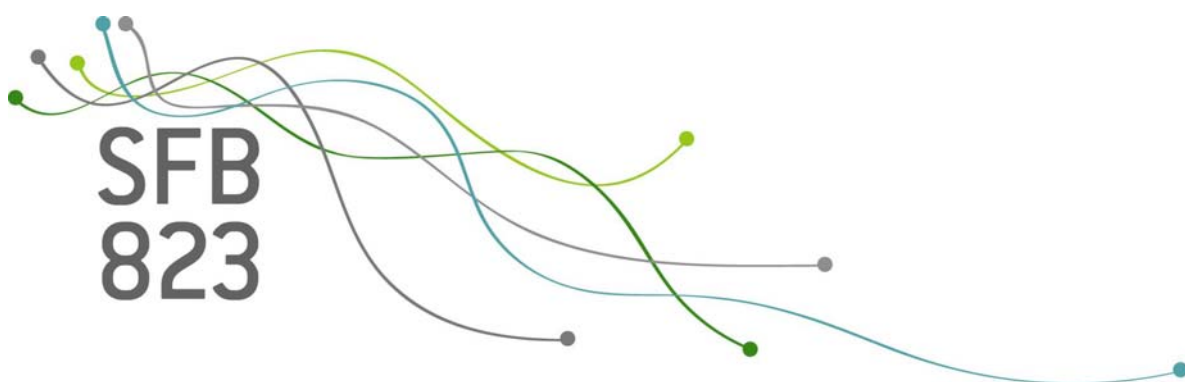# Rank-based tests of the cointegrating rank in semiparametric error correction models

Marc Hallin, Ramon van den Akker,
Bas J. M. Werker

Rank-based Tests of the Cointegrating Rank in Semiparametric Error Correction Models

Marc Hallin[a,b,c,1,*], Ramon van den Akker[c], Bas J.M. Werker[c,d,2]

[a]*ECARES, Université Libre de Bruxelles*
[b]*ORFE, Princeton University*
[c]*Econometrics group, CentER, Tilburg University*
[d]*Finance group, CentER, Tilburg University*

Abstract

This paper introduces rank-based tests for the cointegrating rank in an Error Correction Model with i.i.d. elliptical innovations. The tests are asymptotically distribution-free, and their validity does not depend on the actual distribution of the innovations. This result holds despite the fact that, depending on the alternatives considered, the model exhibits a non-standard Locally Asymptotically Brownian Functional (LABF) and Locally Asymptotically Mixed Normal (LAMN) local structure—a structure which we completely characterize. Our tests, which have the general form of Lagrange multiplier tests, depend on a reference density that can freely be chosen, and thus is not restricted to be Gaussian as in traditional quasi-likelihood procedures. Moreover, appropriate choices of the reference density are achieving the semiparametric efficiency bounds. Simulations show that our asymptotic analysis provides an accurate approximation to finite-sample behavior. Our results are based on an extension, of independent interest, of two abstract results on the convergence of statistical experiments and the asymptotic linearity of statistics to the context of, possibly non-stationary, time series.

Keywords: Cointegration model, Cointegration rank, Elliptical densities, Error correction model, Lagrange multiplier test, Local Asymptotic Brownian Functional, Local Asymptotic Mixed Normality, Local Asymptotic Normality, Multivariate ranks, non-Gaussian Quasi-Likelihood Procedures.

JEL codes: C14, C32.

1. Introduction

Since the introduction of cointegration by Granger (1981) and Engle and Granger (1987) an extensive literature on inference for cointegration models has emerged. Traditionally, Gaussian likelihood or moment-based procedures are used for estimation and testing problems concerning, e.g., the cointegrating rank or the cointegrating vectors; see, e.g., Stock (1987), Johansen and Juselius (1990), Johansen (1988, 1991, 1995), Phillips (1991) and Reinsel and Ahn (1992). Moreover, some papers describe (part of) the asymptotic structure of cointegrating models from a statistical point of view, notably Phillips (1991), Jeganathan (1997) and Hodgson (1998a, 1998b). We unify and extend these results by providing a complete characterization of the limiting experiments (à la Le Cam) arising within non-seasonal cointegrating Error Correction Models (ECMs) with independent and identically elliptically distributed innovations.

The main contribution of this paper is twofold. First of all, we use the limiting experiments to derive the form of locally and asymptotically optimal tests about the cointegrating rank. These optimal tests take the cointegrating vectors, the short term dynamics, and the exact form of the error distribution as nuisance parameters and turn out to be of the Lagrange Multiplier type. Secondly, we introduce rank-based versions of these statistics, i.e., asymptotically equivalent statistics that are based on a multivariate notion of ranks for the innovations.

The limiting experiments that may arise are described in Proposition 2.1. For special cases, most notably when the cointegrating rank is supposed to be known and under the absence of a linear time trend, the model turns out to be Locally Asymptotically Mixed Normal (LAMN) as in Hodgson (1998b). However, we demonstrate the *statistical* consequences of linear time trends in the model. These lead, both within and outside the cointegrating space, to specific directions in which the model is asymptotically Locally Asymptotically Normal (LAN) with convergence rate $T^{3/2}$. Specifically for inference

---

*Corresponding author

Email addresses: mhallin@ulb.ac.be (Marc Hallin), R.vdnAkker@TilburgUniversity.edu (Ramon van den Akker), Werker@TilburgUniversity.edu (Bas J.M. Werker)

about the cointegrating rank, we exploit these directions to construct locally and asymptotically optimal tests (optimality here is to be understood in the sense of stringency) that turn out to be of the Lagrange Multiplier type.

There does not seem to be consensus in the literature about the role of linear time trends in cointegration models; see, e.g., Saikkonen and Lütkepohl (2000). The main reason for this is that existing inference procedures are non-similar, i.e., their limiting distributions under the null hypothesis depend on the presence of a trend. This leads some people to reject inference procedures that derive their asymptotic properties from the linear trend. However, the standard cointegrated VAR as in our Section 1.1 does not exclude this behavior. Without explicit additional restrictions on inference procedures, asymptotically optimal inference procedures will achieve $T^{3/2}$ rates. The exact form and validity of such restrictions, in our view, is an empirical question and the answer may differ between applications. We already note here that the inference procedures we propose are similar under the null hypothesis. Furthermore, one may also interpret the present paper as using a particular form of asymptotic analysis to suggest inference procedures; see also the discussion in Hallin, Van den Akker, and Werker (2011). The simulations in Section 5 then show that their finite sample properties in many cases compare favorably to existing tests and thus provide a useful additional toolkit to the econometrician.

As mentioned before, our second contribution lies in the fact that our tests are rank-based. We do not restrict attention to rank-based tests a priori, but we prove that rank-based versions of the locally and asymptotically most stringent tests exist in the cointegration model with elliptically and identically distributed innovations. The use of such rank-based tests offers several advantages. First of all, the rank-based tests are asymptotically distribution-free. Thus, critical values do not depend on the actual distribution of the innovations. Without estimated nuisance parameters this distribution-freeness would even hold exactly in finite samples. In general, asymptotically, it provides a form of robustness that stabilizes finite sample sizes (see Section 5). Secondly, as an easy corollary of our analysis, we obtain a Chernoff-Savage result showing that Gaussian rank-based procedures uniformly improve over the pseudo-Gaussian ones.

However, probably most strikingly, the rank-based procedures enjoy a quasi- or

pseudo-likelihood property, not only when based on Gaussian scores but irrespective of the actual *reference density* used in the analysis. That is, imposing incorrectly a density g in the analysis, while the actual underlying density is $f \neq g$ does not invalidate our tests in terms of their asymptotic size. This is in sharp contrast with classical procedures. There, incorrectly imposing a Gaussian innovation density, and using appropriately adapted critical values, also leads to valid tests. However, incorrectly imposing, say, a t-distribution for the innovations leads to incorrectly sized tests and, in terms of estimation, inconsistent estimators. This drawback does not hold for our rank-based procedures that always are appropriately sized whatever the reference density g used and whatever the true underlying density f. This allows the researcher to select, e.g., a fat-tailed innovation density when innovations are suspected to be fat-tailed. This will lead to an increased local power of the test, without invalidating it when the assumed reference density differs from the actual one.

It may be informative to compare our contribution to recent work in Boswijk, Jansson, and Nielsen (2012). That paper improves upon likelihood ratio tests by a more careful description of the relevant alternatives. Formally, their alternatives are specified as cones instead of linear subspaces, much like a univariate test for a unit root generally uses the one-sided alternative that the autocorrelation parameter is strictly smaller than unity. That paper uses Gaussian likelihoods. We do not consider these more general cone-like specifications of the alternative. Both approaches are complementary and, thus, could possibly be integrated.

Before we formally introduce the model to be considered, a few remarks are in place. First, while we ignore the possibility of seasonality or linear trends in ΔX_t, these could be added and analyzed using the same techniques. This is largely due to the fact that the present paper contains some technical innovations that are of independent interest, and extend to a more general context. First, Appendix C contains a general result to obtain Local Asymptotic Quadraticity (in the sense of Jeganathan (1995)) of time-series models. The conditions we impose are serial extensions of the conditions needed to prove LAN for models with independent and identically distributed observations. Therefore, they are close to minimal. Secondly, when nuisance parameters are present and need to be estimated, the *asymptotic linearity* of the statistics of interest has to be established,

so that the effect of using estimated initial parameters can be made explicit; for that purpose, we generalize (Appendix C) a method of proof proposed by Van der Vaart (1988) for statistics based on independent observations to a time series context.

1.1. The model

We consider finite realizations $X^{(T)} := (X_1, \ldots, X_T)'$ from a p-dimensional time series $\{X_t \mid t \in \mathbb{N}\}$ generated by the kth order vector autoregressive model written in *error-correction form* (ECM)

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \mu + \varepsilon_t, \quad t \in \mathbb{N}, \tag{1.1}$$

where Δ is first-order differencing, $X_{1-k}, \ldots, X_0$ are deterministic starting values, $\Pi \in \mathbb{R}^{p \times p}$, $\Gamma := (\Gamma_1, \ldots, \Gamma_{k-1}) \in \mathbb{R}^{p \times (k-1)p}$ and $\mu \in \mathbb{R}^p$ are parameters, and $\{\varepsilon_t\}$ is an i.i.d. sequence of elliptically distributed innovations (centered at the origin) with density $\mathfrak{f}$. We shall assume that each component of $\{X_t\}$ is integrated of order one at most.

The assumptions we impose on this model are of two types: assumptions on the density $\mathfrak{f}$ of ε_t (the innovation density) and assumptions on the parameters μ, Γ, and Π.

1.2. Innovation densities

We assume throughout this paper that the innovations are elliptically distributed. Thus, their density $\mathfrak{f}$ satisfies the following assumption.

Assumption 1. *(Elliptical symmetry) There exists a $p \times p$ symmetric positive definite matrix Σ and a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\int_0^\infty z^{p-1} f(z) \mathrm{d}z = 1$ (the* radial density*) such that*

$$\mathfrak{f}(e) = \mathfrak{f}_{\Sigma, f}(e) := \frac{1}{\omega_p \sqrt{\det \Sigma}} f\left(\|e\|_\Sigma\right), \quad e \in \mathbb{R}^p, \tag{1.2}$$

where ω_p denotes the $(p-1)$-dimensional Lebesgue measure of the unit sphere $\mathcal{S}^{p-1}$ in $\mathbb{R}^p$ and $\|e\|_\Sigma := (e'\Sigma^{-1}e)^{1/2}$.

Clearly, f and Σ in (1.2) are only identified up to a scale factor; the requirement that $\int_0^\infty z^{p-1} f(z) \mathrm{d}z = 1$ thus is an identification constraint that does not imply any loss of generality (both Σ and f, moreover, are nuisance parameters in our inferential problems of interest). Under Assumption 1, the radial distance $\|\varepsilon_t\|_\Sigma$ at $z \in \mathbb{R}_+$ has density $\tilde{f}_p(z) := z^{p-1} f(z)$; write $\tilde{F}_p$ for the corresponding distribution function. Note

that, contrary to $\tilde{f}_p$, and despite of its name, the radial density f is not a probability density, as it does not integrate to one. Denoting by $\Sigma^{1/2}$ the symmetric square root of Σ, still under Assumption 1, $U_t^\varepsilon := \Sigma^{-1/2}\varepsilon_t/\|\varepsilon_t\|_\Sigma$ is uniformly distributed over the unit sphere $\mathcal{S}^{p-1}$ and independent of $\|\varepsilon_t\|_\Sigma$. In particular, it has mean $\mathrm{E}U_t^\varepsilon = 0$ and covariance matrix $\mathrm{Var}U_t^\varepsilon = I_p/p$.

On the radial density f itself we impose the following smoothness assumption.

Assumption 2. *(Radial density)*

   (a) *The radial density f is absolutely continuous with a.e. derivative f', that is, there exists a function f' such that for all $0 \leq a < b$*

$$f(b) - f(a) = \int_a^b f'(z)\mathrm{d}z.$$

   (b) *The radial Fisher information*

$$\mathcal{I}_p(f) := \mathrm{E}\phi_f^2(\|\varepsilon_1\|_\Sigma) = \int_0^1 \phi_f^2\left(\tilde{F}_p^{-1}(u)\right)\mathrm{d}u,$$

      where $\phi_f := -f'/f$ denotes the so-called location score associated with f, is finite.

   (c) *$f(z) > 0$ for all $z \in \mathbb{R}_+$.*

Observe that the location score for the p-variate density $\mathfrak{f}$ is given by

$$-\mathrm{grad}_e \log \mathfrak{f}(e) = \phi_f(\|e\|_\Sigma)\Sigma^{-1/2}U \text{ a.e.,} \quad \text{with} \quad U = \Sigma^{-1/2}e/\|e\|_\Sigma, \quad e \in \mathbb{R}^p.$$

This explains our terminology for ϕ_f and $\mathcal{I}_p(f)$. Also, it is possible to weaken Assumption 2 into an assumption of quadratic mean differentiability for the mapping $e \mapsto f^{1/2}(\|e\|)$; this is mainly of theoretical interest, though, and we refer to Section 1 of Hallin and Paindaveine (2002a) for details.

Let $\mathcal{F}$ denote the set of all radial densities satisfying Assumption 2. Note that the existence of moments for ε_t is completely determined by the existence of the corresponding moments for the radial density, and denote by $\mathcal{F}_2$ the subset of $\mathcal{F}$ corresponding to radial densities with finite second moment, i.e.,

$$\mathcal{F}_2 := \left\{f \in \mathcal{F} \mid \int \zeta^2 \tilde{f}_p(\zeta)\mathrm{d}\zeta = \int z^{p+1}f(z)\mathrm{d}z < \infty\right\}.$$

1.3. Parameter restrictions

In the present paper we are interested in the case that $\{X_t\}$ is integrated of order one ($I(1)$), in the sense that each component of the process is at most $I(1)$, and has no seasonal unit roots. The number of cointegrating relationships is denoted by r and can either be specified or unspecified. The required restrictions on the parameters are well known (see, e.g., Johansen (1995) or Juselius (2006)); for the sake of completeness, and to set notation, we briefly recall them here.

The characteristic polynomial $A_{\Gamma,\Pi}$ associated with the ECM equation (1.1) is

$$A_{\Gamma,\Pi}(z) := (1-z)I_p - \Pi z - \sum_{j=1}^{k-1} \Gamma_j(1-z)z^j, \quad z \in \mathbb{C}.$$

A (non-seasonal) unit root implies that Π is singular, since then $0 = |A_{\Gamma,\Pi}(1)| = |\Pi|$, with rank r, $0 \le r \le p-1$. We then can write Π as a product $\Pi = \alpha\beta'$ for $p \times r$ matrices α and β of rank r; in case $r = 0$, we define $\alpha = \beta = 0_{p \times p}$. Also, let $\alpha_\perp$ and $\beta_\perp$ be $p \times (p-r)$ matrices of rank $p-r$ satisfying $\alpha'\alpha_\perp = 0_{r \times (p-r)} = \beta'\beta_\perp$; for $r = 0$, we define $\alpha_\perp = I_p = \beta_\perp$. Note that the matrices $\alpha, \alpha_\perp, \beta$, and $\beta_\perp$ are not uniquely defined (unless $r = 0$). That lack of identifiability plays little role in the sequel. In case it does, we mention it explicitly.

These parameter restrictions are formalized as follows.

Assumption 3. *Using the above notation, the matrices $\Pi = \alpha\beta'$ and Γ in (1.1) are such that*

(a) $\operatorname{rank}\Pi < p$;

(b) *if* $|A_{\Gamma,\Pi}(z)| = 0$ *then* $|z| > 1$ *or* $z = 1$;

(c) *the* $(p-r) \times (p-r)$ *matrix* $\Psi_{\Gamma,\Pi} := \alpha'_\perp \left(I_p - \sum_{j=1}^{k-1} \Gamma_j \right) \beta_\perp$ *is non-singular.*

Assumption 3(b) excludes explosive behavior and seasonal unit roots in the process $\{X_t\}$ and Assumption 3(c) is equivalent to the requirement that $A_{\Gamma,\Pi}$ has exactly $(p-r)$ unit roots.

Under Assumption 3, the following version of the Granger-Johansen representation theorem (see Theorem 3.2 in Nielsen (2009)) holds: letting

$$C_{\Gamma,\Pi} := \beta_\perp \Psi_{\Gamma,\Pi}^{-1} \alpha'_\perp, \tag{1.3}$$

the process $\{X_t\}$ generated by the ECM equation (1.1) admits the representation

$$X_t = (C_{\Gamma,\Pi}\mu)t + C_{\Gamma,\Pi}\sum_{s=1}^{t}\varepsilon_s + Y_t + a_{\mu,\Gamma,\Pi}, \tag{1.4}$$

where

(i) the deterministic starting value $a_{\mu,\Gamma,\Pi}$ is

$$a_{\mu,\Gamma,\Pi} = C_{\Gamma,\Pi}\left(\left(I_p - \sum_{j=1}^{k-1}\Gamma_j\right)X_0 + \sum_{\ell=0}^{k-2}\sum_{j=\ell+1}^{k-1}\Gamma_j\Delta X_{-\ell}\right); \tag{1.5}$$

(ii) the process Y_t is given by $Y_t = \Upsilon V_t$ with the $p \times (r + (k-1)p)$ matrix

$$\Upsilon := \left((I_p - C_{\Gamma,\Pi}(I_p - \sum_{j=1}^{k-1}\Gamma_j))\beta(\beta'\beta)^{-1} \quad -C\sum_{j=1}^{k-1}\Gamma_j \quad \cdots \quad -C\sum_{j=k-1}^{k-1}\Gamma_j\right),$$

and the process $V_t' := ((\beta'X_t)', (\Delta X_t)', \ldots, (\Delta X_{t-k+2})')$ satisfying

$$V_t = \Xi V_{t-1} + \begin{pmatrix} \beta' \\ I_p \\ 0_{(k-2)p \times p} \end{pmatrix}\varepsilon_t,$$

where all eigenvalues of Ξ are all less than 1 in absolute value.

Also observe, for later reference, that $C_{\Gamma,\Pi}\mu$ is an element of the column space of $\beta_\perp$.

Let Θ denote the set of admissible values of the parameter $\vartheta := (\mu, \Gamma, \Pi)$, i.e., those values of ϑ where $\mu \in \mathbb{R}^p$ is unrestricted, but Γ and Π are such that Assumption 3 holds. Write $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$ or $\mathrm{P}^{(T)}_{\mu,\Gamma,\Pi;\Sigma,f}$ for the distribution of $(X_1, \ldots, X_T)$ generated by the ECM equation (1.1), conditional on the starting values $X_{-k+1}, \ldots, X_0$, under the Euclidean parameter values $(\vartheta; \Sigma) = (\mu, \Gamma, \Pi; \Sigma)$ and the infinite-dimensional parameter f.

For all ϑ and $\Sigma > 0$, define the residuals

$$\epsilon_t = \epsilon_t(\vartheta) := \Delta X_t - \Pi X_{t-1} - \sum_{j=1}^{k-1}\Gamma_j\Delta X_{t-j} - \mu \tag{1.6}$$

and the corresponding *sphericized* unit vectors (playing the role of *multivariate signs*)

$$U_t = U_t(\vartheta; \Sigma) := \frac{1}{\|\epsilon_t(\vartheta)\|_\Sigma}\Sigma^{-1/2}\epsilon_t(\vartheta), \tag{1.7}$$

with the convention that $U_t := 0$ in case $\epsilon_t = 0$. Note that, if ϑ is the true parameter value, that is, under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, $\epsilon_t(\vartheta)$ and $U_t(\vartheta; \Sigma)$ coincide with ε_t and U_t^ε, respectively.

1.4. Rank-based inference

Throughout, (Σ, f) plays the role of a nuisance, ϑ being the parameter of interest. In principle, (Σ, f) either should be included in the list of parameters indexing local limiting experiments, or it should enter the definition of tangent spaces in the treatment of semiparametric efficiency. However, as we shall see in Section 4, the testing problems we consider are adaptive with respect to (Σ, f). As a consequence, we do not have to discuss paths with respect to Σ and f in our local experiments.

To motivate the statistics we propose, we now consider the model in which Σ is known and f is unknown, i.e. the semiparametric model

$$\mathcal{P}^{(T)} := \{ \mathrm{P}^{(T)}_{\vartheta;\Sigma,f} |\ \vartheta \in \Theta,\ f \in \mathcal{F}_2 \}.$$

Semiparametric efficiency bounds (in the sense of Bickel et al. (1993)), whether at some chosen *reference density* $g \in \mathcal{F}_2$ or uniformly over all densities $g \in \mathcal{F}_2$, provide the relevant optimality concept and represent the best performance one can hope for. The reason for treating Σ here as known is that this allows us to motivate our statistics by invariance arguments. Indeed, general results by Hallin and Werker (2003) indicate that if

(a) the parametric fixed-f submodels $\mathcal{P}^{(T)}_f := \{ \mathrm{P}^{(T)}_{\vartheta;\Sigma,f} |\ \vartheta \in \Theta \}$, $f \in \mathcal{F}$, are Locally Asymptotically Normal (LAN, with respect to ϑ), with *central sequence* $\Delta^{(T)}_{\vartheta;\Sigma,f}$, and

(b) the nonparametric fixed-ϑ submodels $\mathcal{P}^{(T)}_\vartheta := \{ \mathrm{P}^{(T)}_{\vartheta;\Sigma,f} |\ f \in \mathcal{F} \}$, $\vartheta \in \Theta$, are generated by some group of transformations $\mathcal{G}^{(T)}_{\vartheta;\Sigma}$ acting on $X^{(T)}$, with maximal invariant $M^{(T)}_{\vartheta;\Sigma}$ (typically, a combination of residual ranks and signs),

then, inference procedures reaching, as $T \to \infty$, the semiparametric efficiency bounds at $\mathrm{P}^{(T)}_{\vartheta;\Sigma,g}$ for given $g \in \mathcal{F}$ can be based on the conditional expectation

$$\underset{\sim}{\Delta}^{(T)}_{\vartheta;\Sigma,g} := \mathrm{E}_{\vartheta;\Sigma,g} \left[\Delta^{(T)}_{\vartheta;\Sigma,g} |\ M^{(T)}_{\vartheta;\Sigma} \right] \tag{1.8}$$

(where $\mathrm{E}_{\vartheta;\Sigma,g}$ stands for expectation under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,g}$). In other words, $\underset{\sim}{\Delta}^{(T)}_{\vartheta;\Sigma,g}$, at ϑ and g, is a version of the semiparametrically efficient central sequence for the semiparametric model $\mathcal{P}^{(T)}$. The fact that it is measurable with respect to the maximal invariant

9

of a generating group implies that, contrary to the traditional semiparametrically efficient central sequence resulting from tangent space projection, it is distribution-free under $\mathcal{P}_{\vartheta}^{(T)}$. Inference based on $\underset{\approx}{\Delta}_{\vartheta;\Sigma,g}^{(T)}$ thus remains valid irrespective of the actual density f, and g here plays the role of a *reference density*, to be chosen by the researcher. Still without affecting validity, the same reference density also can be substituted with an estimator $\hat{f}^{(T)}$ of f; we refer to Hallin and Werker (2003) for details.

In the present context the maximal invariant $M_{\vartheta;\Sigma}^{(T)}$ happens to be the T-tuple of multivariate signs $U_t(\vartheta;\Sigma)$ defined in (1.7), along with the ranks $R_t^{(T)}(\vartheta;\Sigma)$ of the residual moduli $\|\epsilon_t(\vartheta)\|_\Sigma$ (see (1.6)). Those multivariate signed ranks have been successfully used in a series of papers by Hallin and Paindaveine (2002a, 2002b, 2005a, 2005b), Hallin, Oja, and Paindaveine (2006) and Hallin, Paindaveine, and Verdebout (2010, 2012) for various problems (hypothesis testing and point estimation) in multivariate analysis and multivariate time series. All these papers consider models for which (a) is satisfied. However, for the cointegration model (a) is not satisfied as only specific submodels are LAN (see Section 2). Invoking adaptivity arguments, we show that it is nevertheless possible to develop efficient tests using signed ranks.

Finally, we note that ranks have been used in the context of cointegration before by Breitung and Gouriéroux (1997) and Breitung (2001). Those papers, however, consider testing for nonlinear forms of cointegration, while we focus on optimal inference within linear semiparametric cointegration models.

1.5. Outline of the paper

The remainder of this paper is organized as follows. Section 2 unifies and generalizes results on the (local) limit experiments resulting from the ECM model (1.1). These limit experiments allow us to rely on existing theory about (optimal) asymptotic inference, see, e.g., Strasser (1985), Le Cam (1986), Le Cam and Yang (1990), or Van der Vaart (2000). It turns out that the central sequence relevant for testing hypotheses on the cointegrating rank is of a specific form. In Section 3 we provide a rank-based version of that central sequence, and study its properties under fixed distributions and under local alternatives. In particular, we pay attention to the effect of using aligned ranks, i.e., those based on estimated residuals instead of innovations. These results are subsequently used in Section 4 to derive asymptotically distribution-free most-stringent rank-based tests about

10

the cointegrating rank; for the sake of completeness, and the purpose of comparisons, we also provide the explicit form of the corresponding optimal pseudo- or quasi-Gaussian tests. Section 5 provides a small simulation study that shows that our asymptotic analysis indeed provides a decent approximation to the actual finite-sample distributions. Finally, various appendices contain some results of independent interest and the most technical proofs. Appendices A and B establish the main results about the various limiting experiments and the asymptotic behavior of our rank-based statistics, respectively. Appendix C generalizes two known abstract results on likelihood ratio expansions and asymptotic linearity of statistics for models with independent observations to, possibly non-stationary, time series models. This also leads to results on the asymptotic behavior of statistics under local alternatives that are not of the classical Gaussian shift type. Finally, additional and more detailed simulation results are provided in Appendix D.

