SFB 823

Group invariance, likelihood ratio tests, and the incidental parameter problem in a high-dimensional linear model

Marc Hallin, Marcelo J. Moreira, Alexei Onatski

Discussion Paper

Nr. 1/2013

SFB 823

Group Invariance, Likelihood Ratio Tests, and the Incidental Parameter Problem in a High-Dimensional Linear Model

Marc Hallin[a], Marcelo J. Moreira[b], and Alexei Onatski[c]
[a]*ECARES, Université libre de Bruxelles, Belgium and ORFE, Princeton University, USA*
[b]*Fundação Getulio Vargas, Rio de Janeiro, Brazil*
[c]*University of Cambridge, UK*

Abstract

This paper considers a linear panel data model with reduced rank regressors and interactive fixed effects. The leading example is a factor model where some of the factors are observed, some others not. Invariance considerations yield a maximal invariant statistic whose density does not depend on incidental parameters. It is natural to consider a likelihood ratio test based on the maximal invariant statistic. Its density can be found by using as a prior the unique invariant distribution for the incidental parameters. That invariant distribution is least favorable and leads to minimax optimality properties. Combining the invariant distribution with a prior for the remaining parameters gives a class of admissible tests. A particular choice of distribution yields the *spiked* covariance model of Johnstone (2001). Numerical simulations suggest that the maximal invariant likelihood ratio test outperforms the standard likelihood ratio test. Tests which are not invariant to data transformations (i) are uniquely represented as randomized tests of the maximal invariant statistic and (ii) do not solve the incidental parameter problem.

Keywords: Panel data models, factor models, incidental parameters, invariance, integrated likelihood, minimax, likelihood ratio test.

JEL Classification Numbers: C12, C44.

1. Introduction

Inference in large statistical/econometric models has recently generated a most active area of research. A major feature of such models is the so-called *incidental parameter problem*—the description of which can be traced back to the seminal paper by Neyman and Scott (1948). A major consequence of that problem is that such classical and commonly used procedures as maximum likelihood estimation or likelihood ratio testing, when the number of nuisance parameters is large, fail to provide adequate inference methods. This is the case in a number of econometric models, such as panel, weak instrumental variable, and factor models, to mention only a few. In dynamic linear panel data models, Arellano and Bond (1991) introduce GMM estimators, Hahn and Kuersteiner (2002) propose a bias-corrected OLS estimator, Lancaster (2002) uses noninformative priors, Chamberlain and Moreira (2009) and Bai (2012) advocate for a correlated-random effects estimator, and Moreira (2009) relies on invariance arguments to propose a maximum likelihood estimation method. Bekker (1994) shows that, in many weak instrumental variable models, the limited information maximum likelihood estimator (LIML) is consistent and asymptotically normal under i.i.d. errors, while Chioda and Jansson (2009) derive attainable efficiency bounds, and Hausman *et al.* (2012) propose a jacknife estimator to accommodate for heteroskedastic errors. In factor models, Chamberlain and Rothschild (1983) extend the approximate model of Ross (1976), Bai and Ng (2002) propose new methods to determine the number of factors in large models, and Bai (2009) obtains the asymptotic distribution of an interactive-effects estimator. See Lancaster (2000) and Arellano (2003) for a general discussion of the incidental parameter problem in econometric models.

This paper applies classical decision-theoretical principles to a simple linear panel data model. The model is invariant to the action of groups of transformations of the sample space; such groups, acting on the sample space, induce groups of transformations of the parameter space. This symmetry yields an invariant measure for the incidental parameters. When the group is *compact*, this measure can be transformed into an invariant probability measure. This methodology yields minimax optimality results where the invariant distribution is *least favorable*. Integrating out the likelihood with respect to the invariant measure gives the distribution of the maximal invariant. When the group is not compact, the integrated likelihood approach yields a likelihood ratio statistic for the maximal invariant. For the groups considered here, we can still obtain a minimax result by approximating the invariant measure by sequences of probability measures. Most of our theoretical results are only expository. Ferguson (1967) describes the decision theory framework, Eaton (1989) discusses the role of invariance in statistical models, and Lehmann and Romano (2005) outline the main concepts in hypothesis testing. Our contribution is to point out that maximal invariant likelihood ratio tests have optimality properties, whereas the standard likelihood ratio test does not.

Our results are illustrated by four examples. The leading example is a factor model with i.i.d. errors and reduced-rank regressors as in Moon and Weidner (2012); those regressors can be interpreted as observed factors as in Fama and French (1993). After transformations in the parameter and sample spaces, we obtain structural parameters (γ, σ) and incidental parameters (β, λ, ω). We are interested in γ and σ, which have fixed dimension irrespective of the sample size. The dimensions of β and λ increase with the cross-sectional dimension, while the dimension of ω grows with the time series length. We show that the model is invariant under a group which induces, on the parameter space, a group that acts *transitively* on the incidental parameters. The maximal invariant does not depend on the incidental parameters at all and the group characterizes an invariant (probability) measure. This distribution is least favorable, and yields a minimax optimality resolving the incidental parameter problem. Using the invariant distribution as a prior leads to an integrated likelihood method. The integrated likelihood yields the distribution for the maximal invariant, so it coincides with the marginal maximal invariant likelihood approach.

A problem of particular interest that can be efficiently handled by this approach is that of testing for the presence of latent factors in the data. Such a test may be useful, for example, in auction studies where one suspects that heterogeneous bidders' valuations are affected by common factors that are known to the bidders but unobserved by the econometrician (Athey and Haile (2007, Section 6)). Recent signal processing literature (Nadakuditi and Edelman (2008) and Nadakuditi and Silverstein (2010)) points out the existence of a fundamental limit on the power of factor detection procedures based on standard likelihood ratio tests: such procedures have trivial power in the detection of factors when the signal-to-noise ratio lies below a "fundamental impossibility threshold." Onatski *et al.* (2011) show that, in contrast with this, the likelihood ratio test based on a maximal invariant statistic has non-trivial asymptotic power, even below the "impossibility threshold." In Section 4, we show that the power of such a test remains non-trivial in finite samples. Moreover, we find that this power is very close to the finite-sample power envelope.

The *spiked* covariance model of Johnstone (2001) arises from a distributional assumption for the factors in Example 1. This distribution can be decomposed into the invariant probability measure for ω and a distribution on γ (which has fixed dimension). Hence the *spiked* covariance model automatically protects against the incidental parameter problem for large time series, with no apparent robustness issues. It is then natural to impose the invariant probability measure for λ as well, which again yields the minimax optimality and the maximal invariant likelihood test. In Example 2, we consider the special case of testing the null hypothesis that the errors' covariance matrix is identity against the *spiked* covariance alternative. We show that tests which are not invariant to orthogonal transformations of cross-sectional observations are uniquely represented as randomized tests based on the maximal invariant statistic.

Our first two examples are relatively simple, but the principle of maximal invariant likelihood ratio tests is applicable to more general models. For example, Chamberlain and Rothschild (1983) introduce approximate factor models with correlated errors. When the covariance matrix of the vectorized error matrix has a Kronecker product structure, the same group of transformations as in Examples 1 and 2 applies. The distribution of the maximal invariant statistic now depends on the eigenvalues of the error covariance matrix. However, the maximal invariant likelihood ratio test avoids maximization over a large number of eigenvectors of the error covariance matrix. This may bring considerable power gains with respect to the standard likelihood ratio test.

The last application, given in Example 4, involves the null hypothesis under which the covariance matrices from two different populations are equal. This setup can be particularly useful for testing peer effects, such as exam performance in different classrooms or schools; see Graham (2008). Using invariance arguments yields considerable dimension reduction. Under the null hypothesis that the covariance matrices are equal, the action group acts transitively. Hence likelihood ratio tests are pivotal (distribution-free under the null), and controlling size is trivial. Under the alternative, we are able to reduce the parameter dimensionality by a half. This reduction can be important in applications with large clusters.

The paper is organized as follows. Section 2 presents the panel data model and derives a canonical representation for it. Section 3.1 describes invariance groups acting on the sample space, and the induced groups acting on the parameter spaces, respectively. Section 3.2 discusses invariance properties of the likelihood ratio tests and introduces likelihood ratio tests based on maximal invariants. Section 3.3 studies optimal invariant tests, and points out their close connection to the likelihood ratio tests based on maximal invariant. Section 3.4 shows that a Gaussian assumption on the factors yields the *spiked* covariance model of Johnstone (2001). Section 3.5 briefly discusses a few other examples and their relation with the factor model. Section 4 compares (still in the factor model of Example 1) the finite-sample performances of the maximal invariant likelihood ratio test with those of the standard likelihod ratio test. Section 5 concludes, and the Appendix provides the proofs of theoretical results.

2. The Model

Consider the linear factor model under which the observation consists of a matrix $Y \in \mathbb{R}^{p\times n}$ satisfying a model of the form

$$Y = \widetilde{\beta} X + \widetilde{\lambda} \widetilde{f}' + \Sigma^{1/2}(\sigma) W \tag{1}$$

where $X \in \mathbb{R}^{j\times n}$ is a full row rank matrix of observed factors or regressors, and $\widetilde{f} \in \mathbb{R}^{n\times k}$ a matrix of unobserved factors. The unobserved $l \times n$ disturbance matrix $W \sim N(0, I_n \otimes I_l)$ affects the observed cross-sectional values of Y through $\Sigma^{1/2}(\sigma) W$, where $\Sigma^{1/2}(\sigma)$ is the symmetric square root of a symmetric positive definite $n \times n$ matrix $\Sigma(\sigma)$; the parameters $\sigma \in \mathbb{R}^r$ and $\widetilde{\beta} \in \mathbb{R}^{p\times j}$, and the factor loadings $\widetilde{\lambda} \in \mathbb{R}^{p\times k}$ are unspecified. Throughout, we assume $j, p \leq n$.

In all examples below, we treat $\widetilde{\beta}$ as a nuisance parameter. It is convenient then to apply a one-to-one transformation of the model that simplifies the derivation of our results. Consider therefore the polar decomposition

$$X = \begin{pmatrix} \rho & 0 \end{pmatrix} q'$$

of X, where $\rho = (XX')^{1/2}$ is the unique symmetric square root matrix of XX' and $q \in \mathcal{O}_n$, the group of $n \times n$ orthogonal matrices. Partitioning q into $(q_1\ q_2)$ with $q_1 \in \mathbb{R}^{n\times j}$ and $q_2 \in \mathbb{R}^{n\times(n-j)}$, let

$$X = \begin{pmatrix} \rho & 0 \end{pmatrix} \begin{pmatrix} q_1' \\ q_2' \end{pmatrix} = \rho q_1';$$

we can rotate Y and work with $Z = Yq$. Because W has a spherical distribution,

$$Z = Yq \overset{d}{=} \begin{pmatrix} \widetilde{\beta}\rho & 0 \end{pmatrix} + \widetilde{\lambda}\widetilde{f}'q + \Sigma^{1/2}(\sigma) W.$$

Partitioning $(f_1', f_2') = f' = \widetilde{f}'q = (\widetilde{f}'q_1, \widetilde{f}'q_2)$ conformably with q ($f_1 \in \mathbb{R}^{j\times k}$, $f_2 \in \mathbb{R}^{(n-j)\times k}$) yields

$$Z \overset{d}{=} \begin{pmatrix} \widetilde{\beta}\rho + \widetilde{\lambda} f_1' & \widetilde{\lambda} f_2' \end{pmatrix} + \Sigma^{1/2}(\sigma) W. \tag{2}$$

Write $\beta = \widetilde{\beta}\rho + \widetilde{\lambda} f_1'$ and let $\lambda\gamma\omega'$ be a singular value decomposition of $\widetilde{\lambda} f_2'$, where λ belongs to the set $\mathcal{F}_{k,p}$ of $p \times k$ matrices with mutually orthogonal unit-length columns, ω similarly belongs to $\mathcal{F}_{k,n-j}$, and γ is a $k \times k$ diagonal matrix with non-increasing elements along the diagonal; this defines a natural embedding of γ into $\mathbb{R}^k$. The model for $Z = (Z_1, Z_2)$ (with $Z_1 \in \mathbb{R}^{p\times j}$ and $Z_2 \in \mathbb{R}^{p\times(n-j)}$) is

$$\begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \overset{d}{=} \begin{pmatrix} \beta & \lambda\gamma\omega' \end{pmatrix} + \Sigma^{1/2}(\sigma) \begin{pmatrix} W_1 & W_2 \end{pmatrix} \tag{3}$$

and will be considered the *canonical form* of (1).

