - 1^\tau(y_i))[p_{n,i}(y_i) - p_{0,i}(y_i)]dy_i = \frac{1}{\sqrt{n}}E_{P_{0,i}}[g(y_i,\theta_0)f_i^p(y_i)] + \sum_{j=1}^4 I_{j,i},$$

where $I_{1,i} := -\frac{1}{\sqrt{n}}\int g(y_i,\theta_0)(1-1^\tau(y_i))f_i^p(y_i)dy_i$, $I_{2,i} = \int g(y_i,\theta_0)1^\tau(y_i)R_{n,i}^p(y_i)dy_i$, $I_{3,i} = \int g(y_i,\theta_0)(1-1^\tau(y_i))p_{n,i}(y_i)dy_i$ and $I_{4,i} = -\int g(y_i,\theta_0)(1-1^\tau(y_i))p_{0,i}(y_i)dy_i$. We bound the sample averages of these terms. We have $\frac{1}{n}\sum_{i=1}^n I_{1,i} = o(\frac{1}{\sqrt{n}})$, from $|I_{1,i}| \leq \frac{1}{\sqrt{n}}E_{P_{0,i}}[\|g(y_i,\theta_0)\|^{2r}]^{1/(2r)}$ $E_{P_{0,i}}[(f_i^p(y_i))^{2r}]^{1/(2r)}E_{P_{0,i}}[1-1^\tau(y_i)]^{1/\bar{r}}$ and $E_{P_{0,i}}[1-1^\tau(y_i)] = P_{0,i}[\|g(y_i,\theta_0)\| \geq \tau_n] \leq \frac{E_{P_{0,i}}[\|g(y_i,\theta_0)\|^{2r}]}{\tau_n^{2r}}$. From Lemma 1(b), we get $\frac{1}{n}\sum_{i=1}^n I_{2,i} = o(\frac{1}{\sqrt{n}})$. Moreover, $|I_{3,i}| \leq E_{P_{n,i}}[\|g(y_i,\theta_0)\|^{2r}]^{1/(2r)}E_{P_{n,i}}[1-1^\tau(y_i)]^{1/q} \leq \frac{E_{P_{n,i}}[\|g(y_i,\theta_0)\|^{2r}]}{\tau_n^{2r/q}} = O(\frac{1}{\tau_n^{2r-1}})$, where $1/(2r)+1/q = 1$, which yields $\frac{1}{n}\sum_{i=1}^n I_{3,i} = o(\frac{1}{\sqrt{n}})$ because $\tau_n \gg n^{\frac{1}{2(2r-1)}}$. Similarly $\frac{1}{n}\sum_{i=1}^n I_{4,i} = o(\frac{1}{\sqrt{n}})$. Hence, we get the condition $0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{n}}\left(E_{P_{0,i}}[g(y_i,\theta_0)f_i^p(y_i)] + E_{P_{0,i}}[\frac{\partial g(y_i,\theta_0)}{\partial\theta'}]h\right) + o(\frac{1}{\sqrt{n}})$, and the conclusion follows. Q.E.D.

Proof of Proposition 1: We parallel the arguments in the proof of Theorem 12.2.3 in LR. Define $\xi_{n,i} := \sqrt{\frac{q_{n,i}(x_i)}{q_{0,i}(x_i)}} - 1$ and use $\log(1+y) = y - \frac{1}{2}y^2 + y^2 r(y)$, with $r(y)\to 0$ if $y \to 0$. Then, by following the so-called Le Cam's square root trick, we expand:

$$\log L_{n,f} = 2\sum_{i=1}^n \log(1+\xi_{n,i}) = 2\sum_{i=1}^n \xi_{n,i} - \sum_{i=1}^n \xi_{n,i}^2 + 2\sum_{i=1}^n \xi_{n,i}^2 r(\xi_{n,i}). \tag{B.6}$$

We get the asymptotic behaviour of the three terms in the RHS of (B.6) in four steps. We use the q.m.d. $\sqrt{q_{n,i}(x)} - \sqrt{q_{0,i}(x)} = \frac{1}{2\sqrt{n}}f_i^q(x)\sqrt{q_{0,i}(x)} + R_{n,i}^q(x)$ with $\sum_{i=1}^n\int[R_{n,i}^q(x)]^2dx = o(1)$ from Assumption 3. Now, (i) $\sum_{i=1}^n E_{Q_{0,i}}[\xi_{n,i}] = -\frac{1}{8}\sigma(f)^2 + o(1)$. To show this equation, we use: $\sum_{i=1}^n E_{Q_{0,i}}[\xi_{n,i}] = \sum_{i=1}^n\int\left(\sqrt{\frac{q_{n,i}(x)}{q_{0,i}(x)}} - 1\right)q_{0,i}(x)dx = -\frac{1}{2}\sum_{i=1}^n\int\left(\sqrt{q_{n,i}(x)} - \sqrt{q_{0,i}(x)}\right)^2 dx$. The latter quantity converges to $-\frac{1}{8}\sigma(f)^2$, where $\sigma(f)^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n E_{Q_{0,i}}[(f_i^q(x_i))^2]$ as $n\to\infty$. (ii) $\sum_{i=1}^n(\xi_{n,i} - E_{Q_{0,i}}[\xi_{n,i}]) = \frac{1}{2}Z_{n,f} + o_{P_0^n}(1)$. Indeed, write $\xi_{n,i} = \frac{1}{2\sqrt{n}}f_i^q(x_i) + \frac{1}{\sqrt{n}}r_{n,i}$, where $r_{n,i} := \sqrt{n}\frac{R_{n,i}^q(x_i)}{\sqrt{q_{0,i}(x_i)}}$. Then, $\sum_{i=1}^n(\xi_{n,i} - E_{Q_{0,i}}[\xi_{n,i}]) = \frac{1}{2}Z_{n,f} + \frac{1}{\sqrt{n}}\sum_{i=1}^n(r_{n,i} - E_{Q_{0,i}}[r_{n,i}])$. We have $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(r_{n,i} - E_{Q_{0,i}}[r_{n,i}]\right) = o_{P_0^n}(1)$, since this term has zero mean and vanishing variance, due to $\frac{1}{n}\sum_{i=1}^n E_{Q_{0,i}}[r_{n,i}^2] = \sum_{i=1}^n\int[R_{n,i}^q(x)]^2dx = o(1)$. (iii) $\sum_{i=1}^n \xi_{n,i}^2 = \frac{1}{4}\sigma(f)^2 + o_{P_0^n}(1)$. To show this equation, we use $\sum_{i=1}^n\xi_{n,i}^2 = \frac{1}{4n}\sum_{i=1}^n[f_i^q(x_i)]^2 + \frac{1}{n}\sum_{i=1}^n r_{n,i}^2 + \frac{1}{n}\sum_{i=1}^n f_i^q(x_i)r_{n,i}$. The first term in the RHS converges to $\frac{1}{4}\sigma(f)^2$ under P_0^n, while the second and third terms are $o_{P_0^n}(1)$, since $\frac{1}{n}\sum_{i=1}^n E_{Q_{0,i}}[r_{n,i}^2] = o(1)$. (iv) Finally, $\sum_{i=1}^n\xi_{n,i}^2 r(\xi_{n,i}) = o_{P_0^n}(1)$ follows by extending the arguments in LR, p. 491. From Equation (B.6) and Steps (i)-(iv), Part (a)

follows. The Lyapunov CLT and Part (a) yield Part (b). Then, Part (c) follows from Equation (8) and Corollary 12.3.1 in LR. To get Part (d), we project $Z_{n,f}$ orthogonally onto $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(y_i, \theta_0) = \frac{1}{\sqrt{n}}\sum_{m=1}^{J_n}(\sum_{i\in B_m} g(y_i, \theta_0))$, and get covariance $E_{P_0^n}\left[(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(y_i, \theta_0))Z_{n,f}\right]$
$= \frac{1}{n}\sum_{m=1}^{J_n} E_{P_0^m}\left[(\sum_{i\in B_m} g(y_i, \theta_0))f_m(Y_m S_m^{-1})\right] = \frac{1}{n}\sum_{m=1}^{J_n}\sum_{i\in B_m} E_{P_0^m}\left[g(y_i, \theta_0)f_m(Y_m S_m^{-1})\right] = \frac{1}{n}\sum_{m=1}^{J_n}\sum_{i\in B_m} E_{P_0^m}\left[g(y_i, \theta_0)E_{P_0^m}[f_m(Y_m S_m^{-1})|y_i]\right] = \frac{1}{n}\sum_{i=1}^{n} E_{P_{0,i}}\left[g(y_i, \theta_0)f_i^p(y_i)\right]$, where we use the Law of iterated expectation and the definition of f_i^p in Lemma 1(c). By taking the limit for $n \to \infty$, and using Lemma 2, we get $\lim_{n\to\infty} E_{P_0^n}\left[(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(y_i, \theta_0))Z_{n,f}\right] = -J_0 h$. Then, representing $Z_{n,f}$ as the sum of its orthogonal projection onto $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(y_i, \theta_0)$ plus the remainder term, we get $Z_{n,f} = h'Z_n^* + Z_{n,\perp} + o_{P_0^n}(1)$. Asymptotic normality of Z_n^* follows from Assumption 5. Moreover, by orthogonality, we have $\sigma(f)^2 = h'\Sigma_0^{-1}h + \sigma_\perp^2$, where $E_{P_0^n}[(Z_n^\perp)^2] = \sigma_\perp^2 + o(1)$, and we deduce the stated asymptotic representation of the log likelihood ratio. Q.E.D.

Proof of Proposition 2: As in the proof of Theorem 13.5.4 in LR, we argue by contradiction. Thus, suppose there exists a subsequence n_j such that:

$$\lim_{j\to\infty} \inf\{\beta_{n_j,f}(\phi_{n_j}) \; : \; h'A'(A\Sigma_0 A')^{-1}Ah \geq \lambda_{nc}^2\} > 1 - F_{\chi^2(r,\lambda_{nc}^2)}(c_{r,1-\alpha}). \tag{B.7}$$

From (12), there exists a further subsequence such that, for any h, $\beta_{n_{j_m},f}(\phi_{n_{j_m}}) \to \beta(h) := E_h[\tilde{\phi}(\tilde{Z})]$, for a test $\tilde{\phi}$ in the Gaussian experiment $\tilde{Z} \sim \mathcal{N}(h, \Sigma_0)$. Then, Inequality (B.7) implies that, for any h such that $h'A'(A\Sigma_0 A')^{-1}Ah \geq \lambda_{nc}^2$, we have $\beta(h) > 1 - F_{\chi^2(r,\lambda_{nc}^2)}(c_{r,1-\alpha})$, so that

$$\inf\{\beta(h) \; : \; h'A'(A\Sigma_0 A')^{-1}Ah \geq \lambda_{nc}^2\} > 1 - F_{\chi^2(r,\lambda_{nc}^2)}(c_{r,1-\alpha}). \tag{B.8}$$

The strict inequality in (B.8) contradicts the power of the maximin test of the linear hypothesis $Ah = 0$ being equal to, and not above, $1 - F_{\chi^2(r,\lambda_{nc}^2)}(c_{r,1-\alpha})$ in the Gaussian experiment $\tilde{Z} \sim \mathcal{N}(h, \Sigma_0)$, as stated in the next lemma.

Lemma 3 *Consider the Gaussian experiment $\tilde{Z} \sim \mathcal{N}(h, \Sigma_0)$ with unknown mean $h \in \mathbb{R}^p$ and given covariance matrix Σ_0, and consider the null hypothesis $Ah = 0$ against the alternative $h'A'(A\Sigma_0 A')^{-1}Ah \geq \lambda_{nc}^2$, where A is a full row-rank matrix of rank r, and $\lambda_{nc}^2 > 0$ is a constant.*

We have that (a) the test $\phi(\tilde{Z}) = \{\tilde{Z}'A'(A\Sigma_0 A')^{-1}A\tilde{Z} > c_{r,1-\alpha}\}$ *is maximin at level* α*, for* $\alpha \in (0,1)$*, and (b) the maximin power of the above test is* $1 - F_{\chi^2(r,\lambda^2_{nc})}(c_{r,1-\alpha})$.

Then, Inequality (B.7) cannot be true and the conclusion follows. Q.E.D.

Proof of Lemma 3: Define the random vector $Z = (Z_1', Z_2')'$ in $\mathbb{R}^r \times \mathbb{R}^{p-r}$, where $Z_1 = A\tilde{Z}$ and $Z_2 = B\Sigma_0^{-1}\tilde{Z}$, with a $(p-r)\times p$ full row-rank matrix B such that $AB' = 0$, i.e., the rows of B span the kernel of A. Then, $Z \sim \mathcal{N}(\eta, \Omega)$, where $\eta = (\eta_1', \eta_2')'$ with $\eta_1 = Ah$ and $\eta_2 = B\Sigma_0^{-1}h$, and matrix Ω is block-diagonal with diagonal blocks $\Omega_{11} = A\Sigma_0 A'$ and $\Omega_{22} = B\Sigma_0^{-1}B'$. The two Gaussian components Z_1 and Z_2 are independent since $Cov(Z_1, Z_2) = AB' = 0$. For a given Σ_0, testing the linear hypothesis $Ah = 0$ in the Gaussian experiment $\tilde{Z} \sim \mathcal{N}(h, \Sigma_0)$ is tantamount to testing $\eta_1 = 0$ in the Gaussian experiment $Z \sim \mathcal{N}(\eta, \Omega)$. Vector Z_1 is a sufficient statistic for η_1, so that we can focus on tests built on $Z_1 \sim \mathcal{N}(\eta_1, \Omega_{11})$. Problem 8.29 in LR states that the test which rejects when $Z'\Omega_{11}^{-1}Z > c_{r,1-\alpha}$ is maximin for testing $\eta_1 = 0$ against $\eta_1'\Omega_{11}^{-1}\eta_1 \geq \lambda^2_{nc}$ at level α, and the maximin power is $1 - F_{\chi^2(r,\lambda^2_{nc})}(c_{r,1-\alpha})$. Since $Z'\Omega_{11}^{-1}Z = \tilde{Z}'A'(A\Sigma_0 A')^{-1}A\tilde{Z}$ and $\eta_1'\Omega_{11}^{-1}\eta_1 = h'A'(A\Sigma_0 A')^{-1}Ah$, the conclusion follows. Q.E.D.

C Spectral Characterisation of Spherical Models

Lemma 4 characterizes spherical models through a plateau in the spectrum of $\Sigma = FF' + V_\varepsilon$.

Lemma 4 *A symmetric positive-definite* $T \times T$ *matrix* Σ *admits the representation* $\Sigma = FF' + \bar{\sigma}^2 I_T$*, with a* $T \times k$ *matrix* F *and* $\bar{\sigma}^2 > 0$*, if, and only if, the* $T - k$ *smallest eigenvalues of* Σ *are equal to* $\bar{\sigma}^2$.

Corollary 2 *The spherical models with* k *latent factors are a strict subset of the general specifications, if, only if,* $k \leq T - 2$.

In other words, we can always write any matrix Σ symmetric and positive-definite as $FF' + \bar{\sigma}^2 I_T$ with a $T \times k$ matrix F, for $k = T - 1$ (or $k = T$), while it is not the case with the restriction

$k \leq T-2$. Hence, it is meaningful to talk about a spherical model only for $k \leq T-2$. Importantly, some models with $V_\varepsilon \neq \bar{\sigma}^2 I_T$ admit a spherical representation with a larger number of factors. For instance, suppose that $\Sigma = FF' + V_\varepsilon$, where the diagonal elements of V_ε are all equal to $\bar{\sigma}^2$ except for date t, where $V_{\varepsilon,tt} = \sigma_1^2$, with $\sigma_1^2 > \bar{\sigma}^2$. Then, we have $V_\varepsilon = \bar{\sigma}^2 I_T + \eta^2 e_t e_t'$, with $\eta^2 := \sigma_1^2 - \bar{\sigma}^2$. It follows that $\Sigma = \tilde{F}\tilde{F}' + \bar{\sigma}^2 I_T$, with $\tilde{F} := [F : \eta e_t]C$ and C an orthogonal matrix such that $\tilde{F}'\tilde{F}$ is diagonal. Hence, we have a model with $k+1$ factors and spherical errors.