2. Limit experiments

In this section, we review, extend, and unify existing results on the (local) limit experiments (à la Le Cam) induced by the ECM model (1.1). We are interested in those limit experiments for several reasons. First, asymptotic efficiencies (parametric and semiparametric) of tests and estimators are defined via those limits. Secondly, the limit experiments allow us to perform local power calculations by applying Le Cam's third lemma. Third, limit experiments often suggest how efficient procedures can be constructed. Finally, simulations show that these limit experiments indeed yield good approximations of the actual finite-sample distributions of the statistics of interest.

Two main results on limit experiments for the ECM model are available in the literature. The first one is for the so-called triangular models studied by Phillips (1991) and Jeganathan (1997); the second one deals with cointegration models with known cointegrating rank and no trend—see Hodgson (1998b). The triangular models lead to limiting experiments of the Locally Asymptotically Brownian Functional (LABF) form. Fairly little is known about optimal inference in that type of experiment; see, however, Gushchin (1996), Jansson (2008), Müller (2012), and Ploberger and Phillips (2012) for some results. In this section, we identify exactly which local parameter directions lead, in the ECM model with elliptical noise, to this LABF behavior. Hodgson (1998b), in the

case of an ECM model with known cointegrating rank and without trend, derives, under an assumption of central symmetry on the innovation density $\mathfrak{f}$, a Locally Asymptotically Mixed Normal (LAMN) limiting behavior. Our results extend Hodgson (1998b) in that they also allow for inference on the cointegrating rank in the presence of a linear time trend. Possibly somewhat surprisingly, the LAMN property is lost in the presence of such a trend (see the discussion below Corollary 2.1).

Both the LABF and LAMN limit experiments are special cases of Locally Asymptotically Quadratic (LAQ) experiments, see Jeganathan (1995), which often arise in the context of (nearly) non-stationary time series (see, however, Drost, Van den Akker, and Werker (2009) for a unit root problem leading to a non-LAQ limit experiment). Proposition 2.1 states that the ECM model is LAQ.

2.1. Locally Asymptotically Quadratic limits

To derive the local limiting experiments, we analyze the (limiting) behavior of likelihood ratios for local perturbations $\vartheta^{(T)} := (\mu^{(T)}, \Gamma^{(T)}, \Pi^{(T)})$ of $\vartheta = (\mu, \Gamma, \Pi) \in \Theta$. Note that, since Θ is not an open set, it may happen that $\vartheta^{(T)}$ does not belong to Θ and possibly corresponds, for instance, with explosive root alternatives. The results we are deriving nevertheless also hold for such alternatives. The model we are investigating is, in that sense, slightly larger than the one parametrized by Θ; this is, however, common practice in the cointegration literature—see, e.g. Johansen (1995) on Gaussian maximum likelihood estimators.

We parametrize those local perturbations in a way that allows us to identify exactly which parameter directions (which local subexperiments) lead to LAN, LAMN, or LABF behavior[3]. The inference problems we discuss in this paper turn out to be *adaptive* with respect to Σ, so that local alternatives for Σ need not be considered, and adaptive with respect to f, so that semiparametric and parametric efficiency coincide (see Sections 4).

Our local perturbations are defined as follows. Building on the factorization $\Pi = \alpha\beta'$, where α and β are full rank $p \times r$ matrices, we define local (in the vicinity of $\vartheta = (\mu, \Gamma, \Pi)$) alternatives $\vartheta^{(T)}$ of the form $(\mu^{(T)}, \Gamma^{(T)}, \Pi^{(T)})$, with

$$\mu^{(T)} = \mu_m^{(T)} := \mu + T^{-1/2}m, \qquad \Gamma^{(T)} = \Gamma_G^{(T)} := \Gamma + T^{-1/2}G, \qquad (2.1)$$

---

[3]We are grateful to Peter Boswijk for his concrete suggestions at this point.

and

$$\Pi^{(T)} = \Pi^{(T)}_{A,b,B,d,D} := \alpha^{(T)}_A \beta^{(T)'}_{b,B} + T^{-3/2}\alpha_\perp d(C_{\Gamma,\Pi}\mu)' + T^{-1}\alpha_\perp D\beta'_\perp, \qquad (2.2)$$

with local parameters $m \in \mathbb{R}^p$, $G_1, \ldots, G_{k-1} \in \mathbb{R}^{p\times p}$, $d \in \mathbb{R}^{p-r}$, $D \in \mathbb{R}^{(p-r)\times(p-r)}$, and $\alpha^{(T)}_A$ and $\beta^{(T)}_B$ of the form

$$\alpha^{(T)} = \alpha^{(T)}_A := \alpha + T^{-1/2}A, \qquad \beta^{(T)} = \beta^{(T)}_{b,B} := \beta + T^{-3/2}(C_{\Gamma,\Pi}\mu)b' + T^{-1}\beta_\perp B', \quad (2.3)$$

for $A \in \mathbb{R}^{p\times r}$, $b \in \mathbb{R}^r$, and $B \in \mathbb{R}^{r\times(p-r)}$.

The local perturbations for μ, Γ and α are standard root-T perturbations, and lead to asymptotically normal behavior. The local alternatives for the cointegrating vectors β modify the cointegrating space without essentially affecting the cointegrating dimension r. These alternatives are specified (at rate $T^{3/2}$, and provided that $C_{\Gamma,\Pi}\mu \neq 0$) by the r-dimensional vector b and (at slower rate T) by the $r\times(p-r)$-dimensional matrix B. As $C_{\Gamma,\Pi}\mu$ lies in the column space of $\beta_\perp$, b and B are not separately identified. Therefore, we throughout restrict B to be linearly constrained such that $\beta_\perp B' \perp (C_{\Gamma,\Pi}\mu)$. Also note that α and β are not separately identified, as only their product $\Pi = \alpha\beta'$ is, which implies that the interpretation of the local parameters A, b, and B depends on the factorization adopted for Π and the chosen versions for $\alpha_\perp$ and $\beta_\perp$. This, however, has no further consequences for the remainder of the paper.

The constraint $\beta_\perp B' \perp (C_{\Gamma,\Pi}\mu)$ also illustrates the rationale for splitting the local perturbations of β into b and B. From (1.4) we see that, under $P_{\vartheta;\Sigma,f}$, $\beta'X_t$ is stationary and has zero drift (as $\beta'C_{\Gamma,\Pi}\mu = 0$). For local perturbations $\beta^{(T)}$ of β induced by B (at rate T), this still holds (under $P_{\vartheta;\Sigma,f}$); indeed, $\beta^{(T)'}X_t$ remains stationary since the additional effect of B on the drift is $T^{-1}B\beta'_\perp C_{\Gamma,\Pi}\mu = 0$. On the other hand, the local perturbations $\beta^{(T)}$ of β induced (at rate $T^{3/2}$) by b lead (under $P_{\vartheta;\Sigma,f}$, and provided that $C_{\Gamma,\Pi}\mu \neq 0$) to non-zero drifts in $\beta^{(T)'}X_t$, of magnitude $T^{-3/2}b(C_{\Gamma,\Pi}\mu)'C_{\Gamma,\Pi}\mu = T^{-3/2}b|C_{\Gamma,\Pi}\mu|^2$. These effects on the drifts also explain the different localizing rates for b and B.

In order to study local perturbations that affect the cointegrating rank, we similarly use the local parameters d and D in (2.2). For the same reason as above, and with the same interpretation, we throughout linearly restrict D by imposing (whenever $C_{\Gamma,\Pi}\mu \neq 0$) that $\beta_\perp D' \perp (C_{\Gamma,\Pi}\mu)$.

13

This local parametrization $\Pi^{(T)}_{A,b,B,d,D}$ allows us (see Proposition 2.1 below) to derive a LAQ property that covers the ECM model with specified cointegration rank (corresponding to the restriction $d = 0 = D$) as well as the ECM model with unspecified cointegration rank (corresponding to local alternatives with $d \neq 0$, in case $C_{\Gamma,\Pi}\mu \neq 0$ or $D \neq 0$). Note that the parametrization is such that both $d \neq 0$ and $D \neq 0$ increase the rank of Π; in case $d \neq 0$ and $D = 0$, however, the rank of Π increases by one unit exactly.

In order to obtain convenient expressions for the *central sequence* $\Delta^{(T)}_{\vartheta}$ and the finite-sample *Fisher information* $J^{(T)}_{\vartheta}$ featuring in our LAQ result, let us introduce some partial sum processes: for all $u \in [0,1]$ and $\vartheta \in \Theta$, define (with $\epsilon_t = \epsilon_t(\vartheta)$, $U_t = U_t(\vartheta, \Sigma)$ and $Y_t = Y_t(\vartheta)$ from (1.4))

$$W^{(T)}_{\epsilon}(u) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[uT]}\epsilon_t, \quad W^{(T)}_{\phi}(u) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[uT]}\Sigma^{-1/2}U_t\phi_f(\|\epsilon_t\|_{\Sigma}), \tag{2.4}$$

$$W^{(T)}_{\Delta X \otimes \phi, j}(u) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[uT]}\Delta X_{t-j} \otimes \Sigma^{-1/2}U_t\phi_f(\|\epsilon_t\|_{\Sigma}), \quad j = 1,\ldots,k-1, \tag{2.5}$$

and

$$W^{(T)}_{Y \otimes \phi}(u) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[uT]}Y_{t-1} \otimes \Sigma^{-1/2}U_t\phi_f(\|\epsilon_t\|_{\Sigma}), \tag{2.6}$$

where $A \otimes BC$ stands for $A \otimes (BC)$ – see, e.g., Magnus and Neudecker (1988). Lemma A.1 establishes the limiting Brownian behavior, under $P^{(T)}_{\vartheta;\Sigma,f}$, of these partial sums, and provides explicit expressions for their covariance structure. These results are not surprising, but we include a proof for completeness.

Associated with each parameter μ, $\Gamma_1, \ldots, \Gamma_{k-1}$, and α (hence, with each local parameter m, $G_1, \ldots, G_{k-1}$, and A) are components of the central sequence denoted by $\Delta^{(T)}_{\mu}, \Delta^{(T)}_{\Gamma_1}, \ldots, \Delta^{(T)}_{\Gamma_{k-1}}$, and $\Delta^{(T)}_{\alpha}$, respectively. For the matrix-valued parameters, that central sequence should be interpreted as that of the corresponding vectorized matrix. Thus, for instance, $\Delta^{(T)}_{\alpha}$ actually is the central sequence for $\text{vec}(\alpha)$ perturbed into $\text{vec}(\alpha) + T^{-1/2}\text{vec}(A)$; whenever possible, however, we avoid such heavy notation as $\Delta^{(T)}_{\text{vec}(\alpha)}$. As we split the local parameters associated with β' and $\beta'_{\perp}$ into (b, B) and (d, D), respectively, we write $\Delta^{(T)}_{b}$, $\Delta^{(T)}_{B}$, $\Delta^{(T)}_{d}$, and $\Delta^{(T)}_{D}$ for the corresponding ele-

14

ments of the central sequence[4]. This, at the same time, also distinguishes the parameters of interest from the nuisances μ, $\Gamma_1, \ldots, \Gamma_{k-1}$, and α.

The central sequence $\Delta_\vartheta^{(T)} = (\Delta_\mu^{(T)\prime}, \Delta_{\Gamma_1}^{(T)\prime}, \ldots, \Delta_{\Gamma_{k-1}}^{(T)\prime}, \Delta_\alpha^{(T)\prime}, \Delta_b^{(T)\prime}, \Delta_B^{(T)\prime}, \Delta_d^{(T)\prime}, \Delta_D^{(T)\prime})'$ and finite-sample Fisher information $J_\vartheta^{(T)}$ appearing in Proposition 2.1 below are

$$\Delta_\vartheta^{(T)} := \sum_{t=1}^{T} \begin{pmatrix} Z_{Tt}^{(1)}(\vartheta) \otimes I_p \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha' \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha'_\perp \end{pmatrix} \Sigma^{-1/2} U_t(\vartheta) \phi_f(\|\epsilon_t(\vartheta)\|_\Sigma), \tag{2.7}$$

and

$$J_\vartheta^{(T)} := \frac{\mathcal{I}_p(f)}{p} \sum_{t=1}^{T} \begin{pmatrix} Z_{Tt}^{(1)} Z_{Tt}^{(1)\prime} \otimes \Sigma^{-1} & Z_{Tt}^{(1)} Z_{Tt}^{(2)\prime} \otimes \Sigma^{-1}\alpha & Z_{Tt}^{(1)} Z_{Tt}^{(2)\prime} \otimes \Sigma^{-1}\alpha_\perp \\ & Z_{Tt}^{(2)} Z_{Tt}^{(2)\prime} \otimes \alpha'\Sigma^{-1}\alpha & Z_{Tt}^{(2)} Z_{Tt}^{(2)\prime} \otimes \alpha'\Sigma^{-1}\alpha_\perp \\ & & Z_{Tt}^{(2)} Z_{Tt}^{(2)\prime} \otimes \alpha'_\perp \Sigma^{-1}\alpha_\perp \end{pmatrix}, \tag{2.8}$$

respectively, where

$$Z_{Tt}^{(1)} = Z_{Tt}^{(1)}(\vartheta) := \begin{pmatrix} T^{-1/2} \\ T^{-1/2}\Delta X_{t-1} \\ \vdots \\ T^{-1/2}\Delta X_{t-k+1} \\ T^{-1/2}\beta' X_{t-1} \end{pmatrix} \quad \text{and} \quad Z_{Tt}^{(2)} = Z_{Tt}^{(2)}(\vartheta) := \begin{pmatrix} T^{-3/2}(C_{\Gamma,\Pi}\mu)' X_{t-1} \\ T^{-1}\beta'_\perp (X_{t-1} - (t-1)C_{\Gamma,\Pi}\mu) \end{pmatrix}.$$

$$\tag{2.9}$$

Remark 2.1. While a formal derivation of the central sequence is given in the proof of Proposition 2.1 below, the form of $\Delta_\vartheta^{(T)}$ in (2.7) also follows by pointwise differentiation of the log likelihood. Such differentiation does not yield the terms $-(t-1)C_{\Gamma,\Pi}\mu$ in the lower entries of $Z_{Tt}^{(2)}(\vartheta)$. However, these terms, corresponding to B and D, vanish in the expansion of the likelihood ratio due to the imposed orthogonality

$$\text{vec}(B)' \left(\beta'_\perp C_{\Gamma,\Pi}\mu \otimes \alpha' \right) = \text{vec}\left(\alpha B \beta'_\perp C_{\Gamma,\Pi}\mu \right)' = 0$$

and the corresponding condition for D. $\qquad\qquad\square$

The central sequence $\Delta_\vartheta^{(T)}$ can also be expressed in terms of the partial sum processes (2.4)-(2.6):

$$\Delta_\mu^{(T)} = W_\phi^{(T)}(1), \quad \Delta_{\Gamma_j}^{(T)} = W_{\Delta X \otimes \phi, j}^{(T)}(1), \quad 1 \le j \le k-1, \quad \Delta_\alpha^{(T)} = (\beta' \otimes I_p)W_{Y \otimes \phi}^{(T)}(1), \tag{2.10}$$

$$\Delta_b^{(T)} = |C_{\Gamma,\Pi}\mu|^2 \int_0^1 \text{id}_T(u-)\text{d}(\alpha' W_\phi^{(T)})(u) + T^{-1/2} \int_0^1 (C_{\Gamma,\Pi}\mu)'(C_{\Gamma,\Pi} W_\epsilon^{(T)}(u-))\text{d}(\alpha' W_\phi^{(T)})(u)$$

---

[4]All those central sequence components depend on ϑ, Σ and f, which we omit for notational simplicity. The subscripts μ, Γ_1, ... in $\Delta_\mu^{(T)}$, $\Delta_{\Gamma_1}^{(T)}$, ... thus only play the role of labels. So do the subscripts b, B, ... in $\Delta_b^{(T)}$, $\Delta_B^{(T)}$, ..., which should not be misinterpreted as indicating any functional relation between a local parameter as b and a central sequence component $\Delta_b^{(T)}$.

$$+ T^{-1}((C_{\Gamma,\Pi}\mu)' \otimes \alpha')W_{Y\otimes\phi}^{(T)}(1) + T^{-1}(C_{\Gamma,\Pi}\mu)'a_{\Gamma,\Pi,\mu}\alpha'W_\phi^{(T)}(1), \tag{2.11}$$

$$\Delta_B^{(T)} = \int_0^1 \beta_\perp' C_{\Gamma,\Pi}W_\epsilon^{(T)}(u-) \otimes I_r \mathrm{d}(\alpha'W_\phi^{(T)})(u) + T^{-1}(\beta_\perp' \otimes \alpha')W_{Y\otimes\phi}^{(T)}(1)$$
$$+ T^{-1}(\beta_\perp' a_{\Gamma,\Pi,\mu} \otimes \alpha'W_\phi^{(T)}(1)), \tag{2.12}$$

$$\Delta_d^{(T)} = |C_{\Gamma,\Pi}\mu|^2 \int_0^1 \mathrm{id}_T(u-)\mathrm{d}(\alpha_\perp'W_\phi^{(T)})(u) + T^{-1/2}\int_0^1 (C_{\Gamma,\Pi}\mu)'(C_{\Gamma,\Pi}W_\epsilon^{(T)}(u-))\mathrm{d}(\alpha_\perp'W_\phi^{(T)})(u)$$
$$+ T^{-1}((C_{\Gamma,\Pi}\mu)' \otimes \alpha_\perp')W_{Y\otimes\phi}^{(T)}(1) + T^{-1}(C_{\Gamma,\Pi}\mu)'a_{\Gamma,\Pi,\mu}\alpha_\perp'W_\phi^{(T)}(1), \tag{2.13}$$

$$\Delta_D^{(T)} = \int_0^1 \beta_\perp' C_{\Gamma,\Pi}W_\epsilon^{(T)}(u-) \otimes I_{p-r}\mathrm{d}(\alpha_\perp'W_\phi^{(T)})(u) + T^{-1}(\beta_\perp' \otimes \alpha_\perp')W_{Y\otimes\phi}^{(T)}(1)$$
$$+ T^{-1}(\beta_\perp' a_{\Gamma,\Pi,\mu} \otimes \alpha_\perp'W_\phi^{(T)}(1)), \tag{2.14}$$

where id_T denotes the cadlag function $\mathrm{id}_T(u) := [uT]/T$, which converges to the identity function $\mathrm{id}(u) := u$ on $[0,1]$.

We are able now to formulate the main result of this section, which provides a complete characterization of the possible limiting local experiments in the ECM model (1.1). In essence, the proof of this result consists of two steps. The first step establishes a quadratic expansion of the log-likelihood ratio, where the central sequence appears in the first-order term and the finite-sample Fisher information in the second-order one. The second step provides the joint limiting behavior, in distribution, of the central sequence and the finite-sample Fisher information. In view of the representations above, that asymptotic distribution follows from a standard application of the literature on convergence to stochastic integrals. We refer the reader to Appendix A for the proof. The proof of the first step follows from an application of Proposition C.4, of independent interest, which generalizes LAN results based on differentiability in quadratic mean (DQM) for models with independent observations to the time series context with possibly non-LAN limits.

Proposition 2.1. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, and $f \in \mathcal{F}_2$. Consider a sequence of perturbations*

$$h_T' := (m_T', (\mathrm{vec}\, G_T)', (\mathrm{vec}\, A_T)', b_T', (\mathrm{vec}\, B_T)', d_T', (\mathrm{vec}\, D_T)'),$$

with h_T bounded, $\beta_\perp B_T' \perp (C_{\Gamma,\Pi}\mu)$, and $\beta_\perp D_T' \perp (C_{\Gamma,\Pi}\mu)$, which defines a local parameter sequence $\vartheta^{(T)}$, see (2.1)-(2.3). Then we have, under $\mathrm{P}_{\vartheta;\Sigma,f}^{(T)}$, as $T \to \infty$,

$$\log \frac{\mathrm{dP}_{\vartheta^{(T)};\Sigma,f}^{(T)}}{\mathrm{dP}_{\vartheta;\Sigma,f}^{(T)}} = h_T'\Delta_\vartheta^{(T)} - \frac{1}{2}h_T'J_\vartheta^{(T)}h_T + o_P(1). \tag{2.15}$$

Under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, $(\Delta^{(T)}_{\vartheta}, J^{(T)}_{\vartheta})$ *converges in distribution*[5] *to* (Δ, J) *satisfying*

$$\mathrm{E}\exp\left(h'\Delta - h'Jh/2\right) = 1,$$

where $\Delta' := (\Delta'_{\mu}, \Delta'_{\Gamma_1}, \ldots, \Delta'_{\Gamma_{k-1}}, \Delta'_{\alpha}, \Delta'_{b}, \Delta'_{B}, \Delta'_{d}, \Delta'_{D})$ *with, for* $j = 1, \ldots, k-1$,

$$\Delta_{\mu} := W_{\phi}(1), \quad \Delta_{\Gamma_j} := W_{\Delta X \otimes \phi, j}(1), \quad \Delta_{\alpha} := (\beta' \otimes I_p)W_{Y \otimes \phi}(1),$$

and

$$\Delta_b := |C_{\Gamma,\Pi}\mu|^2 \int_0^1 u\,\mathrm{d}(\alpha'W_{\phi})(u), \qquad \Delta_B := \int_0^1 (\beta'_{\perp}C_{\Gamma,\Pi}W_{\epsilon}(u) \otimes I_r)\,\mathrm{d}(\alpha'W_{\phi})(u),$$

$$\Delta_d := |C_{\Gamma,\Pi}\mu|^2 \int_0^1 u\,\mathrm{d}(\alpha'_{\perp}W_{\phi})(u), \qquad \Delta_D := \int_0^1 (\beta'_{\perp}C_{\Gamma,\Pi}W_{\epsilon}(u) \otimes I_{p-r})\,\mathrm{d}(\alpha'_{\perp}W_{\phi})(u),$$

respectively. As for J, *it decomposes into blocks in a conform way: the block corresponding to* $(\Delta'_{\mu}, \Delta'_{\Gamma_1}, \ldots, \Delta'_{\Gamma_{k-1}}, \Delta'_{\alpha}, \Delta'_{b}, \Delta'_{d})'$ *is given by the covariance matrix of this random vector (that part of* J *thus is deterministic, and only depends on* ϑ*); the block corresponding to* $(\Delta'_{B}, \Delta'_{D})'$ *is random (its distribution does not involve the local parameters, and only depends on* ϑ*). The full definition of* J *is explicitly provided in Appendix A.1.*

Remark 2.2. Proposition 2.1 and Le Cam's first Lemma (see, e.g., Lemma 6.2 in Van der Vaart (2000)) jointly imply that the sequences of probability measures $\mathrm{P}^{(T)}_{\vartheta^{(T)};\Sigma,f}$ and $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, $T \in \mathbb{N}$, are contiguous. Consequently, in expressions like (2.15), we do not have to worry whether o_P's or O_P's are taken at the null or at local alternatives of the form $\mathrm{P}^{(T)}_{\vartheta^{(T)};\Sigma,f}$. Throughout, this consequence of contiguity is used without further mention. □

Remark 2.3. The Brownian motions $C_{\Gamma,\Pi}W_{\epsilon}$ and $\alpha'W_{\phi}$ are independent, since

$$\mathrm{E}\left[C_{\Gamma,\Pi}\varepsilon_t \left(\alpha'\Sigma^{-1/2}U_t\phi_f(\|\varepsilon_t\|_{\Sigma})\right)'\right] = \mathrm{E}\left[C_{\Gamma,\Pi}I_p\alpha\right] = 0$$

in view of the fact that $\mathrm{E}\|\varepsilon_t\|_{\Sigma}\phi_f(\|\varepsilon_t\|_{\Sigma}) = p$ and $C_{\Gamma,\Pi}\alpha = 0$ (see (1.3)). □

Remark 2.4. Note that the Fisher Information matrix in Proposition 2.1, both in finite-sample form $J^{(T)}_{\vartheta}$ and in the limit J, is somewhat involved. However, while its structure is used to classify the various limiting experiments, the exact forms are not needed in the rank-based test statistics of Sections 4. □

2.2. Special cases

Proposition 2.1 provides a complete picture of all possible local experiments in the ECM model (1.1). More precisely, it identifies how the local parameters m, G, A, b, B, d and D lead to various types of limiting experiments. From the form of the corresponding

---

[5]Throughout, we use the notation $(A_T, B_T) \overset{d}{\to} (A, B)$ as shorthand for $(\mathrm{vec}(A_T)', \mathrm{vec}(B_T)')' \overset{d}{\to} (\mathrm{vec}(A)', \mathrm{vec}(B)')'$.

central sequence in Proposition 2.1, we see that the behavior induced by D is of the LABF type. Putting $D = 0$ leads to the following corollary, Part (iii) of which will be used in Section 4 to construct optimal tests for the cointegrating rank.

Corollary 2.1. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, and $f \in \mathcal{F}_2$. Then,*

(i) *the sequences of local subexperiments associated with perturbations of the form (2.1)-(2.3) with $\alpha'_\perp m = 0$ and $D = 0$ are Asymptotically Mixed Normal whenever $\alpha'_\perp \mu = 0$;*

(ii) *the sequences of local subexperiments associated with perturbations of the form (2.1)-(2.3) with $m = 0$, $G = 0$, $A = 0$, $d = 0$, and $D = 0$ are Asymptotically Mixed Normal (with respect to the remaining local parameters b and B);*

(iii) *the sequences of local subexperiments associated with perturbations of the form (2.1)-(2.3) with $B = 0$ and $D = 0$ are Asymptotically Normal.*

Remark 2.5. Part (i) of this corollary is equivalent to a previous LAMN result by Hodgson (1998b). Without imposing $\alpha'_\perp \mu = 0$ and, for the corresponding local parameter $\alpha'_\perp m = 0$, the sequence of local subexperiments corresponding to $D = 0$ is *not* LAMN. This somewhat surprising result is due to the fact that, conditionally on $J_{\mu,\Gamma,\alpha,b,B,d}$, $\Delta_{\mu,\Gamma,\alpha,b,B,d}$ is no longer zero-mean Gaussian due to, for instance, the non-zero correlation between W_ϵ and W_ϕ. A detailed study of the statistical consequences of this phenomenon, and thereby the construction of optimal tests on the cointegrating vectors, is beyond the scope of the present paper. Note that, in Part (ii) above, with respect to the local parameters b and B only, we do obtain LAMN as $C_{\Gamma,\Pi}W_\epsilon$, which is the only source of randomness in $J_{b,B}$, and $\alpha'W_\phi$ are mutually independent (Remark 2.3).