Example 1. For our purposes, the leading example is a simple linear factor model with some observed factors X, and possibly some unobserved ones $\widetilde{f}$:

$$Y|X \overset{d}{=} \widetilde{\beta}X + \widetilde{\lambda}\widetilde{f}' + \sigma W. \tag{4}$$

Here, W is $p \times n$, and the unobserved errors have variance σ^2. We are interested in testing for the presence of the unobserved factor, that is,

$$H_0 : \|\widetilde{f}\| = 0 \text{ against } H_1 : \|\widetilde{f}\| > 0, \tag{5}$$

where $\|\widetilde{f}\| = \left(\mathrm{tr}(\widetilde{f}'\widetilde{f})\right)^{1/2}$ is the Frobenius norm of $\widetilde{f}$. The canonical form of (4) is

$$\begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \overset{d}{=} \begin{pmatrix} \beta & \lambda\gamma\omega' \end{pmatrix} + \sigma W.$$

The factor component f_1 in $\beta = \widetilde{\beta}\rho + \widetilde{\lambda}f_1'$ is not identified, and so the problem becomes that of testing

$$H_0 : \|\gamma\| = 0 \text{ against } H_1 : \|\gamma\| > 0. \tag{6}$$

Example 2. In the model with observed factors X and unrestricted $p \times p$ covariance matrix Σ

$$Y = \widetilde{\beta}X + \Sigma^{1/2}W, \tag{7}$$

consider the problem of testing

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_0 : \Sigma \neq \Sigma_0. \tag{8}$$

Without loss of generality, we can assume $\Sigma_0 = I_p$, otherwise we can work with the transformed variable $\Sigma_0^{-1/2}Y$.

Example 3. Assuming that, in Example 1, $\tilde{f} \sim N\left(0, \sigma^2 I_k \otimes I_n\right)$, model (4) can be written as

$$Y \mid X \overset{d}{=} \tilde{\beta}X + \Sigma^{1/2}W,$$

where $\Sigma = \sigma^2(I_p + \tilde{\lambda}\tilde{\lambda}')$ and $W \sim N(0, I_n \otimes I_p)$. A modification of this model, allowing for cross-sectional and temporal covariances in the disturbance matrix, is (for simplicity, we restrict to the case $X = 0$ of unobserved factors)

$$Y = \Sigma^{1/2}W\Omega^{1/2}. \tag{9}$$

The covariance matrix for Y is $\Phi = \Omega \otimes \Sigma$. The eigenvalues τ_i, $i = 1, \ldots, np$ of Φ are the np products of an eigenvalue of Ω with an eigenvalue of Σ. We could test whether $r_{(1)}$, the largest eigenvalue of Φ, is relatively important to the other ones. For example,

$$H_0 : r_{(1)} \leq a r_{(2)} \text{ against } H_1 : r_{(1)} > a r_{(2)}$$

for some given constant $a \geq 1$. This testing problem can be seen as generalizing (from an i.i.d. to a correlated noise context) the problem of detecting the presence of factors.

Example 4. Consider two independent random samples with m-dimensional normal observations $Y_{1,i} \overset{iid}{\sim} N_m(\widetilde{\beta}_1, \Sigma_1)$ and $Y_{2,i} \overset{iid}{\sim} N_m(\widetilde{\beta}_2, \Sigma_2)$, $i = 1, \ldots, n$, respectively. We want to test whether the covariance matrices in those two samples are the same, that is,

$$H_0 : \Sigma_1 = \Sigma_2 \text{ against } H_1 : \Sigma_1 \neq \Sigma_2. \tag{10}$$

We can stack the observations into a $p \times n$ matrix Y whose i-th column is $\left(Y'_{1,i}, Y'_{2,i}\right)'$, with dimension $p = 2m$. The model is

$$Y = \begin{bmatrix} \widetilde{\beta}_1 \\ \widetilde{\beta}_2 \end{bmatrix} 1'_n + \begin{bmatrix} \Sigma_1^{1/2} & 0 \\ 0 & \Sigma_2^{1/2} \end{bmatrix} W. \tag{11}$$

The matrix X here is just the vector $1'_n$, but we could also easily accommodate explanatory variables, as well as a larger number $K > 2$ of random samples. This model is a special case of (1) without unobserved factors nor covariates ($k = 0$, $j = 1$), where $\widetilde{\beta}$ is the $2m \times 1$ vector stacking $\widetilde{\beta}_1$ on top of $\widetilde{\beta}_2$, $\sigma = (\Sigma_1, \Sigma_2)$, and

$$\Sigma(\sigma) = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}.$$

The $n \times n$ orthogonal matrix q here has first column $q_1 = n^{-1/2} 1_n$; the other columns are arbitrarily defined (under the constraint that q is orthogonal). The first column Z_1 of $Z = Yq$ is then the empirical mean of Y's rows, multiplied by a factor $\sqrt{n}$; Z_2 is $2p \times (n-1)$, and consists of the remaining $(n-1)$ columns. Clearly, Z_2 coincides with the last $(n-1)$ columns of $\Sigma^{1/2}(\sigma)Wq$, which in turn equals $\Sigma^{1/2}(\sigma)W$ in distribution. To summarize, the distribution of Z_2 is the same as that of the $2m \times (n-1)$ matrix $(Y_2 - \widetilde{\beta}, \ldots, Y_n - \widetilde{\beta})$.

3. Invariance and likelihood ratio tests

3.1. Group invariance and linear factor model

Denote by $\mathrm{P}_{\beta,\gamma,\sigma,\lambda,\omega}$ or P_ϑ the distribution of Z in (3) under parameter value $\vartheta = (\beta, \gamma, \sigma, \lambda, \omega)$, where $\vartheta \in \Theta = \mathbb{R}^{p \times j} \times \mathbb{R}^k \times \mathbb{R}^r \times \mathcal{F}_{k,p} \times \mathcal{F}_{k,n-j}$, and let $\mathcal{P} = \{\mathrm{P}_\vartheta | \, \vartheta \in \Theta\}$. Throughout, β will be a nuisance parameter, the parameter of interest being $\theta = (\gamma, \sigma, \lambda, \omega)$: all testing problems here are thus of the form

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \Theta_1. \tag{12}$$

The linear model (1) and the testing problem (12) enjoy natural invariance properties, which are better described in reference to the canonical form (3). We use the notation G for a general group of transformations acting on the observation or sample space, with elements g and the binary operation $\circ$. If the group G is such that the distribution $\mathrm{P}_\vartheta^{gZ}$ of gZ under ϑ still belongs to $\mathcal{P}$, there exists a unique parameter value—$\bar{g}\vartheta$, say, such that $\mathrm{P}_\vartheta^{gZ} = \mathrm{P}_{\bar{g}\vartheta}$: $\mathcal{P}$ is said to be *invariant* under the action of G. The mappings $\vartheta \mapsto \bar{g}\vartheta$ yield another group of transformations $\bar{G}$, the *induced group*, acting on the parameter space Θ. We refer to Eaton (1989) or Chapter 6 of Lehmann and Romano (2005) for details.

Consider, for example, the group $G_1 = \{g_b | \, b \in \mathbb{R}^{p \times j}\}$, where

$$g_b Z = g_b (Z_1, Z_2) = (Z_1 + b, Z_2),$$

hence $g_{b_2} \circ g_{b_1} = g_{b_1 + b_2}$. That group is a group of translations acting on the first j columns of the observation space. The action of g_b on Z yields a shifted distribution for Z_1, with mean $\beta + b$, while the distribution of Z_2 remains the same: it follows that $\mathcal{P}$ is invariant under G_1, with $\bar{g}_b(\beta, \gamma, \sigma, \lambda, \omega) = (\beta + b, \gamma, \sigma, \lambda, \omega)$. Clearly, $\bar{G}_1$ acts transitively on Θ. *Maximal invariants* for G_1 and $\bar{G}_1$ are Z_2 and $\theta = (\gamma, \sigma, \lambda, \omega)$, respectively. The distribution of an invariant statistic only depends on the maximal invariant of the induced group (see Lehmann and Romano (2005, Theorem 6.3.2)); the distribution of Z_2 thus only depends on θ: denote it as P_θ, and write p_θ for the corresponding probability density.

A testing problem of the form (12) is invariant under G_1, in the sense that $\bar{g}_b \Theta_0 = \Theta_0$ and $\bar{g}_b \Theta_1 = \Theta_1$ for all b. When a testing problem is invariant under a group, it seems natural to restrict attention to tests that are themselves invariant under that group. Such tests are measurable functions of the maximal invariant. When making inference on the parameter of interest $\theta = (\gamma, \sigma, \lambda, \omega)$, invariance arguments thus suggest to look only at Z_2-measurable procedures.

Depending on the invariance features of Θ_0 and Θ_1, the testing problems considered here involve various groups G, all admitting G_1 as a subgroup. Theorem 6.2.2 of Lehmann and Romano (2005, p. 217) tells us that, in such cases, Z_2 rather than Z can be considered as the observation; the corresponding family of distributions is $\mathcal{P}_2 = \{\mathrm{P}_\theta | \, \theta \in \mathbb{R}^k \times \mathbb{R}^r \times \mathcal{F}_{k,p} \times \mathcal{F}_{k,n-j}\}$. Invariant (with respect to G, hence also with respect to G_1) inference on θ in the presence of observed factors is thus the same as invariant inference on θ without observed factors, albeit with a smaller sample size $n - j$ in the canonical model (3).

Now, assume that all the elements g of G are differentiable mappings, and denote by χ_g their Jacobian determinant. It follows from traditional results on continuous functions of random variables that, under parameter value θ, the density of gZ_2 at gz is $\mathrm{p}_\theta(z)\chi_g(z)$, $z \in \mathbb{R}^{p\times(n-j)}$. On the other hand, invariance of $\mathcal{P}_2$ under G implies $\mathrm{P}_\theta^{gZ_2} = \mathrm{P}_{\bar{g}\theta}$, where $\mathrm{P}_\theta^{gZ_2}$ denotes the distribution of gZ_2 under parameter value θ, so that the density under θ of gZ_2, evaluated at gz, is of the form $\mathrm{p}_{\bar{g}\theta}(gz)$. Therefore,

$$\mathrm{p}_{\bar{g}\theta}(gz) = \mathrm{p}_\theta(z)\chi_g(z) \quad z - \text{a.e.}, \tag{13}$$

hence, replacing θ with $\bar{g}^{-1}\theta$, we obtain the identity

$$\mathrm{p}_\theta(gz) = \mathrm{p}_{\bar{g}^{-1}\theta}(z)\chi_g(z) \quad z - \text{a.e.} \tag{14}$$

Identity (14) is quite general and actually holds for any parametric family $\mathcal{P} = \{\mathrm{P}_\theta\}$ of absolutely continuous distributions with densities p_θ invariant under a group G with differentiable elements $g{:}z \mapsto gz$. Let us illustrate this identity in the context of Example 1 (Examples 2-4 are treated in Section 3.5) and the group G with elements $g_{s,c,d}$, $(s, c, d) \in \mathbb{R}^+ \times \mathcal{O}_p \times \mathcal{O}_{n-j}$, where

$$g_{s,c,d}Z_2 = s\, cZ_2 d', \quad \text{hence} \quad g_{s_1,c_1,d_1} \circ g_{s_2,c_2,d_2} = g_{s_1 s_2, c_1 c_2, d_1 d_2}. \tag{15}$$

Under the canonical model (3),

$$g_{s,c,d}Z_2 \overset{d}{=} c\lambda\ (s\gamma)\ \omega' d' + (s\sigma)\ W_2.$$

The distribution $\mathrm{P}_\theta^{g_{s,c,d}Z_2}$ of $g_{s,c,d}Z_2$ under parameter value θ is thus $\mathrm{P}_{s\gamma, s\sigma, c\lambda, d\omega}$, so that

$$\bar{g}_{s,c,d}(\gamma, \sigma, \lambda, \omega) = (s\gamma, s\sigma, c\lambda, d\omega),$$

which defines the action of the induced group $\bar{G}$. This transformation leaves the ratio γ/σ unchanged, and thus preserves the testing problem (6).