Proof of Lemma 4: Suppose first that $\Sigma = FF' + \bar{\sigma}^2 I_T$, for a $T \times k$ matrix F and $\bar{\sigma}^2 > 0$. Let $\mu_j \geq 0$, for $j = 1, ..., k$, be the eigenvalues of matrix FF'. Then, the eigenvalues of Σ are $\delta_j = \mu_j + \bar{\sigma}^2$, for $j = 1, ..., k$, and $\delta_j = \bar{\sigma}^2$, for $j = k+1, ..., T$. Conversely, let $\Sigma = \sum_{j=1}^T \delta_j P_j$ be the spectral decomposition of Σ with eigenvalues δ_j ranked in decreasing order, and eigenprojectors P_j, and suppose that $\delta_j = \bar{\sigma}^2 > 0$, for $j = k+1, ..., T$. Then, we have $\Sigma = \sum_{j=1}^k \delta_j P_j + \bar{\sigma}^2 \sum_{j=k+1}^T P_j = \sum_{j=1}^k (\delta_j - \bar{\sigma}^2) P_j + \bar{\sigma}^2 \sum_{j=1}^T P_j$. We have $\sum_{j=1}^T P_j = I_T$. Moreover, $\sum_{j=1}^k (\delta_j - \bar{\sigma}^2) P_j$ is a positive semi-definite matrix of rank k and hence we can write it as $\sum_{j=1}^k (\delta_j - \bar{\sigma}^2) P_j = FF'$, for a $T \times k$ matrix F. Thus, we get $\Sigma = FF' + \bar{\sigma}^2 I_T$. Q.E.D.

D Characterisation of FA-GMM Estimators

Appendix E in the Online Appendix (OA) provides the detailed proofs of technical Lemmas 5-8 below supporting the computations of this appendix.

D.1 Unconstrained FA-GMM Estimator

The FA-GMM estimator is $\hat{\theta} = (\hat{\mu}', \hat{\vartheta}')' = \arg\min_{\theta \in \Theta} Q_n(\theta)$, where $Q_n(\theta) = \hat{g}_n(\theta)' \hat{V}_g^{-1} \hat{g}_n(\theta)$. The condition $\theta \in \Theta$ imposes the normalization constraint $F' V_\varepsilon^{-1} F$ being diagonal, that we rewrite as $h(\vartheta) = 0$, where $h(\vartheta)$ is the $\rho \times 1$ vector stacking the unique off-diagonal elements $\{(F' V_\varepsilon^{-1} F)_{i,j}\}_{i<j}$, with $\rho := k(k-1)/2$.

a) Consistency. From Assumptions 1, 2, A.1 and A.2, $\hat{V}_y \xrightarrow{p} V_y^0 = \Sigma(\vartheta_0) = F_0 F_0' + V_\varepsilon^0$ and $\bar{y} \xrightarrow{p} \mu_0$

(see Lemma 5 in FGS). Then, $\hat{g}_n(\theta) \xrightarrow{p} g_\infty(\theta) := [(\mu_0 - \mu)', vech(\Sigma(\vartheta_0) - \Sigma(\vartheta) + \mu_0\mu_0' - \mu\mu')]'$ uniformly in $\theta \in \Theta$. By the consistency of $\hat{V}_g$ under Assumptions 1-2, A.1-A.4, and A.7 shown in Section D.3, we get that the criterion $Q_n(\theta)$ converges in probability uniformly to $Q_\infty(\theta) := g_\infty(\theta)' V_g^{-1} g_\infty(\theta)$. The global identification condition in Assumption A.3 implies that Q_∞ is uniquely minimized over compact set Θ at true parameter value θ_0. Then, standard results for extremum estimators (see e.g. Newey and McFadden (1994)) yield consistency of $\hat{\theta}$.

b) Asymptotic normality. The First-Order Conditions (FOC) for the FA-GMM estimator yield

$$\frac{\partial \hat{g}_n(\hat{\theta})'}{\partial \theta} \hat{V}_g^{-1} \hat{g}_n(\hat{\theta}) = 0, \qquad h(\hat{\vartheta}) \ = 0. \tag{D.1}$$

The Lagrange multipliers vector associated to the factor normalization is equal to zero, because the criterion $Q_n(\theta)$ is invariant under rotations of the columns of F. Let us define matrix $J_0 :=$ plim $\frac{\partial \hat{g}_n(\theta_0)}{\partial \theta'} = [J_{\mu,0} : J_{\vartheta,0}]$ in block form, and let the $p_\vartheta \times \rho$ matrix $H := \frac{\partial h(\vartheta_0)'}{\partial \vartheta}$ be full-rank, with $p_\vartheta := (k+1)T$. We apply the mean-value theorem to system (D.1) around $\tilde{\theta}_0$. By the consistency of estimator $\hat{\theta}$ proved in part (a), and the normalization $h(\tilde{\vartheta}_0) = 0$, we get:

$$\begin{aligned} J_0' V_g^{-1} \sqrt{n} \hat{g}_n(\tilde{\theta}_0) + J_0' V_g^{-1} J_0 \sqrt{n}(\hat{\theta} - \tilde{\theta}_0) &= o_p(1), && \text{(D.2)} \\ H' \sqrt{n}(\hat{\vartheta} - \tilde{\vartheta}_0) &= o_p(1). && \text{(D.3)} \end{aligned}$$

Let L_H be a full-rank $p \times \tilde{p}$ matrix, such that $M_H := I_p - H(H'H)^{-1}H' = L_H L_H'$ and $L_H' L_H = I_{\tilde{p}}$, namely L_H is the matrix of standardized eigenvectors of M_H for the $\tilde{p} := p - \rho$ unit eigenvalues. Then, using $J_0 \sqrt{n}(\hat{\theta} - \tilde{\theta}_0) = J_{\mu,0} \sqrt{n}(\hat{\mu} - \mu_0) + J_{\vartheta,0} L_H \sqrt{n} L_H' (\hat{\vartheta} - \tilde{\vartheta}_0) + o_p(1)$ from (D.3), and inserting this equation into (D.2), we get the asymptotic expansion:

$$\begin{pmatrix} \sqrt{n}(\hat{\mu} - \mu_0) \\ \sqrt{n} L_H' (\hat{\vartheta} - \tilde{\vartheta}_0) \end{pmatrix} = -(\tilde{J}_0' V_g^{-1} \tilde{J}_0)^{-1} \tilde{J}_0' V_g^{-1} \sqrt{n} \hat{g}_n(\tilde{\theta}_0) + o_p(1), \tag{D.4}$$

where matrix $\tilde{J}_0 := [J_{\mu,0} : J_{\vartheta,0} L_H]$ is full column-rank under the local identification condition in Assumption A.4, as we show next.

Lemma 5 *Matrix $\tilde{J}_0$ is full column-rank if, and only if, Assumption A.4 holds.*

To interpret Equation (D.4), note that there exists a smooth one-to-one change of parameters $\vartheta = \phi(h, \eta)$ locally around ϑ_0, where $h \in \mathbb{R}^{\rho}$ is the parameter vector that corresponds to rotations of the latent factor matrix, with true value $h_0 = 0$ under our normalization, and vector $\eta \in \mathbb{R}^{\tilde{p}}$ parametrizes the remaining degrees of freedom, with true value η_0. Let $L_H := \frac{\partial \phi(0,\eta_0)}{\partial \eta'}$. Then, $L_H' H = 0$. We can choose this change of parameters such that $L_H' L_H = I_{\tilde{p}}$. Then, we have $L_H'(\vartheta - \vartheta_0) = \eta - \eta_0$ locally around ϑ_0. Hence, the lower block in (D.4) is the asymptotic expansion for the estimator of the free parameters invariant to factor rotations, and the full-rank of matrix $\tilde{J}_0$ corresponds to the standard GMM local identification condition with the transformed parameters. By using $\sqrt{n}\hat{g}_n(\tilde{\theta}_0) \Rightarrow \mathcal{N}(0, V_g)$ under Assumptions 1-2, A.1, A.2, A.5, and A.7 (see Section D.3 for the proof), we get $(\sqrt{n}(\hat{\mu} - \mu_0)', [\sqrt{n} L_H'(\hat{\vartheta} - \tilde{\vartheta}_0)]')' \Rightarrow \mathcal{N}(0, \tilde{\Sigma}_0)$, where $\tilde{\Sigma}_0 := (\tilde{J}_0' V_g^{-1} \tilde{J}_0)^{-1} = \begin{pmatrix} \tilde{\Sigma}_{\mu\mu,0} & \tilde{\Sigma}_{\mu\eta,0} \\ \tilde{\Sigma}_{\eta\mu,0} & \tilde{\Sigma}_{\eta\eta,0} \end{pmatrix}$ in a block form. Then, the Asymptotic Variance (AV) is

$$AV \begin{pmatrix} \sqrt{n}(\hat{\mu} - \mu_0) \\ \sqrt{n}(\hat{\vartheta} - \tilde{\vartheta}_0) \end{pmatrix} = \begin{pmatrix} \tilde{\Sigma}_{\mu\mu,0} & \tilde{\Sigma}_{\mu\eta,0} L_H' \\ L_H \tilde{\Sigma}_{\eta\mu,0} & L_H \tilde{\Sigma}_{\eta\eta,0} L_H' \end{pmatrix}. \tag{D.5}$$

The asymptotic variance of block $\sqrt{n}(\hat{\vartheta} - \tilde{\vartheta}_0)$ in (D.5) is degenerate because of the factor normalization. The asymptotic variances of components $\sqrt{n}vec(\hat{F} - F_0)$ and $\sqrt{n}diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon)$ are the upper-left $(kT) \times (kT)$ and lower-right $T \times T$ blocks $\Sigma_F = (L_H \tilde{\Sigma}_{\eta\eta,0} L_H')_{11}$ and $\Sigma_{V_\varepsilon} = (L_H \tilde{\Sigma}_{\eta\eta,0} L_H')_{22}$. Let us now obtain explicitly matrices J_0 and H. We compute the partial derivatives of the moment vector, and use Lemma 6 below as well as $vech(V_\varepsilon) = (E_{diag,T}) diag(V_\varepsilon)$, where

$E_{diag,T} := \frac{1}{\sqrt{2}}[I_T \; : \; 0_{T \times \frac{T(T-1)}{2}}]'$. We get: $J_0 = -\begin{pmatrix} I_T & 0_{T \times (kT)} & 0_{T \times T} \\ \frac{\partial vech(\mu\mu')}{\partial \mu'} & \frac{\partial vech(FF')}{\partial vec(F)'} & \frac{\partial vech(V_\varepsilon)}{\partial diag(V_\varepsilon)'} \end{pmatrix}$

$= -\begin{pmatrix} I_T & 0_{T \times (kT)} & 0_{T \times T} \\ A_T'(\mu \otimes I_T) & A_T'(F \otimes I_T) & E_{diag,T} \end{pmatrix}$. Here, we use the link $vec(Z) = A_T vech(Z)$, where A_T is the $T^2 \times \frac{1}{2}T(T+1)$ duplication matrix (Magnus and Neudecker (2007)) suited to our definition of the half-vectorization operator $vech$. With e_i being the ith unit vector in dimension T, it is given by $A_T = \left[\sqrt{2}(e_1 \otimes e_1) : \cdots : \sqrt{2}(e_T \otimes e_T) : \{e_i \otimes e_j + e_j \otimes e_i\}_{i<j}\right]$.

Lemma 6 *Let C be a $T \times s$ matrix. Then, we get* $\frac{\partial vech(CC')}{\partial vec(C)'} = A_T'(C \otimes I_T)$.

Concerning matrix H, let us write $h(\vartheta) = E_{off,k} vech(F'V_\varepsilon F)$, where $E_{off,k} = [0_{\rho \times k} : I_\rho]$ is the matrix that selects the lower $\rho \times 1$ block in a $(k+\rho) \times 1$ vector. Moreover, $F' = [f_1 : \cdots : f_T]$, $vec(F')' = (f_1', ..., f_T')$, $vech(F'V_\varepsilon^{-1}F) = \sum_{s=1}^T \frac{1}{V_{\varepsilon,ss}} vech(f_s f_s')$. Then, using Lemma 6, we get $\frac{\partial vech(F'V_\varepsilon^{-1}F)}{\partial f_t'} = \frac{1}{V_{\varepsilon,tt}} A_k'(f_t \otimes I_k)$, which yields $\frac{\partial vech(F'V_\varepsilon^{-1}F)}{\partial vec(F')'} = A_k'[(F'V_\varepsilon^{-1}) \otimes I_k]$. The chain rule gives: $\frac{\partial vech(F'V_\varepsilon^{-1}F)}{\partial vec(F)'} = \frac{\partial vech(F'V_\varepsilon^{-1}F)}{\partial vec(F')'} \frac{\partial vec(F')}{\partial vec(F)'} = A_k'[(F'V_\varepsilon^{-1}) \otimes I_k]K_{T,k} = A_k' K_{k,k}[(F'V_\varepsilon^{-1}) \otimes I_k]K_{T,k} = A_k'[I_k \otimes (F'V_\varepsilon^{-1})]$. Moreover, using $vech(F'V_\varepsilon^{-1}F) = \sum_{s=1}^T \frac{1}{2V_{\varepsilon,ss}} A_k'(f_s \otimes f_s)$, we get $\frac{\partial vech(F'V_\varepsilon^{-1}F)}{\partial V_{\varepsilon,tt}} = -\frac{1}{2V_{\varepsilon,tt}^2} A_k'(f_t \otimes f_t)$, which yields: $\frac{\partial vech(F'V_\varepsilon^{-1}F)}{\partial diag(V_\varepsilon)'} = -\frac{1}{2} A_k'[(F'V_\varepsilon^{-1}) \otimes_{\text{col}} (F'V_\varepsilon^{-1})]$, where $C \otimes_{\text{col}} C := [(c_1 \otimes c_1) : \cdots (c_T \otimes c_T)]$ denotes 'columnwise' Kronecker product of a matrix $C = [c_1 : \cdots : c_T]$. By putting all together, we get: $H' = \frac{\partial h(\vartheta_0)}{\partial \vartheta'} = E_{off,k} A_k' \left[I_k \otimes (F'V_\varepsilon^{-1}) \ : \ -\frac{1}{2}(F'V_\varepsilon^{-1}) \otimes_{\text{col}} (F'V_\varepsilon^{-1}) \right]$.

D.2 Constrained FA-GMM Estimator under Sphericity

The constrained FA-GMM estimator is $\hat{\theta}^c = (\hat{\mu}^{c\prime}, \hat{\vartheta}^{c\prime})' = \arg\min_{\theta \in \Theta} Q_n(\theta)$ subject to $a(\theta) = 0$, where $a(\theta) := L_{1_T}' diag(V_\varepsilon)$ and $M_{1_T} = L_{1_T} L_{1_T}'$. Consistency of $\hat{\theta}^c$ under the null hypothesis follows from the arguments in Section D.1 by rewriting the constraint as $\theta \in \Theta_0 = \{\theta \in \Theta : a(\theta) = 0\}$, i.e. a compact set that contains the true value.