Most relevant for the present paper is Part (iii) above. This (pure) LAN result, and the corresponding nonstandard rates $T^{3/2}$, do not seem to have been noted in the literature. We will exploit this result in Section 4 to derive optimal tests for the cointegrating rank.

3. Rank-based inference

The complete characterization of the possible limit experiments for the cointegration model (1.1) allows us to derive locally and asymptotically efficient tests for hypotheses on the cointegrating rank. As explained in the introduction, we will base these tests on optimal rank-based procedures derived in the LAN limit experiment associated with some reference density g and parameter perturbations satisfying $B = 0 = D$, i.e., in the situation of Corollary 2.1(iii)—except for the fact that g needs not be equal to f. This approach offers several advantages, similar to the rank-based unit root tests proposed in

Hallin, Van den Akker, and Werker (2011). First, we can use standard LAN projection results in order to derive the scores to be used in the rank-based procedures. Secondly, the restriction to $D = 0$ is innocuous as far as validity, i.e., the asymptotic size of our tests, is concerned. Concerning (local) powers, they easily follow (see part (iv) of Proposition 3.1) from an application of Le Cam's third lemma. Such approach is also feasible for local perturbations with nonzero D (hence, in a non-LAN context), but leads to fairly uninformative and complicated expressions. With respect to nonzero B we show that our test is actually adaptive. Thirdly, our procedures are valid for arbitrary reference density g (satisfying Assumption 5 below, which is extremely mild), not only the Gaussian one. This is in contrast to classical QMLE procedures that generally loose their size or consistency properties when based on non-Gaussian reference densities. As for the the power of our procedures, it obviously does depend on the chosen reference density, and choices that are close to the actual density lead to more powerful tests.

Finally, our procedures are asymptotically distribution-free. More precisely, would the signs and ranks they are based on be computed from *exact residuals* (namely, under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, our procedures would be strictly distribution-free. In practice, however, ϑ and Σ are not (fully) specified under the null hypotheses of interest, and estimated residuals $\|\epsilon_t(\hat{\vartheta}^{(T)})\|_{\hat{\Sigma}^{(T)}}$, yielding *aligned signs* and *aligned ranks*, will be substituted for the exact ones. The resulting procedures then remain asymptotically distribution-free, in the sense that the relevant *aligned test statistics* are asymptotically equivalent to the (unimplementable) *exact* and strictly distribution-free ones.

3.1. Ranks and signs as maximal invariants

As explained in Section 1.4, our rank-based semiparametric approach relies on two structural assumptions related with (a) the nature of fixed-f local experiments, and (b) the existence of a generating group for fixed-ϑ submodels. Our rank-based tests then are achieving semiparametric efficiency at the selected reference density g. The nature of local experiments involved has been studied in Section 2. In this section, we briefly explain which ranks and which signs enter into the picture, via assumption (b), as maximal invariants.

Fixing ϑ and Σ (recall that Σ, at this stage, is considered a fixed quantity), consider

the group $\mathcal{G}^{(T)}_{\vartheta;\Sigma},\circ$ of transformations $g^{\mathfrak{m}}_{\vartheta;\Sigma}$ acting on $\mathbb{R}^{pT}$ and indexed by $\mathfrak{m} \in \mathfrak{M}$, with

$$\mathfrak{M} := \left\{ \mathfrak{m} : \mathbb{R}^+ \to \mathbb{R}^+ \mid \mathfrak{m}(0) = 0, \ \lim_{z\to\infty} \mathfrak{m}(z) = \infty, \ \mathfrak{m} \text{ continuous, monotone increasing} \right\},$$

where $g^{\mathfrak{m}}_{\vartheta;\Sigma}$ is mapping the series $X^{(T)}$ with p-dimensional observations X_t onto the transformed series $X^{\mathfrak{m};(T)} := g^{\mathfrak{m}}_{\vartheta;\Sigma}(X^{(T)})$ with p-dimensional observations

$$X^{\mathfrak{m}}_t := (I_p + \Pi)X^{\mathfrak{m}}_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X^{\mathfrak{m}}_{t-j} + \mu + \mathfrak{m}(\|\epsilon_t(\vartheta)\|_\Sigma)\Sigma^{1/2}U_t(\vartheta;\Sigma), \quad t = 1,\dots,T$$

(a recursive definition, with the k deterministic starting values remaining unchanged). Letting $\mathfrak{m} = \tilde{G}_p^{-1}\circ\tilde{F}_p$, it is easy to see that the joint distribution of $X^{(T)}$ is $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$ if and only if the joint distribution of $g^{\mathfrak{m}}_{\vartheta;\Sigma}(X^{(T)})$ is $\mathrm{P}^{(T)}_{\vartheta;\Sigma,g}$, so that $\mathcal{G}^{(T)}_{\vartheta;\Sigma},\circ$ is a generating group for $\mathcal{P}^{(T)}_\vartheta$, with maximal[6] invariant $(U_1(\vartheta;\Sigma),\dots,U_T(\vartheta;\Sigma); R_1^{(T)}(\vartheta;\Sigma),\dots,R_T^{(T)}(\vartheta;\Sigma))$, where $R_t^{(T)}(\vartheta;\Sigma)$ denotes the rank of $\|\epsilon_t(\vartheta)\|_\Sigma$ among $\|\epsilon_1(\vartheta)\|_\Sigma,\dots,\|\epsilon_n(\vartheta)\|_\Sigma$.

3.2. Rank-based inference

The statistics of interest in Section 4 are functions of

$$S_f^{(T)}(\vartheta;\Sigma) \quad := \quad T^{-1/2}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)U_t(\vartheta;\Sigma)\phi_f\left(\|\epsilon_t(\vartheta)\|_\Sigma\right), \tag{3.1}$$

with $\epsilon_t(\vartheta)$ and $U_t(\vartheta;\Sigma)$ defined in (1.6) and (1.7). The rank-based version of this statistic, for reference density g, is

$$\underset{\sim}{S}_g^{(T)}(\vartheta;\Sigma) \quad := \quad T^{-1/2}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)U_t(\vartheta;\Sigma)\phi_g\left(\tilde{G}_p^{-1}\left(\frac{R_t^{(T)}(\vartheta;\Sigma)}{T+1}\right)\right). \tag{3.2}$$

As $\sum_{t=1}^{T}(t/(T+1) - 1/2) = 0$, $\underset{\sim}{S}_g^{(T)}(\vartheta;\Sigma)$ is centered (under $\mathrm{P}_{\vartheta;\Sigma,f}$). A straightforward adaptation of traditional results on linear rank statistics (see, e.g., Hájek and Sidák (1967), Section V.1.6) shows that the so-called *approximate score* form (3.2) of $\underset{\sim}{S}_g^{(T)}$ is

---

[6]For the maximal invariance of $(U_1(\vartheta;\Sigma),\dots,U_T(\vartheta;\Sigma); R_1^{(T)}(\vartheta;\Sigma),\dots,R_T^{(T)}(\vartheta;\Sigma))$, it is required to show that two points x and y in the observation space belong to the same orbit of $\mathcal{G}^{(T)}_{\vartheta;\Sigma}$ if and only if they yield the same U_i's and the same R_i's. That belonging to the same orbit implies having the same U_i's and the same R_i's is straightforward. The converse (two points x and y sharing the same U_i's and the same R_i's belong to the same orbit) is shown to hold by exhibiting a transformation that belongs to the group and maps x onto y. Such a transformation can be constructed exactly as in the traditional one-dimensional signed-rank case—see Lehmann and Romano (2005, p.242)—by considering the (monotone, due to their common ranking) piecewise linear function $\mathfrak{m} \in \mathfrak{M}$ that maps $d_1(x)$ to $d_1(y)$, ..., and $d_T(x)$ to $d_T(y)$, with linear interpolation in between.

asymptotically equivalent (i.e., the difference is $o_P(1)$ under $P_{\vartheta;\Sigma,g}$) to the *exact score form* (cf. Equation 1.8)

$$T^{-1/2} \sum_{t=1}^{T} \left(\frac{t}{T+1} - \frac{1}{2} \right) E_{\vartheta;\Sigma,g} \left[U_t(\vartheta;\Sigma) \phi_g \left(\|\epsilon_t(\vartheta)\|_\Sigma \right) \Big| U_t(\vartheta;\Sigma), \ R_t^{(T)}(\vartheta;\Sigma) \right]. \quad (3.3)$$

The main results we need on the asymptotic behavior of $\mathcal{S}_g^{(T)}(\vartheta;\Sigma)$ are summarized in Proposition 3.1 below. As already mentioned, since in practice Σ is unknown, the statistic $\mathcal{S}_g^{(T)}(\vartheta;\hat{\Sigma}^{(T)})$, where $\hat{\Sigma}^{(T)}$ is an estimator of Σ, will be used rather than $\mathcal{S}_g^{(T)}(\vartheta,\Sigma)$ itself; the statements in Proposition 3.1 accordingly are in terms of $\mathcal{S}_g^{(T)}(\vartheta;\hat{\Sigma}^{(T)})$. Some mild assumptions, however, are to be made on the estimator $\hat{\Sigma}^{(T)}$.

Assumption 4. *The sequence $\hat{\Sigma}^{(T)}$ is such that*

   (i) *for some $a \in \mathbb{R}_+$, $T^{1/2}(\hat{\Sigma}^{(T)} - a\Sigma)$ is $O_P(1)$ under $P_{\vartheta;\Sigma,f}^{(T)}$, as $T \to \infty$;*

   (ii) *$\hat{\Sigma}^{(T)}$ is a measurable function of the $\epsilon_t(\vartheta)$'s, and invariant under their permutations and reflections with respect to the origin.*

All concepts of scatter considered in the literature, among which the empirical covariance matrix, Tyler (1987)'s robust estimator, as well as the R-estimators of Hallin, Oja, and Paindaveine (2006) (when computed from the $\epsilon_t(\vartheta)$'s) satisfy Assumption 4.

For all results on rank-based statistics, we moreover need the following classical assumption on the reference density g.

Assumption 5. *The (radial) reference density g satisfies Assumption 1, and is such that ϕ_g is the difference of two continuous and monotone increasing functions.*

Finally, in Proposition 3.1 (iv) and (v) below, we are making an assumption of *local asymptotic discreteness.* This concept is well known for root-T consistent estimators, but needs a refinement in order to handle the T and $T^{3/2}$ consistency rates in the cointegrating vectors β. Moreover, the analysis is complicated by the fact that these rates are associated with directions that themselves depend on an unknown parameter, namely $C_{\Gamma,\Pi}\mu$. We therefore formulate the definition of local asymptotic discreteness in a somewhat nonstandard form. A rate-optimal estimator $\hat{\theta}^{(T)} = \left(\hat{\mu}^{(T)}, \hat{\Gamma}^{(T)}, \hat{\alpha}^{(T)} \hat{\beta}^{(T)\prime} \right)$ of $\theta = \left(\mu, \Gamma, \Pi \right)$ is called *locally asymptotically discrete* if it satisfies the following assumption[7].

---

[7]As α and β as well as $\hat{\alpha}^{(T)}$ and $\hat{\beta}^{(T)}$ are not uniquely identified, we implicitly impose that it is possible to select versions of these objects such that the assumption holds.

Assumption 6. *The estimation errors $T^{1/2}\big(\hat{\mu}^{(T)} - \mu\big)$, $T^{1/2}\big(\hat{\Gamma}^{(T)} - \Gamma\big)$, $T^{1/2}\big(\hat{\alpha}^{(T)} - \alpha\big)$, $T\big(\hat{\beta}^{(T)} - \beta\big)$, and $T^{3/2}\big(\hat{\beta}^{(T)} - \beta\big)' C_{\Gamma,\Pi}\mu$ all are $O_P(1)$, under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$ (rate optimality). Moreover, for any $M > 0$, the number of distinct possible values of these estimation errors in balls of radius M, centered at the origin, is eventually bounded as $T \to \infty$.*

Remark 3.1. In standard situations, where all parameters are estimable at rate $T^{1/2}$, any $T^{1/2}$-consistent estimator can be turned into a locally asymptotically discrete one by simply rounding each element to the closest point in the grid $\big\{kT^{-1/2} : k \in \mathbb{Z}\big\}$. Such rounding does not affect $T^{1/2}$-consistency and leads to the desired discreteness. In our model, the same approach can be adopted for the parameters μ, Γ, and α, but not for β. Therefore, we introduce a refined discretization algorithm for β in order to obtain the desired discreteness while preserving the $T^{3/2}$ consistency property of $\hat{\beta}^{(T)'} C_{\Gamma,\Pi}\mu$ (as an estimator of $\beta' C_{\Gamma,\Pi}\mu = 0$).

To be precise, consider an estimator $\hat{\beta}^{(T)}$ satisfying the consistency properties as in the statement of Assumption 6. For example, the reduced-rank regression estimator discussed in Johansen (1995) would do, see his Lemma 13.2. In a first step, round each element of $\hat{\beta}^{(T)}$ to a $T^{-3/2}$-grid, that is, to the closest point in $\big\{kT^{-3/2} : k \in \mathbb{Z}\big\}$. In order not to clutter notation, we still denote this rounded estimator by $\hat{\beta}^{(T)}$ and observe that this first step preserves the T and $T^{3/2}$ consistency properties of the original estimator. It also ensures that $T^{3/2}\big(\hat{\beta}^{(T)} - \beta\big)' C_{\Gamma,\Pi}\mu$, as $T \to \infty$, only has a bounded number of possible values within balls of radius M. However, $T\big(\hat{\beta}^{(T)} - \beta\big)$ in general still will take an unbounded number (of the order of $T^{1/2}$, to be precise) of possible values over such balls. We, therefore, apply a second discretization step, of order T, to each of the columns of $\hat{\beta}^{(T)}$, in such a way that the $T^{3/2}$ consistency of $\hat{\beta}^{(T)'} C_{\Gamma,\Pi}\mu$ is preserved. Choose $j = 1,\ldots,r$ and consider column $\hat{\beta}_j^{(T)}$. We now essentially project $\hat{\beta}_j^{(T)}$ sequentially $(p - r - 1)$ times on the closest of a series of parallel hyperplanes generated by the $(p - r - 1)$ columns of $\beta_\perp$ that are orthogonal to $C_{\Gamma,\Pi}\mu$ and at distance of order T^{-1}. This general idea is complicated by the fact that both $\beta_\perp$ and $C_{\Gamma,\Pi}\mu$ must be estimated, while the grid to be used is not allowed to be random.

Concretely, for this second discretization of $\hat{\beta}_j^{(T)}$, we note that, on the basis of Assumption 6, one can readily construct a $T^{1/2}$-consistent estimator of rank $(p - r - 1)$ for $\beta_\perp - (C_{\Gamma,\Pi}\mu)(C_{\Gamma,\Pi}\mu)'\beta_\perp/|C_{\Gamma,\Pi}\mu|^2$. Choose vectors $\hat{A}_l$, $l = 1,\ldots,\,p-r-1$, that generate the same column space. Subsequently, we discretize each of the elements of $\hat{A}_l$ on a $T^{-1/2}$-grid. These discretized columns are, again for notational convenience, denoted by $\hat{A}_l$. Note that $\hat{A}_l' C_{\Gamma,\Pi}\mu = O_P(T^{-1/2})$, under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$. Now, perform in fact $(p - r - 1)$ discretization steps on $\hat{\beta}_j^{(T)}$ by sequentially projecting on the closest of the hyperplanes $kT^{-1}\hat{A}_l + \big[\hat{A}_l\big]_\perp$, $l = 1,\ldots,\,p-r-1$. The relevant insight is that such projections will not affect the $T^{3/2}$ consistency of $\hat{\beta}_j^{(T)'} C_{\Gamma,\Pi}\mu$, as they affect $\hat{\beta}_j^{(T)}$ by a quantity $\eta\hat{A}_l$ for some $\eta \in (-1/T, 1/T)$ which, still under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, entails $\eta\hat{A}_l' C_{\Gamma,\Pi}\mu = O_P(T^{-3/2})$. Moreover, after all of these $(p - r - 1)$ projections have been carried out, the resulting $T\big(\hat{\beta}^{(T)} - \beta\big)$ (again, for notational simplicity, we keep the same notation for the discretized estimator as for the original one) only takes a bounded number of possible values over balls of radius M, due to the earlier $T^{1/2}$-discretization of $\hat{A}_l$, $l = 1,\ldots,p-r-1$.

It should be insisted, though, that local asymptotic discreteness is needed in formal

asymptotic statements, but has no practical implications. Certainly, one should not bother to discretize estimators in practice: see pages 125 or 188 of Le Cam and Yang (1990) for a discussion on this point.

The proof of the following proposition is given in Appendix B.

Proposition 3.1. *Let Assumptions 1-5 hold, $\vartheta \in \Theta$ and $f \in \mathcal{F}_2$. Consider a sequence of perturbations*

$$h'_T := \Big(m'_T, (\operatorname{vec} G_T)', (\operatorname{vec} A_T)', b'_T, (\operatorname{vec} B_T)', d'_T, (\operatorname{vec} D_T)' \Big), \tag{3.4}$$

with h_T bounded, $\beta_\perp B'_T \perp (C_{\Gamma,\Pi}\mu)$, and $\beta_\perp D'_T \perp (C_{\Gamma,\Pi}\mu)$, which defines a local parameter sequence $\vartheta^{(T)}$ by (2.1)-(2.3). Then (all limits below are taken for $T \to \infty$),

(i) *(asymptotic representation) under $\mathrm{P}^{(T)}_{\vartheta^{(T)};\Sigma,f}$,*

$$\underset{\sim}{S}{}^{(T)}_g(\vartheta^{(T)}; \hat{\Sigma}^{(T)}) = T^{-1/2} \sum_{t=1}^{T} \Big(\frac{t}{T+1} - \frac{1}{2} \Big) U_t(\vartheta^{(T)}; \Sigma) \phi_g \Big(\tilde{G}_p^{-1} \big(\tilde{F}_p \big(\|\epsilon_t(\vartheta^{(T)})\|_\Sigma \big) \big) \Big)$$
$$+ o_P(1) =: \underline{S}{}^{(T)}_g(\vartheta^{(T)}; \Sigma) + o_P(1); \tag{3.5}$$

(ii) *(asymptotic normality under the null) under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, $\underline{S}{}^{(T)}_g(\vartheta; \hat{\Sigma}^{(T)})$ is asymptotically normal, with mean zero and variance $\frac{1}{12p}\mathcal{I}_p(g)I_p$;*

(iii) *(asymptotic normality under alternatives) if also $B_T = 0 = D_T$, $\underline{S}{}^{(T)}_g(\vartheta; \hat{\Sigma}^{(T)})$ is asymptotically normal under $\mathrm{P}^{(T)}_{\vartheta^{(T)};\Sigma,f}$, with mean*

$$\frac{1}{12p}\mathcal{I}_p(f,g)|C_{\Gamma,\Pi}\mu|^2 \Sigma^{-1/2}(\alpha b + \alpha_\perp d), \tag{3.6}$$

where

$$\mathcal{I}_p(f,g) := \int_0^1 \phi_f(\tilde{F}_p^{-1}(u))\phi_g(\tilde{G}_p^{-1}(u))\mathrm{d}u, \tag{3.7}$$

and variance $\frac{1}{12p}\mathcal{I}_p(g)I_p$;

(iv) *(asymptotic linearity) we have, under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$,*

$$\underline{S}{}^{(T)}_g(\vartheta^{(T)}; \hat{\Sigma}^{(T)}) - \underline{S}{}^{(T)}_g(\vartheta; \hat{\Sigma}^{(T)}) = -\mathcal{I}_p(f,g)\frac{1}{pT^{1/2}} \sum_{t=1}^{T} \Big(\frac{t}{T+1} - \frac{1}{2} \Big) v_{Tt} + o_P(1), \tag{3.8}$$

with $v_{Tt} = (v_{Tt,1}, \ldots, v_{Tt,p})'$, where ($e_j$ stands for the jth unit vector in $\mathbb{R}^p$'s canonical basis)

$$v_{Tt,j} := h'_T \begin{pmatrix} Z^{(1)}_{Tt}(\vartheta) \otimes \Sigma^{-1/2}e_j \\ Z^{(2)}_{Tt}(\vartheta) \otimes \alpha'\Sigma^{-1/2}e_j \\ Z^{(2)}_{Tt}(\vartheta) \otimes \alpha'_\perp\Sigma^{-1/2}e_j \end{pmatrix}, \quad j = 1, \ldots, p,$$

and $Z^{(1)}_{Tt}$, $Z^{(2)}_{Tt}$ are defined in (2.9);

23

(v) *the asymptotic linearity property (3.8) remains valid if $\vartheta^{(T)}$ is replaced with a locally asymptotically discrete (in the sense of Assumption 6) random sequence $\hat{\vartheta}^{(T)}$.*

Remark 3.2. Observe that Part (iii) of the above proposition assumes $B_T = 0 = D_T$. The reason is that nonzero B_T perturbations lead to mixed normal limits and nonzero D_T perturbations to limits as in continuous time Ornstein-Uhlenbeck processes. As a result, the limiting distribution of $\underline{S}_g^{(T)}(\vartheta^{(T)}; \hat{\Sigma}^{(T)})$ is no longer a shifted Gaussian distribution. While exact calculations are possible, they do not necessarily provide much insight and, therefore, are omitted. Note that the use of aligned ranks, of course, implies that the estimators used may deviate from the actual data generating process in the B and/or D directions. Thus, the assumptions $B_T = 0$ and $D_T = 0$ are not made in parts (iv) and (v) of the proposition. □

From Proposition 3.1, we see that the rank-based statistic $\underline{S}_g^{(T)}$ is well-behaved even under completely misspecified reference density. Under correctly specified reference density $(g = f)$, the rank-based statistic $\underline{S}_g^{(T)}$ is equivalent to the parametric statistic $\underline{S}_f^{(T)} = S_f^{(T)}$, hence has the same properties as the latter, which justifies its interpretation as a rank-based version of $S_f^{(T)}$. Note that the asymptotic representation result in Part (i) shows that estimating the scatter matrix Σ has no asymptotic impact on $\underline{S}_g^{(T)}(\vartheta^{(T)}; \hat{\Sigma}^{(T)})$, neither under the null nor under contiguous alternatives.

4. Testing the cointegrating rank

In this section, we use the limiting behavior of local experiments in the ECM model (1.1), as obtained in Section 2, along with the asymptotic results on rank-based statistics of Section 3, to construct a locally and asymptotically optimal test about the cointegrating rank r. Since the radial density f remains unspecified, this testing problem is of a semiparametric nature. Nevertheless, we show that the tests we are proposing, which are asymptotically distribution-free under the null hypothesis, achieve *parametric* efficiency in the subexperiments described in Corollary 2.1(iii), under correctly specified reference densities $g = f$.

4.1. Local experiments and efficient central sequences

Formally, we consider the problem of testing the null hypothesis $H : r = r_0$ that the cointegrating rank is r_0 against the alternative $H' : r = r_0 + 1$. This global hypothesis (expressed in terms of the global parameter ϑ) yields, at any ϑ satisfying H, a local form that involves the local parameters, namely, $H : d = 0 = D$ and $H' : d \neq 0$ or $D \neq 0$,

with the parameters m, G, A, b, and B, the scatter matrix Σ, and the radial density f, playing the role of nuisance parameters.

As the local asymptotic behavior of the ECM model with respect to D is of the LABF type, for which no precise optimality results exist, we actually seek optimality, at any ϑ satisfying H, against alternatives of the (local) form $H' : d \neq 0$, $D = 0$, that is, in constrained local experiments of the form considered in Corollary 2.1(iii). Subsequently, we show that adaptivity is achieved with respect to B and Σ. Since, moreover, f is unspecified, we consider constrained local experiments associated with some chosen reference density g. Finally, in order to avoid degenerate Fisher information, we restrict to those values of ϑ such that $C_{\Gamma,\Pi}\mu \neq 0$. While their optimality properties are restricted to those subexperiments, restricted ϑ values and chosen radial density g, we stress that the validity of our tests extends to arbitrary ϑ and f (since they are rank-based).

Building on standard projection results, the following lemma provides the parametrically efficient central sequence for testing $H : r = r_0$ in the LAN situation of Corollary 2.1(iii).

Lemma 4.1. *Let $\vartheta \in \Theta$ be such that $r = r_0$, $C_{\Gamma,\Pi}\mu \neq 0$, $\Sigma > 0$, and $f \in \mathcal{F}_2$. Then,*

(i) *in the Gaussian shift limit experiment (with $B = 0$ and $D = 0$) considered in Corollary 2.1(iii), the efficient score for d, when the local parameters m, G, A, and b are treated as nuisances, is*

$$\Delta_d^{\star} = |C_{\Gamma,\Pi}\mu|^2 \int_{u=0}^{1} \left(u - \frac{1}{2}\right) d\left[(I_p - P_\alpha)\alpha_\perp\right]' W_\phi(u), \qquad (4.1)$$

where P_α denotes the (non-orthogonal) projection matrix $P_\alpha = \alpha \left(\alpha'\Sigma^{-1}\alpha\right)^{-1} \alpha'\Sigma^{-1}$, with the convention that $P_\alpha = 0_{p \times p}$ (and $\alpha_\perp$ the identity) in case $r_0 = 0$;

(ii) *for the corresponding local sequence of cointegration models, under Assumptions 1–3, a version of that efficient central sequence is*

$$\begin{aligned}
\Delta_d^{(T)\star} &= |C_{\Gamma,\Pi}\mu|^2 \int_{u=0}^{1} \left(\mathrm{id}_T(u) - \frac{1}{2}\right) d\left[(I_p - P_\alpha)\alpha_\perp\right]' W_\phi^{(T)}(u) \\
&= |C_{\Gamma,\Pi}\mu|^2 \left[(I_p - P_\alpha)\alpha_\perp\right]' \Sigma^{-1/2} S_f^{(T)}.
\end{aligned} \qquad (4.2)$$

Proof. The efficient score for d is obtained as the residual of the regression, in the covariance structure $J_{\mu,\Gamma,\alpha,b,d}$, of the score for d on that for the nuisances m, G, A, and b. Let us show that this residual is indeed $\Delta_d^{\star}$ given in (4.1). First, observe

$$\Delta_d - \Delta_d^{\star} = |C_{\Gamma,\Pi}\mu|^2 \int_{u=0}^{1} u\, d\left[\alpha_\perp - \left[(I_p - P_\alpha)\alpha_\perp\right]\right]' W_\phi(u)$$

25

$$+ |C_{\Gamma,\Pi}\mu|^2 \frac{1}{2} [(I_p - P_\alpha)\,\alpha_\perp]'\,W_\phi(1).$$

The components of the first term are in the space spanned by the components of the score for b, the components of the second term in the space spanned by the components of the score for m. As $\int_{u=0}^{1}(u - 1/2)\,\mathrm{d}u = 0$, (4.1) moreover is orthogonal to $W_{\Delta X \otimes \phi}(1)$ and $W_{Y \otimes \phi}(1)$, thus to the scores induced by m, G, and A. Finally, (4.1) is also orthogonal to the scores induced by b, as $[(I_p - P_\alpha)\,\alpha_\perp]'\,W_\phi(u)$ and $\alpha' W_\phi(u)$ are independent (their covariance vanishes due to the fact that $[(I_p - P_\alpha)\,\alpha_\perp]'\,\Sigma^{-1}\alpha = 0$). In case $r_0 = 0$, the result follows along the same lines for the scores induced by m and G, while those induced by A and b need not be considered. This establishes Part (i) of the lemma; Part (ii) is an immediate consequence. $\qquad\square$

Remark 4.1. From the above proof, we see that, in order to get the efficient central sequence (4.2), it is sufficient to regress the central sequence for d with respect to those for m and b. In other words, the efficient central sequence (4.2) for d when m, G, A, and b are nuisance parameters is the same as if only m and b were nuisances and G and A were known. In this respect, the model is *adaptive* to perturbations of the Γ_j's and α.