The probability density of Z_2 at $z \in \mathbb{R}^{p\times(n-j)}$ when $\theta = (\gamma, \sigma, \lambda, \omega)$ is

$$\mathrm{p}_\theta\,(z) = \left(2\pi\sigma^2\right)^{-p(n-j)/2} \exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}(z - \lambda\gamma\omega')'(z - \lambda\gamma\omega')\right).$$

For $g = g_{s,c,d}$, the left-hand side of (13) yields $\mathrm{p}_{\bar{g}\theta}\,(gz) = \mathrm{p}_{s\gamma,s\sigma,c\lambda,d\omega}\,(sczd')$. Hence,

$$\begin{aligned}
\mathrm{p}_{\bar{g}\theta}\,(gz) &= \left(2\pi s^2\sigma^2\right)^{-p(n-j)/2} \exp\left(-\frac{1}{2s^2\sigma^2}\mathrm{tr}(s\,czd' - s\,c\lambda\gamma\omega' d')'(s\,czd' - s\,c\lambda\gamma\omega' d')\right) \\
&= \left(2\pi s^2\sigma^2\right)^{-p(n-j)/2} \exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}(z - \lambda\gamma\omega')'(z - \lambda\gamma\omega')\right) \\
&= \mathrm{p}_{\bar{g}\theta}\,(gz)\,\chi_g(z)
\end{aligned}$$

with $\chi_g(z) = s^{-p(n-j)}$. This Jacobian $\chi_g(z)$ depends on $g = g_{s,c,d}$ only through the multiplicative constant s.

3.2. Group invariance and likelihood ratio tests

Irrespective of any invariance considerations, a natural idea, in a Gaussian testing problem of the form (12), consists in performing a *likelihood ratio test* (LRT). That LRT either could be based on the full observation $Z = (Z_1, Z_2)$, or on Z_2 only. The following proposition establishes the remarkable fact that, for any problem of the form (12), (14) implies the invariance under G of the Z_2-based LRT statistic; the validity of this proposition actually extends, well beyond Example 1, to any problem for which (14) holds.

Proposition 3.1 *Suppose that the group G preserves the null and alternative hypotheses in (12), and that (14) holds for all $g \in G$. Then, the Z_2-based likelihood ratio test statistic*

$$LR(Z_2) = 2\left[\max_{\theta\in\Theta_1} \ln \mathrm{p}_\theta\left(Z_2\right) - \max_{\theta\in\Theta_0} \ln \mathrm{p}_\theta\left(Z_2\right)\right] \tag{16}$$

is invariant under G, that is, $LR(gZ_2) = LR(Z_2)$ for any $g \in G$.

In Example 1, the same invariance property also holds for the Z-based LRT statistic

$$LR(Z_1, Z_2) = 2\left[\max_{\vartheta\in\mathbb{R}^{p\times j}\times\Theta_1} \ln \mathrm{p}_\vartheta\left(Z_1, Z_2\right) - \max_{\vartheta\in\mathbb{R}^{p\times j}\times\Theta_0} \ln \mathrm{p}_\vartheta\left(Z_1, Z_2\right)\right],$$

based on the entire sample $Z = (Z_1\, Z_2)$. Indeed, (i) Z_1 and Z_2 are independent and (ii) the maximum likelihood estimator of β is Z_1 itself, under $\theta \in \Theta_0$ as well as under $\theta \in \Theta_1$, so that $LR(Z_1, Z_2) = LR(Z_2)$.

Proposition 3.1 has crucial implications. The distribution of any invariant statistic only depends on the maximal invariant of $\bar{G}$ in the parameter space. Since the group $\bar{G}$ here preserves the ratio γ/σ and acts *transitively* on $(\sigma, \lambda, \omega)$, γ/σ is maximal invariant for $\bar{G}$. The distribution of $LR(Z_2)$ thus depends on θ only through γ/σ, which can mislead us to think that the incidental parameter problem is solved. This is an illusion, though, as the maximum likelihood estimator for $(\sigma, \lambda, \omega)$, hence the Z_2-based LRT, has poor properties when p and n are large; see Onatski (2012). Instead of the Z_2-based LRT, we therefore suggest to use the LRT based on a maximal invariant (for G) statistic.

In order to determine such maximal invariants for the group G defined in (15), it is convenient to notice that G is the group generated by the union of two subgroups, G_2 and G_0, where G_2, with elements $g_{c,d}$, is the (compact) group of orthogonal transformations acting on Z_2, namely,

$$g_{c,d} Z_2 = c Z_2 d', \quad (c, d) \in \mathcal{O}_p \times \mathcal{O}_{n-j}, \tag{17}$$

while G_0, with elements g_s, $s \in \mathbb{R}_0^+$, is a group of scale transformations: $g_s Z_2 = s Z_2$. The following result then is straightforward.

Proposition 3.2 *(i) A maximal invariant for G_2 is the collection $T\left(Z_2\right) = \left(l_{(1)}, l_{(2)}, \ldots, l_{(p)}\right)$ of ordered eigenvalues of $Z_2 Z_2'$.*

(ii) A maximal invariant for G is the collection $M\left(Z_2\right) = \left(l_{(2)}/l_{(1)}, \ldots, l_{(p)}/l_{(1)}\right)$ of ratios of ordered eigenvalues of $Z_2 Z_2'$.

Since $T(Z_2)$ is maximal invariant for G_2, its distribution $\mathrm{P}^T_{\gamma,\sigma}$ only depends on (γ, σ), which is maximal invariant for $\bar{G}_2$. Inference based on $T\left(Z_2\right)$ thus completely eliminates the impact of λ and ω. This is quite desirable when λ and ω (as well as β) are nuisance parameters, as in the testing problem (6). Consider the $T(Z_2)$-based likelihood ratio test statistic for (6)

$$LR(T(Z_2)) = 2\left[\max_{(\gamma,\sigma)\in\mathbb{R}^k\times\mathbb{R}_0^+} \ln \mathrm{p}^T_{\gamma,\sigma}\left(T\left(Z_2\right)\right) - \max_{\sigma\in\mathbb{R}_0^+} \ln \mathrm{p}^T_{0,\sigma}\left(T\left(Z_2\right)\right)\right], \tag{18}$$

where $\mathrm{p}^T_{\gamma,\sigma}\left(t\right)$ denotes the density of $T(Z_2)$ under parameter value $\theta = (\gamma, \sigma, \lambda, \omega)$ evaluated at $t \in \mathbb{R}^p$. The same argument as in Proposition 3.1 shows that the distribution of $LR(T(Z_2))$ does not depend on σ, leading to an exact test for (6).

Alternatively, for the same testing problem (6), we could consider the $M(Z_2)$-based likelihood ratio test statistic

$$LR(M(Z_2)) = 2\left[\max_{\gamma/\sigma\in\mathbb{R}^k} \ln \mathrm{p}^M_{\gamma/\sigma}(M(Z_2)) - \ln \mathrm{p}^M_0(M(Z_2))\right], \tag{19}$$

where $\mathrm{p}^M_{\gamma/\sigma}(m)$ is the density, under parameter value $\theta = (\gamma, \sigma, \lambda, \omega)$ of the maximal invariant (for G) $M(Z_2)$ evaluated at $m \in \mathbb{R}^{p-1}$. Section 3.3 provides two optimality results which support the use of both likelihood ratio tests.

3.3. Finite-sample optimality issues

The group of transformations $\bar{G}$ acts *transitively* on $(\sigma, \lambda, \omega) \in \mathbb{R}^+_0 \times \mathcal{F}_{k,p} \times \mathcal{F}_{k,n-j}$. Hence, the power function of an invariant test $\phi(M(Z_2))$ depends only on γ/σ. This suggests that a $(M(Z_2))$-based LRT might be minimax (with respect to the nuisance parameters $(\sigma, \lambda, \omega)$). The group G is not compact; however, the subgroup G_2 is, and the induced group $\bar{G}_2$ acts transitively on $(\lambda, \omega) \in \mathcal{F}_{k,p} \times \mathcal{F}_{k,n-j}$. Because G_2 is compact, Theorem 2.2 of Eaton (1989, p. 25) implies that there exists a unique G_2-invariant probability measure on $\mathcal{F}_{k,p} \times \mathcal{F}_{k,n-j}$, which is a product measure of the form $\mathrm{H} = \mathrm{H}_p \times \mathrm{H}_{n-j}$, the product of the invariant probability measures H_p on $\mathcal{F}_{k,p}$ and H_{n-j} on $\mathcal{F}_{k,n-j}$, respectively.

Consider the integrated (with respect to H) likelihood

$$\mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(Z_2) = \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{p}_{\gamma,\sigma,\lambda,\omega}(Z_2)\, \mathrm{H}(\mathrm{d}\lambda \times \mathrm{d}\omega). \tag{20}$$

The following proposition establishes the close relation between $\mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(Z_2)$ and the density of the maximal invariant $T(Z_2) = \left(l_{(1)}, l_{(2)}, \ldots, l_{(p)}\right)$ (recall that $\min(n, p) = p$). Denote by $\eta(T(Z_2))$ the $p \times (n-j)$ matrix with diagonal elements $l^{1/2}_{(1)}, l^{1/2}_{(2)}, \ldots, l^{1/2}_{(\min(p,n-j))}, 0, \ldots, 0$ and off-diagonal elements zero.

Proposition 3.3 *(i) The integrated likelihood (20) is such that* $\mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(Z_2) = \mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(\eta(T(Z_2)))$.

(ii) The density, evaluated at $t \in \mathbb{R}^p$, of $\mathrm{P}^T_{\gamma,\sigma}$ with respect to the measure μT^{-1}, where μ is the Lebesgue measure on $\mathbb{R}^{p\times(n-j)}$, is $\mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(\eta(t))$.

The invariant distribution H also has an appealing decision-theoretic interpretation. Suppose we are interested in making inference on (γ, σ). The possibly high-dimensional parameters λ and ω are only nuisance parameters. Let $\mathcal{A}$ be the decision space, and denote by $L(\gamma, \sigma; a)$ the loss incurred when decision $a \in \mathcal{A}$ is taken while the parameter value is γ, σ. We observe Z_2 and make a decision $\delta(Z_2)$. For simplicity, we only consider nonrandomized decision rules $\delta : Z_2 \to \mathcal{A}$ and convex loss functions; we also omit measurability conditions. The risk function associated with a decision rule δ is

$$R(\gamma, \sigma, \lambda, \omega; \delta) = \int_{\mathbb{R}^{p\times(n-j)}} L(\gamma, \sigma; \delta(Z_2))\, \mathrm{p}_{\gamma,\sigma,\lambda,\omega}(z)\ \mu(\mathrm{d}z).$$

Because G_2 acts transitively on (λ, ω) but does not affect (γ, σ), we use the invariant probability measure for (λ, ω) and an arbitrary weight function W for (γ, σ) in the evaluation of a decision rule δ: a decision rule δ^* is optimal within a class $\mathcal{C}$ if $\delta^* \in \mathcal{C}$ and

$$\delta^* = \arg\min_{\delta\in\mathcal{C}} \int_{\mathbb{R}^k\times\mathbb{R}^r\times\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} R(\gamma, \sigma, \lambda, \omega; \delta)\ \mathrm{H}(\mathrm{d}\lambda \times \mathrm{d}\omega)\, W(\mathrm{d}\gamma \times \mathrm{d}\sigma)\ . \tag{21}$$

The following proposition shows that the optimal decision rule δ^* which solves (21) is minimax.