By the mean-value theorem, and consistency of $\hat{\theta}^c$ under the null hypothesis, we get from the FOC with $\hat{\lambda}$ being the $(T-1) \times 1$ vector of Lagrange multipliers for the constraint $a(\theta) = 0$:

$$J_{\mu,0}' V_g^{-1} \sqrt{n} \hat{g}_n(\tilde{\theta}_0) + J_{\mu,0}' V_g^{-1} J_0 \sqrt{n}(\hat{\theta}^c - \tilde{\theta}_0) = o_p(1), \tag{D.6}$$

$$J_{\vartheta,0}' V_g^{-1} \sqrt{n} \hat{g}_n(\tilde{\theta}_0) + J_{\vartheta,0}' V_g^{-1} J_0 \sqrt{n}(\hat{\theta}^c - \tilde{\theta}_0) + \sqrt{n} \frac{\partial a(\theta_0)'}{\partial \vartheta} \hat{\lambda} = o_p(1), \tag{D.7}$$

$$H' \sqrt{n}(\hat{\vartheta}^c - \tilde{\vartheta}_0) = o_p(1), \tag{D.8}$$

$$\frac{\partial a(\theta_0)}{\partial \vartheta'} \sqrt{n}(\hat{\vartheta}^c - \tilde{\vartheta}_0) = o_p(1), \tag{D.9}$$

where $\frac{\partial a(\theta_0)}{\partial \mu'} = 0$ and $\frac{\partial a(\theta_0)}{\partial \vartheta'} = \left[0_{(T-1) \times (Tk)} \ : \ L_{1_T}' \right]$, and we separate in Equations (D.6) and (D.7) the FOC for the μ and ϑ components. We use $J_0 \sqrt{n}(\hat{\theta}^c - \tilde{\theta}_0) = J_{\mu,0} \sqrt{n}(\hat{\mu}^c - \mu_0) + \tilde{J}_{\vartheta,0} \sqrt{n} L_H'(\hat{\vartheta}^c -$

$\tilde{\theta}_0) + o_p(1)$ in (D.6) and (D.7), where $\tilde{J}_{\vartheta,0} := J_{\vartheta,0} L_H$, and $\sqrt{n}(\hat{\vartheta}^c - \vartheta_0) = L_H \sqrt{n} L_H'(\hat{\vartheta}^c - \tilde{\vartheta}_0) + o_p(1)$ in (D.9), and left-multiply (D.7) times L_H', so that we can write an expression in a block form:

$$\begin{bmatrix} \tilde{J}_0' V_g^{-1} \tilde{J}_0 & \tilde{A} \\ \tilde{A}' & 0_{(T-1)\times(T-1)} \end{bmatrix} \begin{bmatrix} \begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \mu_0) \\ \sqrt{n} L_H'(\hat{\vartheta}^c - \tilde{\vartheta}_0) \end{pmatrix} \\ \sqrt{n}\hat{\lambda} \end{bmatrix} = - \begin{bmatrix} \tilde{J}_0' V_g^{-1} \sqrt{n} \hat{g}_n(\tilde{\theta}_0) \\ 0 \end{bmatrix} + o_p(1), \tag{D.10}$$

where $\tilde{A}' = \left[0_{(T-1)\times T} \ : \ \frac{\partial a(\theta_0)}{\partial \vartheta'} L_H \right] = \left[0_{(T-1)\times T} \ : \ L_{1_T}' \tilde{L}_H \right]$ and $\tilde{L}_H$ is the lower $T \times \tilde{p}$ block of L_H. The derivation of the joint asymptotic distribution of the constrained FA-GMM estimator and the Lagrange multiplier vector continues along the lines of standard theory (see e.g. Newey and McFadden (1994)). By inversion of the block matrix on the LHS of Equation (D.10), we have: $\begin{bmatrix} \begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \mu_0) \\ \sqrt{n} L_H'(\hat{\vartheta}^c - \tilde{\vartheta}_0) \end{pmatrix} \\ \sqrt{n}\hat{\lambda} \end{bmatrix} \Rightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} (I-P)'\tilde{\Sigma}_0(I-P) & \Gamma_0 \\ \Gamma_0' & (\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1} \end{bmatrix} \right)$, where $\Gamma_0 = (I-P)'\tilde{\Sigma}_0 \tilde{A}(\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}$, $\tilde{\Sigma}_0 := (\tilde{J}_0' V_g^{-1} \tilde{J}_0)^{-1}$ is the asymptotic variance of the unconstrained FA-GMM estimator, and $P = \tilde{A}(\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\tilde{A}'\tilde{\Sigma}_0$ is the oblique projection matrix onto the columns of matrix $\tilde{A}$ w.r.t. the scalar product induced by matrix $\tilde{\Sigma}_0$. In particular, the asymptotic variance of the Lagrange multipliers vector is $AV(\sqrt{n}\hat{\lambda}) = (\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1} = (L_{1_T}' \tilde{L}_H \tilde{\Sigma}_{\eta\eta,0} \tilde{L}_H' L_{1_T})^{-1} = (L_{1_T}' AV[\sqrt{n} diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon)] L_{1_T})^{-1}$, where $AV[\sqrt{n} diag(\hat{V}_\varepsilon - \tilde{V}_\varepsilon)] = (L_H \tilde{\Sigma}_{\eta\eta,0} L_H')_{22}$ is the lower $T \times T$ block of matrix $L_H \tilde{\Sigma}_{\eta\eta,0} L_H'$ from (D.5).

D.3 Feasible CLT for the Sample Orthogonality Vector

The next lemma establishes the asymptotic normality of the sample orthogonality vector in the FA model, namely that Assumption 5 holds in our FA-GMM setting.

Lemma 7 *Under Assumptions 1-2, A.1, A.2, A.5, and A.7, as* $n \to \infty$ *we have* $\frac{1}{\sqrt{n}} \sum_{i=1}^n g(y_i, \tilde{\theta}_0) \Rightarrow \mathcal{N}(0, V_g)$, *with* $V_g = \begin{pmatrix} q_{\mathcal{B}} V_\varepsilon & [(Q_{\mathcal{B},1}' \mathcal{F}_0') \otimes V_\varepsilon] A_T \\ A_T'[(\mathcal{F}_0 Q_{\mathcal{B},1}) \otimes V_\varepsilon] & A_T' \left[(\mathcal{F}_0 Q_{\mathcal{B}} \mathcal{F}_0') \otimes V_\varepsilon \right] A_T + \Omega_Z \end{pmatrix}$, *where* $\mathcal{F}_0 :=$

$[\mu_0 : F_0]$, $Q_{\mathcal{B}} := \lim \frac{1}{n}\mathcal{B}'\check{\Sigma}\mathcal{B}$ *with* $\mathcal{B} = [1_n : \beta]$, $q_{\mathcal{B}} := \lim \frac{1}{n}1_n'\check{\Sigma}1_n$, *and* $Q_{\mathcal{B},1} := \lim \frac{1}{n}\mathcal{B}'\check{\Sigma}1_n$. *Moreover,* $\Omega_Z = \frac{1}{4}[A_T'(V_\varepsilon^{1/2} \otimes V_\varepsilon^{1/2})A_T](D + \kappa I_{\frac{T(T+1)}{2}})[A_T'(V_\varepsilon^{1/2} \otimes V_\varepsilon^{1/2})A_T]$ *is the asymptotic variance of the* $vech$ *of* $Z_n = \sqrt{n}(\frac{1}{n}\varepsilon\varepsilon' - E[\frac{1}{n}\varepsilon\varepsilon'])$.

The asymptotic variance matrix V_g involves the FA parameters vector θ, asymptotic variance matrix Ω_Z, and matrix $Q_{\mathcal{B}}$. The PML estimator of FGS, for example, yields a preliminary consistent estimator of θ. We can follow FGS, Section E.5., to get a consistent estimator of Ω_Z. We establish below a consistent estimator of $Q_{\mathcal{B}}$. Then, we get a consistent estimator $\hat{V}_g$ by plug-in.[19]

Let $\hat{\Psi}_{\mathcal{B}} := \frac{1}{n}\sum_m \sum_{i,j \in B_m}(\hat{\mathcal{B}}_i\hat{\mathcal{B}}_j') \otimes (\hat{\varepsilon}_i\hat{\varepsilon}_j')$, where $\hat{\mathcal{B}}_i = (1, \hat{\beta}_i')'$ with $\hat{\beta}_i = (\hat{F}'\hat{V}_\varepsilon^{-1}\hat{F})^{-1}\hat{F}'\hat{V}_\varepsilon^{-1}(y_i - \hat{\mu})$ and $\hat{\varepsilon}_i = M_{\hat{F},\hat{V}_\varepsilon}(y_i - \hat{\mu})$, where $\hat{\mu}$, $\hat{F}$, $\hat{V}_\varepsilon$ are preliminary root-n consistent estimators, e.g. from PML or FA-GMM with identity weighting matrix. Further, let $\hat{B}_{\mathcal{B}} := (N_{\hat{F},\hat{V}_\varepsilon} \otimes M_{\hat{F},\hat{V}_\varepsilon})A_T\hat{\Omega}_Z A_T'$ $(N_{\hat{F},\hat{V}_\varepsilon}' \otimes M_{\hat{F},\hat{V}_\varepsilon}')$, where $N_{F,V_\varepsilon} := \begin{pmatrix} 0_{1\times T} \\ (F'V_\varepsilon^{-1}F)^{-1}F'V_\varepsilon^{-1} \end{pmatrix}$. In Lemma 8 below, we show:

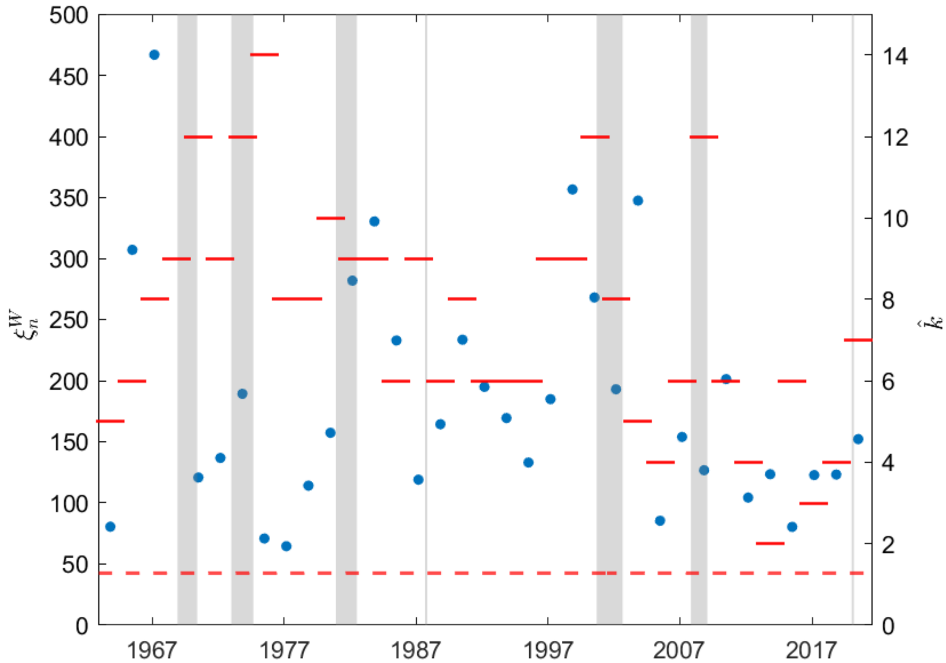
$$\hat{\Psi}_{\mathcal{B}} - \hat{B}_{\mathcal{B}} = Q_{\mathcal{B}} \otimes (M_{F,V_\varepsilon}V_\varepsilon) + o_p(1). \tag{D.11}$$

By half-vectorizing this equation, solving for $Q_{\mathcal{B}}$ by Least Squares projection and using a consistent estimator for $M_{F,V_\varepsilon}V_\varepsilon$, we get a consistent estimator for $Q_{\mathcal{B}}$.

Lemma 8 *Define the symmetric matrix* $\hat{Q}_{\mathcal{B}}$ *via* $vech(\hat{Q}_{\mathcal{B}}) = (\hat{P}'\hat{P})^{-1}\hat{P}'vech(\hat{\Psi}_{\mathcal{B}} - \hat{B}_{\mathcal{B}})$, *where* $\hat{P} = \frac{1}{2}A_{T(k+1)}'\left(I_{k+1} \otimes \left[(K_{T,k+1} \otimes I_T) \otimes (I_{k+1} \otimes vec(M_{\hat{F},\hat{V}_\varepsilon}\hat{V}_\varepsilon))\right]\right)A_{T(k+1)}$. *Then, under Assumptions 1-2, A.1-A.4, and A.7, estimator* $\hat{Q}_{\mathcal{B}}$ *is consistent for* $Q_{\mathcal{B}}$.

[19] A nonparametric strategy based on the averaging $\frac{1}{n}\sum_m \hat{g}_{m,n}\hat{g}_{m,n}'$ with $\hat{g}_{m,n} = \sum_{i \in B_m} g(y_i, \widehat{\theta})$ does not work here since the within-block averages $\bar{\beta}_{m,n} = \frac{1}{b_{m,n}}\sum_{i \in B_m}\beta_i$ and $\bar{V}_{\beta,m,n} - I_k = \frac{1}{b_{m,n}}\sum_{i \in B_m}(\beta_i\beta_i' - I_k)$ do not necessarily vanish despite the full-sample normalizations $\bar{\beta} = 0$ and $\tilde{V}_\beta = I_k$. It again exemplifies the difficulty to allow for cross-sectional dependence.

Figure 1: We display the values for the statistic ξ_n^W for the subperiods from January 1963 to December 2021. We partition the sample into 35 nonoverlapping windows of $T = 20$ months. The red horizontal segments indicate the estimated number $\hat{k}$ of factors. The red dotted line corresponds to the critical value of a $\chi^2(19)$ distribution at significance level $5\%/35$. Grey shaded vertical bars flag bear market phases.



ONLINE APPENDIX

Optimal Maximin GMM Tests for Sphericity in Latent Factor Analysis of Short Panels

Alain-Philippe Fortin, Patrick Gagliardini, and Olivier Scaillet

Appendix E provides the detailed proofs of technical Lemmas 5-8 supporting the computations of Appendix D. In Appendix F, we outline an asymptotically equivalent FA-GMM estimator, which is easier to compute numerically. It relies on modified moment restrictions. We also explain how we can compute numerically the unconstrained and constrained estimators via Newton-Raphson and zigzag algorithms and provide a numerical study of their estimation performance. Appendix G gathers the Monte Carlo results on size and power for the LM and LR tests. Appendix H makes the link with the panel model of Chamberlain (1992) Section 4 and discusses how we can incorporate second-order moment information in sets of orthogonality restrictions for that model as in Arellano and Bonhomme (2012) Section 3.4 and our FA setting. Appendix I gives the detailed proof of Proposition 3.

E Proofs of Lemmas 5-8

Proof of Lemma 5: From Lemma 6 in FGS, Assumption A.4 is equivalent to non-singularity of matrix $L_H' \frac{\partial^2 L_0(\vartheta_0)}{\partial\vartheta\partial\vartheta'} L_H$, where $L_0(\vartheta) = -\frac{1}{2}\log|\Sigma(\vartheta)| - \frac{1}{2}Tr(V_y^0\Sigma(\vartheta)^{-1}) =: \mathscr{L}(\Sigma(\vartheta))$ is the population criterion of PML. We have $\frac{\partial^2 L_0(\vartheta_0)}{\partial\vartheta\partial\vartheta'} = \frac{\partial vech(\Sigma(\vartheta_0))'}{\partial\vartheta} \frac{\partial^2\mathscr{L}(\Sigma(\vartheta_0))}{\partial vech(\Sigma)\partial vech(\Sigma)'} \frac{\partial vech(\Sigma(\vartheta_0))}{\partial\vartheta'}$. To compute the second-order derivatives, we use matrix differentials (Magnus and Neudecker (2007), Chapter 6). We have $d\mathscr{L} = -\frac{1}{2}Tr(\Sigma^{-1}d\Sigma) + \frac{1}{2}Tr(V_y^0\Sigma^{-1}(d\Sigma)\Sigma^{-1})$ and $d^2\mathscr{L} = \frac{1}{2}Tr(\Sigma^{-1}(d\Sigma)\Sigma^{-1}(d\Sigma)) - Tr(V_y^0\Sigma^{-1}(d\Sigma)\Sigma^{-1}(d\Sigma)\Sigma^{-1})$. By evaluating the differential at $\Sigma = \Sigma(\vartheta_0) = V_y^0$, we get $d^2\mathscr{L} = -\frac{1}{2}Tr((d\Sigma)(V_y^0)^{-1}(d\Sigma)(V_y^0)^{-1}) = -\frac{1}{2}vec(d\Sigma)'[(V_y^0)^{-1}\otimes(V_y^0)^{-1}]vec(d\Sigma)$, which yields $\frac{\partial^2\mathscr{L}(\Sigma(\vartheta_0))}{\partial vech(\Sigma)\partial vech(\Sigma)'} = -A_T'[(V_y^0)^{-1}\otimes(V_y^0)^{-1}]A_T$. Thus, matrix $L_H'\frac{\partial^2 L_0(\vartheta_0)}{\partial\vartheta\partial\vartheta'}L_H = -L_H'\frac{\partial vech(\Sigma(\vartheta_0))'}{\partial\vartheta}$

$A_T'[(V_y^0)^{-1}\otimes(V_y^0)^{-1}]A_T\frac{\partial vech(\Sigma(\vartheta_0))}{\partial\vartheta'}L_H$ is non-singular if, and only if, matrix $\frac{\partial vech(\Sigma(\vartheta_0))}{\partial\vartheta'}L_H$ is full rank. Moreover, $\tilde{J}_0=-\begin{pmatrix} I_T & 0_{T\times(k+1)T} \\ \frac{\partial vech(\mu\mu')}{\partial\mu'} & \frac{\partial vech(\Sigma(\vartheta_0))}{\partial\vartheta'}L_H \end{pmatrix}$ is full-rank if, and only if, $\frac{\partial vech(\Sigma(\vartheta_0))}{\partial\vartheta'}L_H$ is full rank. The conclusion follows. Q.E.D.