Similarly, the efficient central sequence (4.2) is also orthogonal to the score for B as, again, $[(I_p - P_\alpha)\,\alpha_\perp]'\,W_\phi(u)$ and $\alpha' W_\phi(u)$ are independent. This suggest that the model is also adaptive with respect to the local parameter B, a claim which is indeed verified by Proposition 4.2 below. This holds despite the fact that, as indicated in Corollary 2.1, the model is not jointly LAMN with respect to m, G, A, b, B, and d.

4.2. Optimal rank-based tests

Lemma 4.1 is the basis for the construction of locally and asymptotically optimal rank-based tests about the cointegration rank.

Proposition 4.2. *Let Assumptions 1-5 hold, $\vartheta \in \Theta$ be such that $r = r_0$, $\Sigma > 0$, and $f \in \mathcal{F}_2$. Consider the rank-based statistic $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ in (3.2) and the quadratic form*

$$\underline{Q}_g^{(T)}(\vartheta, \Sigma) := \frac{12p}{\mathcal{I}_p(g)} \underline{S}_g^{(T)}(\vartheta; \Sigma)' \Sigma^{-1/2} \Big[(I_p - P_\alpha)\,\alpha_\perp\Big]\Big[\alpha_\perp'\Big(\Sigma^{-1} - \Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}\Big)\alpha_\perp\Big]^{-1}$$
$$\times \Big[(I_p - P_\alpha)\,\alpha_\perp\Big]'\Sigma^{-1/2}\underline{S}_g^{(T)}(\vartheta; \Sigma), \qquad (4.3)$$

with $\underline{Q}_g^{(T)}(\vartheta, \Sigma) = (12p/\mathcal{I}_p(g))\left|\underline{S}_g^{(T)}(\vartheta; \Sigma)\right|^2$ in case $r_0 = 0$. Then,

(i) *$\underline{Q}_g^{(T)}(\vartheta, \Sigma)$ under $\mathrm{P}_{\vartheta;\Sigma;f}^{(T)}$ has a limiting $\chi_{p-r_0}^2$ distribution;*

(ii) *under $\mathrm{P}_{\vartheta^{(T)};\Sigma;f}$, where $\vartheta^{(T)}$ is as in (2.1)–(2.3) with $D = 0$, $\underline{Q}_g^{(T)}(\vartheta, \Sigma)$ has a limiting non-central $\chi_{p-r_0}^2$ distribution with noncentrality parameter*

$$\frac{\mathcal{I}_p(f,g)^2}{12p\mathcal{I}_p(g)} |C_{\Gamma,\Pi}\mu|^4\, d'\alpha_\perp'\Sigma^{-1}\Big[(I_p - P_\alpha)\,\alpha_\perp\Big]\Big[\alpha_\perp'\Big(\Sigma^{-1} - \Sigma^{-1}\alpha\,(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}\Big)\alpha_\perp\Big]^{-1}$$
$$\times \Big[(I_p - P_\alpha)\,\alpha_\perp\Big]'\Sigma^{-1}\alpha_\perp d, \qquad (4.4)$$

for $r_0 > 0$ and $\left(\mathcal{I}_p(f,g)^2/12p\mathcal{I}_p(g)\right)|C_{\Gamma,\Pi}\mu|^4\,\|d\|_\Sigma^2$ in case $r_0 = 0$;

26

(iii) the limiting distributions above remain valid if the statistic $Q_g^{(T)}(\vartheta, \Sigma)$ is computed on the basis of estimators $\hat{\Sigma}^{(T)}$ and $\hat{\vartheta}^{(T)}$ satisfying Assumptions 4 and 6, respectively, and the constraint that $r = r_0$;

(iv) for $C_{\Gamma, \Pi}\mu \neq 0$ and $f = g$, a locally and asymptotically z-level most stringent test of $H : d = 0 = D$ against $H' : d \neq 0, D = 0$ consists in rejecting H whenever $Q_g^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ exceeds the $(1 - z)$-quantile of a chi-square distribution with $p - r_0$ degrees of freedom. That test is parametrically efficient under radial density $f = g$.

Proof. Part (i) follows immediately from Proposition 3.1(ii). Part (ii) follows from an application of Le Cam's third lemma as in Proposition 3.1(iii). Note that, without loss of generality, we may assume $B = 0$, as the score associated with B depends on $\alpha' W_\phi$, which is independent of $[(I_p - P_\alpha)\alpha_\perp]' W_\phi$. Thus, applying Le Cam's third lemma as in Theorem 6.6 in Van der Vaart (2000) and using Girsanov's theorem, then leads to the same distribution under the null as under local alternatives generated by B since the sharp bracket of the score for B and $[(I_p - P_\alpha)\alpha_\perp]' W_\phi$ vanishes. As a result, we obtain, for the asymptotic mean, under $P_{\vartheta^{(T)}; \Sigma; f}$, of $[(I_p - P_\alpha)\alpha_\perp]' \Sigma^{-1/2} S_g^{(T)}(\vartheta; \Sigma)$,

$$\frac{1}{12p}\mathcal{I}_p(f, g)|C_{\Gamma, \Pi}\mu|^2 [(I_p - P_\alpha)\alpha_\perp]' \Sigma^{-1}(\alpha b + \alpha_\perp d)$$

$$= \frac{1}{12p}\mathcal{I}_p(f, g)|C_{\Gamma, \Pi}\mu|^2 [(I_p - P_\alpha)\alpha_\perp]' \Sigma^{-1}\alpha_\perp d.$$

Part (iii) is a consequence of Proposition 3.1(iii)-(v), due to the premultiplication of $S_g^{(T)}(\vartheta; \Sigma)$ by $[(I_p - P_\alpha)\alpha_\perp]' \Sigma^{-1/2}$. More precisely, the shift in (3.8) vanishes with respect to the local parameter B as $[(I_p - P_\alpha)\alpha_\perp]' \Sigma^{-1}\alpha = 0$. Shifts due to d and D are not possible, as a cointegrating rank r_0 is imposed as a constraint in the construction of the estimator. Concerning the remaining local parameters m, G, A, and b, premultiplying the shift (3.6) by $[(I_p - P_\alpha)\alpha_\perp]' \Sigma^{-1/2}$ yields zero as well. Finally, Part (iv) follows from the standard construction of (conditionally) most stringent tests in LAN experiments with nuisance parameters using Lemma 4.1 and the observation that, in view of Part (iii), equally stringent tests can be constructed when Σ and B are unknown. $\square$

Remark 4.2. If the reduced rank regression method is used, as in Johansen (1995), to estimate ϑ under the null, the rank-based test statistics $Q_g^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ are invariant with respect to non-singular linear transformations of the data.

The rank-based test statistic $Q_g^{(T)}(\vartheta, \Sigma)$ in Proposition 4.2 has several quite desirable properties. First of all, it is exactly distribution-free under the null, due to its dependence on the innovation ranks only. In particular, note that any scalar factor in the innovation scatter matrix Σ would cancel out of the statistic. Also, the statistic does not depend on the versions chosen for α and β or $\alpha_\perp$ and $\beta_\perp$. As a result, this statistic has a constant size over the null hypothesis. This result carries over to the asymptotic size when considering the aligned-rank version $Q_g^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$. Note that the actual radial density f needs

not be estimated for this.

The local and asymptotic power of the test, determined by the noncentrality parameter (4.4), does depend on the actual underlying density f, the scatter matrix Σ, the drift $C_{\Gamma,\Pi}\mu$ and the column spaces of α and $\alpha_\perp$. Again, this asymptotic power result is not affected by the estimation of ϑ and Σ (under Assumptions 4 and 6). The power of the test gets larger as the reference density g gets closer, as measured by $\mathcal{I}_p(f,g)$, to the actual density f. Also, the power increases with larger values of $|C_{\Gamma,\Pi}\mu|$. For $C_{\Gamma,\Pi}\mu = 0$, the test has asymptotically no power at the rate $T^{3/2}$ in the local alternatives we consider—but note that no test ever would. However, it may very well have power against alternatives at rate T.

Part (iv) of Proposition 4.2 asserts that, for well-chosen reference density, the rank-based test achieves the parametric lower bound in case $C_{\Gamma,\Pi}\mu \neq 0$. In that case, the limiting experiment (still, with $B = 0 = D$) is LAN, so that the concept of efficiency is well defined. The notion that the power of the rank test increases as g gets closer to f suggests the use of a pre-estimated density $\hat{f}$ instead of the fixed reference density g. This idea has been pursued in other contexts, see, e.g., Hallin and Werker (2003), and is equally applicable in the present setting. Incidentally this shows, as was to be expected, that the inference problem is adaptive with respect to f as well.

Quite remarkably, the dependence on f of local powers (the noncentrality parameters (4.4)) is entirely characterized by the scalar *cross-information quantity* $\mathcal{I}_p(f,g)$ defined in (3.7). Those cross-information quantities are exactly the same as in the location problems considered in Hallin and Paindaveine (2002a). As a result, the Chernoff-Savage property established in their Proposition 6 also holds here for the normal-score or van der Waerden version of the test described in Proposition 3.1. The latter relies on the quadratic form $Q_\phi^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ based on $\underset{\sim}{S}_\phi^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$, where

$$\underset{\sim}{S}_\phi^{(T)}(\vartheta; \Sigma) := T^{-1/2} \sum_{t=1}^{T} \left(\frac{t}{T+1} - \frac{1}{2} \right) U_t(\vartheta; \Sigma) \left(F_{\chi_p^2}^{-1}\left(\frac{R_t^{(T)}(\vartheta; \Sigma)}{T+1} \right) \right)^{1/2} \qquad (4.5)$$

(here, $F_{\chi_p^2}$ stands for the chi-square distribution function with p degrees of freedom). This means that the asymptotic relative efficiencies under radial density f (ARE_f), with respect to the pseudo- or quasi-Gaussian methods (based on Proposition 4.3 below), of the van der Waerden rank-based tests (based on (4.5)), are always larger than or equal

28

to one, irrespective of the actual unknown radial density f; equality is achieved in the Gaussian case only, that is, for Gaussian $\mathfrak{f}$, meaning a root chi-square (p degrees of freedom) radial density f.

4.3. Optimal Gaussian tests

Although we throughout avoid Gaussian assumptions and emphasize the nonparametric nature of the problem, Gaussian procedures—more precisely, the pseudo- or quasi-Gaussian ones—remain a classical benchmark. Particularizing Lemma 4.1 to the case of a Gaussian f immediately yields a parametric Gaussian version of Proposition 4.2, with an optimal Gaussian test of the Lagrange multiplier type. The following proposition provides a description of that test and summarizes its properties. Proofs are a simplified version of those of the previous section (no f-dependent asymptotic representation of the test statistic is needed here), and details are left to the reader.

Proposition 4.3. *Let $\vartheta \in \Theta$ be such that $r = r_0$, $C_{\Gamma,\Pi}\mu \neq 0$, $\Sigma > 0$, and assume that Assumption 3 holds.*

(i) Consider the Gaussian case $\mathfrak{f}(e) = (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp(-\frac{1}{2}\|e\|_{\Sigma}^2)$, that is,

$$f(z) = 2^{1-p/2}\Gamma^{-1}(p/2)\exp(-z^2/2) \quad and \quad \phi_f(z) = z. \tag{4.6}$$

Then, the efficient central sequence (4.2) of Lemma 4.1(ii) becomes

$$\Delta_{d\dagger}^{(T)\star} = |C_{\Gamma,\Pi}\mu|^2 \left[(I_p - P_\alpha)\,\alpha_\perp\right]' \Sigma^{-1/2} S_\dagger^{(T)}, \tag{4.7}$$

with

$$S_\dagger^{(T)} := S_\dagger^{(T)}(\vartheta, \Sigma) := T^{-1/2}\sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2}\right)\Sigma^{-1/2}\epsilon_t(\vartheta).$$

Consider the quadratic form

$$Q_\dagger^{(T)}(\vartheta, \Sigma) := 12 S_\dagger^{(T)}(\vartheta, \Sigma)' \Sigma^{-1/2} \left[(I_p - P_\alpha)\,\alpha_\perp\right]\left[\alpha_\perp'\left(\Sigma^{-1} - \Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}\right)\alpha_\perp\right]^{-1}$$
$$\times \left[(I_p - P_\alpha)\,\alpha_\perp\right]' \Sigma^{-1/2} S_\dagger^{(T)}(\vartheta, \Sigma), \tag{4.8}$$

with $Q_\dagger^{(T)}(\vartheta, \Sigma) = 12\left|S_\dagger^{(T)}(\vartheta, \Sigma)\right|^2$ in case $r_0 = 0$. Then,

(ii) for any $f \in \mathcal{F}_2$, $Q_\dagger^{(T)}(\vartheta, \Sigma)$ under $\mathrm{P}_{\vartheta;\Sigma;f}^{(T)}$ has an asymptotic $\chi_{p-r_0}^2$ distribution (this distribution is exact for f given in (4.6));

(iii) under $\mathrm{P}_{\vartheta^{(T)};\Sigma;f}$, where $\vartheta^{(T)}$ is as in (2.1)–(2.3) with $D = 0$ and $f \in \mathcal{F}_2$, $Q_\dagger^{(T)}(\vartheta, \Sigma)$ has a limiting non-central $\chi_{p-r_0}^2$ distribution with noncentrality parameter

$$\frac{\mathcal{I}_p^2(f,g)}{12p^2}|C_{\Gamma,\Pi}\mu|^4\, d'\alpha_\perp'\Sigma^{-1}\left[(I_p - P_\alpha)\,\alpha_\perp\right]\left[\alpha_\perp'\left(\Sigma^{-1} - \Sigma^{-1}\alpha\left(\alpha'\Sigma^{-1}\alpha\right)^{-1}\alpha'\Sigma^{-1}\right)\alpha_\perp\right]^{-1}$$

$$\times \left[\left(I_p - P_\alpha \right) \alpha_\perp \right]' \Sigma^{-1} \alpha_\perp d, \qquad (4.9)$$

for $r_0 > 0$, and $\left(\mathcal{I}_p^2(f, g)/12p^2 \right) |C_{\Gamma,\Pi}\mu|^4 \, \|d\|_\Sigma^2$ in case $r_0 = 0$;

(iv) the limiting $\chi_{p-r_0}^2$ and non-central $\chi_{p-r_0}^2$ distributions in (ii) and (iii) remain asymptotically valid for $Q_\dagger^{(T)}(\hat\vartheta^{(T)}, \hat\Sigma^{(T)})$, where the estimators $\hat\Sigma^{(T)}$ and $\hat\vartheta^{(T)}$ satisfy Assumptions 4 and 6, respectively, and the constraint that $r = r_0$;

(v) when $C_{\Gamma,\Pi}\mu \neq 0$, the Gaussian test rejecting H whenever $Q_\dagger^{(T)}(\hat\vartheta^{(T)}, \hat\Sigma^{(T)})$ exceeds the $(1 - z)$-quantile of a chi-square distribution with $p - r_0$ degrees of freedom is locally and asymptotically z-level most stringent, and reaches parametric efficiency, against alternatives of the local form $H' : d \neq 0, \ D = 0$ with Gaussian radial density f (4.6).

The tests described in Part (v) of Proposition 4.3 are Lagrange multiplier counterparts of Johansen's likelihood ratio based tests (Johansen (1991), Johansen (1995)); both qualify as pseudo-Gaussian tests, as their validity extends to all elliptical radial densities g with finite second-order moments.

5. Monte Carlo study

This section reports the results of a small Monte Carlo study to corroborate our asymptotic analysis and to assess the finite-sample performances of the pseudo-Gaussian statistic and the proposed rank-based statistics. We compare their performances to those of Johansen's LR-based tests, i.e., the *maxeig* (maximum eigenvalue) and the *trace* test. These LR-based tests are often used in empirical work and, as documented by Hubrich, Lütkepohl, and Saikkonen (2001), have comparatively good performances. The study is implemented in MATLAB 7.14 and the code is available upon request. We use reduced-rank regression[8] to determine the residuals and to estimate α. The pseudo-Gaussian test uses the empirical covariance matrix of the residuals to estimate Σ and the rank-based tests use Tyler's estimator of scatter[9]. Throughout we will use a 5% level. For the pseudo-Gaussian statistic and the rank-based tests we use the critical values implied by their asymptotic χ^2-distributions and for the *maxeig* and *trace* test we use Matlab's tabulated values (computed using methods described in MacKinnon, Haug, and Michelis (1999)).

---

[8]We use MATLAB's Econometrics Toolbox to compute the reduced-rank regression estimators as well as the *maxeig* and *trace* statistics.

[9]The iterative scheme of Randles (2000, p.1267) is used to compute Tyler's estimator of scatter (using the Frobenius norm and 10^{-10} as tolerance).

The DGPs we consider are based on Hubrich, Lütkepohl, and Saikkonen (2001) and Toda (1994). These DGPs are described by

$$\Delta X_t = \begin{pmatrix} 0_{(p-1)\times 1} \\ 1 \end{pmatrix} + \begin{pmatrix} \phi I_r & 0_{r\times(p-r)} \\ 0_{(p-r)\times r} & 0_{(p-r)\times(p-r)} \end{pmatrix} X_{t-1} + \varepsilon_t, \tag{5.1}$$

where ϕ is a scalar. Note that $\alpha' = \begin{pmatrix} \phi I_r & 0_{r\times(p-r)} \end{pmatrix}$ and $\beta' = \begin{pmatrix} I_r & 0_{r\times(p-r)} \end{pmatrix}$. Also note that, for $\phi \neq 0$, the cointegrating rank of $\{X_t\}$ equals r. These DGPs can be considered as 'canonical forms', since interesting models can be obtained by nonsingular linear transformations of the data; see Hubrich, Lütkepohl, and Saikkonen (2001) and Toda (1994) for details. Restricting to the canonical forms is legitimate because the pseudo-Gaussian and the rank-based tests we consider, as well as the LR-based *maxeig* and *trace* test, are all invariant with respect to such transformations.

In the simulation study we consider $p \in \{2, 3, 5\}$ and $\phi = -.3$. As innovation distributions we consider $\mathcal{N}(0, \Sigma_p)$, $t_3(\Sigma_p)$ and $t_{10}(\Sigma_p)$, where $\Sigma_p \in \{I_p, \Sigma_{p,c}\}$. The matrices $\Sigma_{p,c}$, yielding correlation between the components of ε_t, are given by

$$\Sigma_{2,c} = \begin{pmatrix} 1 & 0.8 \\ & 1 \end{pmatrix}, \quad \Sigma_{3,c} = \begin{pmatrix} 1 & 0.4 & 0.8 \\ & 1 & 0 \\ & & 1 \end{pmatrix}, \quad \Sigma_{5,c} = \begin{pmatrix} 1 & 0.4 & 0.8 & 0 & 0 \\ & 1 & 0 & 0 & 0 \\ & & 1 & 0 & 0 \\ & & & 1 & 0 \\ & & & & 1 \end{pmatrix}.$$

The choices for $p = 2$ and $p = 3$ are taken from Hubrich, Lütkepohl, and Saikkonen (2001) ($p = 5$ was not considered in that study). Also in line with the setup in Hubrich, Lütkepohl, and Saikkonen (2001), we use $X_0 = 0_{p\times 1}$ as initial values and use 50 periods to 'warm up' the process (so we use $X_{51}, \ldots, X_{50+T}$ as observations). As sample sizes we consider $T \in \{100, 250, 500\}$. To conserve space we only present the results for $p = 5$ with $\Sigma_5 = \Sigma_{5,c}$ in the main text; the results for $p = 2, 3$ and $p = 5$ with $\Sigma_5 = I_5$ are organized in Appendix D.

To illustrate the convergence to a χ^2-distribution under the null hypothesis, Figure 1 presents a scaled histogram of simulated values of the rank-based tests $Q_g^{(T)}$, $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 5$, $r_0 = 0$, $\Sigma_5 = I_5$, and $T = 100$ along with their limiting χ_5^2-density. Table 1 presents the size of the pseudo-Gaussian test $Q_{\dagger}^{(T)}$,
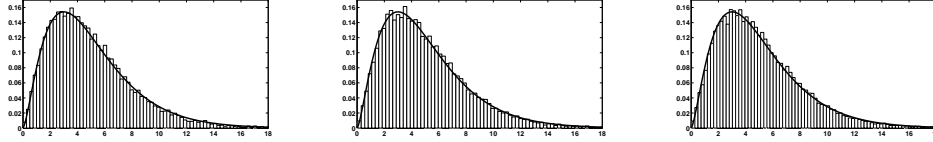
31

Figure 1: Simulated (25,000 replications) finite-sample, $T = 100$, distributions of the rank-tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for $p = 5$, $r_0 = 0$, $\Sigma_5 = I_5$ compared to their limiting χ_5^2-distribution.

the sizes of the rank-based tests $Q_g^{(T)}$, $g \in \{\phi, t_3, t_{10}\}$, as well as the sizes of the *maxeig* and *trace* tests under (5.1) for $p = 5$, $r_0 \in \{0, \ldots, p-1\}$ and $\Sigma_5 = \Sigma_{5,c}$. The sizes for $\Sigma_5 = I_5$ are reported in Table 4 in Appendix D. The size of the rank-based statistics is indeed more stable than for the classical *maxeig* and *trace* statistics. For smaller sample size ($T = 100$), the rank-based statistics are somewhat undersized for larger r_0, while the classical counterparts are somewhat oversized for small r_0. For larger sample sizes, all statistics are well sized.

To assess the finite-sample powers of the tests, we consider alternatives of the form[10]

$$\Pi_h^{(T)} = \Pi + T^{-3/2} \alpha_\perp d (C_{\Gamma, \Pi} \mu)', \tag{5.2}$$

with $d = -h1_{(p-r) \times 1}$ in (2.2). From Proposition 4.2 we know that, under these local alternatives, the rank tests asymptotically follow a non-central χ^2-distribution. As an illustration of this weak convergence result, Figure 2 compares the finite-sample distribution of the rank-based test $Q_\phi^{(T)}$ under (5.2) with $h = 3$ for $p = 5$, $r_0 = 0$, $f = \mathcal{N}(0, I_5)$, and $T = 500$, to the theoretically appropriate non-central χ^2-distribution. For completeness, we also plot the limiting null distribution once more. Again, we conclude that our asymptotic analysis provides a close approximation to these finite sample experiments.

Figures 3-7 present the power curves of the various tests for $p = 5$ and $r_0 = 0, \ldots, 4$. Each figure shows the power, as function of h in (5.2), for the sample sizes $T \in \{100, 250, 500\}$ and the innovation distributions $f \in \{\mathcal{N}(0, \Sigma_{5_c}), t_3(\Sigma_{5_c}), t_{10}(\Sigma_{5_c})\}$. The results for the case $\Sigma_5 = I_5$ are presented in Figures 18-22 in Appendix D. As mentioned before this appendix also contains the simulation results for $p = 2, 3$.

First of all we note that the simulation results corroborate the asymptotic result that the power of the rank-based tests is maximal in case $f = g$. The figures also show that

---

[10]We use the version $\alpha_\perp' = \begin{pmatrix} 0_{(p-r) \times r} & I_{(p-r) \times (p-r)} \end{pmatrix}$.