Proposition 3.4 *The optimal solution δ^* in (21) is minimax; more precisely,*

$$\delta^* = \arg\min_{\delta\in\mathcal{C}} \sup_{\lambda\in\mathcal{F}_{k,p},\, \omega\in\mathcal{F}_{k,n-j}} \int_{\mathbb{R}^k\times\mathbb{R}^r} R(\gamma, \sigma, \lambda, \omega; \delta)\ W(\mathrm{d}\gamma \times \mathrm{d}\sigma),$$

and depends on Z_2 only through $T(Z_2)$. Furthermore, H *is the* least favorable distribution *for (λ, ω).*

Proposition 3.4 implies that restricting to G_2-invariant (equivalently, T-measurable) decision rules does not affect the minimax solution of the decision problem. The proof is standard and uses the fact that G_2 is compact (implying a finite Haar measure)[1].

The $T(Z_2)$-based *LRT* (18) replaces the arbitrary choice of a weight function W by a maximization over γ and σ. This procedure may not be admissible, but it does avoid the incidental parameter problem. Analogously, the $M(Z_2)$-based *LRT* test (19) replaces the arbitrary W by the invariant measure on σ and a maximization over γ/σ. The resulting test may not be admissible for general one-sided alternatives involving γ/σ.

3.4. The unconditional factor model

Instead of the conditional (on the factors f) approach that has been adopted so far, consider the unconditional approach under which

$$\widetilde{f}|X \overset{d}{=} N(X'\widetilde{\pi}, \Sigma \otimes I_n), \tag{22}$$

where $\widetilde{\pi} \in \mathbb{R}^{j\times k}$ and the positive definite matrix Σ are unspecified. In the rotated model (2), where $f = q'\widetilde{f}$, we obtain

$$f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \overset{d}{=} N\left(\begin{pmatrix} \pi \\ 0 \end{pmatrix}, \Sigma \otimes I_n\right),$$

where $\pi = \rho\widetilde{\pi}$. Elementary algebraic manipulations yield

$$\begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \overset{d}{=} N\left(\begin{pmatrix} \overline{\beta} & 0 \end{pmatrix}, \sigma^2 I_n \otimes (I_p + VhV')\right),$$

where $\overline{\beta} = \widetilde{\beta}\rho + \widetilde{\lambda}\pi'$, $V \in \mathcal{F}_{k,p}$, and h is the $k\times k$ diagonal matrix of ordered eigenvalues $h_j \geq 0$ of the standardized covariance matrix $\sigma^{-2}\Sigma$.

As in the conditional model, we first apply the translation group G_1 to eliminate Z_1. This yields Z_2 as a maximal invariant, with

$$Z_2 \overset{d}{=} N\left(0, \sigma^2 I_{n-j} \otimes (I_p + VhV')\right), \tag{23}$$

that is, a *spiked* covariance model in the sense of Johnstone (2001). The same invariance properties as in (17) hold for (23): the maximal invariant for G_2 remains $T(Z_2)$, the distribution of which depends only on σ and h; see Onatski *et al.* (2012).

The distributional assumption (22) on the factors $\widetilde{f}$ at first sight may look too restrictive. The following proposition shows that this assumption actually is related to the invariant distribution for an incidental parameter. In that sense, the unconditional model (23) for Z_2 is close to the model described by the integrated likelihood (20) considered in the conditional model context.

To formulate this, let us introduce some notation. For any full-rank matrix $x \in \mathbb{R}^{(n-j)\times k}$ with $k < n-j$, let $\varphi\gamma\omega'$ be the unique singular value decomposition of x' such that the first row of $\varphi \in \mathcal{O}_k$ has positive elements, γ is the $k\times k$ diagonal matrix of ordered singular values of x, and $\omega \in \mathcal{F}_{k,n-j}$. The matrices φ and γ are uniquely determined by $x'x$, so we write $\varphi = \varphi(x'x)$ and $\gamma = \gamma(x'x)$. In contrast, the matrix ω depends on the full matrix x, and so we write $\omega = \omega(x)$. Further, let uhu' be a spectral decomposition of Σ/σ^2, and denote by $\mathrm{P}^{f_2'f_2}_{u,h,\sigma}$ the probability measure on $\mathbb{R}^{k\times k}$ of the random matrix $f_2'f_2$.

Proposition 3.5 *Assume that $k < n-j$. In the unconditional model (19),*

(i) the density function of Z_2 at $z \in \mathbb{R}^{p\times(n-j)}$ is

$$\mathrm{q}_{\sigma,V,h}(z) = \int_{\mathcal{F}_{k,n-j}} \int_{\mathbb{R}^{k\times k}} \mathrm{p}_{\gamma(y),\sigma,Vu'\varphi(y),\omega}(z)\, \mathrm{H}_{n-j}(\mathrm{d}\omega)\, \mathrm{P}^{f_2'f_2}_{u,h,\sigma}(\mathrm{d}y); \tag{24}$$

[1]It is possible to extend this proposition to accommodate also the non-compact group G_0 by applying the Hunt-Stein theorem; see Lehmann and Romano (2005, p. 333). However, for the purpose of obtaining tests with good power for large p and n, invariance with respect to G_0 is unnecessary as long as the dimension r of σ^2 (in most cases, $r = 1$) remains small irrespective of the sample size n.

(ii) defining $\mathrm{q}^{\mathrm{H}}_{h,\sigma}(z) = \int_{\mathcal{F}_{k,p}} \mathrm{q}_{\sigma,V,h}(z)\mathrm{H}_p(\mathrm{d}V)$, *we have*

$$\mathrm{q}^{\mathrm{H}}_{h,\sigma}(z) = \int_{\mathbb{R}^{k\times k}} \mathrm{p}^{\mathrm{H}}_{\gamma(y),\sigma}(z)\, \mathrm{P}^{f_2' f_2}_{u,h,\sigma}(\mathrm{d}y)\,;$$

therefore $\mathrm{q}^{\mathrm{H}}_{h,\sigma}(\eta(t))$ *is the density of* $T = T(Z_2)$ *at* $T = t$.

The density $\mathrm{q}^{\mathrm{H}}_{h,\sigma}(\eta(t))$ is related to (20), which gives the likelihood for $T(Z_2)$ in the conditional model. The unconditional model density $\mathrm{q}_{\sigma,V,h}$ implicitly integrates the parameter ω out of the conditional model $\mathrm{p}_{\gamma,\sigma,\lambda,\omega}$ with respect to the uniform invariant distribution on $\mathcal{F}_{k,n-j}$. Further, after a reparameterization $\tilde{\lambda} \mapsto Vu'$ and $\lambda \mapsto Vu'\varphi$, it integrates out $u'\varphi\gamma$ with respect to a Wishart distribution on $\mathbb{R}^{k\times k}$. So, the starting point of our approach differs from the one adopted by Johnstone (2001). The conditional model treats the factor loadings as "fixed" effect parameters. The minimax optimality yields the least favorable distribution. The unique invariant distribution on $\omega \in \mathcal{F}_{k,n-j}$ can be combined with a family of distributions for $u'\varphi\gamma \in \mathbb{R}^{k\times k}$ to obtain the unconditional, *spiked* covariance model.

Proposition 3.5 has some implications for testing the presence of factors for large n and p. First, the correlated factor assumption imposes a distributional assumption on $\varphi\gamma^2\varphi' = f_2' f_2$ for invariant tests. The parameters φ and γ have fixed dimension, irrespective of the sample size, so that inference on the unconditional model as in Onatski *et al.* (2011) is similar to the conditional model. Second, a standard likelihood ratio test based on the unconditional model avoids the incidental parameter problem for large n by imposing a minimal distributional assumption. Third, the unconditional model still entails maximization over $\lambda \in \mathcal{F}_{k,p}$, so it does not fully solve the high-dimensionality problem for large p. Therefore, we suggest to integrate out the incidental parameter to obtain a minimax result.

3.5. Further Examples

The finite-sample theory of Sections 3.1-3.4 can be extended to other models when both p and n are large. The main ingredients in our analysis (maximal invariants in the sample and parameter spaces, Haar measure, and minimax tests) remain valid for appropriate groups, which are specific to each particular problem. We illustrate our approach in the context of Examples 2-4; in all of them, β is a nuisance parameter, and we therefore consider the canonical model (3).

Example 2 (cont.): Example 2 could be extended to the problem of testing *sphericity*, that is, $H_0 : \Sigma = \sigma^2 I$ against $H_1 : \Sigma \neq \sigma^2 I$ with unspecified $\sigma > 0$—the problem considered in Ledoit and Wolf (2002), Srivastava (2005), Schott (2006), Bai *et al.* (2009), Chen *et al.* (2010), Cai and Ma (2012)—by including the additional group G_0 of scale transformation and $M(Z_2)$ instead of $T(Z_2)$ as maximal invariant. For simplicity, however, we only consider the specified-σ case, and the group G_2, acting on Z_2, described in (17). Under model (9),

$$g_{c,d}Z_2 = cZ_2 d' = c\Sigma^{1/2}W_2 d', \quad (c,d) \in \mathcal{O}_p \times \mathcal{O}_{n-j},$$

the induced group $\bar{G}_2$, now acting on Σ (rather than on $(\gamma,\sigma,\lambda,\omega)$, as in Section 3.1), is characterized by

$$\bar{g}_{c,d}\Sigma = c\Sigma d', \quad (c,d) \in \mathcal{O}_p \times \mathcal{O}_{n-j},$$

and the testing problem (8) (with $\Sigma_0 = I_p$) is clearly invariant under G_2. It follows from Proposition 3.2(ii) that a maximal invariant for G_2 is the ordered collection $T(Z_2) = (l_{(1)},\ldots,l_{(p)})$ of eigenvalues of $Z_2 Z_2'$. A maximal invariant for $\bar{G}_2$ is the ordered collection $v_{(1)},\ldots,v_{(p)}$ of eigenvalues of the covariance matrix Σ. The group G_2 moreover yields unique (up to a positive multiplicative constant) *left- and right-invariant measures*. A minimax result can be obtained along the same lines as in Proposition 3.4.

Invariance under left-hand orthogonal transformations of the form $g_{c,I_{n-j}}$, $c \in \mathcal{O}_p$ reduces the $p(p+1)/2$-dimensional parameter space of Σ to the p-dimensional space of the maximal invariant $v_{(1)},\ldots,v_{(p)}$. Right-hand orthogonal transformations of the form $g_{I_p,d}$, $d \in \mathcal{O}_{n-j}$, do not yield any further reduction in the parameter space. So, we may not need to require tests to be invariant to right-hand orthogonal transformations. However, as shown by the following lemma, tests which are not invariant under right-hand orthogonal transformations are, in fact, randomized $T(Z_2)$-measurable tests which are invariant under both left- and right-hand orthogonal transformations.

Lemma 3.1 *For each test ϕ invariant to left-hand orthogonal transformations, there exists (up to zero-measure sets) a unique randomized test ϕ^* measurable with respect to the maximal invariant $T(Z_2)$, hence invariant under both left- and right-hand orthogonal transformations, which has the same power function as ϕ, that is, such that $\mathrm{E}_\Sigma \phi^* = \mathrm{E}_\Sigma \phi$ for all Σ.*

Most tests in the literature are based on the eigenvalues $T(Z_2)$ of the sample covariance matrix, but not all of them. An example of a test which is invariant under left-hand but not under right-hand orthogonal transformations is the test proposed by Cai and Ma (2012), based on a statistic of the form $\psi\left(Z_2' Z_2\right)$. The following proposition provides a representation result for such tests.