Proof of Lemma 6: Let us write $C=[c_1:\cdots:c_s]$, $vec(C)'=(c_1',...,c_s')$ and $CC'=\sum_{j=1}^s c_jc_j'$, where the c_j are the columns of matrix C. Then, $vech(CC')=\frac{1}{2}A_T'vec(CC')=\frac{1}{2}\sum_{j=1}^s A_T'(c_j\otimes c_j)$, and $\frac{\partial vech(CC')}{\partial c_j'}=\frac{1}{2}A_T'\frac{\partial(c_j\otimes c_j)}{\partial c_j'}=\frac{1}{2}A_T'[I_T\otimes c_j+c_j\otimes I_T]=A_T'(c_j\otimes I_T)$. For the latter equality, we use that $A_T'(I_T\otimes c_j)=A_T'K_{T,T}(I_T\otimes c_j)K_{T,1}=A_T'(c_j\otimes I_T)$ by the properties of the commutation matrices. The conclusion follows. Q.E.D.

Proof of Lemma 7: We use $\bar{y}=\mu_0+\frac{1}{\sqrt{n}}u_n$ and $\hat{V}_y=\tilde{V}_y+\frac{1}{\sqrt{n}}\Psi_y+o_p(\frac{1}{\sqrt{n}})$ under Assumptions 1, 2, A.1, A.2, where $u_n:=\sqrt{n}\bar{\varepsilon}$, $\tilde{V}_y=F_0F_0'+\tilde{V}_\varepsilon=\Sigma(\tilde{\vartheta}_0)$ and $\Psi_y=W_nF_0'+F_0W_n'+Z_n$, with $W_n=\frac{1}{\sqrt{n}}\varepsilon\beta$ and $Z_n=\sqrt{n}(\frac{1}{n}\varepsilon\varepsilon'-\tilde{V}_\varepsilon)$, and $\tilde{\vartheta}_0=(vec(F_0)',diag(\tilde{V}_\varepsilon)')'$ for $\tilde{V}_\varepsilon:=\frac{1}{n}E[\varepsilon\varepsilon']$ (see FGS Lemma 5). Then, we get the asymptotic expansion $\sqrt{n}\hat{g}_n(\tilde{\theta}_0)=[u_n',vech(W_nF_0'+F_0W_n'+Z_n+u_n\mu_0'+\mu_0u_n')']'+o_p(1)=[u_n',vech(\mathcal{W}_n\mathcal{F}_0'+\mathcal{F}_0\mathcal{W}_n'+Z_n)']'+o_p(1)$, where $\mathcal{W}_n:=\frac{1}{\sqrt{n}}\varepsilon\mathcal{B}=[u_n:W_n]$. Under Assumptions 1-2, A.1, A.2, A.5, and A.7, by Lemmas 2 and 7 in FGS generalized to accommodate any $r>1$ replacing the condition $r=2$ (see below for details), we have $u_n\Rightarrow u$, $\mathcal{W}_n\Rightarrow\mathcal{W}$ and $Z_n\Rightarrow Z$, where vectors $vec(\mathcal{W})$ and $vech(Z)$ are jointly Gaussian, and u is the upper $T\times 1$ block of $vec(\mathcal{W})$. Then, we get $\sqrt{n}\hat{g}_n(\tilde{\theta}_0)\Rightarrow\mathcal{N}(0,V_g)$, where V_g is the variance of Gaussian vector $[u',vech(\mathcal{W}\mathcal{F}_0'+\mathcal{F}_0\mathcal{W}'+Z)']'$.

Let us now characterize matrix V_g explicitly. We have $vech(\mathcal{W}\mathcal{F}_0')=\frac{1}{2}A_T'vec(\mathcal{W}\mathcal{F}_0')=\frac{1}{2}A_T'(\mathcal{F}_0\otimes I_T)vec(\mathcal{W})$, and $vech(\mathcal{F}_0\mathcal{W}')=\frac{1}{2}A_T'vec(\mathcal{F}_0\mathcal{W}')=\frac{1}{2}A_T'(I_T\otimes\mathcal{F}_0)vec(\mathcal{W}')=\frac{1}{2}A_T'(I_T\otimes\mathcal{F}_0)K_{T,k}vec(\mathcal{W})=\frac{1}{2}A_T'K_{T,T}(I_T\otimes\mathcal{F}_0)K_{T,k}vec(\mathcal{W})=\frac{1}{2}A_T'(\mathcal{F}_0\otimes I_T)vec(\mathcal{W})$, using properties of the commutation matrices $K_{p,q}$ and matrix A_T. Thus, we have $vech(\mathcal{W}\mathcal{F}_0'+\mathcal{F}_0\mathcal{W}'+Z)=A_T'(\mathcal{F}_0\otimes I_T)vec(\mathcal{W})+vech(Z)$. Moreover, $vech(Z)=\frac{1}{2}[A_T'(V_\varepsilon^{1/2}\otimes V_\varepsilon^{1/2})A_T]vech(\mathcal{Z})$, where $\mathcal{Z}:=V_\varepsilon^{-1/2}ZV_\varepsilon^{-1/2}$. Under Assumptions 1-2, A.1, A.2, A.5, and A.7, Lemmas 1 and 7 in FGS imply that $vec(\mathcal{W})\sim\mathcal{N}(0,\Omega_\mathcal{W})$ and $vech(\mathcal{Z})\sim\mathcal{N}(0,\Omega_\mathcal{Z})$ are mutually independent, with

$\Omega_{\mathcal{W}} = Q_{\mathcal{B}} \otimes V_\varepsilon$ and $\Omega_{\mathcal{Z}} = D + \kappa I_{\frac{T(T+1)}{2}}$, which yields the statement in Lemma 7.

To conclude, we detail here the verification of the Liapunov condition in the CLT for Z_n. Let $\mathcal{Z}_n := V_\varepsilon^{-1/2} Z_n V_\varepsilon^{-1/2} = \frac{1}{\sqrt{n}} \sum_{m=1}^{J_n} \zeta_{m,n}$, where the elements of the $T \times T$ simmetric matrix $\zeta_{m,n}$ are $\zeta_{m,n}^{ts} = \sum_{i \in B_m} (w_{i,t} w_{i,s} - 1_{t,s}) \check{\sigma}_{ii} + \sum_{i,j \in B_m, i \neq j} w_{i,t} w_{j,s} \check{\sigma}_{ij} =: \zeta_{m,n}^{a,ts} + \zeta_{m,n}^{b,ts}$. Moreover, $V[vech(\mathcal{Z}_n)] = \frac{1}{n} \sum_{m=1}^{J_n} V[vech(\zeta_{m,n})] = D_n + \kappa_n I_{\frac{T(T+1)}{2}} = \Omega_n$. The multivariate Liapunov condition is $\|\Omega_n^{-1/2}\|^{2r} \frac{1}{n^r} \sum_{m=1}^{J_n} E[\|vech(\zeta_{m,n})\|^{2r}] = o(1)$, for $r > 1$. Because the eigenvalues of matrix Ω_n are bounded away from zero uniformly in n under Assumption A.5, and we have $\|x\|^{2r} \leq k^{r-1} \sum_{j=1}^k |x_j|^{2r}$ for a vector $x \in \mathbb{R}^k$ and $r > 1$, by the triangular inequality it is enough to show that: (a) $\frac{1}{n^r} \sum_{m=1}^{J_n} E[|\zeta_{m,n}^{a,ts}|^{2r}] = o(1)$, and (b) $\frac{1}{n^r} \sum_{m=1}^{J_n} E[|\zeta_{m,n}^{b,ts}|^{2r}] = o(1)$, for any t, s.

To prove (a), we write $\zeta_{m,n}^{a,ts} = \sum_{i \in B_m} \zeta_i$ as a sum of independent variables $\zeta_i := (w_{i,t} w_{i,s} - 1_{t,s}) \check{\sigma}_{ii}$ with mean zero and $E[|\zeta_i|^{2r}] < \infty$ for $r > 1$ (we omit indices t, s to ease notation). The Marcinkiewicz and Zygmund (1937) inequality yields $E[|\sum_{i \in B_m} \zeta_i|^{2r}] \leq C_r (\sum_{i \in B_m} E[|\zeta_i|^2])^r$, for a constant C_r that depends on $r > 1$ only. Under our assumptions, we get $E[|\sum_{i \in B_m} \zeta_i|^{2r}] = O(b_{m,n}^r)$ uniformly in m, which yields $\frac{1}{n^r} \sum_{m=1}^{J_n} E[|\zeta_{m,n}^{a,ts}|^{2r}] = O(\frac{1}{n^r} \sum_{m=1}^{J_n} b_{m,n}^r) = o(1)$ from Assumption 2 d).

Let us now show bound (b). If the random variables $w_{i,t}$ are symmetric, it can be established by bounds for higher-order moments of U-statistics in McConnell and Taqqu (1986). We extend to possibly non-symmetric variables next. First, by using the equality $\sum_{i,j \in B_m, i \neq j} w_{i,t} w_{j,s} \check{\sigma}_{ij}$ $= \frac{1}{2} \left(\sum_{i,j \in B_m, i \neq j} (w_{i,t} + w_{i,s})(w_{j,t} + w_{j,s}) \check{\sigma}_{ij} - \sum_{i,j \in B_m, i \neq j} w_{i,t} w_{j,t} \check{\sigma}_{ij} - \sum_{i,j \in B_m, i \neq j} w_{i,s} w_{j,s} \check{\sigma}_{ij} \right)$, it is enough to get bounds for $E[|\sum_{i,j \in B_m, i \neq j} w_i w_j \check{\sigma}_{ij}|^{2r}]$ for either $w_i = w_{i,t}$, or $w_i = w_{i,t} + w_{i,s}$, for any t, s with $t \neq s$. Note that those variables have zero mean and variance $\sigma^2 = E[w_i^2]$ that is either 1, or 2. Second, we use the decoupling result in de la Pena (1992), Theorem 1, which yields $E[|\sum_{i,j \in B_m, i \neq j} w_i w_j \check{\sigma}_{ij}|^{2r}] \leq 8^{2r} E[|\sum_{i,j \in B_m, i \neq j} w_i \tilde{w}_j \check{\sigma}_{ij}|^{2r}]$, where the $\tilde{w}_j$ are independent copies of the w_j. Next, we use an argument similar to the proof of Proposition 2.1 in Giné, Latala and Zinn (2000). Write $\sum_{i,j \in B_m, i \neq j} w_i \tilde{w}_j \check{\sigma}_{ij} = \sum_{i \in B_m} \zeta_i$, where $\zeta_i := w_i (\sum_{j \in B_m, j \neq i} \tilde{w}_j \check{\sigma}_{ij})$, and $E[|\sum_{i,j \in B_m, i \neq j} w_i \tilde{w}_j \check{\sigma}_{ij}|^{2r}] = E\left[\tilde{E}[|\sum_{i \in B_m} \zeta_i|^{2r}]\right]$ by the law of iterated expectation, where $\tilde{E}[\cdot]$ denotes expectation conditional on $\{\tilde{w}_j,\ j = 1, ..., n\}$. The ζ_i are independent conditional on $\{\tilde{w}_j,\ j = 1,...,n\}$, with conditional expectation $\tilde{E}(\zeta_i) = 0$ and conditional second moment $\tilde{E}(\zeta_i^2) = \sigma^2(\sum_{j\in B_m, j\neq i}\tilde{w}_j\check{\sigma}_{ij})^2$. Then, the Marcinkiewicz and Zygmund (1937) inequality yields $\tilde{E}[|\sum_{i\in B_m}\zeta_i|^{2r}] \leq C_r\sigma^{2r}(\sum_{i\in B_m}(\sum_{j\in B_m,j\neq i}\tilde{w}_j\check{\sigma}_{ij})^2)^r$. From the discrete Hölder inequality we have $(\sum_{i\in B_m}(\sum_{j\in B_m,j\neq i}\tilde{w}_j\check{\sigma}_{ij})^2)^r \leq (b_{m,n})^{r-1}\sum_{i\in B_m}(\sum_{j\in B_m,j\neq i}\tilde{w}_j\check{\sigma}_{ij})^{2r}$. Now, by taking the expectation with respect to variables $\{\tilde{w}_j,\ j = 1,...,n\}$, and using once more the Marcinkiewicz and Zygmund (1937) inequality, we get $E[(\sum_{i\in B_m}(\sum_{j\in B_m,j\neq i}\tilde{w}_j\check{\sigma}_{ij})^2)^r] \leq (b_{m,n})^{r-1}\sum_{i\in B_m}E[(\sum_{j\in B_m,j\neq i}\tilde{w}_j\check{\sigma}_{ij})^{2r}] \leq (b_{m,n})^{r-1}C_r\sigma^{2r}\sum_{i\in B_m}(\sum_{j\in B_m,j\neq i}\check{\sigma}_{ij}^2)^r$. By combining the inequalities obtained so far, we get $E[|\sum_{i,j\in B_m,i\neq j}w_iw_j\check{\sigma}_{ij}|^{2r}] \leq C_r^2(8\sigma^2)^{2r}(b_{m,n})^{r-1}$ $\sum_{i\in B_m}(\sum_{j\in B_m,j\neq i}\check{\sigma}_{ij}^2)^r$. Because $\sum_{j\in B_m,j\neq i}\check{\sigma}_{ij}^2 = O((b_{m,n})^{\check{\delta}})$ uniformly in i, we get $E[|\sum_{i,j\in B_m,i\neq j}w_iw_j\check{\sigma}_{ij}|^{2r}] = O(b_{m,n}^{r(1+\check{\delta})})$. Hence, $\frac{1}{n^r}\sum_{m=1}^{J_n}E[|\zeta_{m,n}^{b,ts}|^{2r}] = O(\frac{1}{n^r}\sum_{m=1}^{J_n}b_{m,n}^{r(1+\check{\delta})}) = o(1)$ from Assumption 2 d). Q.E.D.

Proof of Lemma 8: Consider the infeasible estimator $\Psi_{\mathcal{B}} = \frac{1}{n}\sum_m\sum_{i,j\in B_m}(\mathcal{B}_i\mathcal{B}_j')\otimes(\varepsilon_i\varepsilon_j')$, where $\mathcal{B}_i := (1,\beta_i')'$. Using $E[\varepsilon_i\varepsilon_j'] = \check{\sigma}_{ij}V_\varepsilon$, the expectation is $E[\Psi_{\mathcal{B}}] = \frac{1}{n}\sum_m\sum_{i,j\in B_m}(\mathcal{B}_i\mathcal{B}_j')\check{\sigma}_{i,j}\otimes V_\varepsilon = Q_{\mathcal{B}}\otimes V_\varepsilon + o(1)$ by Assumption A.7. Moreover, $V[\Psi_{\mathcal{B}}] = o(1)$ by using $\frac{1}{n^2}\sum_m b_{m,n}^{2(1+\check{\delta})} = o(1)$. Thus, $\Psi_{\mathcal{B}} = Q_{\mathcal{B}}\otimes V_\varepsilon + o_p(1)$. Now, consider the feasible estimator $\hat{\Psi}_{\mathcal{B}}$ obtained from $\Psi_{\mathcal{B}}$ after replacing $\mathcal{B}_i$ with $\hat{\mathcal{B}}_i$ and ε_i with $\hat{\varepsilon}_i$. We use root-n consistency of PML or FA-GMM estmates with identity weighting under Assumptions 1-2, A.1-A.4. We have, $\hat{\mathcal{B}}_i = \mathcal{B}_i + \eta_i + O_p(\frac{1}{\sqrt{n}})$ and $\hat{\varepsilon}_i = \tilde{\varepsilon}_i + O_p(\frac{1}{\sqrt{n}})$, where $\tilde{\varepsilon}_i := M_{F,V_\varepsilon}\varepsilon_i$, $\eta_i = N_{F,V_\varepsilon}\varepsilon_i$. We control the effect of the remainder terms at order $O_p(\frac{1}{\sqrt{n}})$ by the condition $\frac{1}{n^{3/2}}\sum_m b_{m,n}^2 = o(1)$. Then, $\hat{\Psi}_{\mathcal{B}} = \frac{1}{n}\sum_m\sum_{i,j\in B_m}(\mathcal{B}_i\mathcal{B}_j')\otimes(\tilde{\varepsilon}_i\tilde{\varepsilon}_j') + \frac{1}{n}\sum_m\sum_{i,j\in B_m}(\eta_i\mathcal{B}_j')\otimes(\tilde{\varepsilon}_i\tilde{\varepsilon}_j') + \frac{1}{n}\sum_m\sum_{i,j\in B_m}(\mathcal{B}_i\eta_j')\otimes(\tilde{\varepsilon}_i\tilde{\varepsilon}_j') + \frac{1}{n}\sum_m\sum_{i,j\in B_m}(\eta_i\eta_j')\otimes(\tilde{\varepsilon}_i\tilde{\varepsilon}_j') + o_p(1)$. The first term equals $(I_{k+1}\otimes M_{F,V_\varepsilon})\Psi_{\mathcal{B}}(I_{k+1}\otimes M_{F,V_\varepsilon}')$, and converges to $Q_{\mathcal{B}}\otimes(M_{F,V_\varepsilon}V_\varepsilon M_{F,V_\varepsilon}') = Q_{\mathcal{B}}\otimes(M_{F,V_\varepsilon}V_\varepsilon)$. The second and third terms are $o_p(1)$, because their expectations are nil under the condition $E[w_{i,t}w_{i,s}w_{i,r}] = 0$, for all t,s,r, and their variances are vanishing. In the fourth term $\frac{1}{n}\sum_m\sum_{i,j\in B_m}(\eta_i\eta_j')\otimes(\tilde{\varepsilon}_i\tilde{\varepsilon}_j') = \frac{1}{n}\sum_m(\sum_{i\in B_m}\eta_i\otimes\tilde{\varepsilon}_i)(\sum_{i\in B_m}\eta_i\otimes\tilde{\varepsilon}_i)'$, vector $\eta_i\otimes\tilde{\varepsilon}_i$ has zero expectation, i.e.,$E[\eta_i\otimes\tilde{\varepsilon}_i] = vec(E[\tilde{\varepsilon}_i\eta_i']) = \check{\sigma}_{ii}vec(M_{F,V_\varepsilon}V_\varepsilon N_{F,V_\varepsilon}') = 0$. Thus, $\frac{1}{n}\sum_m(\sum_{i\in B_m}\eta_i\otimes\tilde{\varepsilon}_i)(\sum_{i\in B_m}\eta_i\otimes\tilde{\varepsilon}_i)'$ is a consistent estimator for the asymptotic variance of