Table 1: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, pseudo-Gaussian test $Q_\dagger^{(T)}$ and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 5$, $r_0 \in \{0, \ldots, 4\}$, $\phi = -0.3$, $f \in \{\mathcal{N}(0, \Sigma_{5,c}), t_3(\Sigma_{5,c}), t_{10}(\Sigma_{5,c})\}$. Clearly, for $r_0 = p - 1$ the *maxeig* and *trace* tests are identical.

| Test | Sample size and innovation distribution | | | | | | | | |
| | $T = 100$ | | | $T = 250$ | | | $T = 500$ | | |
	$\mathcal{N}(0, \Sigma_{5,c})$	$t_3(\Sigma_{5,c})$	$t_{10}(\Sigma_{5,c})$	$\mathcal{N}(0, \Sigma_{5,c})$	$t_3(\Sigma_{5,c})$	$t_{10}(\Sigma_{5,c})$	$\mathcal{N}(0, \Sigma_{5,c})$	$t_3(\Sigma_{5,c})$	$t_{10}(\Sigma_{5,c})$
$r_0 = 0$									
maxeig	0.064	0.083	0.065	0.053	0.065	0.055	0.049	0.061	0.054
trace	0.071	0.082	0.072	0.058	0.063	0.058	0.054	0.059	0.055
$Q_\dagger^{(T)}$	0.044	0.038	0.043	0.047	0.047	0.047	0.048	0.047	0.048
$Q_\phi^{(T)}$	0.044	0.043	0.043	0.047	0.049	0.046	0.047	0.048	0.048
$Q_{t_3}^{(T)}$	0.045	0.042	0.046	0.050	0.047	0.047	0.048	0.048	0.049
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.048	0.047	0.047	0.047	0.049
$r_0 = 1$									
maxeig	0.072	0.083	0.077	0.057	0.066	0.062	0.054	0.061	0.055
trace	0.082	0.091	0.084	0.061	0.070	0.065	0.054	0.060	0.057
$Q_\dagger^{(T)}$	0.035	0.033	0.033	0.042	0.041	0.043	0.046	0.046	0.046
$Q_\phi^{(T)}$	0.036	0.041	0.034	0.042	0.046	0.044	0.044	0.048	0.047
$Q_{t_3}^{(T)}$	0.039	0.049	0.040	0.045	0.049	0.045	0.047	0.050	0.047
$Q_{t_{10}}^{(T)}$	0.037	0.045	0.037	0.043	0.049	0.045	0.045	0.049	0.048
$r_0 = 2$									
maxeig	0.029	0.040	0.031	0.053	0.060	0.054	0.051	0.056	0.052
trace	0.036	0.047	0.039	0.053	0.061	0.056	0.052	0.054	0.053
$Q_\dagger^{(T)}$	0.022	0.021	0.022	0.038	0.039	0.041	0.044	0.044	0.045
$Q_\phi^{(T)}$	0.021	0.028	0.021	0.038	0.044	0.040	0.043	0.047	0.044
$Q_{t_3}^{(T)}$	0.030	0.044	0.034	0.041	0.051	0.043	0.045	0.051	0.046
$Q_{t_{10}}^{(T)}$	0.024	0.034	0.026	0.039	0.048	0.043	0.043	0.050	0.045
$r_0 = 3$									
maxeig	0.020	0.024	0.019	0.048	0.054	0.047	0.050	0.053	0.049
trace	0.021	0.027	0.021	0.049	0.054	0.048	0.049	0.054	0.049
$Q_\dagger^{(T)}$	0.017	0.018	0.018	0.039	0.038	0.038	0.042	0.045	0.046
$Q_\phi^{(T)}$	0.016	0.022	0.017	0.038	0.041	0.039	0.042	0.046	0.045
$Q_{t_3}^{(T)}$	0.025	0.036	0.027	0.041	0.048	0.042	0.046	0.049	0.045
$Q_{t_{10}}^{(T)}$	0.019	0.026	0.020	0.039	0.043	0.039	0.044	0.047	0.045
$r_0 = 4$									
maxeig	0.040	0.047	0.038	0.044	0.048	0.045	0.046	0.049	0.050
trace	0.040	0.047	0.038	0.044	0.048	0.045	0.046	0.049	0.050
$Q_\dagger^{(T)}$	0.026	0.025	0.024	0.040	0.040	0.040	0.044	0.046	0.047
$Q_\phi^{(T)}$	0.024	0.026	0.023	0.039	0.042	0.039	0.043	0.047	0.047
$Q_{t_3}^{(T)}$	0.031	0.039	0.032	0.041	0.046	0.044	0.045	0.048	0.048
$Q_{t_{10}}^{(T)}$	0.026	0.036	0.026	0.038	0.044	0.042	0.045	0.048	0.047
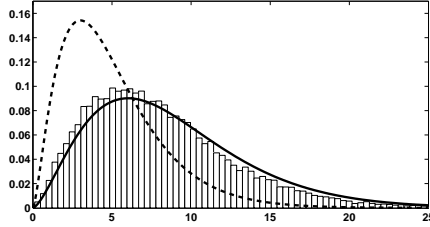
Figure 2: Simulated (25,000 replications) finite-sample, $T = 500$, distribution of the Gaussian rank-based test (4.3), i.e., $g = \phi$, under the alternative $\Pi_3^{(T)}$ for $p = 5$, $r_0 = 0$ and $f = \mathcal{N}(0, I_5)$, compared to its limiting distribution, χ_5^2, under the null hypothesis (dashed line) and its limiting distribution, non-central χ_5^2 with non-centrality parameter 15/4, under the alternative (solid line).

the Gaussian rank-based test dominates the pseudo-Gaussian test, except for Gaussian innovations for which they are equivalent. This is in line with the asymptotic Chernoff-Savage result of Section 4.2. This motivates once more our interest in rank-based tests. If we compare the performances of the pseudo-Gaussian test and the rank-based tests to the performances of the *maxeig* and *trace* test, we can make the following observations. For $T = 250, 500$ the pseudo-Gaussian test and the rank-based tests yield substantial improvements to the *maxeig* test and the *trace* test, while for $T = 100$ there is no uniform conclusion as the *maxeig* test and *trace* test tend to perform better for large values of h. For $r_0 = p - 1$ the pseudo-Gaussian test and rank-based tests do not perform very well for small sample sizes. Summarizing, we conclude that no uniformly valid conclusions are possible and that the rank-based tests, which are easy to compute, nicely complement the *maxeig* and *trace* test and thus are a useful additional toolkit to the econometrician.
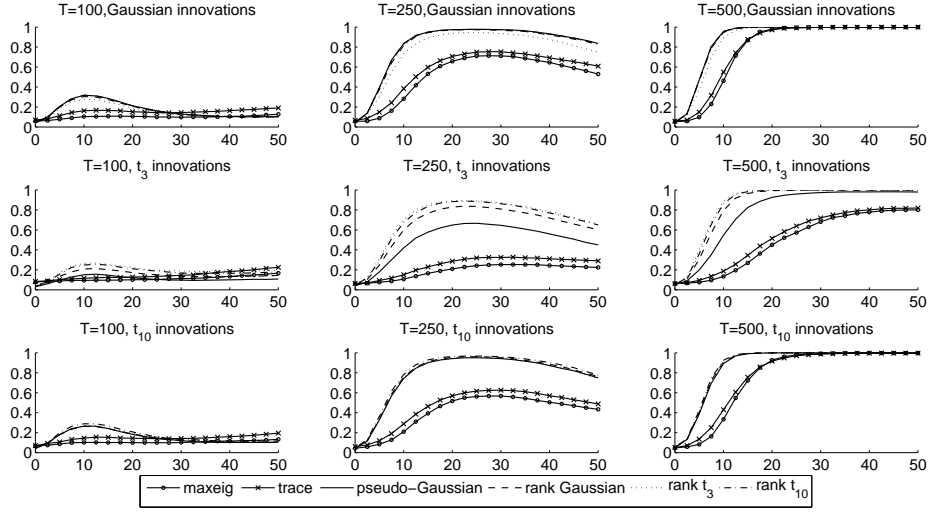
Figure 3: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing <u>$H : r = 0$ versus $H' : r = 1$</u> under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{5,c}), t_3(\Sigma_{5,c}), t_{10}(\Sigma_{5,c})\}$.
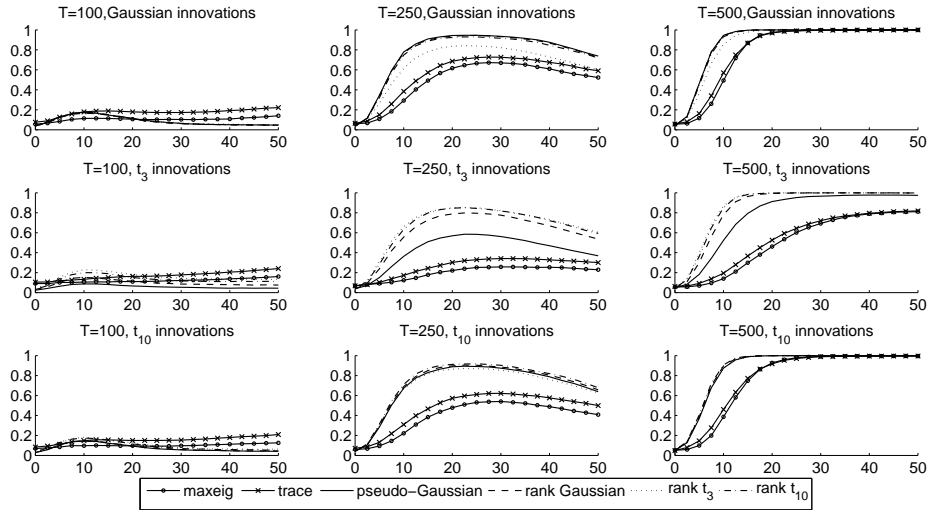

Figure 4: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing <u>$H : r = 1$ versus $H' : r = 2$</u> under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{5,c}), t_3(\Sigma_{5,c}), t_{10}(\Sigma_{5,c})\}$.

35

Figure 5: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 2 \text{ versus } H' : r = 3}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{5,c}), t_3(\Sigma_{5,c}), t_{10}(\Sigma_{5,c})\}$.
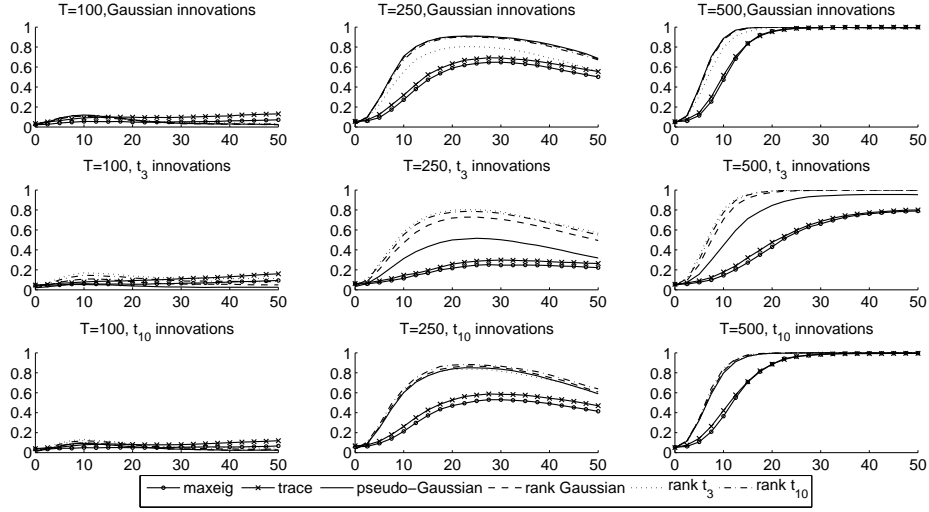

Figure 6: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 3 \text{ versus } H' : r = 4}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{5,c}), t_3(\Sigma_{5,c}), t_{10}(\Sigma_{5,c})\}$.
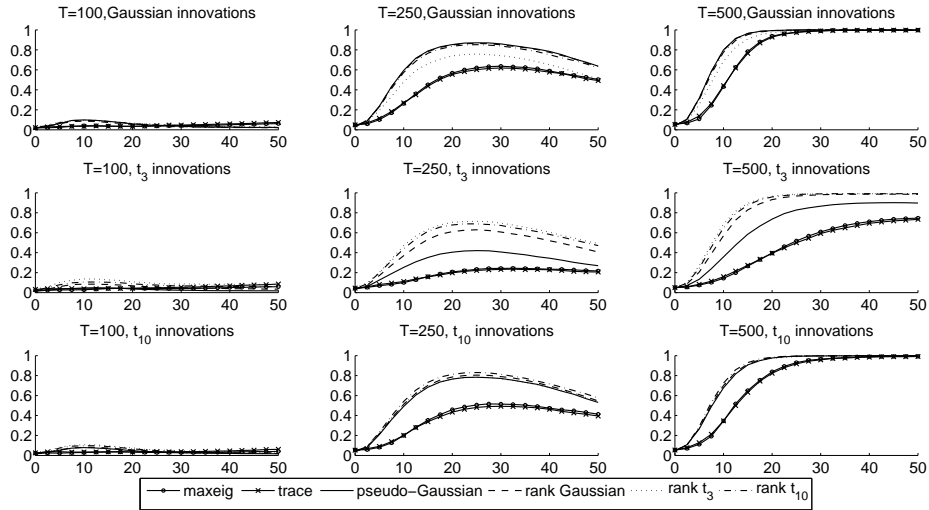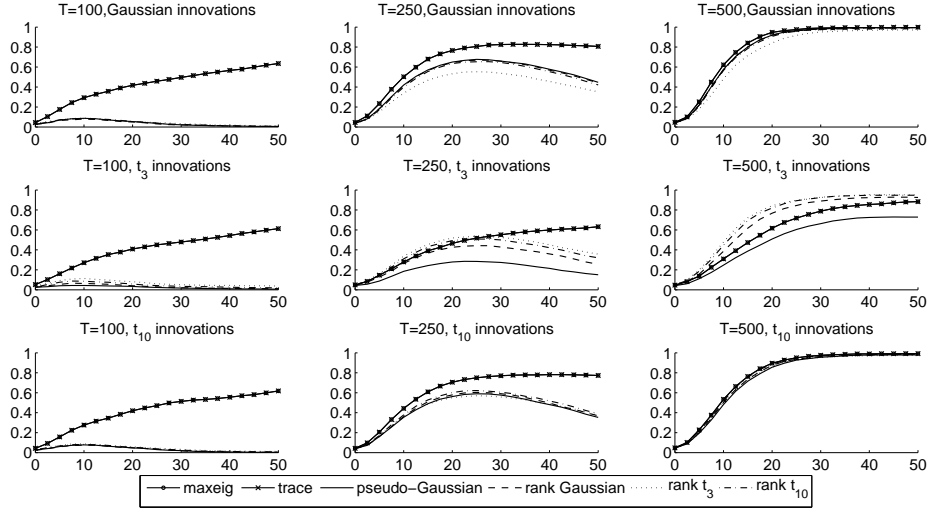
Figure 7: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 4 \text{ versus } H' : r = 5}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{5,c}), t_3(\Sigma_{5,c}), t_{10}(\Sigma_{5,c})\}$.

6. Discussion and Conclusion

We analyze asymptotically optimal testing of the cointegrating rank in Error Correction Models. We show that, in applications where non-zero trends could possibly be expected, standard Locally Asymptotically Normal behavior can be exploited. We propose rank-based versions of the optimal tests to exploit their advantages, in particular, their very stable size and their QMLE property that extends beyond Gaussian reference densities.

By fully characterizing the possible limiting experiments that arise in the Error Correction Model, interesting avenues for future research are discovered. In particular, with respect to the cointegrating vectors, the model is only Locally Asymptotically Mixed Normal in the absence of a linear time trend. Questions on optimal testing for these experiments remain largely unaddressed.

References

BICKEL, P.J., C.A.J. KLAASSEN, Y. RITOV, AND J.A. WELLNER (1993). Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore, MD.

BOSWIJK, H.P., M. JANSSON, AND M.Ø. NIELSEN (2012). Improved Likelihood Ratio Tests for Cointegration Rank in the VAR Model. *Working paper.*

BREITUNG (2001). Rank tests for nonlinear cointegration, *Journal of Business & Economic Statistics* 19, 331–340.

BREITUNG, J., AND C. GOURIÉROUX (1997). Rank Tests for Unit Roots, *Journal of Econometrics* 81, 7–27.

DROST, F.C., C.A.J. KLAASSEN, AND B.J.M. WERKER (1997). Adaptive estimation in time series models, *Annals of Statistics* 25, 786–818.

DROST, F.C., R. VAN DEN AKKER, AND B.J.M. WERKER (2009). The asymptotic structure of nearly unstable non negative integer-valued AR(1) models, *Bernoulli* 15, 297–324.

ENGLE, R.F. AND C.W.J. GRANGER (1981). Co-integration and error correction: representation, estimation, and testing, *Econometrica* 55, 251-276.

GRANGER, C.W.J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* 16, 121–130.

GUSHCHIN, A.A. (1996). Asymptotic optimality of parameter estimators under the LAQ Condition, *Theory of Probability and its Applications* 40, 261–272.

HÁJEK, J. AND Z. ŠIDÁK (1967). Theory of Rank Tests. Academic Press, New York.

HALL, P. AND C.C. HEYDE (1980). Martingale limit theory and its application. Academic Press, New York.

HALLIN, M., H. OJA, AND D. PAINDAVEINE (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *The Annals of Statistics* 34, 2757–2789.

HALLIN, M. AND D. PAINDAVEINE (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks, *The Annals of Statistics* 30, 1103–1133.

HALLIN, M. AND D. PAINDAVEINE (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence, *Bernoulli* 8, 787–815.

HALLIN, M. AND D. PAINDAVEINE (2005a). Asymptotic linearity of serial and nonserial multivariate signed rank statistics, *Journal of Statistical Planning and Inference* 136, 1–32.

HALLIN, M. AND D. PAINDAVEINE (2005b). Affine-invariant aligned rank tests for the multivariate general linear model with ARMA errors, *Journal of Multivariate Analysis* 93, 122–163.

HALLIN, M. AND D. PAINDAVEINE (2006a). Asymptotic linearity of serial and nonserial multivariate signed rank statistics, *Journal of Statistical Planning and Inference* 136, 1–32.

HALLIN, M., D. PAINDAVEINE, AND T. VERDEBOUT (2010). Optimal rank-based testing for principal components. *The Annals of Statistics* 38, 3245–3299.

HALLIN, M., D. PAINDAVEINE, AND T. VERDEBOUT (2012). Optimal rank-based tests for common principal components. *Bernoulli* to appear.

HALLIN, M., R. VAN DEN AKKER, AND B.J.M. WERKER (2011). A class of simple semiparametrically efficient rank-based unit root tests. *Journal of Econometrics* 163, 200–214.

HALLIN, M. AND B.J.M. WERKER (2003). Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli* 9, 137–165.

HANSEN, B.E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory* 8, 489–500.

HODGSON, D.J. (1998a). Adaptive estimation of cointegrating regressions with ARMA errors, *Journal of Econometrics* 85, 231–267.

HODGSON, D.J. (1998b). Adaptive estimation of error correction models, *Econometric Theory* 14, 44–69.

HUBRICH, K., H. LÜTKEPOHL, AND P. SAIKKONEN (2001). A review of systems cointegration tests, *Econometric Reviews* 20, 247–318.

JACOD, J. AND A.N. SHIRYAEV (2002). Limit Theorems for Stochastic Processes. Berlin: Springer.

JANSSON, M. (2008). Semiparametric power envelopes for tests of the unit root hypothesis, *Econometrica* 76), 1103–1142.

JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models, *Econometric Theory* 11, 818–887.

JEGANATHAN, P. (1997). On asymptotic inference in linear cointegrated time series systems, *Econometric Theory* 13 692–745.

Johansen, S. (1988). Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231-254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551–1580.

Johansen, S. (1995). Likelihood-based inference in cointegrated vector auto-regressive models. New York: Oxford University Press.

Johansen, S. and K. Juselius (1990). Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, *Oxford Bulletin of Economics and Statistics* 52, 169–210.

Juselius, K., (2006). The cointegrated VAR model: methodology and applications. Oxford: Oxford University Press.

Karatzas, I. and S.E. Shreve (1991). Brownian motion and stochastic calculus. Springer-Verlag, second edition.

Kreiss, J-P. (1987). On adaptative estimation in stationary ARMA processes, *Annals of Statistics* 15, 112–133.

Le Cam, L. (1986). Asymptotic methods in statistical decision theory. New York: Springer-Verlag.

Le Cam, L. and G.L. Yang (1990). Asymptotics in statistics - some basic concepts. New York: Springer.

Lehmann, E.L. and J.P. Romano (2005). Testing statistical hypothesis. Springer.

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999). Numerical distribution functions of Likelihood Ratio tests for cointegration, *Journal of Applied Econometrics* 14, 563–577.

Magnus, J.R. and H. Neudecker (1988). Matrix Differential Calculus with Applications in Statistics and Econometrics. New York: John Wiley & Sons.

Müller, U.K. (2012). Efficient tests under a weak convergence assumption, *Econometrica* 79, 395–435.

Nielsen, B. (2009). Test for cointegration rank in general vector autoregressions, *Working paper*.

Phillips, P.C.B. (1991). Optimal inference in cointegrated systems, *Econometrica* 59, 283–306.

Ploberger, W. and P.C.B. Phillips (2012). Optimal estimation under nonstandard conditions, *Journal of Econometrics* 169, 258–265.

Randles, R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test, *Journal of the American Statistical Association* 95, 1263–1268.

Reinsel, G.C. and S.K. Ahn (1992). Vector AR models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis* 13, 353–375.

Saikkonen, P. and H. Lütkepohl (2000). Testing for the cointegrating rank of a VAR process with an intercept, *Econometric Theory* 16, 373–406.

Stock, J.H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* 55 (1987), 1035–1056.

Strasser, H. (1985). Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory. Walter de Gruyter.

Toda, H.Y. (1994). Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present, *Review of Economics and Statistics* 76, 66–79.

Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter, *The Annals of Statistics* 15, 234–251.

Van der Vaart, A.W. (1988). Statistical Estimation in Large Parameter Spaces. CWI tract 44. Amsterdam: CWI.

van der Vaart, A.W. (2000). Asymptotic statistics. Cambridge: Cambridge University Press.

Technical Appendix to "Rank-based Tests of the Cointegrating Rank in Semiparametric Error Correction Models"

A. Proof of Proposition 2.1

In this appendix we present the proof of Proposition 2.1. Section A.1 introduces the limiting Brownian motion corresponding to the partial sum processes (2.4)-(2.6) and exploits this Brownian motion to provide the full definition of J. And Section A.2 contains the proof of Proposition 2.1. This proof exploits Proposition C.4 which provides high-level conditions for quadratic expansions of log likelihood ratios.

Throughout this appendix we evaluate, unless mentioned otherwise, expectations, O_P's, and o_P's under $P_{\vartheta;\Sigma,f}$. We repeatedly use the filtrations $\mathbb{F}^{(T)} := \left(\mathcal{F}_u^{(T)}, \, u \in [0,1] \right)$, $T \in \mathbb{N}$, defined by $\mathcal{F}_u^{(T)} := \sigma(\varepsilon_t, \, t \in \mathbb{N} : \, t \leq [uT]), \, u \in [0,1]$. The angle-bracket process $\left\langle A_i^{(T)}, A_j^{(T)} \right\rangle (u)$ and the straight-bracket process $\left[A_i^{(T)}, A_j^{(T)} \right] (u)$ are now well-defined for all $\mathbb{F}^{(T)}$-adapted locally square-integrable martingales and semimartingales $A_i^{(T)}$, respectively (see, e.g., Jacod and Shiryaev (2002)). If $A_i^{(T)}$, $i = 1, 2$, are square-integrable martingales of the form $A_i^{(T)}(u) = \sum_{t=1}^{[uT]} I_{Tt}^{(i)}$ with $I_{Tt}^{(i)}$ $\mathcal{F}_t$-measurable, we have $\left[A_1^{(T)}, A_2^{(T)} \right] (u) = \sum_{t=1}^{[uT]} I_{Tt}^{(1)} I_{Tt}^{(2)'}$ and $\left\langle A_1^{(T)}, A_2^{(T)} \right\rangle (u) = \sum_{t=1}^{[uT]} \mathrm{E} \left[I_{Tt}^{(1)} I_{Tt}^{(2)'} \mid \mathcal{F}_{t-1} \right]$. Recall that for a square-integrable martingale with continuous sample paths the angle-brackets and straight-brackets coincide.

A.1. Limiting behavior partial sum processes

First we note that, under Assumptions 1-3 with $\vartheta \in \Theta$ and $f \in \mathcal{F}_2$, the vector autoregressive process $V_t' = ((\beta' X_t)', \Delta X_t', \ldots, \Delta X_{t-k+2}')$ from the Granger-Johansen representation (1.4) can be given a stationary starting value under $P_{\vartheta;\Sigma,f}$. Let us denote the probability measure under which $\{V_t\}$ is stationary by $P_{\vartheta;\Sigma,f}^\star$.

Let $m = p + p + (k-1)p^2 + p^2$ and summarize the partial sum processes (2.4)-(2.6) in the process $\mathcal{W}^{(T)}$ taking values in $D_{\mathbb{R}^m}[0,1]$, i.e. $\mathcal{W}^{(T)'} = (W_\epsilon^{(T)'}, W_\phi^{(T)'}, W_{\Delta X \otimes \phi}^{(T)'}, W_{Y \otimes \phi}^{(T)'})$ with $W_{\Delta X \otimes \phi}^{(T)'} = (W_{\Delta X \otimes \phi, 1}^{(T)'}, \ldots, W_{\Delta X \otimes \phi, k-1}^{(T)'})$.

Let $\mathcal{W}' = (W_\epsilon', W_\phi', W_{\Delta X \otimes \phi}', W_{Y \otimes \phi}')$ denote a m-variate Brownian motion with vari-

ance per unit of time given by:

$$\mathrm{Var}\begin{pmatrix} W_\epsilon(1) \\ W_\phi(1) \end{pmatrix} = \begin{pmatrix} \frac{1}{p}\int_0^\infty r^{p+1}f(r)\mathrm{d}r\ \Sigma & I_p \\ I_p & \frac{1}{p}\mathcal{I}_p(f)\Sigma^{-1} \end{pmatrix},$$

$$\mathrm{Var}\begin{pmatrix} W_{\Delta X\otimes\phi}(1) \\ W_{Y\otimes\phi}(1) \end{pmatrix} = \frac{1}{p}\mathcal{I}_p(f)\mathrm{E}^\star_{\vartheta;\Sigma,f}\begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix}\begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix}' \otimes \Sigma^{-1},$$

and

$$\mathrm{Cov}\left(\begin{pmatrix} W_{\Delta X\otimes\phi} \\ W_{Y\otimes\phi} \end{pmatrix}, \begin{pmatrix} W_\epsilon(1) \\ W_\phi(1) \end{pmatrix}\right) = \begin{pmatrix} 1_{(k-1)\times1}\otimes C_{\Gamma,\Pi}\mu \\ \mathrm{E}^\star_{\vartheta;\Sigma,f}Y_0 \end{pmatrix} \otimes \begin{pmatrix} I_p & \frac{1}{p}\mathcal{I}_p(f)\Sigma^{-1} \end{pmatrix}.$$

The following lemma shows that $\mathcal{W}^{(T)}$ weakly converges to $\mathcal{W}$. The proof, an application of a functional central limit theorem for arrays of martingale differences, is included for completeness.