Proposition 3.6 *Consider a test ϕ rejecting the null hypothesis $H_0 : \Sigma = I_p$ whenever the statistic $\psi\left(Z_2' Z_2\right)$ exceeds the critical value c_α, where c_α is such that $\mathrm{E}_{\Sigma=I_p}\phi = \alpha$. Then, the test $\phi^*\left(T\right) = 1 - F_T\left(c_\alpha\right)$, where F_t is the distribution function of ψ conditional on $T(Z_2) = t$, has the same power function as ϕ.*

Example 3 (cont.): This example involves testing whether one eigenvalue or several of them lie above the bulk of the remaining ones, such as $H_0 : r_{(1)} \leq a r_{(2)}$ against $H_1 : r_{(1)} > a r_{(2)}$ for some given constant $a \geq 1$. This testing problem can be linked to the methods of estimation of the number of factors based on the evidence of a separation of some of the largest eigenvalues from the bulk (Hallin and Liska (2007), Onatski (2010), Alessi *et al.* (2010), and Ahn and Horenstein (2012)).

The model (9) does not contain additional regressors. The relevant group again is G_2, acting on $Y = Z_2$, with maximal invariant $T(Z_2)$. The maximal invariant in the parameter space is the (ordered) collections of the p eigenvalues of Σ and n eigenvalues of Ω. The $T\left(Z_2\right)$-based LRTs are minimax for hypotheses involving the eigenvectors of Σ and the eigenvectors of Ω. The $T\left(Z_2\right)$-based LRTs use information not only from a few of the largest sample eigenvalues, but from all of them. We leave it to future research to compare the $T\left(Z_2\right)$-based LRTs to the standard LRT and modifications thereof (e.g., Moon and Weidner (2012)) when both p and n tend to infinity.

Example 4 (cont.): In the testing problem (10), the relevant group is $G = \{g_{c,d} | \, c \in \mathcal{G}_m^+, \, d \in \mathcal{O}_{n-1}\}$, where $\mathcal{G}_m^+$ is the set of $m \times m$ lower triangular matrices with positive diagonal elements, and

$$g_{c,d} Z_2 = \left(I_2 \otimes c\right) Z_2 d',$$

so that

$$g_{c,d} Z_2 \overset{d}{=} \begin{bmatrix} c\Sigma_1^{1/2} & 0 \\ 0 & c\Sigma_2^{1/2} \end{bmatrix} W_2 d'.$$

The action on (Σ_1, Σ_2) of the induced group $\bar{G} = \{\bar{g}_{c,d} | \, c \in \mathcal{G}_m^+, \, d \in \mathcal{O}_{n-1}\}$ is thus

$$\bar{g}_{c,d}\left(\Sigma_1, \Sigma_2\right) = \left(c\Sigma_1 c', c\Sigma_2 c'\right). \tag{25}$$

The probability density of Z_2 under $\theta = (\Sigma_1, \Sigma_2)$, evaluated at z, is

$$\mathrm{p}_\theta\left(z\right) = (2\pi)^{-p(n-1)/2} \left|\Sigma_1\right|^{-(n-1)/2} \left|\Sigma_2\right|^{-(n-1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(\sigma) Z_2 Z_2'\right]\right\}.$$

For $g = g_{c,d}$, the density under $\bar{g}\theta$ of gZ_2 evaluated at gz is

$$\begin{aligned} \mathrm{p}_{\bar{g}_{c,d}}\left(g_{c,d} z\right) &= \mathrm{p}_{c\Sigma_1 c', c\Sigma_2 c'}\left((I_2 \otimes c) z d'\right) \\ &= (2\pi)^{p(n-1)/2} \left|c\Sigma_1 c'\right|^{-(n-1)/2} \left|c\Sigma_2 c'\right|^{-(n-1)/2} \\ &\quad \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\left(I_2 \otimes c'^{-1}\right)\Sigma^{-1}(\sigma)\left(I_2 \otimes c^{-1}\right)\left(I_2 \otimes c\right) Z_2 d' d Z_2' (I_2 \otimes c')\right]\right\}. \end{aligned} \tag{26}$$

We can partition each $c \in \mathcal{G}_m^+$ as

$$c = \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix},$$

where $c_{11} \in \mathcal{G}_{m-1}^{+}$, c_{21} is a $1 \times (m-1)$ vector, and c_{22} a scalar. Hence, c^{-1} in (26) is

$$c^{-1} = \begin{bmatrix} c_{11}^{-1} & 0 \\ -c_{22}^{-1} c_{21} c_{11}^{-1} & c_{22}^{-1} \end{bmatrix}.$$

From (13),

$$\mathrm{p}_{\bar{g}\theta}(gz) = \mathrm{p}_{\theta}(z)\,\chi_g(z) \quad \text{with} \quad \chi_g(z) = |c|^{-2(n-1)} = \left(\prod_{i=1}^{m} c_{ii}\right)^{-2(n-1)}, \tag{27}$$

where c_{ii} is the (i,i)-entry of $c \in \mathcal{G}_m^{+}$.

We now obtain a maximal invariant statistic for G (assume $m \le 2(n-1)$).

Proposition 3.7 *Define the $m \times 2(n-1)$ matrix $Z_2^{+} = [Z_{21}, Z_{22}]$ where Z_{21} and Z_{22} are the first and last m rows of Z_2, respectively*[2]*. Write Z_2^{+} as $Z_2^{+} = c_2 w_2'$, where $c_2 \in \mathcal{G}_m^{+}$ and $w_2 \in \mathcal{F}_{m,2(n-1)}$, and partition w_2' into two $m \times (n-1)$ submatrices: $w_2' = [w_{21}', w_{22}']$. Maximal invariants in the sample and parameter spaces are*

$$\begin{bmatrix} w_{21}' w_{21} & w_{21}' w_{22} \\ w_{21}' w_{22} & w_{22}' w_{22} \end{bmatrix} \qquad \text{and} \qquad \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2},$$

respectively.

Invariance under G has several important implications for testing procedures. First, the invariance argument reduces the composite null hypothesis in the original problem to a simple one: controlling size is thus straightforward if we restrict to invariant tests (actually, it is important to be invariant under the subgroup $G = \{g_{c,I_{n-1}} | \, c \in \mathcal{G}_m^{+}\}$ only; invariance to right-hand multiplication by orthogonal matrices simply entails data reduction). Second, the invariance argument reduces the dimension of the parameter space by one half. This reduction may yield considerable power gains for the LRT based on the maximal invariant. For example, consider testing

$$H_0 : \Sigma_2 = \Sigma_1 \text{ against } H_1 : \Sigma_2 = \xi\Sigma_1, \tag{28}$$

where ξ is some fixed number and Σ_1 is unspecified. The group of transformations $\bar{G}$ acts transitively on both the null hypothesis and the alternative. The optimal invariant test then is given by the following proposition.

Proposition 3.8 *An α-level uniformly most powerful invariant test for (28) is the test rejecting H_0 whenever*

$$\frac{\displaystyle\int \exp\left\{-\frac{1}{2}\mathrm{tr}\left(c\left[Z_{21}Z_{21}' + \xi^{-1}Z_{22}Z_{22}'\right]c'\right)\right\}\left(\prod_{i=1}^{m} c_{ii}\right)^{-2(n-1)} \mathrm{d}c}{\displaystyle\int \exp\left\{-\frac{1}{2}\mathrm{tr}\left(c\left[Z_{21}Z_{21}' + Z_{22}Z_{22}'\right]c'\right)\right\}\left(\prod_{i=1}^{m} c_{ii}\right)^{-2(n-1)} \mathrm{d}c} > c_{\alpha},$$

where c_{α} is chosen so that the size under H_0 is α.

This test is the LRT based on the maximal invariant for G, and has constant risk function. We can then obtain a minimax result as an application of the Hunt-Stein theorem, even though G is not a compact group.

Finally, recall that Proposition 3.1 shows that the standard LRT procedure for (28) is invariant. Hence, the maximal invariant-based LRT dominates the standard LRT for the alternative $H_1 : \Sigma_2 = \xi\Sigma_1$. This statement can easily be generalized to alternatives of the form $H_1 : \Sigma_2 = \Sigma_1^{1/2}\Omega\Sigma_1^{1/2}$, where Ω is an arbitrary fixed positive definite $m \times m$ matrix.

[2] Z_2^{+} "stacks horizontally" the $m \times (n-1)$ upper and lower halves of Z_2; while Z_2 is $2m \times (n-1)$, Z_2^{+} is $m \times 2(n-1)$.

4. Monte Carlo Simulations

This section uses Monte Carlo simulations to compare the finite sample power functions of the standard LRT on one hand, and of the LRT based on $T(Z_2)$ and on $M(Z_2)$ on the other hand. We focus on testing

$$H_0 : h = 0 \text{ against } H_1 : h > 0,$$

when the data $Z_2 \sim N\left(0, \sigma^2 I_n \otimes (I_p + VhV')\right)$ satisfy the *spiked* covariance model with scalar h. As explained above, Z_2 can be interpreted as a part of the canonically transformed data generated from an unconditional factor model.

We consider the cases of specified and unspecified σ^2. When σ^2 is specified, we set it to $\sigma^2 = 1$ without loss of generality. In such a case, we compare the standard LRT with the LRT based on the maximal invariant statistic $T(Z_2)$. The corresponding test statistic is defined as

$$LR_T = 2\left[\max_{h \in \mathbb{R}_0^+} \ln q_h^T(T(Z_2)) - \ln q_0^T(T(Z_2))\right],$$

where $q_h^T(t)$ denotes the density of $T(Z_2)$ under parameter value h evaluated at $t \in \mathbb{R}^p$.

In the case where σ^2 is left unspecified, we compare the standard LRT with the LRT based on the maximal invariant statistic $M(Z_2)$. The corresponding test statistic is defined as

$$LR_M = 2\left[\max_{h \in \mathbb{R}_0^+} \ln q_h^M(M(Z_2)) - \ln q_0^M(M(Z_2))\right],$$

where $q_h^M(m)$ denotes the density of $M(Z_2)$ under parameter value h evaluated at $m \in \mathbb{R}^{p-1}$.

The computation of LR_T requires a numerical evaluation of the density $q_h^T(T(Z_2))$. As Proposition 3.5 shows, this density can, in principle, be obtained by numerical integration

$$q_h^T(T(Z_2)) = \int_{\mathbb{R}^{k\times k}} \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} p_{\gamma(y),\sigma,Vu'\varphi(y),\omega}(Z_2)\, \mathrm{H}(\mathrm{d}V \times \mathrm{d}\omega)\, p_{u,h,\sigma}^{f_2' f_2}(\mathrm{d}y).$$

However, this approach is impractical in situations where n and/or p are large. Therefore, we use a computationally simpler formula (Onatski *et al.* (2011, formula 2.9)) which expresses the ratio $q_h^T(T(Z_2))/q_0^T(T(Z_2))$ as the contour integral

$$(hn/2)^{-(p-2)/2}(1+h)^{(p-n-2)/2}\,\Gamma(p/2)\,\frac{1}{2\pi i}\oint_{\mathcal{K}} \exp\left(\frac{n}{2}\frac{h}{1+h}z\right)\prod_{j=1}^{p}\left(z - l_{(j)}/n\right)^{-\frac{1}{2}}\mathrm{d}z,$$

where $\mathcal{K}$ is an arbitrary contour in the complex plane that starts at $-\infty$, encircles counter-clockwise the eigenvalues $l_{(1)}/n, ..., l_{(p)}/n$ of $Z_2 Z_2'/n$, and goes back to $-\infty$. For the computation of LR_M, we use a similar contour integral formula (Onatski *et al.* (2011, formula 2.10)).