$\frac{1}{\sqrt{n}}\sum_i \eta_i \otimes \tilde{\varepsilon}_i = vec(\frac{1}{\sqrt{n}}\sum_i \tilde{\varepsilon}_i \eta_i') = vec(M_{F,V_\varepsilon} Z_n N_{F,V_\varepsilon}') = (N_{F,V_\varepsilon} \otimes M_{F,V_\varepsilon}) A_T vech(Z_n)$. Thus, we have $\frac{1}{n}\sum_m \sum_{i,j\in B_m} (\eta_i \eta_j') \otimes (\tilde{\varepsilon}_i \tilde{\varepsilon}_j') = B_{\mathcal{B}} + o_p(1)$, where $B_{\mathcal{B}} := (N_{F,V_\varepsilon} \otimes M_{F,V_\varepsilon}) A_T \Omega_Z A_T' (N_{F,V_\varepsilon}' \otimes M_{F,V_\varepsilon}')$. A consistent estimator of the latter matrix is $\hat{B}_{\mathcal{B}} := (N_{\hat{F},\hat{V}_\varepsilon} \otimes M_{\hat{F},\hat{V}_\varepsilon}) A_T \hat{\Omega}_Z A_T' (N_{\hat{F},\hat{V}_\varepsilon}' \otimes M_{\hat{F},\hat{V}_\varepsilon}')$. By combining the above results, we get equation (D.11). By applying the half-vectorization operator, and the rule for the vec of a kronecker product, we get $vech[Q_{\mathcal{B}} \otimes (M_{F,V_\varepsilon} V_\varepsilon)] = P vech(Q_{\mathcal{B}})$ where $P = \frac{1}{2} A_{T(k+1)}' (I_{k+1} \otimes [(K_{T,k+1} \otimes I_T) \otimes (I_{k+1} \otimes vec(M_{F,V_\varepsilon} V_\varepsilon))]) A_{T(k+1)}$. Matrix P is non-singular. The conclusion follows. Q.E.D.

F Practicable Asymptotically Equivalent FA-GMM Estimator

F.1 Modified Moment Restrictions

We first show that, by considering the cross-sectional variance instead of the cross-sectional second moment in the moment restrictions, we obtain an asymptotically equivalent FA-GMM estimator which is easier to compute. Indeed, we can write the lower block of sample moment vector $\hat{g}_n(\theta)$ as $vech(\frac{1}{n}\sum_i y_i y_i' - \Sigma(\vartheta) - \mu\mu') = vech(\hat{V}_y - \Sigma(\vartheta) + \bar{y}\bar{y}' - \mu\mu') = vech[\hat{V}_y - \Sigma(\vartheta)] + vech[(\bar{y} - \mu)\mu' + \mu(\bar{y} - \mu)'] + vech[(\bar{y} - \mu)(\bar{y} - \mu)']$. Moreover, we have $vech[(\bar{y} - \mu)\mu' + \mu(\bar{y} - \mu)'] = A_T'(\mu \otimes I_T)(\bar{y} - \mu)$. Then, we get:

$$\begin{aligned} \hat{g}_n(\theta) &= \begin{pmatrix} I_T & 0_{T\times\frac{T(T-1)}{2}} \\ A_T(\mu \otimes I_T) & I_{\frac{T(T-1)}{2}\times\frac{T(T-1)}{2}} \end{pmatrix} \begin{pmatrix} \bar{y} - \mu \\ vech[\hat{V}_y - \Sigma(\vartheta)] \end{pmatrix} + \hat{r}_n(\theta) \\ &=: D(\theta)\hat{g}_n^*(\theta) + \hat{r}_n(\theta), \end{aligned}$$

where $\hat{r}_n(\theta) := (0_{T\times 1}', vech[(\bar{y} - \mu)(\bar{y} - \mu)']')'$. Term $\hat{r}_n(\theta)$ yields an asymptotically negligible component in the estimator, since $\sqrt{n}\hat{r}_n(\theta_0) = o_p(1)$ and $\frac{\partial \hat{r}_n(\theta_0)}{\partial \theta'} = o_p(1)$. Thus, asymptotically, moment vector $\hat{g}_n(\theta)$ is a parameter-dependent non-singular linear transformation of moment vector $\hat{g}_n^*(\theta)$. Hence, the FA-GMM estimator $\hat{\theta}$ is asymptotically equivalent to

$$\hat{\theta}^* = \arg\min_{\theta\in\Theta} \hat{g}_n^*(\theta)'(\hat{V}_g^*)^{-1}\hat{g}_n^*(\theta), \tag{F.1}$$

where $\hat{V}_g^*$ is a consistent estimator of V_g^*, i.e., the asymptotic variance of $\sqrt{n}\hat{g}_n^*(\tilde{\theta}_0)$. We have the link $V_g = D(\theta_0)V_g^* D(\theta_0)'$.

The two subvectors that build $\hat{g}_n^*(\theta)$ depend separately on parameters μ and ϑ, respectively. Moreover, the upper subvector yields an exactly identified set of moment restrictions for parameter μ. It has important implications for the computation of estimator $\hat{\theta}^*$. First, the FOC for parameter μ yields $\frac{\partial \hat{g}_n^*(\theta)'}{\partial \mu}(\hat{V}_g^*)^{-1}\hat{g}_n^*(\theta) = -(\hat{V}_g^*)^{11}(\bar{y}-\mu) - (\hat{V}_g^*)^{12} vech[\hat{V}_y - \Sigma(\vartheta)] = 0$, where the upper indices indicate blocks of the inverse matrix $(\hat{V}_g^*)^{-1}$. Then, we concentrate out parameter μ to get $\mu = \bar{y} + [(\hat{V}_g^*)^{11}]^{-1}(\hat{V}_g^*)^{12} vech[\hat{V}_y - \Sigma(\vartheta)]$. The concentrated criterion becomes $\hat{g}_n^*(\theta)'(\hat{V}_g^*)^{-1}\hat{g}_n^*(\theta) = vech[\hat{V}_y - \Sigma(\vartheta)]' \left((\hat{V}_g^*)^{22} - (\hat{V}_g^*)^{21}[(\hat{V}_g^*)^{11}]^{-1}(\hat{V}_g^*)^{12} \right) vech[\hat{V}_y - \Sigma(\vartheta)] = vech[\hat{V}_y - \Sigma(\vartheta)]'(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\vartheta)]$, where $\hat{V}_{g,22}^*$ is the lower-right block of $\hat{V}_g^*$. Thus, the components of estimator $\hat{\theta}^*$ simplify to

$$\hat{\mu}^* = \bar{y} + [(\hat{V}_g^*)^{11}]^{-1}(\hat{V}_g^*)^{12} vech[\hat{V}_y - \Sigma(\hat{\vartheta}^*)], \tag{F.2}$$

$$\hat{\vartheta}^* = \arg\min_{\vartheta \in \mathcal{T}} vech[\hat{V}_y - \Sigma(\vartheta)]'(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\vartheta)], \tag{F.3}$$

where $\mathcal{T}$ is a compact subset of $\{\vartheta : h(\vartheta) = 0\}$. Hence, estimator $\hat{\vartheta}^*$ is obtained by minimizing a quadratic form of $vech[\hat{V}_y - \Sigma(\vartheta)]$, and estimator $\hat{\mu}^*$ is obtained by plug-in without optimisation.

F.2 Numerical Computation of the Unconstrained FA-GMM Estimator

Let us now discuss the numerical computation of the estimate $\hat{\vartheta}^*$. As for optimisation of the PML FA criterion, we face two options: the Newton-Raphson (NR) algorithm for the full parameter vector ϑ, and the zigzag algorithm alternating among its components $vec(F)$ and $diag(V_\varepsilon)$. We start with the first option, and leave the second one for Subsection E.2.2.

F.2.1 Newton-Raphson Algorithm

The FOC w.r.t. parameter vector ϑ is $M(\vartheta)'(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\vartheta)] = 0$, where $M(\vartheta) := \frac{\partial vech[\Sigma(\vartheta)]}{\partial \vartheta'} = [A_T'(F \otimes I_T) \;:\; E_{diag,T}]$, and the normalization constraint is $h(\vartheta) = 0$. We expand the

FOC and the constraint around a 'guess' $\vartheta_{(0)}$ that meets the constraints, i.e., $h(\vartheta_{(0)}) = 0$,[20] and get the linearized conditions $M(\vartheta_{(0)})'(\hat{V}^*_{g,22})^{-1} vech[\hat{V}_y - \Sigma(\vartheta_{(0)})] - M(\vartheta_{(0)})'(\hat{V}^*_{g,22})^{-1} M(\vartheta_{(0)})(\vartheta - \vartheta_{(0)}) = 0$ and $H(\vartheta_{(0)})'(\vartheta - \vartheta_{(0)}) = 0$, where $H(\vartheta) := \frac{\partial h(\vartheta)'}{\partial \vartheta}$ (see end of Section D.1 for a characterisation of matrix $H(\vartheta)$). Let $L_H(\vartheta)$ be a full-rank $p \times (p - \rho)$ matrix, such that $I_p - H(\vartheta)(H(\vartheta)'H(\vartheta))^{-1} H(\vartheta)' = L_H(\vartheta)L_H(\vartheta)'$ and $L_H(\vartheta)'L_H(\vartheta) = I_{p-\rho}$. Then, as in Section D.1, the solution of the linearized equation under constraint is

$$\begin{aligned}\vartheta &= \vartheta_{(0)} + L_H(\vartheta_{(0)}) \left(L_H(\vartheta_{(0)})' M(\vartheta_{(0)})'(\hat{V}^*_{g,22})^{-1} M(\vartheta_{(0)}) L_H(\vartheta_{(0)}) \right)^{-1} \\ &\quad \times L_H(\vartheta_{(0)})' M(\vartheta_{(0)})'(\hat{V}^*_{g,22})^{-1} vech[\hat{V}_y - \Sigma(\vartheta_{(0)})]. \qquad \text{(F.4)}\end{aligned}$$

Then, we obtain the updated estimate $\vartheta_{(1)}$ from ϑ after rotating the factor matrix F such that $F'V_\varepsilon^{-1}F$ is diagonal.[21] The NR algorithm iterates this procedure until a convergence criterion is met.

F.2.2 Zigzag Algorithms

A second option is the analogue of a zigzag algorithm (Magnus and Neudecker (2007), Hautsch et al. (2023)). It consists in alternating the computation of an estimate of V_ε for given F, and an estimate of F for given V_ε, until a convergence criterion is met. For the former step, we use that $vech(\hat{V}_y - \Sigma(\vartheta)) = vech(\hat{V}_y - FF') - (E_{diag,T}) diag(V_\varepsilon)$ (see Section D.1 for a characterisation of matrix $E_{diag,T}$). Thus, we get the closed-form solution $diag(V_\varepsilon) = (E'_{diag,T}(\hat{V}^*_{g,22})^{-1} E_{diag,T})^{-1}$ $E'_{diag,T}(\hat{V}^*_{g,22})^{-1} vech(\hat{V}_y - FF')$ by GLS.[22] For the latter step, the problem consists in minimizing the criterion $vech(\hat{V}_y - FF' - V_\varepsilon)'(\hat{V}^*_{g,22})^{-1} vech(\hat{V}_y - FF' - V_\varepsilon)$ w.r.t. the $T \times k$ matrix F such that $F'V_\varepsilon^{-1}F = diag$, for given V_ε. We use that $vech(\hat{V}_y - FF' - V_\varepsilon) = D(V_\varepsilon) vech(\hat{\Xi} - UU')$,

[20] We can for example use the FA estimates obtained by the zigzag routine (Magnus and Neudecker (2007), p. 407) applied to the Gaussian PML criterion as in FGS.

[21] This rotation is needed because the constraint $h(\vartheta) = 0$ is implemented only at first-order, and not exactly, in the NR updating step.

[22] For given F corresponding to the PML factor estimates, we can use such a solution to initialise the NR algorithm, but we need then to rotate the factor starting values to satisfy the standardisation $h(\vartheta_{(0)}) = 0$. In our Monte Carlo results, we have not found much improvement over initialising directly with PML estimates.

where $\hat{\Xi} := V_\varepsilon^{-1/2}\hat{V}_y V_\varepsilon^{-1/2} - I_T$ and $U := V_\varepsilon^{-1/2}F$ and $D(V_\varepsilon) := \frac{1}{2}A_T'(V_\varepsilon^{1/2} \otimes V_\varepsilon^{1/2})A_T$. By direct calculation, we can check that $D(V_\varepsilon)$ is a diagonal $\frac{T(T+1)}{2} \times \frac{T(T+1)}{2}$ matrix, with diagonal elements $V_{\varepsilon,11}, \cdots V_{\varepsilon,TT}, \{\sqrt{V_{\varepsilon,ii}V_{\varepsilon,jj}}\}_{i<j}$. Then, the problem becomes

$$\min_{U \in \mathbb{R}^{T\times k}:\, U'U=diag} vech(\hat{\Xi} - UU')'\hat{\Omega}^* vech(\hat{\Xi} - UU'), \tag{F.5}$$

where $\hat{\Omega}^* = D(V_\varepsilon)(\hat{V}_{g,22}^*)^{-1}D(V_\varepsilon)$. In an alternative parameterization, let $U = \mathcal{U}\Gamma^{1/2}$ where $\mathcal{U}$ is such that $\mathcal{U}'\mathcal{U} = I_k$ and $\Gamma = diag(\gamma)$ for $\gamma \in \mathbb{R}_+^k$. Thus, we can also formulate the minimization problem as:

$$\min_{\mathcal{U} \in \mathbb{R}^{T\times k}:\, \mathcal{U}'\mathcal{U}=I_k} \min_{\gamma \in \mathbb{R}_+^k} vech(\hat{\Xi} - \mathcal{U}\Gamma\mathcal{U}')'\hat{\Omega}^* vech(\hat{\Xi} - \mathcal{U}\Gamma\mathcal{U}'). \tag{F.6}$$

The two formulations (F.5) and (F.6) lead to different NR steps, that we detail below. We use the terminology zigzag1 and zizag2 to differentiate them in the numerical study of Section E.5.