Lemma A.1. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, and $f \in \mathcal{F}_2$. Then we have, in $D_{\mathbb{R}^m}[0,1]$ and under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$,*

$$\mathcal{W}^{(T)} \Rightarrow \mathcal{W} \tag{A.1}$$

and, still under $\mathrm{P}_{\vartheta;\Sigma,f}$, we have

$$\left\langle \mathcal{W}^{(T)}, \mathcal{W}^{(T)} \right\rangle(1) = \left[\mathcal{W}^{(T)}, \mathcal{W}^{(T)} \right](1) + o_P(1) = \mathrm{Var}\left(\mathcal{W}(1)\right) + o_P(1). \tag{A.2}$$

Proof. Introduce $U_t^{(1)} = (\varepsilon_t', (\Sigma^{-1/2}U_t^\varepsilon\phi_f(\|\varepsilon_t\|_\Sigma))')'$, $U_t^{(2)} = v_{t-1}\otimes\Sigma^{-1/2}U_t^\varepsilon\phi_f(\|\varepsilon_t\|_\Sigma)$ with $v_{t-1} = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', Y_{t-1}')'$. Note that $\mathcal{W}^{(T)} = (\mathcal{W}_1^{(T)\prime}, \mathcal{W}_2^{(T)\prime})'$ with $\mathcal{W}_i^{(T)}(u) = T^{-1/2}\sum_{t=1}^{[uT]}U_t^{(i)}$ and partition $\mathcal{W} = (\mathcal{W}_1', \mathcal{W}_2')'$ accordingly. An application of Theorem VIII.3.33 in Jacod and Shiryaev (2002) shows that (A.1) and (A.2) hold if, for all $u \in [0,1]$, $i,j \in \{1,2\}$ and $\delta > 0$,

$$\left\langle \mathcal{W}_i^{(T)}, \mathcal{W}_j^{(T)} \right\rangle(u) = \frac{1}{T}\sum_{t=1}^{[uT]}\mathrm{E}\left[U_t^{(i)}U_t^{(j)\prime} \mid \mathcal{F}_{t-1} \right] \xrightarrow{\mathrm{P}} u\,\mathrm{Cov}(\mathcal{W}_i(1), \mathcal{W}_j(1)) \tag{A.3}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\left| U_t^{(i)} \right|^2 1\left\{ \left| U_t^{(i)} \right| > \delta\sqrt{T} \right\} \mid \mathcal{F}_{t-1} \right] \xrightarrow{\mathrm{P}} 0. \tag{A.4}$$

2

We start with (A.3). Fix $u \in [0,1]$. For $i = j = 1$ (A.3) is immediate from the weak law of large numbers (using the independence of U_t^ε and $\|\varepsilon_t\|_\Sigma$ and using the equality $\mathrm{E}\|\varepsilon_t\|_\Sigma \phi_f(\|\varepsilon_t\|_\Sigma) = p$). For $i = j = 2$ we have

$$\frac{1}{T} \sum_{t=1}^{[uT]} \mathrm{E}\left[U_t^{(2)} U_t^{(2)\prime} \mid \mathcal{F}_{t-1} \right] = \frac{1}{p} \mathcal{I}_p(f) \left(\frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} v_{t-1}' \right) \otimes \Sigma^{-1}$$

and for $i = 2$ and $j = 1$ we obtain

$$\frac{1}{T} \sum_{t=1}^{[uT]} \mathrm{E}\left[U_t^{(2)} U_t^{(1)\prime} \mid \mathcal{F}_{t-1} \right] = \frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} \otimes \mathrm{Cov}\left(W_\phi(1), \begin{pmatrix} W_\epsilon(1) \\ W_\phi(1) \end{pmatrix} \right).$$

Because $\{v_t\}$ is a stable vector autoregressive process of the order 1 with finite second moments we have

$$\frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} \xrightarrow{\mathrm{P}} u \mathrm{E}^\star_{\vartheta;\Sigma,f} \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix} \quad \text{and} \quad \frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} v_{t-1}' \xrightarrow{\mathrm{P}} u \mathrm{E}^\star_{\vartheta;\Sigma,f} \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix} \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix}'.$$

Note that $\Delta X_t = C_{\Gamma,\Pi}\mu + C_{\Gamma,\Pi}\varepsilon_t + \Delta Y_t$. We thus have $\mathrm{E}^\star_{\vartheta;\Sigma,f}\Delta X_t = C_{\Gamma,\Pi}\mu$ which completes the verification of (A.3).

Next we verify (A.4). For $i = 1$ (A.4) is immediate (the L_1 norm of the left-hand-side converges to 0 by an application of the dominated convergence theorem). For $i = 2$ we first introduce

$$\phi(M) := \mathrm{E}\left| \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \right|^2 1\left\{ \left| \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \right| \geq M \right\} \qquad (A.5)$$

and note that $\phi(M) \to 0$ as $M \to \infty$ (this essentially is the argument for $i = 1$). This yields

$$\frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\left[\left| U_t^{(2)} \right|^2 1\left\{ \left| U_t^{(2)} \right| > \delta\sqrt{T} \right\} \Big| \mathcal{F}_{t-1} \right] \leq \phi\left(\frac{\delta\sqrt{T}}{\max_{t=1,\dots,T} |v_{t-1}|} \right) \frac{1}{T} \sum_{t=1}^{T} |v_{t-1}|^2 = o_P(1),$$

as we already noted $T^{-1} \sum_{t=1}^{T} v_{t-1} v_{t-1}' = O_P(1)$ and as we have $T^{-1/2} \max_{t=1,\dots,T} |v_{t-1}| = o_P(1)$ (which follows from a combination of (A.4) with Lemma C.3). This concludes the proof. $\square$

Next we use the Brownian motion $\mathcal{W}$ to provide the definition of J. To this end we first introduce an auxiliary process $\tilde{\Delta}$ by

$$\tilde{\Delta}(u) = \begin{pmatrix} W_\phi(u) \\ W_{\Delta X \otimes \phi}(u) \\ (\beta' \otimes I_p) W_{Y \otimes \phi}(u) \\ |C_{\Gamma,\Pi}\mu|^2 \int_0^u s\, \mathrm{d}(\alpha' W_\phi)(s) \\ \int_0^u (\beta_\perp' C_{\Gamma,\Pi} W_\epsilon(s) \otimes I_r)\, \mathrm{d}(\alpha' W_\phi)(s) \\ |C_{\Gamma,\Pi}\mu|^2 \int_0^u s\, \mathrm{d}(\alpha_\perp' W_\phi)(s) \\ \int_0^u (\beta_\perp' C_{\Gamma,\Pi} W_\epsilon(s) \otimes I_{p-r})\, \mathrm{d}(\alpha_\perp' W_\phi)(s) \end{pmatrix}, \quad u \in [0,1],$$

3

which is adapted to the natural filtration generated by $\mathcal{W}$. Note that $\tilde{\Delta}(1) = \Delta$. Now we define J as

$$J := [\tilde{\Delta}, \tilde{\Delta}](1).$$

Using standard calculus rules for quadratic variations we obtain the following representation of J:

$$J = \frac{1}{p}\mathcal{I}_p(f)\begin{pmatrix} j_{11} & j_{12} \otimes \Sigma^{-1}\alpha & j_{12} \otimes \Sigma^{-1}\alpha_\perp \\ & j_{22} \otimes \alpha'\Sigma^{-1}\alpha & j_{22} \otimes \alpha'\Sigma^{-1}\alpha_\perp \\ & & j_{22} \otimes \alpha'_\perp\Sigma^{-1}\alpha_\perp \end{pmatrix},$$

with

$$j_{11} = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{(k-1)p^2} & 0 \\ 0 & 0 & \beta' \otimes I_p \end{pmatrix}\mathrm{Var}\begin{pmatrix} W_\phi(1) \\ W_{\Delta X \otimes \phi}(1) \\ W_{Y \otimes \phi}(1) \end{pmatrix}\begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{(k-1)p^2} & 0 \\ 0 & 0 & \beta' \otimes I_p \end{pmatrix}',$$

$$j_{22} = \begin{pmatrix} \frac{1}{3}|C_{\Gamma,\Pi}\mu|^4 & |C_{\Gamma,\Pi}\mu|^2\left(\int_0^1 u W_\epsilon(u)\mathrm{d}u\right)' C'_{\Gamma,\Pi}\beta_\perp \\ & \beta'_\perp C_{\Gamma,\Pi}\left(\int_0^1 W_\epsilon(u)W'_\epsilon(u)\mathrm{d}u\right) C'_{\Gamma,\Pi}\beta_\perp \end{pmatrix},$$

and

$$j_{12} = \begin{pmatrix} \frac{1}{2}|C_{\Gamma,\Pi}\mu|^2 & \left(\int_0^1 W_\epsilon(u)\mathrm{d}u\right)' C'_{\Gamma,\Pi}\beta_\perp \\ 1_{(k-1)\times 1} \otimes \frac{1}{2}|C_{\Gamma,\Pi}\mu|^2 C_{\Gamma,\Pi}\mu & 1_{(k-1)\times 1} \otimes C_{\Gamma,\Pi}\mu \otimes \left(\int_0^1 W_\epsilon(u)\mathrm{d}u\right)' C'_{\Gamma,\Pi}\beta_\perp \\ \frac{1}{2}|C_{\Gamma,\Pi}\mu|^2 \beta' \mathrm{E}^\star_{\vartheta;\Sigma,f} Y_0 & \beta' \mathrm{E}^\star_{\vartheta;\Sigma,f} Y_0 \left(\int_0^1 W'_\epsilon(u) C'_{\Gamma,\Pi}\beta_\perp \mathrm{d}u\right) \end{pmatrix}.$$

Remark A.1. Note that only the blocks involving B or D are random. Also note that all blocks involving b or d vanish in case $|C_{\Gamma,\Pi}\mu| = 0$.

A.2. Proof

To enhance readability, we split the proof into two parts. In Part A we show

(i) $(\Delta_\vartheta^{(T)}, J_\vartheta^{(T)}) \Rightarrow (\Delta, J)$;

(ii) $\mathrm{E}\exp(h'\Delta - h'Jh/2) = 1$,

and in Part B we show that (2.15) of Proposition 2.1, the quadratic expansion of the log likelihood ratio, holds.

Part A(i)

First we introduce auxiliary processes $\tilde{\Delta}^{(T)}$, $T \in \mathbb{N}$, by[11]

$$\tilde{\Delta}^{(T)}(u) = \begin{pmatrix} W_\phi^{(T)}(u) \\ W_{\Delta X \otimes \phi}^{(T)}(u) \\ (\beta' \otimes I_p) W_{Y \otimes \phi}^{(T)}(u) \\ |C_{\Gamma,\Pi}\mu|^2 \int_0^u \mathrm{id}_T(s-) \mathrm{d}(\alpha' W_\phi^{(T)})(s) \\ \int_0^u (\beta'_\perp C_{\Gamma,\Pi} W_\epsilon(s-) \otimes I_r) \, \mathrm{d}(\alpha' W_\phi^{(T)})(s) \\ |C_{\Gamma,\Pi}\mu|^2 \int_0^u \mathrm{id}_T(s-) \mathrm{d}(\alpha'_\perp W_\phi^{(T)})(s) \\ \int_0^u (\beta'_\perp C_{\Gamma,\Pi} W_\epsilon(s-) \otimes I_{p-r}) \mathrm{d}(\alpha'_\perp W_\phi^{(T)})(s) \end{pmatrix}, \quad u \in [0,1].$$

A combination of Lemma A.1 with Theorem 2.1 in Hansen (1992) (the conditions are trivially met as $\mathcal{W}^{(T)}$ is a martingale with respect to $\mathbb{F}^{(T)}$ and the increments of $W_\phi^{(T)}$ are i.i.d. with finite second moment) yields $\tilde{\Delta}^{(T)} \Rightarrow \tilde{\Delta}$ in $D_{\mathbb{R}^{m^\star}}[0,1]$, where $m^\star$ denotes the number of components of $\tilde{\Delta}^{(T)}$. Using this weak convergence, the identity $[A,B](u) = A(u)B(u) - A(0)B(0) - \int_0^u A(s-)\mathrm{d}B(s) - \int_0^u B(s-)\mathrm{d}A(s)$ and the continuous mapping theorem in combination with Theorem 2.1 in Hansen (1992) (the condition to this theorem is met as $\tilde{\Delta}^{(T)}$ is a martingale with respect to $\mathbb{F}^{(T)}$ and as we have $\sum_{t=1}^T \mathrm{E}|\tilde{\Delta}^{(T)}(t/T) - \tilde{\Delta}^{(T)}((t-1)/T)|^2 = O(1)$) yields,

$$\left(\tilde{\Delta}^{(T)}(1), \left[\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)}\right](1)\right) \Rightarrow \left(\tilde{\Delta}, \left[\tilde{\Delta}, \tilde{\Delta}\right](1)\right) = (\Delta, J). \tag{A.6}$$

A combination of (2.10)-(2.14), which represent the central sequence $\Delta_\vartheta^{(T)}$ in terms of (integrals with respect to) $\mathcal{W}^{(T)}$, with (A.6) and the continuous mapping theorem yields

$$\Delta_\vartheta^{(T)} = \tilde{\Delta}^{(T)}(1) + o_P(1). \tag{A.7}$$

Introduce $\tilde{Z}_{Tt} = (\tilde{Z}_{Tt}^{(1)'}, \tilde{Z}_{Tt}^{(2)'})'$ by $\tilde{Z}_{Tt}^{(1)} = Z_{Tt}^{(1)}$ and

$$\tilde{Z}_{Tt}^{(2)} = \frac{1}{\sqrt{T}} \begin{pmatrix} |C_{\Gamma,\Pi}\mu|^2 \frac{t-1}{T} \\ \beta'_\perp C_{\Gamma,\Pi} W_\epsilon^{(T)}\left(\frac{t-1}{T}\right). \end{pmatrix}$$

Note that

$$\langle \tilde{\Delta}, \tilde{\Delta} \rangle(1) = \frac{\mathcal{I}_p(f)}{p} \sum_{t=1}^T \begin{pmatrix} \tilde{Z}_{Tt}^{(1)} \tilde{Z}_{Tt}^{(1)'}(\vartheta) \otimes \Sigma^{-1} & \tilde{Z}_{Tt}^{(1)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \Sigma^{-1}\alpha & \tilde{Z}_{Tt}^{(1)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \Sigma^{-1}\alpha_\perp \\ & \tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \alpha'\Sigma^{-1}\alpha & \tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \alpha'\Sigma^{-1}\alpha_\perp \\ & & \tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \alpha'_\perp\Sigma^{-1}\alpha_\perp \end{pmatrix}.$$

---

[11]Recall that id_T denotes the cadlag function $\mathrm{id}_T(u) = [uT]/T$ which converges to the identity $\mathrm{id}(u) = u$ on $[0,1]$.

In view of (A.6)-(A.7) and (2.8) the proof of Part A(i) is complete if we show

$$\langle \tilde{\Delta}, \tilde{\Delta} \rangle (1) - \left[\tilde{\Delta}, \tilde{\Delta} \right] (1) = o_P(1) \tag{A.8}$$

and

$$\sum_{t=1}^{T} \tilde{Z}_{Tt} \tilde{Z}'_{Tt} - \sum_{t=1}^{T} Z_{Tt} Z'_{Tt} = o_P(1). \tag{A.9}$$

Indeed (A.9) and (2.8) imply $J_\vartheta^{(T)} = \langle \tilde{\Delta}, \tilde{\Delta} \rangle (1) + o_P(1)$ and (A.7)-(A.8) yield $\left(\Delta_\vartheta^{(T)}, J_\vartheta^{(T)} \right) = \left(\tilde{\Delta}^{(T)}(1), \left[\tilde{\Delta}, \tilde{\Delta} \right] (1) \right) + o_P(1)$.

We first consider (A.8). An application of Theorem 2.23 in Hall and Heyde (1980) shows that (A.8) holds if, for all $\delta > 0$, $i = 1, 2$ and compatible matrices M,

$$\sum_{t=1}^{T} \mathrm{E} \left[|\tilde{Z}_{Tt}^{(i)} \otimes M \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma)|^2 1\{|\tilde{Z}_{Tt}^{(i)} \otimes M \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma)| > \delta\} \mid \mathcal{F}_{t-1} \right] = o_P(1). \tag{A.10}$$

We only show the result for $i = 2$, because for $i = 1$ we already obtained the result for $M = I_p$ in the proof of Lemma A.1 (as $\tilde{Z}_{Tt}^{(1)} = Z_{Tt}^{(1)}$). We easily obtain, with ϕ as defined in (A.5) and using the Frobenius matrix norm, that the left-hand-side of (A.10) is bounded by

$$\|M\|^2 Q_T \phi \left(\frac{\delta \sqrt{T}}{\|M\| \sqrt{Q_T}} \right),$$

with $Q_T = |C_{\Gamma,\Pi} \mu|^4 + p \|\beta'_\perp C_{\Gamma,\Pi}\|^2 \max_{i=1,\ldots,p} \|W_{\epsilon,i}^{(T)}\|_\infty^2$. Since $\|W_\epsilon^{(T),i}\|_\infty \Rightarrow \|W_{\epsilon,i}\|_\infty$, by a combination of Lemma A.1 and the continuous mapping theorem, (A.10) follows.

Finally we demonstrate (A.9). Using $\tilde{Z}_{Tt}^{(1)} = Z_{Tt}^{(1)}$, $\sum_{t=1}^{T} Z_{Tt}^{(1)} Z_{Tt}^{(1)'} = O_P(1)$ (established in the proof of Lemma A.1) and the Cauchy-Schwarz inequality, it follows that it is sufficient to show

$$\sum_{t=1}^{T} \left(\tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'} - Z_{Tt}^{(2)} Z_{Tt}^{(2)'} \right) = o_P(1).$$

This easily follows by invoking the decomposition, implied by the Granger-Johansen representation (1.4),

$$Z_{Tt}^{(2)} = \begin{pmatrix} (C_{\Gamma,\Pi}\mu)' \left(\frac{t-1}{T^{3/2}} C_{\Gamma,\Pi}\mu + \frac{1}{T} C_{\Gamma,\Pi} W_\epsilon^{(T)} \left(\frac{t-1}{T} \right) + \frac{1}{T^{3/2}} (Y_{t-1} + a_{\mu,\Gamma,\Pi}) \right) \\ \beta'_\perp \left(\frac{1}{\sqrt{T}} C_{\Gamma,\Pi} W_\epsilon^{(T)} \left(\frac{t-1}{T} \right) + \frac{1}{T} (Y_{t-1} + a_{\mu,\Gamma,\Pi}) \right) \end{pmatrix},$$

in combination with the stability of the process $\{Y_t\}$, Lemma A.1 and the continuous mapping theorem. This completes the proof of Part A(i).

6

Remark A.2. In Part B of the proof we will exploit

$$\sum_{t=1}^{T} |Z_{Tt}|^2 = O_P(1), \text{ and} \tag{A.11}$$

$$\max_{t=1,\dots,T} |Z_{Tt}| = o_P(1), \tag{A.12}$$

which are implied by a combination of (A.8)-(A.9) with (A.6) and a combination of (A.9)-(A.10) with Lemma C.3 respectively. □

Part A(ii)

All components of the limiting central sequence Δ can be seen as stochastic integrals with respect to linear combinations of $\mathcal{W}$, i.e. $\Delta_i = \int_0^1 \xi_i(u) \mathrm{d}\tilde{\mathcal{W}}_i$ (where $\tilde{\mathcal{W}}_i = a_i' \mathcal{W}$). These integrands ξ_i satisfy

$$\max_i \|\xi_i\|_\infty \leq C(1 + \max_i \|\mathcal{W}_i\|_\infty),$$

for a constant C (depending on ϑ). An application of Corollary 3.5.16 in Karatzas and Shreve (1991) yields $\mathrm{E}\exp(h'\Delta - h'Jh/2) = 1$.

Part B

We use Proposition C.4 to prove the expansion (2.15). To this end we set $\tilde{P}_T = \mathrm{P}^{(T)}_{\vartheta^{(T)};\Sigma,f}$, $P_T = \mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$, and $\mathcal{F}_{Tt} = \mathcal{F}_t$.

For notational convenience we introduce $S_{Tt} = S_{Tt}$, for $T \in \mathbb{N}$ and $t = 1,\dots,T$, by

$$S_{Tt} = \begin{pmatrix} Z_{Tt}^{(1)} \otimes \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \\ Z_{Tt}^{(2)} \otimes \alpha' \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \\ Z_{Tt}^{(2)} \otimes \alpha'_\perp \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \end{pmatrix}.$$

Notice that $\Delta_\vartheta^{(T)} = \sum_{t=1}^T S_{Tt}$ and $J_\vartheta^{(T)} = \sum_{t=1}^T \mathrm{E}\left[S_{Tt} S_{Tt}' | \mathcal{F}_{t-1}\right]$.

In the notation of Proposition C.4 we have,

$$LR_{Tt} = \frac{f(\|\varepsilon_t - w_{Tt}\|_\Sigma)}{f(\|\varepsilon_t\|_\Sigma)}, \tag{A.13}$$

with

$$w_{Tt} = T^{-1/2} m_T + T^{-1/2} \sum_{j=1}^{k-1} G_{T,j} \Delta X_{t-j} + T^{-1/2} A_T \beta' X_{t-1} + T^{-3/2} \alpha b_T (C_{\Gamma,\Pi}\mu)' X_{t-1}$$

$$+ T^{-1} \alpha B_T \beta'_\perp X_{t-1} + T^{-3/2} \alpha_\perp d_T (C_{\Gamma,\Pi}\mu)' X_{t-1} + T^{-1} \alpha_\perp D_T \beta'_\perp X_{t-1}$$

7

$$+ T^{-2} A_T b_T (C_{\Gamma,\Pi} \mu)' X_{t-1} + T^{-3/2} A_T B_T \beta'_\perp X_{t-1}.$$

Exploiting the orthogonality restrictions $\beta_\perp B'_T \perp (C_{\Gamma,\Pi} \mu)$ and $\beta_\perp D'_T \perp (C_{\Gamma,\Pi} \mu)$ we can rewrite the fifth, seventh, and last term of w_{Tt} as:

$$T^{-1} \alpha B_T \beta'_\perp X_{t-1} = T^{-1} \alpha B_T \beta'_\perp (X_{t-1} - (t-1) C_{\Gamma,\Pi} \mu),$$

$$T^{-1} \alpha_\perp D_T \beta'_\perp X_{t-1} = T^{-1} \alpha_\perp D_T \beta'_\perp (X_{t-1} - (t-1) C_{\Gamma,\Pi} \mu), \text{ and}$$

$$T^{-3/2} A_T B_T \beta'_\perp X_{t-1} = T^{-3/2} A_T B_T \beta'_\perp (X_{t-1} - (t-1) C_{\Gamma,\Pi} \mu).$$

Assumption 2 implies, see Hallin and Paindaveine (2002a, Section 1), that the mapping $e \mapsto f^{1/2}(|e|)$ is differentiable in quadratic mean:

$$\frac{\sqrt{f}(\|e - w\|_\Sigma)}{\sqrt{f}(\|e\|_\Sigma)} = 1 + \frac{1}{2} \left[\phi_f(\|e\|_\Sigma) w' \frac{\Sigma^{-1} e}{\|e\|_\Sigma} + r(e, w) \right],$$

where

$$\psi(\delta) = \sup_{w:\, |w| \le \delta} \frac{1}{|w|^2} \mathrm{E} r^2(\varepsilon_1, w) \to 0 \text{ as } \delta \to 0. \tag{A.14}$$

Using the identity $\mathrm{vec}(AXB) = (B' \otimes A) \mathrm{vec}(X)$ we obtain

$$\sqrt{LR_{Tt}} = 1 + \frac{1}{2} (h'_T S_{Tt} + R_{Tt}), \tag{A.15}$$

with

$$R_{Tt} = r(\varepsilon_t, w_{Tt}) + \left(T^{-2} A_T b_T (C_{\Gamma,\Pi} \mu)' X_{t-1} \right)' \phi_f(\|\varepsilon_t\|_\Sigma) \Sigma^{-1/2} U_t$$

$$+ \left(T^{-3/2} A_T B_T \beta'_\perp (X_{t-1} - (t-1) C_{\Gamma,\Pi} \mu) \right)' \phi_f(\|\varepsilon_t\|_\Sigma) \Sigma^{-1/2} U_t. \tag{A.16}$$

So we can conclude that expansion (2.15) holds once we verify the conditions to Proposition C.4.

Condition (a) is immediate as h_T converges by assumption. *Condition (b)*. Square-integrability follows from our assumption $f \in \mathcal{F}_2$. Display (C.2) follows immediately from the independence of U_t and $\phi_f(\|\varepsilon_t\|_\Sigma)$ and $\mathrm{E}\phi_f(\|\varepsilon_t\|_\Sigma) = 0$ (which follows from Assumption 2 and standard arguments (see, e.g., Lemma 7.6 in Van der Vaart (2000)). Next we verify the conditional Lindeberg condition (C.3). We have, for $\delta > 0$ and with ϕ as defined in (A.5),

$$\sum_{t=1}^{T} \mathrm{E} \left[(h'_T S_{Tt})^2 1_{\{|h'_T S_{Tt}| > \delta\}} | \mathcal{F}_{t-1} \right] \le C |h_T|^2 \phi \left(\frac{\delta}{|h_T|_\infty C \max_{t=1,\ldots,T} |Z_{Tt}|} \right) \sum_{t=1}^{T} |Z_{Tt}|^2,$$

8

with $C = 1 + \|\alpha\| + \|\alpha_\perp\|$. Invoking (A.11) and (A.12) yields (C.3). Display (C.4) is immediate as $J_T = J_\vartheta^{(T)} \Rightarrow J$ by Part A(i). *Condition (c)*. We first consider condition (C.5). From (A.11) we obtain

$$\sum_{t=1}^{T} \mathrm{E}\left[\left((T^{-2}A_T b_T (C_{\Gamma,\Pi}\mu)'X_{t-1})' \phi_f(\|\varepsilon_t\|_\Sigma)\Sigma^{-1/2}U_t\right)^2 |\mathcal{F}_{t-1}\right] = o_P(1)$$

and

$$\sum_{t=1}^{T} \mathrm{E}\left[\left(T^{-3/2}A_T B_T \beta_\perp'(X_{t-1} - (t-1)C_{\Gamma,\Pi}\mu)'\phi_f(\|\varepsilon_t\|_\Sigma)\Sigma^{-1/2}U_t\right)^2 |\mathcal{F}_{t-1}\right] = o_P(1).$$

A combination with (A.16) yields

$$\sum_{t=1}^{T} \mathrm{E}\left[R_{Tt}^2|\mathcal{F}_{t-1}\right] \leq 4 \sum_{t=1}^{T} \mathrm{E}\left[r^2(\varepsilon_t, w_{Tt})|\mathcal{F}_{t-1}\right] + 4o_P(1)$$

$$\leq 4\psi \left(\max_{t=1,\dots,T} |w_{Tt}|\right) \sum_{t=1}^{T} |w_{Tt}|^2 + 4o_P(1).$$

Note that we can write w_{Tt} as a linear transformation of Z_{Tt} with bounded coefficients, i.e. $w_{Tt} = A^{(T)}Z_{Tt}$ with $\sup_{T\in\mathbb{N}} \|A^{(T)}\| < \infty$. Hence (A.11) and (A.12) imply

$$\sum_{t=1}^{T} |w_{Tt}|^2 = O_P(1) \text{ and } \max_{t=1,\dots,T} |w_{Tt}| = o_P(1),$$

and we thus have

$$\sum_{t=1}^{T} \mathrm{E}\left[R_{Tt}^2|\mathcal{F}_{t-1}\right] = o_P(1),$$

which shows that (C.5) holds. As we assumed that the radial density f is strictly positive Display (C.6) is immediate. Finally, *Condition (d)* is immediate as the initial values are deterministic.