It is straightforward to show that the standard LRT statistic in the case of specified σ^2 equals

$$LR^s = n\left(l_{(1)}/n - 1 - \log(l_{(1)}/n)\right)$$

if $l_{(1)}/n > 1$, and $LR^s = 0$ otherwise. Here the superscript "s" is used as a reminder that σ^2 is *s*pecified. In the case of unspecified σ^2, we use notation LR^u. It can be shown that

$$LR^u = n(p-1)\log(p-1) - n\log\frac{p}{\sum_{j=1}^{p} l_{(j)}/l_{(1)}} - n(p-1)\log\left(p - \frac{p}{\sum_{j=1}^{p} l_{(j)}/l_{(1)}}\right)$$

if $p/\sum_{j=1}^{p} l_{(j)}/l_{(1)} > 1$, and $LR^u = 0$ otherwise. We omit the derivation of the above two formulae to save space.

h	$n=p=40$			$n=p=200$			$n=p=1000$			$n=p=\infty$		
	LR^s	LR$_T$	N-P	LR^s	LR$_T$	N-P	LR^s	LR$_T$	N-P	LR^s	LR$_T$	N-P
0.00	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.10	0.11	0.11
0.21	0.10	0.11	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.10	0.12	0.13
0.32	0.11	0.11	0.11	0.10	0.10	0.12	0.11	0.12	0.12	0.10	0.13	0.15
0.42	0.11	0.12	0.12	0.10	0.11	0.13	0.12	0.13	0.13	0.10	0.14	0.17
0.53	0.12	0.13	0.14	0.11	0.13	0.14	0.12	0.14	0.15	0.10	0.16	0.19
0.63	0.14	0.15	0.16	0.12	0.14	0.16	0.12	0.16	0.17	0.10	0.18	0.22
0.74	0.16	0.18	0.19	0.14	0.18	0.19	0.13	0.19	0.20	0.10	0.22	0.26
0.84	0.21	0.23	0.23	0.18	0.23	0.24	0.16	0.24	0.25	0.10	0.27	0.31
0.95	0.28	0.31	0.30	0.26	0.32	0.34	0.25	0.35	0.36	0.10	0.39	0.42
1.05	0.37	0.40	0.40	0.42	0.48	0.48	0.52	0.60	0.61	1.00	1.00	1.00
1.16	0.49	0.52	0.51	0.65	0.68	0.68	0.90	0.91	0.92	1.00	1.00	1.00
1.26	0.62	0.64	0.63	0.86	0.87	0.87	1.00	1.00	1.00	1.00	1.00	1.00
1.37	0.73	0.75	0.75	0.97	0.96	0.97	1.00	1.00	1.00	1.00	1.00	1.00
1.47	0.83	0.84	0.84	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
1.58	0.90	0.90	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.68	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.79	0.97	0.97	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.90	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1: Finite sample power functions for the standard LRT (columns LR^s) and LRT based on $T(Z_2)$ (columns LR_T). The finite sample and asymptotic power envelopes are given in columns marked "N-P".

We implement the Monte Carlo exercises for $p = n = 40,\ 200,$ and $1,000$. For each of these sample sizes, we first make 10,000 draws of $Z_2 \sim N(0, I_n \otimes I_p)$, each time computing LR_T, LR^s, LR_M, and LR^u statistics. We take the 9,000-th order statistics of the resulting values of LR_T, LR^s, LR_M, and LR^u as the critical values of the corresponding tests with finite sample size 10%. We also compute the critical values for the tests with finite sample size 5%.

To compute the finite sample powers of the tests, we make 10,000 draws of $Z_2 \sim N(0, I_n \otimes (I_p + VhV'))$ with $V' = (1, 0, ..., 0)$ for each of the 20 equally spaced points on the grid $h = 0 : (2/19) : 2$. For each draw, we compute the values of LR_T, LR^s, LR_M, and LR^u statistics. The finite sample powers of the corresponding tests equal the proportions of the values, out of 10,000, that fall above the corresponding critical values.

Table 1 reports the finite sample power functions of the LRT based on $T(Z_2)$ (columns LR$_T$) and of the standard LRT (columns LR^s) of finite sample size 10% for $p = n = 40, 200,$ and $1,000$. It also reports the values of the finite sample power envelopes for invariant tests (columns N-P, for Neyman-Pearson). By the Neyman-Pearson lemma, the value of such an envelope at h can be obtained by computing the finite sample power of the test that rejects the null hypothesis whenever $q_h^T(T(Z_2))/q_0^T(T(Z_2)) > C$, where C is such that the probability of type-I error equals 10%. We compute this power at the grid points for h using the above contour integral representation of $q_h^T(T(Z_2))/q_0^T(T(Z_2))$.

The asymptotic power envelope as $n = p$ go to infinity as well as the asymptotic power of the LRT and LRT based on $T(Z_2)$ (Onatski *et al.* (2011)) are reported in the last three columns of the table. The results for the 5% size tests are similar, and, to save space, we report them only in the supplementary material file.

We see that, as the sample size increases, both the LRT based on $T(Z_2)$ and the standard LRT have power functions approaching one for $h > 1$. In contrast, when $h < 1$, the power of the standard LRT decreases with the sample size, whereas the power of the LRT based on $T(Z_2)$ remains close to the power envelope. This is consistent with Onatski et al (2012) (see the last three columns of the table), who show that the asymptotic power of the LRT based on $T(Z_2)$, as n, p go to infinity so that $p/n \to c \in (0, \infty)$, is

h	$n=p=40$			$n=p=200$			$n=p=1000$			$n=p=\infty$		
	LR^u	LR$_M$	N-P	LR^u	LR$_M$	N-P	LR^u	LR$_M$	N-P	LR^u	LR$_M$	N-P
0.00	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.11	0.10	0.10	0.09	0.09	0.10	0.10	0.11	0.11	0.10	0.10	0.10	0.10
0.21	0.10	0.10	0.10	0.09	0.10	0.10	0.11	0.11	0.11	0.10	0.10	0.10
0.32	0.10	0.10	0.10	0.09	0.10	0.10	0.11	0.11	0.11	0.10	0.11	0.11
0.42	0.10	0.10	0.10	0.09	0.10	0.10	0.11	0.11	0.11	0.10	0.11	0.12
0.53	0.10	0.11	0.11	0.10	0.10	0.11	0.11	0.12	0.11	0.10	0.12	0.13
0.63	0.11	0.12	0.12	0.10	0.11	0.11	0.11	0.12	0.12	0.10	0.13	0.15
0.74	0.13	0.13	0.13	0.11	0.13	0.13	0.12	0.14	0.14	0.10	0.15	0.17
0.84	0.16	0.16	0.16	0.14	0.16	0.16	0.14	0.17	0.17	0.10	0.19	0.22
0.95	0.21	0.21	0.21	0.22	0.24	0.24	0.22	0.27	0.27	0.10	0.30	0.33
1.05	0.28	0.29	0.29	0.36	0.39	0.39	0.48	0.52	0.53	1.00	1.00	1.00
1.16	0.39	0.39	0.39	0.59	0.61	0.61	0.88	0.88	0.89	1.00	1.00	1.00
1.26	0.51	0.51	0.51	0.82	0.83	0.83	1.00	1.00	1.00	1.00	1.00	1.00
1.37	0.64	0.64	0.64	0.95	0.95	0.96	1.00	1.00	1.00	1.00	1.00	1.00
1.47	0.76	0.76	0.76	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
1.58	0.85	0.85	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.68	0.91	0.91	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.79	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.90	0.97	0.97	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: Finite sample power functions for the standard LRT (columns LR^u) and LRT based on $M(Z_2)$ (columns LR_M). The finite sample and asymptotic power envelopes are given in columns marked "N-P".

close to the asymptotic power envelope, whereas the asymptotic power of any test based on $l_{(1)}$ only, such as the standard LRT, equals the asymptotic size in the region $h \in (0, \sqrt{c})$.

In fact, the finite sample power of the LRT based on $T(Z_2)$ is so close to the corresponding power envelope that the numerical accuracy of the Monte Carlo based on 10,000 replications is sometimes not sufficient to guarantee that the estimates reported in columns LR$_T$ are strictly smaller than the corresponding estimates in columns N-P. For example, for $h = 0.95$ and $n = p = 40$, the estimated power of the LRT based on $T(Z_2)$ is 0.31, whereas the estimated value of the power envelope is 0.30. The standard deviations of the estimates in Table 1 must be no smaller than the standard deviation of the average of 10,000 i.i.d. Bernoulli random variables with 0.5 probability of success, which equals $\sqrt{1/40,000} = 0.005$. The estimates in Table 1 are rounded to the closest number with two digits after the decimal point. Hence, even when the power of the LRT based on $T(Z_2)$ is smaller than the corresponding power envelope by 0.01, we must occasionally see the reported estimates in columns LR$_T$ larger than the corresponding estimates in columns N-P.

Table 2 is an equivalent of Table 1 for the LRT based on $M(Z_2)$ and the standard LRT when σ^2 is left unspecified. It is qualitatively similar to Table 1 although the asymptotic and the finite sample power envelopes are substantially lower for $h < 1$.

5. Conclusion

Previous authors have noted that the LRT may not perform well in finite samples. There are several alternative tests, but they lack optimality motivation and can also have bad power when the number of parameters is large. Muirhead (2005) provides an excellent discussion of likelihood ratio tests and modifications thereof when testing covariance matrices, and on inference problems for principal components.

This paper considers a linear panel data framework which incorporates a few statistical and econometrics models. Group action in the sample space yields a maximal invariant statistic. Its distribution depends on nuisance parameters only through the maximal invariant in the parameter space. The density of the maximal invariant is largely used as means to derive the finite-sample power of several statistics; see Anderson (1984). Here, we propose to use the maximal invariant density to construct a likelihood ratio test. This is our marginal likelihood approach to solve the incidental parameter problem. The group action also yields invariant measures in the parameter space. Integrating the data likelihood for the invariant measures also eliminates the incidental parameters. It also gives the density ratio of the maximal invariant statistic, so the likelihood ratio test based on the marginal distribution of the maximal invariant coincides with the likelihood ratio test based on parameter integration. Combining the invariant measures with prior distributions for the remaining parameters also gives a class of admissible tests.

All results are for finite samples only, but intend to resolve the incidental parameter problem when both panel data dimensions are large. For example, Onatski *et al.* (2011, 2012) show that the maximal invariant likelihood ratio test asymptotically has good power in the *spiked* covariance model, whereas the standard likelihood ratio test has no power at all. It would be interesting to derive the asymptotic behavior of the maximal invariant likelihood ratio test for other models—including the examples of testing sphericity and equality of covariance matrices discussed in this paper. The likelihood ratio representation in terms of Haar measures (as in Proposition 3.8) can simplify the asymptotic derivation. We leave this task for future research.

Acknowledgements. Marc Hallin acknowledges the support of the Sonderforschungsbereich "Statistical modeling of non-linear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, and a Discovery Grant of the Australian Research Council. Marcelo Moreira's research is supported by CNPq and the NSF via grant number SES-0819761.