We use the following notation. For two matrices $B = [b_1 : \ldots : b_m]$ and $C = [c_1 : \ldots : c_m]$, let us define the following variations of the Kronecker product: $B \otimes_< C := [\{b_i \otimes c_j\}_{i<j}]$ and $B \otimes_= C := [\{b_i \otimes c_i\}_{i=1,\ldots,n}]$. We define $B \otimes_\leq C$ and $B \otimes_\geq C$ similarly. Pairs (i,j) are ranked as the indices of a $m \times m$ matrix read row-wise. With this notation, $A_m = \left[\sqrt{2}(I_m \otimes_= I_m) \,:\, (I_m \otimes_< I_m)\right]$. The columns of matrix $B \otimes_\star C$ are a subset of the columns of $B \otimes C$, namely $B \otimes_\star C = (B \otimes C)(I_m \otimes_\star I_m)$, for $\star$ denoting either $<$, $=$, $\leq$, or $\geq$.

For Problem (F.5), we use the vector constraint $h_U(u) = 0$, where $h_U(u) = E_{off,k}vech(U'U)$ and $u = vec(U)$. The gradient is $\frac{\partial h_U(u)}{\partial u'} = E_{off,k}A_k'(I_k \otimes U')$. Hence, we get the $(Tk) \times \frac{k(k-1)}{2}$ matrix $H_U(u) := \frac{\partial h_U'(u)}{\partial u} = (I_k \otimes U)A_k E_{off,k}' = (I_k \otimes U)(I_k \otimes_< I_k) = I_k \otimes_< U$. Then, $L_{H_U}(u)$ is the $(Tk) \times (Tk - \frac{k(k-1)}{2})$ matrix defined by $L_{H_U}(u) := \left[(I_k \otimes_\geq \tilde{U}) \,:\, (I_k \otimes U_\perp)\right]$, where $\tilde{U} := U(U'U)^{-1/2}$ and $U_\perp$ is a $T \times (T-k)$ matrix with orthonormal columns that are orthogonal to the range of U. Indeed, the $Tk - \frac{k(k-1)}{2}$ columns of the matrix $L_{H_U}(u)$ are mutually orthogonal, normalized to length 1, and are orthogonal to any columns of $H_U(u)$, for any u such that $h_U(u) =$

0. Moreover, $M_U(u) = -\frac{\partial vech(\hat{\Xi}-UU')}{\partial vec(U)'} = A_T'(U \otimes I_T)$. Hence, we update

$$u = u_{(0)} + L_{H_U}(u_{(0)}) \left(L_{H_U}(u_{(0)})' M_U(u_{(0)})' \hat{\Omega}^* M_U(u_{(0)}) L_{H_U}(u_{(0)}) \right)^{-1} \times L_{H_U}(u_{(0)})' M_U(u_{(0)})' \hat{\Omega}^* vech[\hat{\Xi} - U_{(0)} U_{(0)}'], \quad \text{(F.7)}$$

and we get $U_{(1)}$ by transforming U to impose $U_{(1)}'U_{(1)} = diag$, i.e., $U_{(1)} = U(U'U)^{-1/2} diag(U'U)^{1/2}$. We can get a more explicit rewriting of the updating equation (F.7) by using

$$M_U(u_{(0)})'\hat{\Omega}^* M_U(u_{(0)}) = (F_{(0)}' \otimes V_\varepsilon^{1/2}) A_T (\hat{V}_{g,22}^*)^{-1} A_T' (F_{(0)} \otimes V_\varepsilon^{1/2}),$$
$$M_U(u_{(0)})'\hat{\Omega}^* vech[\hat{\Xi} - U_{(0)}U_{(0)}'] = (F_{(0)}' \otimes V_\varepsilon^{1/2}) A_T (\hat{V}_{g,22}^*)^{-1} vech(\hat{V}_y - F_{(0)}F_{(0)}' - V_\varepsilon),$$

where $F_{(0)} = V_\varepsilon^{1/2} U_{(0)}$, and $(F_{(0)} \otimes V_\varepsilon^{1/2}) L_{H_U}(u_{(0)}) = \left([F_{(0)} \otimes_\geq (F_{(0)} \Gamma_{(0)}^{-1/2})] \, : \, F_{(0)} \otimes G_{(0)} \right)$, with $G_{(0)} = V_\varepsilon^{1/2} U_{\perp,(0)}$ and $\Gamma_{(0)} = U_{(0)}' U_{(0)}$. Hence, we can write (F.7) as:

$$u = u_{(0)} + L_{H_U}(u_{(0)}) \left(J_{(0)}' (\hat{V}_{g,22}^*)^{-1} J_{(0)} \right)^{-1} J_{(0)}' (\hat{V}_{g,22}^*)^{-1} vech(\hat{V}_y - F_{(0)}F_{(0)}' - V_\varepsilon), \quad \text{(F.8)}$$

where $J_{(0)} = A_T' \left([F_{(0)} \otimes_\geq (F_{(0)} \Gamma_{(0)}^{-1/2})] \, : \, F_{(0)} \otimes G_{(0)} \right)$.

For Problem (F.6), we use the ideas in Manton et al. (2003), namely we solve the inner minimization problem in closed form, and then apply the NR method to the outer minimization after concentration, by parameterizing deviations in matrix $\mathcal{U}$ in terms of the orthogonal complement of its column space. We use $vech(\mathcal{U}\Gamma\mathcal{U}') = \frac{1}{2} A_T' (\mathcal{U} \otimes \mathcal{U}) vec(\Gamma) = \frac{1}{2} A_T' (\mathcal{U} \otimes \mathcal{U})(I_k \otimes_= I_k)\gamma = \frac{1}{2} A_T' (\mathcal{U} \otimes_= \mathcal{U})\gamma =: \mathcal{N}(\mathcal{U})\gamma$. Then, the solution of the inner minimization problem in (F.6) is $\gamma = (\mathcal{N}(\mathcal{U})'\mathcal{N}(\mathcal{U}))^{-1}\mathcal{N}(\mathcal{U})' vech(\hat{\Xi})$. After replacement in (F.6), we get the concentrated criterion value $\phi(\mathcal{U}) = vech(\hat{\Xi})' \left(I_{\frac{T(T+1)}{2}} - \mathcal{P}(\mathcal{U}) \right) vech(\hat{\Xi})$ for the outer minimization problem under constraint $\mathcal{U}'\mathcal{U} = I_k$, where $\mathcal{P}(\mathcal{U}) := \mathcal{N}(\mathcal{U})(\mathcal{N}(\mathcal{U})'\mathcal{N}(\mathcal{U}))^{-1}\mathcal{N}(\mathcal{U})'$. To solve the latter constrained minimization problem by the NR algorithm, we parametrize deviations around $\mathcal{U}_{(0)}$ as $\mathcal{U} = \mathcal{U}_{(0)} + \mathcal{U}_{\perp,(0)}\mathcal{A}$, where $\mathcal{U}_{\perp,(0)}$ is a $T \times (T-k)$ matrix with orthonormal columns which are orthogonal to the range of $\mathcal{U}_{(0)}$, and $\mathcal{A}$ is a $(T-k) \times k$ parameter matrix. Indeed, such parameterization imposes the constraint at first order in $\mathcal{A}$. Let us define $\phi_{(0)}(a) := \phi(\mathcal{U}_{(0)} + \mathcal{U}_{\perp,(0)}\mathcal{A})$ for

$a := vec(\mathcal{A})$. We get a by the NR step

$$a = \left(-\frac{\partial^2 \phi_{(0)}(0)}{\partial a \partial a'}\right)^{-1} \frac{\partial \phi_{(0)}(0)}{\partial a}, \tag{F.9}$$

and then we get $\mathcal{U}_{(1)} = \mathcal{U}(\mathcal{U}'\mathcal{U})^{-1/2}$. To implement the updating rule (F.9), we need the first- and second-order partial derivatives of function $\phi(\mathcal{U}_{(0)} + \mathcal{U}_{\perp,(0)}\mathcal{A})$ w.r.t. a, that we give explicitly next.

Lemma 9 *The first- and second-order partial derivatives are given by:*

$$\begin{aligned}
\frac{\partial \phi_{(0)}(0)}{\partial a} &= -(I_k \otimes \mathcal{U}_{\perp,(0)})'[(I_k \otimes \mathcal{G}_{(0)}) + (\mathcal{H}_{(0)} \otimes I_T)]'[(I_k \otimes_= I_k) \otimes A_T] \\
&\quad \times \left([(\mathcal{N}'_{(0)}\mathcal{N}_{(0)})^{-1}\mathcal{N}'_{(0)} vech(\hat{\Xi})] \otimes [(I - \mathcal{P}_{(0)}) vech(\hat{\Xi})]\right) \\
&= -[(I_k \otimes (\mathcal{G}_{(0)} U_{\perp,(0)})) + (\mathcal{H}_{(0)} \otimes U_{\perp,(0)})]'[(I_k \otimes_= I_k) \otimes A_T] \\
&\quad \times \left([(\mathcal{N}'_{(0)}\mathcal{N}_{(0)})^{-1}\mathcal{N}'_{(0)} vech(\hat{\Xi})] \otimes [(I - \mathcal{P}_{(0)}) vech(\hat{\Xi})]\right)
\end{aligned} \tag{F.10}$$

where $\mathcal{G} := (K_{k,T} \otimes I_T)(I_T \otimes vec(\mathcal{U}))$ and $\mathcal{H} := (I_k \otimes K_{k,T})(vec(\mathcal{U}) \otimes I_k)$, and the index (0) indicates that the quantities are evaluated for $\mathcal{U} = \mathcal{U}_{(0)}$, and:

$$\begin{aligned}
&\frac{\partial^2 \phi_{(0)}(0)}{\partial a \partial a'} = (I_k \otimes \mathcal{U}_{\perp,(0)})' \left(-(I_{kT} \otimes \hat{\xi}_{(0)})' S_3 - S_3'(I_{kT} \otimes \hat{\xi}_{(0)})\right)(I_k \otimes \mathcal{U}_{\perp,(0)}) \\
&+\frac{1}{2}[(I_k \otimes (\mathcal{G}_{(0)}\mathcal{U}_{\perp,(0)})) + (\mathcal{H}_{(0)} \otimes \mathcal{U}_{\perp,(0)})]'[(I_k \otimes_= I_k) \otimes A_T] \\
&\times \Big\{[((\mathcal{N}'_{(0)}\mathcal{N}_{(0)})^{-1}\mathcal{N}'_{(0)}) \otimes (\hat{\eta}_{2,(0)}\hat{\eta}'_{1,(0)})]K_{\frac{T(T+1)}{2},k} - [(\mathcal{N}'_{(0)}\mathcal{N}_{(0)})^{-1} \otimes (\hat{\eta}_{2,(0)}\hat{\eta}'_{2,(0)})] \\
&+(\hat{\eta}_{1,(0)}\hat{\eta}'_{1,(0)}) \otimes (I_{\frac{T(T+1)}{2}} - \mathcal{P}_{(0)}) + [(\hat{\eta}_{1,(0)}\hat{\eta}'_{2,(0)}) \otimes (\mathcal{N}_{(0)}(\mathcal{N}'_{(0)}\mathcal{N}_{(0)})^{-1})]K_{\frac{T(T+1)}{2},k}\Big\} \\
&\times[(I_k \otimes_= I_k) \otimes A_T]'[(I_k \otimes (\mathcal{G}_{(0)}\mathcal{U}_{\perp,(0)})) + (\mathcal{H}_{(0)} \otimes \mathcal{U}_{\perp,(0)})],
\end{aligned} \tag{F.11}$$

where $\hat{\xi} := [(I_k \otimes_= I_k) \otimes A_T](\hat{\eta}_1 \otimes \hat{\eta}_2)$ with $\hat{\eta}_1 := (\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}' vech(\hat{\Xi})$ and $\hat{\eta}_2 := (I_{\frac{T(T+1)}{2}} - \mathcal{P}) vech(\hat{\Xi})$, and $S_3 := [S_1 \otimes (K_{k,T} \otimes I_T)][vec(I_T) \otimes I_{Tk}] + [I_k \otimes (S_2(I_k \otimes K_{k,T}))][(K_{k,Tk} \otimes I_k)(I_{Tk} \otimes vec(I_k))]$, with $S_1 := (I_k \otimes K_{T,k})(vec(I_k) \otimes I_T)$ and $S_2 := (K_{T,k^2T} \otimes I_T)(I_{k^2T} \otimes vec(I_T))$.

F.3 Constrained FA-GMM Estimators

In this subsection, we show that the arguments and numerical methods presented in Subsections E.1 and E.2 extend to the computation of the constrained FA-GMM estimator $\hat{\theta}_c$. First, since imposing the constraint $a(\theta) = 0$ is equivalent to restricting the parameter ϑ to belong to set $\mathcal{T}_c := \{\vartheta : a(\theta) = 0\}$, and it does not involve parameter vector μ, from the arguments in Subsection E.1, we get that $\hat{\theta}_c$ is asymptotically equivalent to a constrained FA-GMM estimator $\hat{\theta}_c^*$ based on moment vector $g_n^*(\theta)$ including the cross-sectional variance, and that the components of vector $\hat{\theta}_c^*$ are computed as:

$$\hat{\mu}_c^* = \bar{y} + [(\hat{V}_g^*)^{11}]^{-1}(\hat{V}_g^*)^{12} vech[\hat{V}_y - \Sigma(\hat{\theta}_c^*)], \tag{F.12}$$

$$\hat{\vartheta}_c^* = \arg\min_{\vartheta \in \mathcal{T}_c} vech[\hat{V}_y - \Sigma(\vartheta)]'(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\vartheta)]. \tag{F.13}$$

Moreover, the FOC for the Lagrangian of the constrained minimization problem defining $\hat{\vartheta}_c^*$, i.e., $-M(\hat{\vartheta}_c^*)'(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\hat{\vartheta}_c^*)] + \frac{\partial a(\hat{\theta}_c^*)'}{\partial \vartheta}\hat{\lambda}^* = 0$, with $a(\theta) = L_{1_T}' diag(V_\varepsilon)$, yields the equation $-E_{diag,T}'(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\hat{\vartheta}_c^*)] + L_{1_T}\hat{\lambda}^* = 0$, that can be used to get the Lagrange-multiplier vector $\hat{\lambda}^* = L_{1_T}' E_{diag,T}(\hat{V}_{g,22}^*)^{-1} vech[\hat{V}_y - \Sigma(\hat{\vartheta}_c^*)]$ as a function of the constrained estimator $\hat{\vartheta}_c^*$.

To compute estimate $\hat{\vartheta}_c^*$ in (F.13), we can use either the NR algorithm for the full vector, or the zigzag algorithm. In the first case, we augment the constraint vector and stack $h(\vartheta)$ and $a(\vartheta)$. We get matrix $H_c(\vartheta) := \left[\frac{\partial h(\vartheta)'}{\partial \vartheta} \; : \; \frac{\partial a(\vartheta)'}{\partial \vartheta}\right]$, and define accordingly matrix $L_{H_c}(\vartheta)$ to span the orthogonal complement of the columns of $H_c(\vartheta)$. Then, the NR update equation is as in (F.4) after replacing $L_H(\vartheta)$ with $L_{H_c}(\vartheta)$. Again, to get $\vartheta_{c,(1)}$, we need to rotate the columns of the factor estimate to meet the nonlinear constraint $F'V_\varepsilon^{-1}F = diag$. If the constraint $a(\theta) = 0$ is linear as in the sphericity test, we do not need to enforce it after the update, as it is automatically imposed in the NR step.

In the zigzag algorithm, the step to get F for given V_ε is unchanged compared to the unconstrained algorithms defined in Subsections E.2.1 and E.2.2, since the constrained vector $a(\theta)$ does not involve parameter F in the sphericity test. In the step aimed at getting V_ε given F, we impose the linear constraint $L_{1_T}' diag(V_\varepsilon) = 0$, which amounts to use the constrained GLS estimate $diag(V_{\varepsilon,c}) = \left[I_T - (E_{diag,T}'(\hat{V}_{g,22}^*)^{-1}E_{diag,T})^{-1}L_{1_T}\left(L_{1_T}'(E_{diag,T}'(\hat{V}_{g,22}^*)^{-1}E_{diag,T})^{-1}L_{1_T}\right)^{-1}L_{1_T}'\right]$ $diag(V_\varepsilon)$ instead ot the unconstrained one $diag(V_\varepsilon) = (E_{diag,T}'(\hat{V}_{g,22}^*)^{-1}E_{diag,T})^{-1}E_{diag,T}'(\hat{V}_{g,22}^*)^{-1}$ $vech(\hat{V}_y - FF')$.