B. Proof of Proposition 3.1

B.1. Proof of Part (i).

Part (i) of the proposition follows as a particular case of the nonserial asymptotic representation result in Proposition 2.1(i) of Hallin and Paindaveine (2006a), with the vector-valued weights $\Sigma^{-1/2}x_{t-i}'K$ replaced with the scalar ones $\left(\frac{t}{T+1} - \frac{1}{2}\right)$.

B.2. Proof of Part (ii).

Part (ii) is a direct consequence of the central limit theorem for weighted sums of independent summands with finite variance and weights satisfying the traditional Noether condition, which directly applies to $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ defined in (3.5) (see Hájek and Sidák (1967), p.153).

B.3. Proof of Part (iii).

The asymptotic normality of $\underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)})$ in Part (iii) results from a classical application of Le Cam's third Lemma (see, for instance, p.90 of Van der Vaart (2000)). Due to contiguity, the asymptotic representation (3.5) also holds under $P_{\vartheta^{(T)};\Sigma,f}^{(T)}$. Hence, the asymptotic mean of $\underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)})$, under local alternatives of the form $P_{\vartheta^{(T)};\Sigma,f}^{(T)}$, is the vector of asymptotic covariances, under $P_{\vartheta;\Sigma,f}^{(T)}$, of the asymptotically joint normal distribution of $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ in (3.5) and the log-likelihood (2.15). Clearly, those covariances involve sums of expectations of products of quantities of the type $U_t^\varepsilon U_{t'}^{\varepsilon\prime}$ and the corresponding summands in the linear part $h'\Delta_\vartheta^{(T)}$ of the approximation of local log-likelihoods; they break into $k+3$ parts, associated with the corresponding subvectors of $\Delta_\vartheta^{(T)}$, which we successively examine.

(a) For the $\Delta_\mu^{(T)}$ part of $\Delta^{(T)}$, the covariance is

$$T^{-1}\sum_{t=1}^T \left(\frac{t}{T+1}-\frac{1}{2}\right)\mathrm{E}\left[\phi_g\left(\tilde{G}_p^{-1}\left(\tilde{F}_p\left(\|\varepsilon_t\|_\Sigma\right)\right)\right)\phi_f\left(\|\varepsilon_t\|_\Sigma\right)U_t^\varepsilon U_t^{\varepsilon\prime}\right]m$$

$$= T^{-1}\sum_{t=1}^T \left(\frac{t}{T+1}-\frac{1}{2}\right)\frac{\mathcal{I}(f,g)}{p}m = 0.$$

(b) For the $\Delta_{\Gamma_i}^{(T)}$, $i=1,\ldots,k-1$ subvectors, we obtain the covariance

$$T^{-1}\sum_{t=1}^T \left(\frac{t}{T+1}-\frac{1}{2}\right)\mathrm{E}\left[\phi_g\left(\tilde{G}_p^{-1}\left(\tilde{F}_p\left(\|\varepsilon_t\|_\Sigma\right)\right)\right)\phi_f\left(\|\varepsilon_t\|_\Sigma\right)U_t^\varepsilon U_t^{\varepsilon\prime}\Sigma^{-1/2}\left(\Delta X'_{t-i}\otimes I_p\right)\right]\mathrm{vec}(G_i)$$

where $\Delta X_{t-i} = C_{\Pi,\Gamma}\mu + C_{\Pi,\Gamma}\varepsilon_{t-i} + \Delta Y_{t-i}$ is independent of ε_t and U_t^ε, yielding

$$T^{-1}\sum_{t=1}^T \left(\frac{t}{T+1}-\frac{1}{2}\right)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\left(\left(C_{\Pi,\Gamma}\mu + \mathrm{E}[\Delta Y_{t-i}]\right)\otimes I_p\right)\mathrm{vec}(G_i) = o(1)$$

since $\sum_{t=1}^T \left(\frac{t}{T+1}-\frac{1}{2}\right) = 0$ and $\mathrm{E}[\Delta Y_{t-i}] = o(1)$ in view of the asymptotic stationarity of $\{Y_t\}$.

10

(c) Turning to Δ_α, the covariance term is

$$T^{-1}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)\mathrm{E}\Big[\phi_g\big(\tilde{G}_p^{-1}\big(\tilde{F}_p\left(\|\varepsilon_t\|_\Sigma\right)\big)\big)\phi_f\big(\|\varepsilon_t\|_\Sigma\big)U_t^\varepsilon U_t^{\varepsilon'}\Sigma^{-1/2}\big((X_{t-1}'\beta)\otimes I_p\big)\Big]\mathrm{vec}(A),$$

which, due to the fact that $X_{t-1}'\beta = Y_{t-1}'\beta$, where Y_{t-1} is independent of ε_t, takes the form

$$T^{-1}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\Big(\big(C_{\Pi,\Gamma}\mu + \mathrm{E}[Y_{t-1}'\beta]\big)\otimes I_p\Big)\mathrm{vec}(A) = o(1)$$

since, again, $\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big) = 0$ and $\{Y_t\}$ is asymptotically stationary.

(d) The perturbation of β' has two parts, one with rate $T^{-3/2}$, the other one with rate T^{-1}. The first one yields a covariance term

$$T^{-2}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)\mathrm{E}\Big[\phi_g\big(\tilde{G}_p^{-1}\big(\tilde{F}_p\left(\|\varepsilon_t\|_\Sigma\right)\big)\big)\phi_f\big(\|\varepsilon_t\|_\Sigma\big)U_t^\varepsilon U_t^{\varepsilon'}\Sigma^{-1/2}\big(X_{t-1}'\otimes\alpha\big)\Big]\mathrm{vec}(b(C_{\Pi\Gamma}\mu)')$$

which, taking into account the Granger representation (1.4) of X_{t-1} and the independence between X_{t-1} and ε_t (hence also U_t^ε), reduces to

$$T^{-2}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\Big(\big((t-1)C_{\Pi,\Gamma}\mu)'+\mathrm{E}[Y_{t-1}']+a_{\Pi,\Gamma,\mu}'\big)\otimes\alpha\Big)\mathrm{vec}(b(C_{\Pi\Gamma}\mu)')$$
$$= A_1 + A_2 + A_3, \quad \text{say.}$$

Now, recalling that $(C'\otimes A)\mathrm{vec}(B) = \mathrm{vec}(ABC)$,

$$\begin{aligned}
A_1 &= T^{-2}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)(t-1)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\big((C_{\Pi,\Gamma}\mu)'\otimes\alpha\big)\mathrm{vec}(b(C_{\Pi\Gamma}\mu)')\\
&= \frac{\mathcal{I}(f,g)}{12p}\Sigma^{-1/2}\mathrm{vec}\big(\alpha b(C_{\Pi,\Gamma}\mu)'(C_{\Pi,\Gamma}\mu)\big) = \frac{\mathcal{I}(f,g)}{12p}\|C_{\Gamma,\Pi}\mu\|^2\Sigma^{-1/2}\alpha b, \quad \text{(B.1)}
\end{aligned}$$

while, in view of $\{Y_t\}$'s asymptotic stationarity,

$$A_2 = T^{-2}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\big(\mathrm{E}[Y_{t-1}']\otimes\alpha\big)\mathrm{vec}(b(C_{\Pi\Gamma}\mu)') = o(1)$$

as $T\to\infty$ and, since $\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big) = 0$,

$$A_3 = T^{-2}\sum_{t=1}^{T}\Big(\frac{t}{T+1}-\frac{1}{2}\Big)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\big(a_{\Pi,\Gamma,\mu}'\otimes\alpha\big)\mathrm{vec}(b(C_{\Pi\Gamma}\mu)') = 0.$$

11

The T^{-1}-perturbation of β' is of the form $T^{-1}B\beta'_{\perp}$, where $B\beta'_{\perp}C_{\Pi,\Gamma}\mu = 0$. Its contribution to the covariance decomposes, as the previous one, into a sum of three terms, A'_1, A'_2 and A'_3, say, with, for the same reasons as above, $A'_2 = o(1)$ and $A'_3 = 0$. As for A'_1,

$$
\begin{aligned}
A'_1 &= T^{-3/2}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)(t-1)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\left((C_{\Pi,\Gamma}\mu)' \otimes \alpha\right)\mathrm{vec}(B\beta'_{\perp}) \\
&= T^{-3/2}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)(t-1)\frac{\mathcal{I}(f,g)}{p}\Sigma^{-1/2}\mathrm{vec}\left(\alpha B\beta'_{\perp}C_{\Pi,\Gamma}\mu\right)
\end{aligned}
$$

where $T^{-3/2}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)(t-1)$ does not converge anymore; the orthogonality condition $B\beta'_{\perp}C_{\Pi,\Gamma}\mu = 0$, however, implies that $A'_1 = 0$.

(e) The reasoning for the $\beta_{\perp}$-part $\Delta^{(T)}_{\beta'_{\perp}}$ of the central sequence is entirely similar, and yields a contribution

$$
\frac{\mathcal{I}(f,g)}{12p}\|C_{\Gamma,\Pi}\mu\|^2\Sigma^{-1/2}\alpha_{\perp}db; \tag{B.2}
$$

details are left to the reader.

The desired result follows from adding the contributions (B.1) and (B.2).

B.4. Proof of Part (iv).

The proof of the asymptotic linearity part (iv) of Proposition 3.1 relies on Proposition C.5, which uses the same quadratic mean differentiability techniques we also use in Proposition C.4, which deals with quadratic approximations of log-likelihood ratios. Those two propositions therefore are regrouped in Appendix C.

In view of the asymptotic equivalence (3.5) and contiguity it is sufficient, in order to establish (3.8), to show that, under $\mathrm{P}^{(T)}_{\vartheta;\Sigma,f}$,

$$
\underline{S}^{(T)}_g(\vartheta^{(T)};\Sigma) - \underline{S}^{(T)}_g(\vartheta;\Sigma) = -\frac{1}{p}\mathcal{I}_p(f,g)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)v_{Tt} + o_P(1). \tag{B.3}
$$

Moreover, in view of Assumption 5, it is sufficient to prove (B.3) for monotone increasing ϕ_g.

Introduce, for $m \in \mathbb{N}$, the truncated versions J_m of $\phi_g \circ \tilde{G}_p^{-1}$

$$
J_m(u) := \begin{cases}
0 & \text{if } 0 \le u \le \frac{1}{m}; \\
\phi_g \circ \tilde{G}_p^{-1}\left(\frac{2}{m}\right) m \left(u - \frac{1}{m}\right) & \text{if } \frac{1}{m} < u \le \frac{2}{m}; \\
\phi_g \circ \tilde{G}_p^{-1}(u) & \text{if } \frac{2}{m} < u \le 1 - \frac{2}{m}; \\
\phi_g \circ \tilde{G}_p^{-1}\left(1 - \frac{2}{m}\right) m \left(1 - \frac{1}{m} - u\right) & \text{if } 1 - \frac{2}{m} < u \le 1 - \frac{1}{m}; \\
0 & \text{if } 1 - \frac{1}{m} \le u \le 1.
\end{cases}
$$

Note that $J_m : [0,1] \to \mathbb{R}$ is uniformly continuous and thus bounded. Moreover, we have

$$
\int_0^1 (\phi_g \circ \tilde{G}_p^{-1}(u))^2 \mathrm{d}u < \infty, \quad \lim_{m \to \infty} J_m(u) = \phi_g \circ \tilde{G}_p^{-1}(u), \quad u \in (0,1), \tag{B.4}
$$

and, using the fact that $\phi_g \circ \tilde{G}_p^{-1}$ is non-decreasing and continuous, for some $m_0 \in \mathbb{N}$,

$$
|J_m(u)| \le \left| \phi_g \circ \tilde{G}_p^{-1}(u) \right|, \quad u \in (0,1), \ m \ge m_0. \tag{B.5}
$$

Next, consider, for $\theta \in \Theta$, the truncated-score statistics

$$
\underline{S}_{g;m}^{(T)}(\theta; \Sigma) := \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t+1}{T} - \frac{1}{2} \right) U_t(\theta, \Sigma) J_m(\tilde{F}_p(\|\epsilon_t(\theta)\|_\Sigma)).
$$

The following lemma relates these truncated-score statistics to their non-truncated counterparts.

Lemma B.1. *For all $\epsilon > 0$,* $\displaystyle \lim_{m \to \infty} \sup_{T \in \mathbb{N}} \sup_{\vartheta \in \Theta} \mathrm{P}_{\vartheta;\Sigma,f} \left(\left| \underline{S}_g^{(T)}(\vartheta; \Sigma) - \underline{S}_{g;m}^{(T)}(\vartheta; \Sigma) \right| > \epsilon \right) = 0.$

The proof of this lemma in turn involves another lemma.

Lemma B.2. *We have, under $\mathrm{P}_{\vartheta;\Sigma,f}$,*

$$
\max_{1 \le t \le T} | \, \|\epsilon_t(\vartheta^{(T)})\|_\Sigma - \|\epsilon_t(\vartheta)\|_\Sigma | = o_P(1), \tag{B.6}
$$

and, for all $\eta > 0$,

$$
\max_{1 \le t \le T} |U_t(\vartheta^{(T)}, \Sigma) - U_t(\vartheta, \Sigma)| 1\{\|\varepsilon_t\|_\Sigma > \eta\} = o_P(1). \tag{B.7}
$$

Proof. Under $\mathrm{P}_{\vartheta;\Sigma,f}$, we have (see the proof of Lemma 4.1 in Hallin and Paindaveine (2006a, pp.29-30) for details),

$$
| \, \|\epsilon_t(\vartheta^{(T)})\|_\Sigma - \|\epsilon_t(\vartheta)\|_\Sigma | \le \|\Sigma^{-1/2}\| |\epsilon_t(\vartheta) - \epsilon_t(\vartheta^{(T)})|
$$

and

$$
|U_t(\vartheta^{(T)}, \Sigma) - U_t(\vartheta, \Sigma)| 1\{\|\varepsilon_t\|_\Sigma > \eta\} \le \frac{2}{\eta} \|\Sigma^{-1/2}\| |\epsilon_t(\vartheta) - \epsilon_t(\vartheta^{(T)})|.
$$

13

Hence, (B.6) and (B.7) follow if we show that

$$\max_{t=1,\ldots,T} |\epsilon_t(\vartheta) - \epsilon_t(\vartheta^{(T)})| = o_P(1). \tag{B.8}$$

We have

$$\epsilon_t(\vartheta^{(T)}) - \varepsilon_t = -T^{-1/2}m_T - \frac{1}{\sqrt{T}}\sum_{j=1}^{k-1}G_j^{(T)}\Delta X_{t-j} - \frac{1}{\sqrt{T}}A_T\beta'Y_{t-1}$$

$$- T^{-3/2}(t-1)|C_{\Gamma,\Pi}\mu|^2\alpha^{(T)}b_T - T^{-1}\alpha^{(T)}b_T(C_{\Gamma,\Pi}\mu)'C_{\Gamma,\Pi}W_\varepsilon^{(T)}\left(\frac{t-1}{T}\right)$$

$$- T^{-3/2}\alpha^{(T)}b_T(C_{\Gamma,\Pi}\mu)'(Y_{t-1} + a_{\mu,\Gamma,\Pi}) - T^{-1}\alpha^{(T)}B_T\beta'_\perp C_{\Gamma,\Pi}W_\varepsilon^{(T)}\left(\frac{t-1}{T}\right)$$

$$- T^{-1}\alpha^{(T)}B_T\beta'_\perp(Y_{t-1} + a_{\mu,\Gamma,\Pi}).$$

As the process $\{V_t\}$ from the Granger-Johansen representation (1.4) is, under $P_{\vartheta;\Sigma,f}$, stable with finite second moments we have, for all $\delta > 0$,

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E_{\vartheta;\Sigma,f}|V_t|^2 1\{|V_t| > \delta\sqrt{T}\} = 0,$$

which implies, by an application of Lemma C.3, that $T^{-1/2}\max_{t=1,\ldots,T}|V_t| = o_P(1)$. We thus have that $T^{-1/2}\max_{t=1,\ldots,T}|\Delta X_{t-j}|$ $(j = 1,\ldots,k-1)$ and $T^{-1/2}\max_{t=1,\ldots,T}|Y_t|$ are $o_P(1)$. Now, an application of Lemma A.1 and the continuous mapping theorem yield $\max_{i=1,\ldots,p}\|W_{\varepsilon,i}^{(T)}\|_\infty = O_P(1)$. This establishes (B.8) and hence concludes the proof of Lemma B.2. □

Proof of Lemma B.1. Let

$$C_{t;m}^{(T)} := U_t^\varepsilon\big(\phi_g\circ\tilde{G}_p^{-1}\circ\tilde{F}_p(\|\varepsilon_t\|_\Sigma) - J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))\big).$$

Then, we have

$$\mathrm{Var}_{\vartheta;\Sigma,f}\big(\underline{S}_g^{(T)}(\vartheta;\Sigma) - \underline{S}_{g;m}^{(T)}(\vartheta;\Sigma)\big) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)^2 E_{\vartheta;\Sigma,f}[C_{t;m}^{(T)}C_{t;m}^{(T)\prime}]$$

$$= \frac{1}{p}I_p E\big(\phi_g\circ\tilde{G}_p^{-1}(V) - J_m(V)\big)^2\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T+1} - \frac{1}{2}\right)^2,$$

where V is uniform over $[0,1]$. Dominated convergence and (B.4)-(B.5) yield that $E[(\phi_g\circ\tilde{G}_p^{-1}(V) - J_m(V))^2]$, which does not depend on T and ϑ, converges to zero as $m \to \infty$. Since $\underline{S}_g^{(T)}(\vartheta;\Sigma)$ and $\underline{S}_{g;m}^{(T)}(\vartheta;\Sigma)$ are both centered under $P_{\vartheta;\Sigma,f}^{(T)}$, an application of the Markov inequality completes the proof of of Lemma B.1. □

We now turn back to the proof of part (iv) of Proposition 3.1. Defining

$$\mathcal{I}_{p,m}(f,g) := \int_0^1 \phi_f(\tilde{F}_p^{-1}(u)) J_m(u) \mathrm{d}u,$$

note that dominated convergence, (B.4) and (B.5) entail

$$\lim_{m\to\infty} \mathcal{I}_{p,m}(f,g) = \mathcal{I}_p(f,g). \tag{B.9}$$

As $T^{-1/2} \sum_{t=1}^T (t/(T+1) - 1/2) v_{Tt} = O_P(1)$ Lemma B.1, contiguity, and (B.9) imply that a sufficient condition for (B.3) to hold is that, for all $m \in \mathbb{N}$,

$$\underline{S}_{g;m}^{(T)}(\vartheta^{(T)};\Sigma) - \underline{S}_{g;m}^{(T)}(\vartheta;\Sigma) = -\frac{1}{p}\mathcal{I}_{p;m}(f,g)\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\frac{t}{T+1} - \frac{1}{2}\right)v_{Tt} + o_P(1). \tag{B.10}$$

Let us show that (B.10) holds componentwise as a consequence of Lemma C.5, with

$$\tilde{Z}_{Tt} = T^{-1/2}\left(\frac{t+1}{T} - \frac{1}{2}\right)U_{t,j}(\vartheta^{(T)},\Sigma)J_m(\tilde{F}_p(\|\epsilon_t(\vartheta^{(T)})\|_\Sigma)),$$

$$Z_{Tt} = T^{-1/2}\left(\frac{t+1}{T} - \frac{1}{2}\right)U_{t,j}(\vartheta,\Sigma)J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)),$$

$\tilde{P}_T = \mathrm{P}_{\vartheta^{(T)};\Sigma,f}^{(T)}$, $P_T = \mathrm{P}_{\vartheta;\Sigma,f}^{(T)}$, and LR_{Tt} as in (A.13).

Note that Conditions (a)-(e) of Proposition C.4, which are required also in Proposition C.5, are satisfied (see Proposition 2.1 and its proof in Appendix A). As J_m is bounded and $|U_t(\theta,\Sigma)| \leq 1$ for all $\theta \in \Theta$, Condition (g) of Proposition C.5 clearly holds; Condition (h) follows straightforwardly.

In order to conclude, thus, we only have to establish that Condition (f) holds true. We have

$$\sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f}\left[\left(\tilde{Z}_{Tt}(LR_{Tt})^{1/2} - Z_{Tt}\right)^2 \mid \mathcal{F}_{T,t-1}\right] \leq 2\sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f}\left[\left(\tilde{Z}_{Tt} - Z_{Tt}\right)^2 \mid \mathcal{F}_{T,t-1}\right]$$

$$+ 2\max_{t=1,\ldots,T}\left((LR_{Tt})^{1/2} - 1\right)^2 \sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f}\left[\tilde{Z}_{Tt}^2 \mid \mathcal{F}_{T,t-1}\right].$$

It follows from (C.12) that $\max_{t=1,\ldots,T} |(LR_{Tt})^{1/2} - 1| = o_P(1)$, and we already noted that $\sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f}\left[\tilde{Z}_{Tt}^2 \mid \mathcal{F}_{T,t-1}\right] = O_P(1)$. Hence Condition (f) holds if we show that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f}\left(D_{t;m}^{(T)}\right)^2 = 0 \tag{B.11}$$

15

with (writing ϵ_t for $\epsilon_t(\vartheta^{(T)})$ and U_t^ϵ for $U_t(\vartheta^{(T)}; \Sigma)$)

$$D_{t;m}^{(T)} := U_{t,j}^\varepsilon J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) - U_{t,j}^\epsilon J_m(\tilde{F}_p(\|\epsilon_t\|_\Sigma)).$$

Let $\eta > 0$ such that $\tilde{F}_p(\eta) < m^{-1}$ and note that $J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))1\{\|\varepsilon_t\|_\Sigma \leq \eta\} = 0$. This yields, for $|D_{t;m}^{(T)}|$, the bound

$$|D_{t;m}^{(T)}| \leq \left| J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) - J_m(\tilde{F}_p(\|\epsilon_t\|_\Sigma)) \right| + \left| J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) \right| \left| U_{t,j}^\varepsilon - U_{t,j}^\epsilon \right| 1\{\|\varepsilon_t\|_\Sigma > \eta\}.$$

Uniform continuity of $J_m \circ \tilde{F}_p$ and (B.6) imply

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f} \left(J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) - J_m(\tilde{F}_p(\|\epsilon_t\|_\Sigma)) \right)^2 = 0.$$

Similarly, (B.7) and bounded convergence yield

$$\frac{1}{T} \sum_{t=1}^T \mathrm{E}_{\vartheta;\Sigma,f} \left| U_{t,j}^\varepsilon - U_{t,j}^\epsilon \right|^2 1\{\|\varepsilon_t\|_\Sigma > \eta\} \leq \mathrm{E}_{\vartheta;\Sigma,f} \max_{t=1,\dots,T} \left| U_{t,j}^\varepsilon - U_{t,j}^\epsilon \right|^2 1\{\|\varepsilon_t\|_\Sigma > \eta\} = o(1).$$

Display (B.11) follows, hence also Condition (f). Lemma C.5 thus applies, which establishes (B.10). Noting that

$$\mathrm{E}_{\vartheta;\Sigma,f} \left[Z_{Tt}(h_T' \Delta_{\vartheta;\Sigma,f}) \mid \mathcal{F}_{t-1} \right] = \frac{1}{p} \mathcal{I}_{p,m}(f,g) \left(\frac{t}{T+1} - \frac{1}{2} \right) v_{Tt,j}$$

completes the proof of Part (iv) of Proposition 3.1. □

B.5. *Proof of Part (v).*

The proof of Part (v) is classical, though the various rates of convergence involved make it more complicated than usual. For given $M > 0$, let $\Theta^{(T)}(M)$ denote the set of possible values of $\vartheta^{(T)}$, as described in the last part of Assumption 6. Observe that the set of sequence $t^{(1)}, t^{(1)}, \dots$ such that $t^{(T)} \in \Theta^{(T)}(M)$ for all $T \in \mathbb{N}$ is a countable set, and that each of those sequences constitutes a sequence of perturbations in the sense of (3.4). The proof of Lemma 4.4 in Kreiss (1987) then applies without changes. The result follows.

C. Quadratic expansions of log-likelihood ratios and a general asymptotic linearity result

Irrespective of the statistical model under study, the derivation of limits of sequences of local experiments often follows along very similar lines, essentially involving differentiability in quadratic mean of (conditional) densities. This appendix establishes two

16

abstract and very general results providing sufficient and nearly necessary conditions for (i) the existence of a quadratic expansion, and (ii) the asymptotic linearity of local log-likelihood ratios[12]; see Sections C.2 and C.3, respectively. Such results have been proved, for models involving independent and identically distributed observations, by, e.g., Bickel, Klaassen, Ritov, and Wellner (1993), Van der Vaart (1988) and Van der Vaart (2000). Here, we extend them to time series models. Unlike their i.i.d. counterparts, our results moreover are not restricted to locally asymptotically normal (LAN) experiments, so that non-stationary time series, for instance, also can be handled.

The results in this appendix thus are of general interest, and are not specific to the cointegration model we are focusing on in the main text. As a result, we do not rely on assumptions or notation introduced elsewhere, and this appendix is self-contained.

C.1. Main notation and some preliminary results.

We first describe the abstract setup and introduce the main notation. For each $T \in \mathbb{N}$, let $(\Omega_T, \mathcal{F}_T)$ be a measurable space on which two probability measures, $\tilde{\mathbb{P}}_T$ and $\mathbb{P}_T$, are defined. Let $\mathcal{F}_{T0} \subset \cdots \subset \mathcal{F}_{TT} \subset \mathcal{F}_T$ be a sequence of increasing σ-fields. Still for $T \in \mathbb{N}$, define the restrictions $\tilde{\mathrm{P}}_T := \tilde{\mathbb{P}}_T|_{\mathcal{F}_{TT}}$ and $\mathrm{P}_T := \mathbb{P}_T|_{\mathcal{F}_{TT}}$ of $\tilde{\mathbb{P}}_T$ and $\mathbb{P}_T$, respectively, to $\mathcal{F}_{TT}$. Using obvious notation, similarly define, for $t = 0, \ldots, T$, the restrictions $\tilde{\mathrm{P}}_{Tt} := \tilde{\mathbb{P}}_T|_{\mathcal{F}_{Tt}}$ and $\mathrm{P}_{Tt} := \mathbb{P}_T|_{\mathcal{F}_{Tt}}$. The Lebesgue decomposition of $\tilde{\mathrm{P}}_{Tt}$ on P_{Tt} (with respect to $\mathcal{F}_{Tt}$) takes the form

$$\tilde{\mathrm{P}}_{Tt}(A) = \int_A L_{Tt} d\mathrm{P}_{Tt} + \tilde{\mathrm{P}}_{Tt}(A \cap N_{Tt}) \qquad A \in \mathcal{F}_{Tt},$$

where $N_{Tt} \in \mathcal{F}_{Tt}$ is such that $\mathrm{P}_{Tt}(N_{Tt}) = 0$ and L_{Tt} is the Radon-Nikodym derivative of that part of $\tilde{\mathrm{P}}_{Tt}$ which is absolutely continuous with respect to P_{Tt}.