6. Appendix: Proofs

Proof of Proposition 3.1. Since $\mathrm{p}_{\bar{g}\theta}\left(gZ_2\right) = \mathrm{p}_{\theta}\left(Z_2\right)\chi_g\left(Z_2\right)$, and since $\bar{g}\Theta_0 = \Theta_0$ and $\bar{g}\Theta_1 = \Theta_1$, the Z_2-based LRT statistic (16) takes the form

$$\begin{aligned} LR(Z_2) &= 2\left[\max_{\theta\in\Theta_1}\ln\left(\mathrm{p}_{\bar{g}\theta}\left(gZ_2\right)/\chi_g\left(Z_2\right)\right) - \max_{\theta\in\Theta_0}\ln\left(\mathrm{p}_{\bar{g}\theta}\left(gZ_2\right)/\chi_g\left(Z_2\right)\right)\right] \\ &= 2\left[\max_{\theta\in\Theta_1}\ln \mathrm{p}_{\bar{g}\theta}\left(gZ_2\right) - \max_{\theta\in\Theta_0}\ln \mathrm{p}_{\bar{g}\theta}\left(gZ_2\right)\right] \\ &= 2\left[\max_{\theta\in\Theta_1}\ln \mathrm{p}_{\theta}\left(gZ_2\right) - \max_{\theta\in\Theta_0}\ln \mathrm{p}_{\theta}\left(gZ_2\right)\right] = LR(gZ_2). \end{aligned}$$

□

Proof of Proposition 3.3. The singular value decomposition for Z_2 is

$$Z_2 = c_2\eta(T\left(Z_2\right))d_2', \quad c_2 \in \mathcal{O}_p,\ d_2 \in \mathcal{O}_{n-j}, \tag{29}$$

where $\eta(T\left(Z_2\right))$ is a rectangular diagonal matrix with diagonal elements the square roots of the ordered eigenvalues of Z_2Z_2'. It follows that

$$\begin{aligned} \mathrm{p}^H_{\gamma,\sigma}\left(Z_2\right) &= \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{p}_{\gamma,\sigma,\lambda,\omega}\left(c_2\eta(T\left(Z_2\right))d_2'\right)\ \mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega) \\ &= \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{p}_{\gamma,\sigma,c_2'\lambda,d_2'\omega}\left(\eta(T\left(Z_2\right))\right)\ \mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega) \\ &= \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{p}_{\gamma,\sigma,\lambda,\omega}\left(\eta(T\left(Z_2\right))\right)\ \mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega) = \mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}\left(\eta(T\left(Z_2\right))\right). \end{aligned} \tag{30}$$

The distribution under $\theta = (\gamma,\sigma,\lambda,\omega)$ of $T = T\left(Z_2\right)$ only depends on γ and σ; for any Borel set B

in $\mathbb{R}^{p\times(n-j)}$,

$$\begin{aligned}\mathrm{P}_{\gamma,\sigma}(T\in B) &= \mathrm{P}_{\gamma,\sigma,\lambda,\omega}(T(Z_2)\in B) = \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{P}_{\gamma,\sigma,\lambda,\omega}(T(Z_2)\in B)\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega)\\ &= \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{P}_{\gamma,\sigma,\lambda,\omega}(Z_2\in T^{-1}(B))\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega)\\ &= \int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}}\int_{T^{-1}(B)} \mathrm{p}_{\gamma,\sigma,\lambda,\omega}(z)\,\mu(\mathrm{d}z)\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega).\end{aligned}$$

Hence, from the Fubini-Tonelli Theorem,

$$\begin{aligned}\mathrm{P}_{\gamma,\sigma}(T\in B) &= \int_{T^{-1}(B)}\int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} \mathrm{p}_{\gamma,\sigma,\lambda,\omega}(z)\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega)\,\mu(\mathrm{d}z)\\ &= \int_{T^{-1}(B)} \mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(z)\,\mu(\mathrm{d}z)\\ &= \int_{T^{-1}(B)} \mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(\eta(T(z)))\,\mu(\mathrm{d}z) = \int_B \mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(\eta(t))\,\mu T^{-1}(\mathrm{d}t),\end{aligned}$$

which completes the proof. □

Proof of Proposition 3.4. The solution δ^* of (21) is the minimizer over $\mathcal{C}$ of

$$\int_{\mathbb{R}^{p\times(n-j)}}\int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}}\int_{\mathbb{R}^k\times\mathbb{R}^r} L(\gamma,\sigma;\delta(z))\,\mathrm{p}_{\gamma,\sigma,\lambda,\omega}(z)\;\mathrm{d}z\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega)\,W(\mathrm{d}\gamma\times\mathrm{d}\sigma).$$

If the loss function is integrable (which is the case in a hypothesis testing context), this expression, by the Tonelli-Fubini Theorem and Proposition 3.3, equals

$$\int_{\mathbb{R}^k\times\mathbb{R}^r}\int_{\mathbb{R}^{p\times(n-j)}} L(\gamma,\sigma;\delta(z))\,\mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(z)\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma)\,\mathrm{d}z =$$
$$\int_{\mathbb{R}^k\times\mathbb{R}^r}\int_{\mathbb{R}^{p\times(n-j)}} L(\gamma,\sigma;\delta(z))\,\mathrm{p}^{\mathrm{H}}_{\gamma,\sigma}(\eta(T(z)))\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma)\,\mathrm{d}z.$$

Hence, δ^* depends on Z_2 only through $T(Z_2)$. It remains to show that δ^* satisfies

$$\delta^* = \arg\min_{\delta\in\mathcal{C}}\sup_{\lambda\in\mathcal{F}_{k,p},\,\omega\in\mathcal{F}_{k,n-j}}\int_{\mathbb{R}^k\times\mathbb{R}^r} R(\gamma,\sigma,\lambda,\omega;\delta)\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma).$$

This, however, follows from the fact that δ^* is a function of the maximal invariant $T(Z_2)$ for G_2 (so that its risk does not depend on (λ,ω)), where G_2 is a compact group (see, for instance, Ferguson (1967, p. 156)). The proof below follows along the same lines as in Chamberlain (2007) and Chamberlain and Moreira (2009). For any decision rule $\delta\in\mathcal{C}$,

$$\begin{aligned}&\sup_{\lambda\in\mathcal{F}_{1,N},\,\omega\in\mathcal{F}_{1,T-J}}\int_{\mathbb{R}^k\times\mathbb{R}^r} R(\gamma,\sigma,\lambda,\omega;\delta)\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma)\\ &\quad\geq \int_{\mathbb{R}^k\times\mathbb{R}^r}\int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} R(\gamma,\sigma,\lambda,\omega;\delta)\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma)\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega)\\ &\quad\geq \int_{\mathbb{R}^k\times\mathbb{R}^r}\int_{\mathcal{F}_{k,p}\times\mathcal{F}_{k,n-j}} R(\gamma,\sigma,\lambda,\omega;\delta^*)\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma)\,\mathrm{H}(\mathrm{d}\lambda\times\mathrm{d}\omega).\end{aligned}$$

Since the risk of δ^* does not depend on (λ,ω), this last expression equals

$$\sup_{\lambda\in\mathcal{F}_{k,p},\,\omega\in\mathcal{F}_{k,n-j}}\int_{\mathbb{R}^k\times\mathbb{R}^r} R(\gamma,\sigma,\lambda,\omega;\delta^*)\;W(\mathrm{d}\gamma\times\mathrm{d}\sigma).$$ □

Proof of Proposition 3.5. (i) Let $\lambda = Vu'\varphi\left(f_2'f_2\right), \gamma = \gamma\left(f_2'f_2\right)$ and $\omega = \omega\left(f_2\right)$. Since $\tilde{\lambda} = Vu'$, $\lambda\gamma\omega$ is a singular value decomposition of $\tilde{\lambda}f_2'$, and since $Z_2 = \tilde{\lambda}f_2' + \sigma W_2$, the density of Z_2 at z conditional on f_2 is

$$\mathrm{p}_{\tilde{\lambda},\sigma,f_2}^{Z_2|f_2}(z) = \mathrm{p}_{\gamma(f_2'f_2),\sigma,Vu'\varphi\left(f_2'f_2\right),\omega(f_2)}(z),$$

whereas the unconditional density is

$$\mathrm{q}_{\sigma,V,h}(z) = \int_{\mathbb{R}^{(n-j)\times k}} \mathrm{p}_{\gamma(x'x),\sigma,Vu'\varphi(x'x),\omega(x)}(z)\, \mathrm{P}_{u,h,\sigma}^{f_2}(\mathrm{d}x), \tag{31}$$

where $\mathrm{P}_{u,h,\sigma}^{f_2}(\mathrm{d}x)$ is the measure on $\mathbb{R}^{(n-j)\times k}$ induced by $f_2 \overset{d}{=} N\left(0, \Sigma \otimes I_{n-j}\right)$.

As shown by James (1954), the distribution of a k-variate normal sample of size $n-j$ can be decomposed into a product of two independent distributions: a uniform distribution on $\mathcal{F}_{k,n-j}$ and a k-dimensional Wishart distribution with $n-j$ degrees of freedom:

$$\mathrm{P}_{u,h,\sigma}^{f_2}(\mathrm{d}x) = \mathrm{H}_{n-j}(\mathrm{d}\omega)\, \mathrm{P}_{u,h,\sigma}^{f_2'f_2}(\mathrm{d}y),$$

where $\omega = \omega(x)$, $\mathrm{H}_{n-j}(\mathrm{d}\omega)$ is the Haar measure on $\mathcal{F}_{k,n-j}$, and $\mathrm{P}_{u,h,\sigma}^{f_2'f_2}$ the (Wishart) measure on $\mathbb{R}^{k\times k}$ induced by $f_2'f_2$. Substituting this into (31) yields (24).

(ii) By the Fubini-Tonelli Theorem,

$$\begin{aligned}
\mathrm{q}_{h,\sigma}^{\mathrm{H}}(z) &= \int_{\mathcal{F}_{k,p}} \mathrm{q}_{\sigma,V,h}(z)\mathrm{H}_p(\mathrm{d}V) \\
&= \int_{\mathcal{F}_{k,n-j}} \int_{\mathbb{R}^{k\times k}} \int_{\mathcal{F}_{k,p}} \mathrm{p}_{\gamma(y),\sigma,Vu'\varphi(y),\omega}(z)\, \mathrm{H}_{n-j}(\mathrm{d}\omega)\, \mathrm{P}_{u,h,\sigma}^{f_2'f_2}(\mathrm{d}y)\, \mathrm{H}_p(\mathrm{d}V) \\
&= \int_{\mathbb{R}^{k\times k}} \mathrm{p}_{\gamma(y),\sigma}^{\mathrm{H}}(z)\, \mathrm{P}_{u,h,\sigma}^{f_2'f_2}(\mathrm{d}y).
\end{aligned}$$

According to Proposition 4.1, $\mathrm{p}_{\gamma(y),\sigma}^{\mathrm{H}}(z) = \mathrm{p}_{\gamma(y),\sigma}^{\mathrm{H}}(\eta(T(z)))$. Therefore, $\mathrm{q}_{h,\sigma}^{\mathrm{H}}(z) = \mathrm{q}_{h,\sigma}^{\mathrm{H}}(\eta(T(z)))$. Following a proof similar to that of Proposition 4.1, we obtain that $\mathrm{q}_{h,\sigma}^{\mathrm{H}}(\eta(t))$ is the density of $T = T(Z_2)$ for the unconditional model. □

Proof of Lemma 3.1. The following traditional notation is used. For any matrix A, denote by $\mathrm{vec}(A)$ the column vector obtained by stacking the columns of A on top of each other. For a symmetric $p\times p$ matrix A, let $\mathrm{vech}(A)$ denote the $(p(p+1)/2)$-dimensional column vector obtained from $\mathrm{vec}(A)$ by eliminating the supradiagonal elements of A. The duplication matrix is the matrix D_p such that $D_p\mathrm{vech}(A) = \mathrm{vec}(A)$. We can then write the density of Z_2 as

$$f_\Sigma(Z_2) = (2\pi)^{-\frac{p(n-j)}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\mathrm{vech}\left(\Sigma^{-1}\right)' D_p' D_p \mathrm{vech}\left(Z_2 Z_2'\right)\right\}. \tag{32}$$

The matrix D_p has full column rank p, so that $D_p'D_p$ is invertible. It follows from Theorem 6.3 of Lehmann and Scheffe (1950) to the density (32) that $\mathrm{vech}(Z_2Z_2')$ is minimal sufficient for Σ. Moreover, by Theorem 4.3.1 of Lehmann and Romano (2005, p. 116), $\mathrm{vech}(Z_2Z_2')$ is complete for the model under which the covariance matrix Σ is unrestricted.