F.4 Proof of Lemma 9

Let us start with the first-order derivatives. We use the notation of differentials (see Magnus and Neudecker (2007), Chapters 5 and 6). The differential of function ϕ is $d\phi = -vech(\hat{\Xi})' d\mathcal{P} vech(\hat{\Xi})$, where, by standard arguments, $d\mathcal{P} = (I_{\frac{T(T+1)}{2}} - \mathcal{P})(d\mathcal{N})(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}' + \mathcal{N}(\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'(I_{\frac{T(T+1)}{2}} - \mathcal{P})$. Then: $d\phi = -2vech(\hat{\Xi})'(I_{\frac{T(T+1)}{2}} - \mathcal{P})(d\mathcal{N})(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi}) = -2\left([(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P})vech(\hat{\Xi})]\right)' vec(d\mathcal{N})$. The differential of function $\mathcal{N}$ is given by $d\mathcal{N} = \frac{1}{2}A_T'[(d\mathcal{U}) \otimes \mathcal{U} + \mathcal{U} \otimes (d\mathcal{U})](I_k \otimes_= I_k)$ and, after vectorization, we get $vec(d\mathcal{N}) = \frac{1}{2}[(I_k \otimes_= I_k) \otimes A_T]' vec[(d\mathcal{U}) \otimes \mathcal{U} + \mathcal{U} \otimes (d\mathcal{U})]$. Next, we use the vectorization of the Kronecker product (see Magnus and Neudecker (2007), p. 48)[23], so that $vec[(d\mathcal{U}) \otimes \mathcal{U} + \mathcal{U} \otimes (d\mathcal{U})] = [(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]vec(d\mathcal{U})$, where $\mathcal{G} := (K_{k,T} \otimes I_T)(I_T \otimes vec(\mathcal{U}))$ is $(kT^2) \times T$ matrix and $\mathcal{H} := (I_k \otimes K_{k,T})(vec(\mathcal{U}) \otimes I_k)$ is $(k^2 T) \times k$ matrix. Thus, we get:

$$\begin{aligned} d\phi &= -\left([(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P})vech(\hat{\Xi})]\right)' [(I_k \otimes_= I_k) \otimes A_T]' \\ &\quad \times [(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]vec(d\mathcal{U}). \end{aligned} \quad \text{(F.14)}$$

By using $d\mathcal{U} = \mathcal{U}_{\perp,(0)} d\mathcal{A}$ and $vec(d\mathcal{U}) = (I_k \otimes \mathcal{U}_{\perp,(0)})da$, we get the gradient vector in (F.10).

Let us now establish the second-order derivatives. To compute the second-order differential $d^2\phi$, we go back to formula (F.14) and compute the differential of the vector function multiplying $vec(d\mathcal{U})$. By the product rule for differentials, we have:

$$\begin{aligned} d^2\phi &= -(vec(d\mathcal{U}))'[(I_k \otimes d\mathcal{G}) + (d\mathcal{H} \otimes I_T)]'[(I_k \otimes_= I_k) \otimes A_T] \\ &\qquad \times \left([(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P})vech(\hat{\Xi})]\right) \\ &\quad -(vec(d\mathcal{U}))'[(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]'[(I_k \otimes_= I_k) \otimes A_T] \\ &\qquad \times \left([d\{(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'\}vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P})vech(\hat{\Xi})]\right) \\ &\quad +(vec(d\mathcal{U}))'[(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]'[(I_k \otimes_= I_k) \otimes A_T] \\ &\qquad \times \left([(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi})] \otimes [(d\mathcal{P})vech(\hat{\Xi})]\right), \end{aligned} \quad \text{(F.15)}$$

[23]It amounts to use the following result: let A and B be $m \times n$ and $p \times q$ matrices, then $vec(A \otimes B) = (I_n \otimes G)vec(A) = (H \otimes I_p)vec(B)$, where $G = (K_{q,m} \otimes I_p)(I_m \otimes vec(B))$ and $H = (I_n \otimes K_{q,m})(vec(A) \otimes I_q)$.

where $d\mathcal{G} = (K_{k,T} \otimes I_T)(I_T \otimes vec(d\mathcal{U}))$ and $d\mathcal{H} = (I_k \otimes K_{k,T})(vec(d\mathcal{U}) \otimes I_k)$ and $d\{(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'\} = -(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'(d\mathcal{N})(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}' + (\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'(I_{\frac{T(T+1)}{2}} - \mathcal{P})$. We obtain the matrix of the second-order partial derivatives of ϕ by writing $d^2\phi$ as a quadratic form in vector $vec(d\mathcal{U})$. To establish this matrix, we need to rewrite the three terms on the RHS of (F.15) using repeatedly the vectorization of matrix products and Kronecker products (footnote 8).

(a) For the first term on the RHS of (F.15), we have successively:

$$
\begin{aligned}
& [(I_k \otimes d\mathcal{G}) + (d\mathcal{H} \otimes I_T)]'[(I_k \otimes_= I_k) \otimes A_T] \left([(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}' vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P}) vech(\hat{\Xi})] \right) \\
=: \; & [(I_k \otimes d\mathcal{G}) + (d\mathcal{H} \otimes I_T)]'\hat{\xi} = vec\left(\hat{\xi}'[(I_k \otimes d\mathcal{G}) + (d\mathcal{H} \otimes I_T)] \right) \\
= \; & (I_{kT} \otimes \hat{\xi})' vec[(I_k \otimes d\mathcal{G}) + (d\mathcal{H} \otimes I_T)] = (I_{kT} \otimes \hat{\xi})' \left[(S_1 \otimes I_{kT^2}) vec(d\mathcal{G}) + (I_k \otimes S_2) vec(d\mathcal{H}) \right] \\
= \; & (I_{kT} \otimes \hat{\xi})' \{ (S_1 \otimes I_{kT^2})[I_T \otimes (K_{k,T} \otimes I_T)] vec(I_T \otimes vec(d\mathcal{U})) \\
& + (I_k \otimes S_2)[I_k \otimes (I_k \otimes K_{k,T})] vec(vec(d\mathcal{U}) \otimes I_k) \} \\
= \; & (I_{kT} \otimes \hat{\xi})' \{ [S_1 \otimes (K_{k,T} \otimes I_T)][vec(I_T) \otimes I_{Tk}] \\
& + [I_k \otimes (S_2(I_k \otimes K_{k,T}))][(K_{k,Tk} \otimes I_k)(I_{Tk} \otimes vec(I_k))] \} \, vec(d\mathcal{U}) \\
=: \; & (I_{kT} \otimes \hat{\xi})' S_3 vec(d\mathcal{U}), \qquad \text{(F.16)}
\end{aligned}
$$

where we define the vector $\hat{\xi} := [(I_k \otimes_= I_k) \otimes A_T] \left([(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}' vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P}) vech(\hat{\Xi})] \right)$ and matrices $S_1 := (I_k \otimes K_{T,k})(vec(I_k) \otimes I_T)$ and $S_2 := (K_{T,k^2T} \otimes I_T)(I_{k^2T} \otimes vec(I_T))$.

(b) For the second term on the RHS of (F.15), we use:

$$\begin{aligned}
& [d\{(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'\}vech(\hat{\Xi})] \otimes [(I_{\frac{T(T+1)}{2}} - \mathcal{P})vech(\hat{\Xi})] \\
= \; & [-(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'(d\mathcal{N})\hat{\eta}_1 + (\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'\hat{\eta}_2] \otimes \hat{\eta}_2 \\
= \; & vec\left[\hat{\eta}_2\left(-(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'(d\mathcal{N})\hat{\eta}_1 + (\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'\hat{\eta}_2\right)'\right] \\
= \; & (I_k \otimes \hat{\eta}_2)vec\left[\left(-(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'(d\mathcal{N})\hat{\eta}_1 + (\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'\hat{\eta}_2\right)'\right] \\
= \; & (I_k \otimes \hat{\eta}_2)vec\left[-\hat{\eta}_1'(d\mathcal{N})'\mathcal{N}(\mathcal{N}'\mathcal{N})^{-1} + \hat{\eta}_2'(d\mathcal{N})(\mathcal{N}'\mathcal{N})^{-1}\right] \\
= \; & \left(-[((\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}') \otimes (\hat{\eta}_2\hat{\eta}_1')]K_{\frac{T(T+1)}{2},k} + [(\mathcal{N}'\mathcal{N})^{-1} \otimes (\hat{\eta}_2\hat{\eta}_2')]\right) vec(d\mathcal{N}) \\
= \; & \frac{1}{2}\left(-[((\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}') \otimes (\hat{\eta}_2\hat{\eta}_1')]K_{\frac{T(T+1)}{2},k} + [(\mathcal{N}'\mathcal{N})^{-1} \otimes (\hat{\eta}_2\hat{\eta}_2')]\right)[(I_k \otimes_= I_k) \otimes A_T]' \\
& \times[(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]vec(d\mathcal{U}) \qquad \text{(F.17)}
\end{aligned}$$

where we define the vectors $\hat{\eta}_1 := (\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi})$ and $\hat{\eta}_2 := (I_{\frac{T(T+1)}{2}} - \mathcal{P})vech(\hat{\Xi})$.

(c) For the third term on the RHS of (F.15), we use:

$$\begin{aligned}
& [(\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}'vech(\hat{\Xi})] \otimes [(d\mathcal{P})vech(\hat{\Xi})] \\
= \; & \hat{\eta}_1 \otimes \left((I_{\frac{T(T+1)}{2}} - \mathcal{P})(d\mathcal{N})\hat{\eta}_1 + \mathcal{N}(\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'\hat{\eta}_2\right) \\
= \; & vec\left((I_{\frac{T(T+1)}{2}} - \mathcal{P})(d\mathcal{N})\hat{\eta}_1\hat{\eta}_1' + \mathcal{N}(\mathcal{N}'\mathcal{N})^{-1}(d\mathcal{N})'\hat{\eta}_2\hat{\eta}_1'\right) \\
= \; & \left[(\hat{\eta}_1\hat{\eta}_1') \otimes (I_{\frac{T(T+1)}{2}} - \mathcal{P}) + [(\hat{\eta}_1\hat{\eta}_2') \otimes (\mathcal{N}(\mathcal{N}'\mathcal{N})^{-1})]K_{\frac{T(T+1)}{2},k}\right] vec(d\mathcal{N}) \\
= \; & \frac{1}{2}\left[(\hat{\eta}_1\hat{\eta}_1') \otimes (I_{\frac{T(T+1)}{2}} - \mathcal{P}) + [(\hat{\eta}_1\hat{\eta}_2') \otimes (\mathcal{N}(\mathcal{N}'\mathcal{N})^{-1})]K_{\frac{T(T+1)}{2},k}\right][(I_k \otimes_= I_k) \otimes A_T]' \\
& \times[(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]vec(d\mathcal{U}). \qquad \text{(F.18)}
\end{aligned}$$

We plug equations (F.16)-(F.18) into (F.15), and exploit the symmetry of the matrix defining the quadratic form, to get:

$$\begin{aligned}
d^2\phi \; = \; & \frac{1}{2}vec(d\mathcal{U})'\left(-(I_{kT} \otimes \hat{\xi})'S_3 - S_3'(I_{kT} \otimes \hat{\xi})\right)vec(d\mathcal{U}) \\
& + \frac{1}{2}vec(d\mathcal{U})'[(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]'[(I_k \otimes_= I_k) \otimes A_T]\Big\{[((\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}') \otimes (\hat{\eta}_2\hat{\eta}_1')]K_{\frac{T(T+1)}{2},k} \\
& -[(\mathcal{N}'\mathcal{N})^{-1} \otimes (\hat{\eta}_2\hat{\eta}_2')] + (\hat{\eta}_1\hat{\eta}_1') \otimes (I_{\frac{T(T+1)}{2}} - \mathcal{P}) + [(\hat{\eta}_1\hat{\eta}_2') \otimes (\mathcal{N}(\mathcal{N}'\mathcal{N})^{-1})]K_{\frac{T(T+1)}{2},k}\Big\} \\
& \times[(I_k \otimes_= I_k) \otimes A_T]'[(I_k \otimes \mathcal{G}) + (\mathcal{H} \otimes I_T)]vec(d\mathcal{U}),
\end{aligned}$$

where $\hat{\xi} = [(I_k \otimes_= I_k) \otimes A_T](\hat{\eta}_1 \otimes \hat{\eta}_2)$ and we can check that matrix $[(\hat{\eta}_1 \hat{\eta}_2') \otimes (\mathcal{N}(\mathcal{N}'\mathcal{N})^{-1})] K_{\frac{T(T+1)}{2},k}$ is the transposed of $[((\mathcal{N}'\mathcal{N})^{-1}\mathcal{N}') \otimes (\hat{\eta}_2 \hat{\eta}_1')] K_{\frac{T(T+1)}{2},k}$ by the properties of the commutation matrix. By using $vec(d\mathcal{U}) = (I_k \otimes \mathcal{U}_{\perp,(0)})da$, we get the Hessian matrix explicitly in (F.11). Q.E.D.

F.5 Numerical Study of the FA-GMM estimators

This section compares the performance of the algorithms developed in Sections E.2.1 (NR) and E.2.2 (zigzag1-2). We investigate both the unconstrained (Section E.1) and constrained (Section E.3) FA-GMM estimators. We set the starting values at the PML estimates of FGS obtained with the zizag routine of Magnus and Neudecker (2007). We also rely on them to compute optimal weighting since they are consistent. Table 2 reports Root Mean Square Error (RMSE), Bias, and Standard Deviation (SD) for the PML and PCA estimators as well as the unconstrained and constrained FA-GMM (NR and zigzag1-2) estimators. We rely on the same simulation design as the one for the MC experiments (Section 5) for the size, power, and local power of the test in a non-gaussian setting (DGP1-3). We compute RMSE as the square root of $E[\|\hat{\theta} - \theta\|^2] = tr(V[\hat{\theta}]) + \|E[\hat{\theta}] - \theta\|^2$, Bias as $\|E[\hat{\theta}] - \theta\|$, and SD as $tr(V[\hat{\theta}])^{1/2}$ for the estimates of $\theta = (\mu', vec(F)', diag(V_\varepsilon)')'$. We use $n = 500$ and $T = 12$. Under sphericity (DGP1), PCA and constrained estimators (con) fare better in terms of RMSE and SD. Under DGP2-3, since sphericity does not hold, the PCA estimators exhibit huge RMSE, Bias, and SD, and this even under a local alternative. It is also the case for constrained NR and zigzag1-2 algorithms, but not for the unconstrained FA-GMM and PML estimators. We observe a slight advantage in terms of RMSE and SD for the unconstrained FA-GMM estimator (NR) for the three DGPs (as expected, at least asymptotically, from optimal weighting delivering efficiency gains w.r.t. PML), and we think it should be the default choice in empirics. The failure of PCA spotted in Table 2 under DGP2-3 should convince empirical researchers to refrain from running PCA in short panels when errors are not spherical. We have observed a deterioration of the performance of the FA-GMM and PML estimates when $T = 6$ for cases, where matrix $M_{F,V_\varepsilon} \odot M_{F,V_\varepsilon}$ is badly conditioned: its smallest singular value is close to zero and local identification is challenged (Assumption A.4). It happens in approximately 17% of the 100 simulated factor paths. When we remove those cases, we observe the same performance as for $T = 12$ (and $T = 24$).

We can also look at computational time for the three algorithms NR and zizag1-2. The NR algorithm needs more iterations than the zigzag routines, but the zigzag routines make NR steps inside each iteration which slows down the computations. The three algorithms give close numerical results, even if NR is slightly better in terms of RMSE and SD, for the unconstrained panel FA-GMM estimators (see Table 2). Their good accuracy shows their relevance for applied work. Our simulation results point to a strong advantage of the NR algorithm in terms of computational speed. Indeed, the average computation time for the NR algorithm is around 5 millisecond, compared to 31-119 milliseconds for the zigzag1-2 algorithms. Relative computational times are even slightly lower for the constrained version of NR w.r.t. zigzag1-2. It explains why we opt for the NR algorithm in our MC experiments (Section 5) and our empirics (Section 6).