Write $U = U(Z_2) = \mathrm{vech}(Z_2Z_2')$. Consider a test ϕ invariant under the orthogonal group $\mathcal{O}_p$. Because the maximal invariant to $\mathcal{O}_p$ is $Z_2'Z_2$, this test is of the form $\phi(Z_2'Z_2)$. Define $\widetilde{\phi}(u) = \mathrm{E}\left(\phi(Z_2'Z_2)\,|U = u\right)$. This conditional expectation does not depend on Σ because $U(Z_2)$ is a sufficient statistic. Since

$$\mathrm{E}_\Sigma\, \phi(Z_2'Z_2) = \mathrm{E}_\Sigma \mathrm{E}\left(\phi(Z_2'Z_2)\,|U\right) = \mathrm{E}_\Sigma\, \widetilde{\phi}(U),$$

the U-measurable test $\widetilde{\phi}(U)$ has the same power function as $\phi(Z_2{}'Z_2)$. Suppose there exists another test $\phi_0(U)$ which has the same power as $\widetilde{\phi}(U)$: $\mathrm{E}_\Sigma \widetilde{\phi}(U) = \mathrm{E}_\Sigma \phi_0(U)$ for all Σ. Then, for all Σ,

$$\mathrm{E}_\Sigma\left[\widetilde{\phi}(U) - \phi_0(U)\right] = 0.$$

Because the statistic U is complete, $\widetilde{\phi}(u) = \phi_0(U)$ almost surely. Hence, $\widetilde{\phi}(U)$ is unique up to zero-measure sets. From Lemma 6.5.1 and Theorem 6.5.3 (i) in Lehmann and Romano (2005), $\widetilde{\phi}(U(Z_2))$ is invariant to transformations by $\mathcal{O}_p$ of Z_2. The maximal invariant for $\mathcal{O}_p$ is $T(Z_2)$; therefore, $\widetilde{\phi}$ is of the form $\widetilde{\phi}(U(Z_2)) = \phi^*(T(Z_2))$. □

Proof of Theorem 3.6. From the definition of ϕ^*, we have

$$\begin{aligned}\phi^*(u) &= \mathrm{E}\left(I\left(\psi\left(Z_2'Z_2\right) > c_\alpha\right)|Z_2Z_2' = u\right)\\ &= \mathrm{P}\left(\psi\left(Z_2'Z_2\right) > c_\alpha|Z_2Z_2' = u\right)\\ &= \mathrm{P}\left(\psi\left(Z_2'Z_2\right) > c_\alpha|(l_{(1)}, l_{(2)}, \ldots, l_{(p)}) = T(Z_2) = t\right)\\ &= = 1 - F_t(c_\alpha)\end{aligned}$$

since conditional (on the invariant $T(Z_2)$) expectations are invariant under $\mathcal{O}_p$ transformations. □

Proof of Proposition 3.7. We use Theorem 6.2.2 of Lehmann and Romano (2005, p. 217) again. The group G is generated by the subgroups $G_m^+ = \{g_{c,I_{n-1}}|\, c \in \mathcal{G}_m^+\}$ and $G_{n-1}^{\mathcal{O}} = \{g_{I_m,d}|\, d \in \mathcal{O}_{n-1}\}$.

On $Z_2^+ = Z_2^+(Z_2)$, $g_{c,I_{n-1}} \in G_m^+$ acts by simple matrix multiplication: $Z_2^+(g_{c,I_{n-1}}Z_2) = cZ_2^+(Z_2)$. Because $m \leq 2n$, the matrix Z_2^+ is a.s. full rank, and can be written uniquely as $Z_2^+ = c_2 w_2'$, where $c_2 \in \mathcal{G}_m^+$ and $w_2 \in \mathcal{F}_{m,2(n-1)}$. The statistic w_2 thus is maximal invariant under G_m^+. Similarly, $g_{I_m,d} \in G_{n-1}^{\mathcal{O}}$ induces on w_2' a transformation of the form

$$w_2'(I_2 \otimes d') = \begin{bmatrix} w_{21}' \\ w_{22}' \end{bmatrix} d'.$$

Using the polar decomposition of $[w_{21} : w_{22}]'$, we thus obtain that the maximal invariant is

$$\begin{bmatrix} w_{21}'w_{21} & w_{21}'w_{22} \\ w_{21}'w_{22} & w_{22}'w_{22} \end{bmatrix}.$$

On $(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}, \Sigma_2)$, the group G_m^+ induces transformations of the form $(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}, c\Sigma_2 c')$, defining an induced group $\bar{G}_m^+$. Because the Cholesky decomposition of Σ_2 is unique, $\bar{G}_m^+$ acts transitively on Σ_2 while preserving $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}$. Therefore, the maximal invariant is $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}$, as claimed. □

Proof of Proposition 3.8. Eaton (1989) shows that G_m^+ is a *topological group*. Hence, there exist unique (up to a constant) *left- and right-invariant measures* to the group G. This group acts *properly*[3] on Z_2, and transitively on (Σ_1, Σ_2) when $\Sigma_2 = \sigma\Sigma_1$.

Following Andersson (1982), the optimal test rejects the null for large values of the density ratio

$$\frac{\displaystyle\int\!\!\int p_{(\Sigma_1,\xi\Sigma_1)}(g_{c,d}Z_2)\left(\prod_{i=1}^{n} c_{ii}\right)^{-2(n-1)} \mathrm{d}c\,\nu(\mathrm{d}d)}{\displaystyle\int\!\!\int p_{(\Sigma_1,\Sigma_1)}(g_{c,d}Z_2)\left(\prod_{i=1}^{n} c_{ii}\right)^{-2(n-1)} \mathrm{d}c\,\nu(\mathrm{d}d)},$$

[3]A mapping ψ from $\mathbb{R}^{d_1}$ to $\mathbb{R}^{d_2}$ is called a *proper mapping* if, for any compact $\mathcal{K} \subset \mathbb{R}^{d_2}$, $\psi^{-1}(\mathcal{K})$ is compact; a group of proper mappings is said to act *properly*.

where ν is the Haar probability measure on $\mathcal{O}_{n-1}$. This density ratio equals

$$\frac{\int\int \exp\Big\{ -\frac{1}{2}\mathrm{tr}\left(\Sigma_1^{-1}c\left[Z_{21}d'dZ_{21}' + \xi^{-1}Z_{22}d'dZ_{22}'\right]c'\right)\Big\}\Big(\prod_{i=1}^{m} c_{ii}\Big)^{-2(n-1)} \mathrm{d}c\,\nu\,(\mathrm{d}d)}{\int\int \exp\Big\{ -\frac{1}{2}\mathrm{tr}\left(\Sigma_1^{-1}c\left[Z_{21}d'dZ_{21}' + Z_{22}d'dZ_{22}'\right]c'\right)\Big\}\Big(\prod_{i=1}^{m} c_{ii}\Big)^{-2(n-1)} \mathrm{d}c\,\nu\,(\mathrm{d}d)}$$

$$= \frac{\int\int \exp\Big\{ -\frac{1}{2}\mathrm{tr}\left(\Sigma_1^{-1}c\left[Z_{21}Z_{21}' + \xi^{-1}Z_{22}Z_{22}'\right]c'\right)\Big\}\Big(\prod_{i=1}^{m} c_{ii}\Big)^{-2(n-1)} \mathrm{d}c\,\nu\,(\mathrm{d}d)}{\int\int \exp\Big\{ -\frac{1}{2}\mathrm{tr}\left(\Sigma_1^{-1}c\left[Z_{21}Z_{21}' + Z_{22}Z_{22}'\right]c'\right)\Big\}\Big(\prod_{i=1}^{m} c_{ii}\Big)^{-2(n-1)} \mathrm{d}c\,\nu\,(\mathrm{d}d)}$$

$$= \frac{\int \exp\Big\{ -\frac{1}{2}\mathrm{tr}\left(\Sigma_1^{-1}c\left[Z_{21}Z_{21}' + \xi^{-1}Z_{22}Z_{22}'\right]c'\right)\Big\}\Big(\prod_{i=1}^{m} c_{ii}\Big)^{-2(n-1)} \mathrm{d}c}{\int \exp\Big\{ -\frac{1}{2}\mathrm{tr}\left(\Sigma_1^{-1}c\left[Z_{21}Z_{21}' + Z_{22}Z_{22}'\right]c'\right)\Big\}\Big(\prod_{i=1}^{m} c_{ii}\Big)^{-2(n-1)} \mathrm{d}c}.$$

Because the group acts transitively on Σ_1, we can choose any value for Σ_1 in the integrals above. In particular, we get the desired result by setting $\Sigma_1 = I_m$. □

7. References

Ahn, S. and Horenstein, A. (2012). Eigenvalue ratio test for the number of factors. *Econometrica*, forthcoming.

Alessi, L., Barigozzi, M. and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters*, **80**, 1806–1813.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons, Inc.

Andersson, S. (1982). Distributions of maximal invariants using quotient measures. *Annals of Statistics*, **10**, 955–961.

Arellano, M. (2003). Panel data econometrics. In M. Arellano, G. Imbens, G. Mizon, A. Pagan and M. Watson (eds.), *Advanced Texts in Econometrics*, Amsterdam: Oxford.

— and Bond, S. R. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**, 277–297.

Athey, S. and Haile, O. (2007). Nonparametric approaches to auctions. In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 6, Elsevier Science, pp. 3487–3965.

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

— (2012). Fixed-effects dynamic panel models, a factor analytical method. *Econometrica*, forthcoming.

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

Bai, Z. D., Jiang, D., Yao, J. F. and Zheng, S. (2009). Corrections to lrt on large-dimensional covariance matrix by rmt. *Annals of Statistics*, **37**, 3822–3840.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variables estimators. *Econometrica*, **62**, 657–681.

Cai, T. T. and Ma, Z. (2012). Optimal hypothesis testing for high-dimensional covariance matrices. *Bernoulli*, forthcoming.

Chamberlain, G. (2007). Decision theory applied to an instrumental variables model. *Econometrica*, **75**, 609–652.

— and Moreira, M. J. (2009). Decision theory applied to a linear panel data model. *Econometrica*, **77**, 107–133.

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

Chen, S. X., Zhang, L. X. and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the Americal Statistical Association*, pp. 810–819.

Chioda, L. and Jansson, M. (2009). Optimal invariant inference when the number of instruments is large. *Econometric Theory*, **25**, 793–805.

Eaton, M. L. (1989). *Group Invariance Applications in Statistics.* Regional Conference Series in Probability and Statistics, Volume 1. Hayward, CA: Institute of Mathematical Statistics.

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

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.*

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica*, **76**, 643–660.

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

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

Hausman, J., Newey, W., Woutersen, T., Chao, J. and Swanson, N. (2012). Testing overidentifying restrictions with many instruments and heteroskedasticity. *Quantitative Economics*, **3**, 279–312.

James, A. (1954). Normal multivariate analysis and the orthogonal group. *Annals of Mathematical Statistics*, **25**, 40–75.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Annals of Statistics*, **29**, 295–327.

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

— (2002). Orthogonal parameters and panel data. *Journal of Econometrics*, **69**, 647–666.

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**, 1081–1102.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd edn., springer.

— and Scheffe, H. (1950). Completeness, similar regions, and unbiased estimation: Part i. *Sankhya: The Indian Journal of Statistics*, **10**, 305–340.

Moon, H. R. and Weidner, M. (2012). Dynamic linear panel data models with interactive fixed effects. Unpublished Manuscript, University of Southern California.

Moreira, M. J. (2009). A maximum likelihood method for the incidental parameter problem. *Annals of Statistics*, **37**, 3660–3696.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*.

Nadakuditi, R. R. and Edelman, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions on Signal Processing*, **56**, 2625–2638.

— and Silverstein, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE Journal of Selected Topics in Signal Processing*, **4**, 468–480.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1–32.

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

— (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, **168**, 244–258.

—, Moreira, M. J. and Hallin, M. (2011). Asymptotic power of sphericity tests for high-dimensional data. Working Paper, Cambridge University.

—, — and — (2012). Asymptotic power of sphericity tests for high-dimensional data and multispiked aternatives. Working Paper, Cambridge University.

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

Schott, J. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *Journal of Multivariate Analysis*, **54**, 175–192.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high-dimensional data. *Journal of the Japan Statistical Society*, **35**, 251–272.