$n = 500$	DGP1			DGP2			DGP3		
$T = 12$	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD
PML	1.33	0.11	1.33	1.96	0.14	1.96	1.56	0.12	1.55
NR	1.32	0.11	1.32	1.95	0.14	1.94	1.54	0.12	1.53
Zigzag1	1.33	0.11	1.32	1.95	0.14	1.94	1.55	0.12	1.55
Zigzag2	1.35	0.11	1.34	1.96	0.15	1.95	1.56	0.13	1.56
PCA	0.65	0.06	0.65	9.54	9.46	0.92	1.71	1.53	0.76
NR con	0.68	0.25	0.64	10.75	10.62	1.44	2.45	2.05	1.12
Zigzag1 con	0.70	0.25	0.65	10.81	8.35	4.82	2.29	2.05	0.98
Zigzag2 con	1.06	0.27	1.02	9.52	7.66	3.95	2.30	1.88	1.31

Table 2: We provide the average Root Mean Square Error (RMSE), Bias, and Standard Deviation (SD), for the different estimators and algorithms, under DGP1-3 and PML, PCA, and FA-GMM approaches. We take averages across 100 different draws of the factor path for $n = 500$, $T = 12$, and two latent factors.

G Monte Carlo Experiments for the LM and LR Tests

This section gathers the Monte Carlo Experiments for the LM and LR tests. Tables 3 and 4 show numbers similar to the entries of Table 1 for the W test.

ξ_n^{LM}	Size (%)			Global Power (%)			Local Power (%)		
T	6	12	24	6	12	24	6	12	24
$n = 500$	6.5	6.2	6.2	93	100	100	30	98	99
	(0.4)	(0.3)	(0.3)	(17.4)	(0.0)	(0.0)	(10.6)	(4.4)	(0.5)
$n = 1000$	5.9	5.8	5.5	97	100	100	27	96	100
	(0.3)	(0.3)	(0.3)	(13.0)	(0.0)	(0.0)	(9.8)	(7.8)	(0.0)
$n = 5000$	5.4	5.3	5.2	100	100	100	23	90	100
	(0.3)	(0.3)	(0.3)	(0.0)	(0.0)	(0.0)	(8.8)	(12.7)	(0.0)

Table 3: For each sample size combination (n, T), we provide the average size, power, and local power in % for the test statistic ξ_n^{LM} under DGP1-3. Nominal size is 5%. In parentheses, we report the standard deviations for size, power, and local power across 100 different draws of the factor path. The number of latent factors is set equal to two.

ξ_n^{LR}	Size (%)			Global Power (%)			Local Power (%)		
T	6	12	24	6	12	24	6	12	24
$n = 500$	6.5	6.2	6.2	93	99	100	30	98	99
	(0.4)	(0.3)	(0.3)	(17.4)	(0.0)	(0.0)	(10.6)	(4.4)	(0.5)
$n = 1000$	5.9	5.8	5.6	97	100	100	27	96	100
	(0.3)	(0.3)	(0.3)	(13.0)	(0.0)	(0.0)	(9.8)	(7.8)	(0.0)
$n = 5000$	5.4	5.3	5.2	100	100	100	23	90	100
	(0.3)	(0.3)	(0.3)	(0.0)	(0.0)	(0.0)	(8.8)	(12.7)	(0.0)

Table 4: For each sample size combination (n, T), we provide the average size, power, and local power in % for the test statistic ξ_n^{LR} under DGP1-3. Nominal size is 5%. In parentheses, we report the standard deviations for size, power, and local power across 100 different draws of the factor path. The number of latent factors is set equal to two.

H Orthogonality Restrictions with Second-Order Moments

This section establishes orthogonality restrictions based on second-order moment restrictions for panel model (4). We build on ideas of Arellano and Bonhommes (2012) Section 3.4, extending the analysis of our FA seting to the more general model (4) and accommodating a random coefficient in the variance (see also footnote 17 in Arellano and Bonhomme (2012)). Specifically, for model $\mathcal{Y}_i = d(z_i, \zeta) + R(z_i, \zeta)\beta_i + \varepsilon_i$, let us assume $E[\varepsilon_i | z_i, \beta_i] = 0$ and $V[\varepsilon_i | z_i, \beta_i, \sigma_i^2] = \sigma_i^2 V_\varepsilon(z_i, \psi)$, where σ_i^2 is a random coefficient and ψ an unknown parameter vector. By taking the Kronecker product of $\mathcal{Y}_i$ with itself, and computing the conditional expectation, we get:

$$\begin{aligned} E[\mathcal{Y}_i \otimes \mathcal{Y}_i | z_i, \beta_i, \sigma_i^2] &= [d(z_i, \zeta) \otimes d(z_i, \zeta)] + [d(z_i, \zeta) \otimes R(z_i, \zeta) + R(z_i, \zeta) \otimes d(z_i, \zeta)]\beta_i \\ &\quad + [R(z_i, \zeta) \otimes R(z_i, \zeta)](\beta_i \otimes \beta_i) + \sigma_i^2 vec(V_\varepsilon(z_i, \psi)). \end{aligned} \tag{G.1}$$

Then, by stacking (G.1) with $E[\mathcal{Y}_i | z_i, \beta_i] = d(z_i, \theta) + R(z_i, \theta)\beta_i$, we get from (4):

$$E(\mathcal{Y}_i^* | z_i, \gamma_i) = d^*(z_i, \theta) + R^*(z_i, \theta)\gamma_i, \tag{G.2}$$

where $\mathcal{Y}_i^* = (\mathcal{Y}_i', (\mathcal{Y}_i \otimes \mathcal{Y}_i)')'$, the vector of random coefficients is $\gamma_i = (\beta_i', (\beta_i \otimes \beta_i)', \sigma_i^2)'$, the augmented parameter vector is $\theta = (\zeta', \psi')'$, and:

$$\begin{aligned} d^*(z_i, \theta) &= \begin{pmatrix} d(z_i, \zeta) \\ d(z_i, \zeta) \otimes d(z_i, \zeta) \end{pmatrix}, \\ R^*(z_i, \theta) &= \begin{pmatrix} R(z_i, \zeta) & 0 & 0 \\ d(z_i, \zeta) \otimes R(z_i, \zeta) + R(z_i, \zeta) \otimes d(z_i, \zeta) & R(z_i, \zeta) \otimes R(z_i, \zeta) & vec(V_\varepsilon(z_i, \psi)) \end{pmatrix}. \end{aligned}$$

The conditional moment restriction in (G.2) is of the type studied in Chamberlain (1992), so that we can use his approach to obtain the orthogonality restrictions to estimate θ and $\phi^* = E[\gamma_i]$. Again, in our FA setting, we can normalize ϕ^*, so that those parameters do not appear in the orthogonality restrictions.

I Proof of Proposition 3

Here, we derive the asymptotic distributions of the unconstrained and constrained FA-GMM estimators as well as the ones of the trinity of FA-GMM test statistics under local alternative hypotheses. We also check the conditions for establishing the Gaussian experiment for testing sphericity in the FA model. Below, probability order $o_p(1)$ and distributional convergence are under the sequence of local alternative hypotheses.

(i) Let us start with the unconstrained FA-GMM estimator and the W statistic. Under the local alternative hypothesis $H_{1,loc}$, we have:

$$\begin{pmatrix} \sqrt{n}(\hat{\mu} - \mu_0) \\ \sqrt{n}L_H'(\hat{\vartheta} - \vartheta_n) \end{pmatrix} = -\tilde{\Sigma}_0 \tilde{J}_0' V_g^{-1} \sqrt{n} \hat{g}_n(\theta_n) + o_p(1), \tag{G.3}$$

similarly as in Equation (D.4), and $\sqrt{n}H'(\hat{\vartheta} - \vartheta_n) = o_p(1)$, and

$$\begin{aligned} \sqrt{n}a(\hat{\theta}) &= \sqrt{n}a(\theta_n) + \frac{\partial a(\theta_n)}{\partial \theta'} \sqrt{n}(\hat{\theta} - \theta_n) + o_p(1) = \delta + \frac{\partial a(\theta_0)}{\partial \theta'} \sqrt{n}(\hat{\theta} - \theta_n) + o_p(1) \\ &= \delta - \tilde{A}' \tilde{\Sigma}_0 \tilde{J}_0' V_g^{-1} \sqrt{n} \hat{g}_n(\theta_n) + o_p(1), \end{aligned} \tag{G.4}$$

from a Taylor expansion and Equation (G.3).

(ii) Let us now consider the constrained FA-GMM estimator and the Lagrange multiplier vector. Under $H_{1,loc}$, Equation (D.9) is modified into $\delta + \frac{\partial a(\theta_0)}{\partial \vartheta'} \sqrt{n}(\hat{\vartheta}^c - \vartheta_n) = o_p(1)$, and Equation (D.10) becomes:

$$\begin{bmatrix} \tilde{J}_0' V_g^{-1} \tilde{J}_0 & \tilde{A} \\ \tilde{A}' & 0_{(T-1)\times(T-1)} \end{bmatrix} \begin{bmatrix} \begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \mu_0) \\ \sqrt{n}L_H'(\hat{\vartheta}^c - \vartheta_n) \end{pmatrix} \\ \sqrt{n}\hat{\lambda} \end{bmatrix} = - \begin{bmatrix} \tilde{J}_0' V_g^{-1} \sqrt{n} \hat{g}_n(\theta_n) \\ \delta \end{bmatrix} + o_p(1). \tag{G.5}$$

By the inversion of the block matrix in the LHS of (G.5), together with Equation (D.8), we get the asymptotic expansions of $\sqrt{n}(\hat{\mu}^c - \mu_0)$, $\sqrt{n}(\hat{\vartheta}^c - \vartheta_n)$ and $\sqrt{n}\hat{\lambda}$ under $H_{1,loc}$:

$$\begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \mu_0) \\ \sqrt{n}L_H'(\hat{\vartheta}^c - \vartheta_n) \end{pmatrix} = -\tilde{\Sigma}_0\tilde{A}(\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\delta - (I-P)'\tilde{\Sigma}_0\tilde{J}_0'V_g^{-1}\sqrt{n}\hat{g}_n(\theta_n) + o_p(1), \quad \text{(G.6)}$$

$$\sqrt{n}H'(\hat{\vartheta}^c - \vartheta_n) = o_p(1),$$

$$\sqrt{n}\hat{\lambda} = (\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\delta - (\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\tilde{A}'\tilde{\Sigma}_0\tilde{J}_0'V_g^{-1}\sqrt{n}\hat{g}_n(\theta_n) + o_p(1). \quad \text{(G.7)}$$

(iii) From Equations (G.4) and (G.7), we get $\sqrt{n}\hat{\lambda} = (\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\sqrt{n}a(\hat{\theta}) + o_p(1)$, i.e., the vectors that build the quadratic forms defining the W and LM statistics are equal up to $o_p(1)$ terms. It then follows that $\xi_n^{LM} = \xi_n^W + o_p(1)$.

(iv) We use $\sqrt{n}\hat{g}_n(\theta_n) \Rightarrow \mathcal{N}(0, V_g)$. From Equation (G.4), we get $\sqrt{n}a(\hat{\theta}) \Rightarrow \mathcal{N}(\delta, \Omega_W)$, with $\Omega_W = \tilde{A}'\tilde{\Sigma}_0\tilde{A}$. Thus, $\xi_n^W \Rightarrow \chi^2(T-1, \delta'\Omega_W^{-1}\delta)$. The same asymptotic distribution applies for the LM statistic ξ_n^{LM} because of the asymptotic equivalence. From Subsection D.1, we have $\tilde{A}'\tilde{\Sigma}_0 A = L_{1_T}'\Sigma_{V_\varepsilon}L_{1_T}$ and $\delta'\Omega_W^{-1}\delta = diag(V_\varepsilon)'L_{1_T}(L_{1_T}'\Sigma_{V_\varepsilon}L_{1_T})^{-1}L_{1_T}'diag(V_\varepsilon)$.

(v) Let us finally consider the LR statistic. By a second-order Taylor expansion around $\hat{\theta}$, we have $Q_n(\hat{\theta}^c) = Q_n(\hat{\theta}) + \frac{1}{2}(\hat{\theta}^c - \hat{\theta})'\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}^c - \hat{\theta})$, where $\bar{\theta}$ is between $\hat{\theta}^c$ and $\hat{\theta}$ componentwise, and the first-order term vanishes because of the FOC of the unconstrained FA-GMM estimator. The second-order partial derivatives matrix of the GMM criterion is given by $\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'} = 2\left[\frac{\partial^2\hat{g}_n(\theta)'}{\partial\theta\partial\theta_j}\hat{V}_g^{-1}\hat{g}_n(\theta)\right]_{j=1,\ldots,(k+2)T} + 2\frac{\partial\hat{g}_n(\theta)'}{\partial\theta}\hat{V}_g^{-1}\frac{\partial\hat{g}_n(\theta)}{\partial\theta'}$. When evaluated at the consistent estimator $\bar{\theta} = \theta_n + o_p(1)$, the first term on the RHS vanishes asymptotically. Then, $\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'} = 2J_0'V_g^{-1}J_0 + o_p(1)$. We get $\xi_n^{LR} = [\sqrt{n}(\hat{\theta}^c - \hat{\theta})]'(J_0'V_g^{-1}J_0)[\sqrt{n}(\hat{\theta}^c - \hat{\theta})] + o_p(1)$. By using $\sqrt{n}(\hat{\theta}^c - \hat{\theta}) = \begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \hat{\mu}) \\ L_H\sqrt{n}L_H'(\hat{\vartheta}^c - \hat{\vartheta}) \end{pmatrix} + o_p(1)$, we get: $\xi_n^{LR} = \begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \hat{\mu}) \\ \sqrt{n}L_H'(\hat{\vartheta}^c - \hat{\vartheta}) \end{pmatrix}'\tilde{\Sigma}_0^{-1}\begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \hat{\mu}) \\ \sqrt{n}L_H'(\hat{\vartheta}^c - \hat{\vartheta}) \end{pmatrix} + o_p(1)$. By taking the difference between (G.3) and (G.6), we get: $\zeta_n^{LR} := \tilde{\Sigma}_0^{-1/2}\begin{pmatrix} \sqrt{n}(\hat{\mu}^c - \hat{\mu}) \\ \sqrt{n}L_H'(\hat{\vartheta}^c - \hat{\vartheta}) \end{pmatrix} = -\tilde{\Sigma}_0^{1/2}\tilde{A}(\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\delta + \tilde{\Sigma}_0^{-1/2}P'\tilde{\Sigma}_0\tilde{J}_0'V_g^{-1}\sqrt{n}\hat{g}_n(\theta_n) + o_p(1) = -\tilde{\Sigma}_0^{1/2}\tilde{A}(\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1}\left[\delta - \tilde{A}'\tilde{\Sigma}_0\tilde{J}_0'V_g^{-1}\sqrt{n}\hat{g}_n(\theta_n)\right] + o_p(1) = -R\zeta_n^W + o_p(1)$, where

$\zeta_n^W := (\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1/2}\sqrt{n}a(\hat{\theta})$ and $R := \tilde{\Sigma}_0^{1/2}\tilde{A}(\tilde{A}'\tilde{\Sigma}_0\tilde{A})^{-1/2}$ with $R'R = I_{T-1}$. Then, it follows that $\xi_n^{LR} = (\zeta_n^{LR})'\zeta_n^{LR} + o_p(1) = (\zeta_n^W)'\zeta_n^W + o_p(1) = \xi_n^W + o_p(1)$. The asymptotic equivalence and step (iv) imply $\xi_n^{LR} \Rightarrow \chi^2(T-1, \delta'\Omega_W^{-1}\delta)$.

(vi) Let us check the conditions for establishing the validity of the Gaussian experiment. The q.m.d. condition in Assumption 3 holds with $f_i^q(x_i) = -\frac{1}{2\bar{\sigma}^2}[\nabla \log \varphi_i\left(\bar{\sigma}^{-1}(x_i - \mathfrak{m}_i(\theta_0))\right)]'\Delta_\varepsilon$ $(\bar{\sigma}^{-1}(x_i - \mathfrak{m}_i(\theta_0)))$ by Assumptions A.8(a)-(b). We use $y_i = \mu_0 + F_0\beta_i + V_{\varepsilon,n}^{1/2}(\sum_{j\in B_m} s_{j,i}w_j)$ to show that the uniform bounds in Assumption 4 follow from Assumptions A.1 and A.2. Further, Lemma 7 shows the CLT in Assumption 5. Besides, the FA-GMM test statistics are asymptotically quadratic forms of $\frac{1}{\sqrt{n}}\sum_i g(y_i, \tilde{\theta}_0)$, so that their subsequence weak limits are independent of $Z_n^\perp$. Then, Assumption 6 holds. Proposition 3 follows. Q.E.D.