The likelihood ratio statistic LR_T for $\tilde{\mathrm{P}}_T$ with respect to P_T is, by definition, L_{TT}. Put $LR_{T0} := L_{T0}$, and define the conditional likelihood ratio contribution of observation t as

$$LR_{Tt} := L_{Tt}/L_{T,t-1}, \quad t = 1, \ldots, T,$$

with the convention $0/0 = 1$. Then, the likelihood ratio statistic LR_T factorizes into

$$LR_T = \prod_{t=0}^{T} LR_{Tt}, \quad \mathrm{P}_T\text{-a.s.}$$

---

[12]Asymptotic linearity will be needed whenever unspecified model parameters are to be replaced, in some statistic of interest, with some preliminary estimator.

17

This factorization follows from the fact that, under P_T, $\{L_{Tt} : 0 \leq t \leq T\}$ is a super-martingale with respect to the filtration $\{\mathcal{F}_{Tt} : 0 \leq t \leq T\}$ (which is easy to check) by repeated application of the following Lemma with $X = L_{Tt}$, $Y = L_{T,t-1}$, and $\mathcal{F} = \mathcal{F}_{T,t-1}$, and $t = 1, \ldots, T$.

Lemma C.1. *Let X be a nonnegative, integrable random variable and Y a $\mathcal{F}$-measurable random variable satisfying $Y \geq \mathrm{E}[X|\mathcal{F}]$. Then, $X1_{\{Y=0\}} = 0$ a.s.*

Proof. This readily follows from the fact that

$$0 \leq \mathrm{E}X1_{\{Y=0\}} = \mathrm{E}\mathrm{E}[X|\mathcal{F}]1_{\{Y=0\}} \leq \mathrm{E}Y1_{\{Y=0\}} = 0. \qquad \square$$

We end this section with two lemmas that are needed in the sequel. The first one is a consequence of Theorem 2.23 and Corollary 3.1 in Hall and Heyde (1980). We refer to Lemma 2.2 in Drost, Klaassen, and Werker (1997) for additional details.

Lemma C.2. *If the square-integrable process $\{X_{Tt} : 1 \leq t \leq T\}$ is adapted to the filtration $(\mathcal{F}_{Tt})_{0 \leq t \leq T}$ and satisfies $\sum_{t=1}^{T} \mathrm{E}\left[X_{Tt}^2 \mid \mathcal{F}_{T,t-1}\right] = o_P(1)$, then,*

$$\sum_{t=1}^{T} X_{Tt}^2 = o_P(1) \quad and \quad \sum_{t=1}^{T} (X_{Tt} - \mathrm{E}[X_{Tt} \mid \mathcal{F}_{T,t-1}]) = o_P(1).$$

The second lemma follows by an application of a result due to Dvoretzky (see the proof of Theorem 2.23 in Hall and Heyde (1980)).

Lemma C.3. *If the process $\{X_{Tt} : 1 \leq t \leq T\}$ is adapted to the filtration $(\mathcal{F}_{Tt})_{0 \leq t \leq T}$ and satisfies, for all $\delta > 0$,*

$$\sum_{t=1}^{T} \mathrm{E}\left[X_{Tt}^2 1_{\{|X_{Tt}|>\delta\}} \mid \mathcal{F}_{T,t-1}\right] = o_P(1),$$

then we have $\max_{t=1,\ldots,T} |X_{Tt}| = o_P(1)$.

C.2. Quadratic expansions of log likelihood ratios

The following proposition provides a very general sufficient condition for the existence of a quadratic expansion of local log likelihood ratios.

Proposition C.4. *Suppose that, for some $k \in \mathbb{N}$, there exist, for each $T \in \mathbb{N}$, $\mathcal{F}_{Tt}$-measurable mappings $S_{Tt} : \Omega_T \to \mathbb{R}^k$ and $R_{Tt} : \Omega_T \to \mathbb{R}$, $t = 1, \ldots, T$, such that the conditional likelihood ratio contribution LR_{Tt} can be written as*

$$LR_{Tt} = \left(1 + \frac{1}{2}(h_T' S_{Tt} + R_{Tt})\right)^2, \tag{C.1}$$

where

18

(a) h_T is a bounded (deterministic) sequence in $\mathbb{R}^k$;

(b) for each $T \in \mathbb{N}$, $\{S_{Tt} : 1 \leq t \leq T\}$ is a P_T-square integrable martingale difference array with respect to the filtration $\{\mathcal{F}_{Tt} : 0 \leq t \leq T\}$, satisfying the conditional Lindeberg condition and with tight squared conditional moments, i.e., such that, under P_T,

$$E_{P_T}[S_{Tt} \mid \mathcal{F}_{T,t-1}] = 0, \quad t = 1, \ldots, T, \qquad (C.2)$$

$$\sum_{t=1}^{T} E_{P_T}\left[(h_T' S_{Tt})^2 \, 1_{\{|h_T' S_{Tt}| > \delta\}} \mid \mathcal{F}_{T,t-1}\right] = o_P(1) \quad \forall \delta > 0, \qquad (C.3)$$

$$J_T = \sum_{t=1}^{T} E_{P_T}[S_{Tt} S_{Tt}' \mid \mathcal{F}_{T,t-1}] = O_P(1); \qquad (C.4)$$

(c) the remainder terms R_{Tt} and the null-sets N_{Tt} from the Lebesgue decomposition of $\tilde{P}_T$ on P_T are sufficiently small, i.e., under P_T,

$$\sum_{t=1}^{T} E_{P_T}[R_{Tt}^2 \mid \mathcal{F}_{T,t-1}] = o_P(1) \qquad (C.5)$$

and

$$\sum_{t=1}^{T} (1 - E_{P_T}[LR_{Tt} \mid \mathcal{F}_{T,t-1}]) = o_P(1); \qquad (C.6)$$

(d) under P_T, $\log LR_{T0} = o_P(1)$.

Then, under P_T, the log likelihood ratio admits the quadratic expansion

$$\log LR_T = h_T' \sum_{t=1}^{T} S_{Tt} - \frac{1}{2} h_T' J_T h_T + o_P(1). \qquad (C.7)$$

Proof. Let $r : 2x \mapsto r(2x) := 2\left(\log(1+x) - x + x^2/2\right)$, and rewrite the log likelihood ratio statistic as

$$\log LR_T = \sum_{t=0}^{T} \log LR_{Tt} = o_P(1) + \sum_{t=1}^{T} h_T' S_{Tt} - \frac{1}{2} h_T' J_T h_T$$

$$+ \frac{1}{4}\left(h_T' J_T h_T - \sum_{t=1}^{T} (h_T' S_{Tt})^2\right) + \sum_{t=1}^{T} (R_{Tt} - E_{P_T}[R_{Tt} \mid \mathcal{F}_{T,t-1}])$$

$$- \frac{1}{4} \sum_{t=1}^{T} R_{Tt}^2 - \frac{1}{2} \sum_{t=1}^{T} h_T' S_{Tt} R_{Tt} + \left(\sum_{t=1}^{T} E_{P_T}[R_{Tt} \mid \mathcal{F}_{T,t-1}] + \frac{1}{4} h_T' J_T h_T\right)$$

$$+ \sum_{t=1}^{T} r(h_T' S_{Tt} + R_{Tt}), \qquad (C.8)$$

where we used Condition (d) to neglect the first term $\log LR_{T0}$. To establish (C.7), we show that the six remainder terms on the right-hand side of (C.8) all converge to zero in probability under P_T.

19

By Theorem 2.23 in Hall and Heyde (1980), Condition (a) and (C.2)-(C.4) we have

$$\sum_{t=1}^{T} (h_T' S_{Tt})^2 - h_T' J_T h_T = o_P(1), \tag{C.9}$$

which shows that the first remainder term is indeed $o_P(1)$.

Since $(L_{Tt})_{0 \leq t \leq T}$ is a P_T-supermartingale, we have $E_{P_T} L R_{Tt} \leq 1$. Since S_{Tt} is also P_T-square integrable, it follows from (C.1), that R_{Tt} is P_T-square integrable. From Lemma C.2 and (C.5), we now immediately obtain

$$\sum_{t=1}^{T} (R_{Tt} - E_{P_T}[R_{Tt} \mid \mathcal{F}_{T,t-1}]) = o_P(1) \text{ and } \sum_{t=1}^{T} R_{Tt}^2 = o_P(1), \tag{C.10}$$

i.e. the second and third remainder term also are negligible.

Next we show that the remainder term $(1/2)\sum_{t=1}^{n} h_T' S_{Tt} R_{Tt}$ vanishes asymptotically. First note that Condition (a) and (C.4) and (C.9) imply $\sum_{t=1}^{T} (h_T' S_{Tt})^2 = O_P(1)$. Combined with (C.10), an application of the Cauchy-Schwarz inequality thus yields the convergence of the fourth remainder term.

To prove the negligibility of the fifth remainder term in (C.8), observe that (C.1), (C.2), (C.4), (C.5), and the Cauchy-Schwarz inequality again, entail

$$\sum_{t=1}^{T} (E_{P_T}[L R_{Tt} \mid \mathcal{F}_{T,t-1}] - 1) = \sum_{t=1}^{T} E_{P_T}[h_T' S_{Tt} \mid \mathcal{F}_{T,t-1}] + \sum_{t=1}^{T} E_{P_T}[R_{Tt} \mid \mathcal{F}_{T,t-1}]$$

$$+ \frac{1}{4}\sum_{t=1}^{T} E_{P_T}\left[(h_T' S_{Tt})^2 \mid \mathcal{F}_{T,t-1}\right] + \frac{1}{4}\sum_{t=1}^{T} E_{P_T}\left[R_{Tt}^2 \mid \mathcal{F}_{T,t-1}\right]$$

$$+ \frac{1}{2}\sum_{t=1}^{T} E_{P_T}\left[(h_T' S_{Tt}) R_{Tt} \mid \mathcal{F}_{T,t-1}\right]$$

$$= \sum_{t=1}^{T} E_{P_T}[R_{Tt} \mid \mathcal{F}_{T,t-1}] + \frac{1}{4} h_T' J_T h_T + o_P(1).$$

Now, the second part of (C.5) implies

$$\sum_{t=1}^{T} E_{P_T}[R_{Tt} \mid \mathcal{F}_{T,t-1}] + \frac{1}{4} h_T' J_T h_T = o_P(1). \tag{C.11}$$

Thus, the fifth remainder term in (C.8) also is negligible.

Turning to the sixth and last remainder term, let us first show that

$$\max_{t=1,\ldots,T} |h_T' S_{Tt} + R_{Tt}| = o_P(1) \quad \text{and} \quad \sum_{t=1}^{T} |h_T' S_{Tt} + R_{Tt}|^3 = o_P(1). \tag{C.12}$$

As (C.3) and (C.5) yield, for $\delta > 0$,

$$\sum_{t=1}^{T} E_{P_T}\left[(h_T' S_{Tt} + R_{Tt})^2 1_{\{|h_T' S_{Tt}+R_{Tt}|>\delta\}} \mid \mathcal{F}_{T,t-1}\right]$$

20

$$\leq 4 \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}} \left[(h'_T S_{Tt})^2 1_{\left\{ |h'_T S_{Tt}| > \delta/2 \right\}} \mid \mathcal{F}_{T,t-1} \right] + 4 \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}} \left[R^2_{Tt} \mid \mathcal{F}_{T,t-1} \right] = o_P(1),$$

the first part of (C.12) follows as an application of Lemma C.3. The second part is obtained from the latter by taking out the maximum (which tends to zero) and by observing that the remaining quadratic term is bounded in probability. In view of the first part of (C.12), indeed, it is sufficient to study the behavior of the final remainder term on the event $\{ |h'_T S_{Tt} + R_{Tt}| \leq 1 \}$. On this set, this remainder term is bounded: using $\left| \log(1+x) - x + \frac{1}{2}x^2 \right| \leq \frac{2}{3}x^3$ for $|x| \leq \frac{1}{2}$, indeed, we obtain

$$\left| \sum_{t=1}^{T} r \left(h'_T S_{Tt} + R_{Tt} \right) \right| \leq \frac{4}{3} \sum_{t=1}^{T} \left(h'_T S_{Tt} + R_{Tt} \right)^3.$$

Convergence to zero is now obtained from the second part of (C.12). This completes the proof of the proposition. $\qquad\square$

C.3. Asymptotic linearity: general result

This appendix provides a sufficient condition for asymptotic linearity which generalizes Proposition A.10 in Van der Vaart (1988) to the case of serially dependent observations and non-LAN limit experiments.

Proposition C.5. *Let, for each $T \in \mathbb{N}$, $\{ \tilde{Z}_{Tt} : 1 \leq t \leq T \}$ and $\{ Z_{Tt} : 1 \leq t \leq T \}$ be a $\tilde{\mathrm{P}}_T$-square integrable martingale difference array, and a $\mathrm{P_T}$-square integrable martingale difference array, respectively. Suppose that the Conditions (a)-(d) of Proposition C.4 hold, as well as the following Conditions (e)-(h):*

(e) $(\sum_{t=1}^{T} S_{Tt}, J_T)$ converges in distribution, under $\mathrm{P_T}$, to a limit (Δ, J) that satisfies, for all $a \in \mathbb{R}^k$, $\mathrm{E}\exp\left(a'\Delta - \frac{1}{2}a'Ja \right) = 1$;

(f) $\displaystyle\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}} \left[\left(\tilde{Z}_{Tt} \sqrt{LR_{Tt}} - Z_{Tt} \right)^2 \mid \mathcal{F}_{T,t-1} \right] = o_P(1)$ under $\mathrm{P_T}$;

(g) $\displaystyle\sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathrm{P}}_T} \left[\tilde{Z}^2_{Tt} \mid \mathcal{F}_{T,t-1} \right] = O_P(1)$ under $\tilde{\mathrm{P}}_T$, and $\displaystyle\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}} \left[Z^2_{Tt} \mid \mathcal{F}_{T,t-1} \right] = O_P(1)$ under $\mathrm{P_T}$;

(h) the conditional Lindeberg condition holds for $\{ \tilde{Z}_{Tt} : 1 \leq t \leq T \}$ under $\tilde{\mathrm{P}}_T$, i.e., for all $\delta > 0$, $\displaystyle\sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathrm{P}}_T} \left[\tilde{Z}^2_{Tt} 1_{\left\{ |\tilde{Z}_{Tt}| > \delta \right\}} \mid \mathcal{F}_{T,t-1} \right] = o_P(1)$ under $\tilde{\mathrm{P}}_T$.

Then, letting $\tilde{I}_T := \displaystyle\sum_{t=1}^{T} \tilde{\iota}_{Tt} := \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}} [(h'_T S_{Tt}) Z_{Tt} \mid \mathcal{F}_{T,t-1}]$, we have, under $\mathrm{P_T}$,

$$\sum_{t=1}^{T} \tilde{Z}_{Tt} = \sum_{t=1}^{T} Z_{Tt} - \tilde{I}_T + o_P(1). \tag{C.13}$$

21

Proof. The proof decomposes into four parts. In Part 1 we show that (C.13) holds if, under P_T,

$$\sum_{t=1}^{T} \tilde{Z}_{Tt}\left(1 - \sqrt{LR_{Tt}}\right) + \frac{1}{2}\tilde{I}_T = o_P(1). \tag{C.14}$$

In Part 2 we show that (C.14) holds in case, under P_T,

$$\sum_{t=1}^{T} \tilde{Z}_{Tt}(h_T' S_{Tt}) - \tilde{I}_T = o_P(1). \tag{C.15}$$

In Part 3 we introduce a new sequence of probability measures (P'_T) and show that it is contiguous to (P_T). Part 4 shows that (C.15) holds under the new sequence (P'_T), hence also under (P_T), which concludes the proof.

Note that Lemma C.4, Condition (e), and Le Cam's first lemma imply that $(\tilde{P}_T)$ and (P_T) are contiguous. It follows that o_P's and O_P's under $(\tilde{P}_T)$ and (P_T) coincide; therefore, in the sequel, we safely can write o_P and O_P without specifying whether $(\tilde{P}_T)$ or (P_T) is the underlying sequence of probability measures.

Part 1. Recalling that $\tilde{I}_T := \sum_{t=1}^{T} \tilde{\iota}_{Tt} := \sum_{t=1}^{T} \mathrm{E}_{P_T}\left[(h_T' S_{Tt})Z_{Tt} \mid \mathcal{F}_{T,t-1}\right]$, we have

$$\sum_{t=1}^{T}\left\{\tilde{Z}_{Tt} - Z_{Tt} + \tilde{\iota}_{Tt}\right\} = \sum_{t=1}^{T}\tilde{Z}_{Tt}\left(1 - \sqrt{LR_{Tt}}\right) + \frac{1}{2}\tilde{I}_T$$

$$+ \sum_{t=1}^{T}\left\{\tilde{Z}_{Tt}\sqrt{LR_{Tt}} - Z_{Tt} - \mathrm{E}_{P_T}\left[\tilde{Z}_{Tt}\sqrt{LR_{Tt}} \mid \mathcal{F}_{T,t-1}\right]\right\}$$

$$+ \sum_{t=1}^{T}\left\{\mathrm{E}_{P_T}\left[\tilde{Z}_{Tt}\sqrt{LR_{Tt}} \mid \mathcal{F}_{T,t-1}\right] + \frac{1}{2}\tilde{\iota}_{Tt}\right\};$$

hence, (C.14) implies (C.13) in case

$$\sum_{t=1}^{T}\left\{\tilde{Z}_{Tt}\sqrt{LR_{Tt}} - Z_{Tt} - \mathrm{E}_{P_T}\left[\tilde{Z}_{Tt}\sqrt{LR_{Tt}} \mid \mathcal{F}_{T,t-1}\right]\right\} = o_P(1) \tag{C.16}$$

and

$$\sum_{t=1}^{T}\left\{\mathrm{E}_{P_T}\left[\tilde{Z}_{Tt}\sqrt{LR_{Tt}} \mid \mathcal{F}_{T,t-1}\right] + \frac{1}{2}\tilde{\iota}_{Tt}\right\} = o_P(1). \tag{C.17}$$

As (C.16) is implied by Condition (f) and Lemma C.2 (recall that $\mathrm{E}_{P_T}[Z_{Tt} \mid \mathcal{F}_{T,t-1}] = 0$), we only need to show that (C.17) holds in order to complete Part 1. We have

$$\sum_{t=1}^{T}\mathrm{E}_{P_T}\left[\tilde{Z}_{Tt}\sqrt{LR}_{Tt} \mid \mathcal{F}_{T,t-1}\right] = \sum_{t=1}^{T}\mathrm{E}_{P_T}\left[Z_{Tt}(1 - \sqrt{LR}_{Tt}) \mid \mathcal{F}_{T,t-1}\right]$$

22

$$+ \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[(\tilde{Z}_{Tt}\sqrt{LR_{Tt}} - Z_{Tt})(1 - \sqrt{LR}_{Tt}) \mid \mathcal{F}_{T,t-1} \right]$$

$$+ \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[\tilde{Z}_{Tt}LR_{Tt} \mid \mathcal{F}_{T,t-1} \right]$$

$$= -\frac{1}{2}\tilde{I}_T - \frac{1}{2}r_T^{(1)} + r_T^{(2)} + r_T^{(3)},$$

with

$$r_T^{(1)} = \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[Z_{Tt}R_{Tt} \mid \mathcal{F}_{T,t-1} \right],$$

$$r_T^{(2)} = \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[(\tilde{Z}_{Tt}\sqrt{LR_{Tt}} - Z_{Tt})(1 - \sqrt{LR}_{Tt}) \mid \mathcal{F}_{T,t-1} \right], \text{ and}$$

$$r_T^{(3)} = \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[\tilde{Z}_{Tt}LR_{Tt} \mid \mathcal{F}_{T,t-1} \right].$$

Starting with $r_T^{(1)}$,

$$|r_T^{(1)}|^2 \leq \left(\sum_{t=1}^{T} \sqrt{\mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[Z_{Tt}^2 \mid \mathcal{F}_{T,t-1} \right]} \sqrt{\mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[R_{Tt}^2 \mid \mathcal{F}_{T,t-1} \right]} \right)^2$$

$$\leq \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[Z_{Tt}^2 \mid \mathcal{F}_{T,t-1} \right] \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[R_{Tt}^2 \mid \mathcal{F}_{T,t-1} \right],$$

so that (C.5) and Condition (g) imply $r_T^{(1)} = o_P(1)$. In the same way (C.4), (C.5) and Condition (f) yield $r_T^{(2)} = o_P(1)$. As for $r_T^{(3)}$, since $\mathrm{E}_{\tilde{\mathrm{P}}_T} \left[\tilde{Z}_{Tt}|\mathcal{F}_{T,t-1} \right] = 0$ we obtain, using (C.5) and Condition (g) again,

$$|r_T^{(3)}|^2 = \left| \sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathrm{P}}_T} \left[\tilde{Z}_{Tt}1_{N_{Tt}}|\mathcal{F}_{T,t-1} \right] \right|^2$$

$$\leq \sum_{t=1}^{T} \mathrm{E}_{\tilde{\mathrm{P}}_T} \left[\tilde{Z}_{Tt}^2|\mathcal{F}_{T,t-1} \right] \sum_{t=1}^{T} (1 - \mathrm{E}_{\mathrm{P}_{\mathrm{T}}} \left[LR_{Tt} \mid \mathcal{F}_{T,t-1} \right]) = o_P(1).$$

Part 2. We have

$$\left| \sum_{t=1}^{T} \tilde{Z}_{Tt}(1 - \sqrt{LR_{Tt}}) + \frac{1}{2}\sum_{t=1}^{T} \tilde{Z}_{Tt}(h_T' S_{Tt}) \right| = \frac{1}{2} \left| \sum_{t=1}^{T} \tilde{Z}_{Tt}R_{Tt} \right| \leq \frac{1}{2} \sqrt{\sum_{t=1}^{T} \tilde{Z}_{Tt}^2} \sqrt{\sum_{t=1}^{T} R_{Tt}^2}.$$

Now, by (C.10), $\sum_{t=1}^{T} R_{Tt}^2 = o_P(1)$ and, by Conditions (g) and (h) and an application of Hall and Heyde (1980, Theorem 2.23), $\sum_{t=1}^{T} \tilde{Z}_{Tt}^2 = O_P(1)$. Hence, (C.14) follows from (C.15).

23

Part 3. For all $T \in \mathbb{N}$, define a new sequence of probability measures $(\mathrm{P}'_{\mathrm{Tt}})_{\mathrm{t}=1}^{\mathrm{T}}$ on $\mathcal{F}_{Tt}$, absolutely continuous with respect to P_{Tt}, with density

$$\frac{\mathrm{dP}'_{\mathrm{Tt}}}{\mathrm{dP}_{\mathrm{Tt}}} := \prod_{s=1}^{t} \sqrt{LR_{Ts}}\; c_{Ts}$$

with, for $s = 1, \ldots, T$, $c_{Ts}^{-1} := \mathrm{E}_{\mathrm{P_T}}\left[\sqrt{LR}_{Ts} \mid \mathcal{F}_{T,s-1}\right]$. Note that the probability that all c_{Ts}^{-1} are strictly positive tends to one, since (C.5) implies

$$\lim_{T \to \infty} \mathrm{P_T}\left(\exists s \in \{1, \ldots, \mathrm{T}\} : c_{Ts}^{-1} = 0\right) \le \lim_{T \to \infty} \mathrm{P_T}\left(\sum_{\mathrm{t}=1}^{\mathrm{T}}(1 - \mathrm{E}_{\mathrm{P_T}}[LR_{Tt}|\mathcal{F}_{T,t-1}]) \ge 1\right) = 0.$$

In the sequel, we thus safely can ignore the event $\{\exists s \in \{1, \ldots, T\} : c_{Ts}^{-1} = 0\}$. Define $\mathrm{P}'_\mathrm{T} := \mathrm{P}'_{\mathrm{TT}}$ and note that $\mathrm{P}'_{\mathrm{Tt}}$ is the restriction of P'_T to $\mathcal{F}_{T,t}$. Because of (C.2), we have $c_{Ts}^{-1} = 1 + \frac{1}{2}\mathrm{E}_{\mathrm{P_T}}\left[R_{Tt} \mid \mathcal{F}_{T,t-1}\right]$. This yields, using an expansion of $\log(1+x)$, (C.5), and (C.11),

$$\sum_{t=1}^{T} \log c_{Tt}^{-1} = -\frac{1}{8} h'_T J_T h_T + o_P(1).$$

Moreover, an application of Lemma C.3 and (C.5) yields $\max_{t=1,\ldots,T} |c_{Tt}^{-1} - 1| = o_P(1)$, and thus also

$$\max_{t=1,\ldots,T} |c_{Tt} - 1| = o_P(1). \tag{C.18}$$

Inserting (C.7) and recalling that $\log LR_{T0} = o_P(1)$, we obtain, under $\mathrm{P_T}$,

$$\log \frac{\mathrm{dP}'_\mathrm{T}}{\mathrm{dP_T}} = \frac{1}{2}\sum_{t=1}^{T} \log LR_{Tt} - \sum_{t=1}^{T} \log c_{Tt}^{-1} + o_P(1) = \frac{1}{2}\sum_{t=1}^{T} h'_T S_{Tt} - \frac{1}{8} h'_T \tilde{I}_T h_T + o_P(1).$$

Condition (e) and Le Cam's first lemma entail that (P'_T) and $(\mathrm{P_T})$ are mutually contiguous. This completes Part 3 of the proof.

Part 4. Let us show that, under the measures (P'_T),

$$\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}'_\mathrm{T}}\left[\tilde{Z}_{Tt}(h'_T S_{Tt}) \mid \mathcal{F}_{T,t-1}\right] = \tilde{I}_T + o_P(1) \tag{C.19}$$

and

$$\sum_{t=1}^{T} \tilde{Z}_{Tt}(h'_T S_{Tt}) = \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}'_\mathrm{T}}\left[\tilde{Z}_{Tt}(h'_T S_{Tt}) \mid \mathcal{F}_{T,t-1}\right] + o_P(1). \tag{C.20}$$

Since $o_P(1)$'s under (P'_T) are $o_P(1)$'s under the contiguous $(\mathrm{P_T})$ too, a combination of these two results yields (C.15) and concludes the proof.

Starting with (C.19), we have

$$\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P}'_\mathrm{T}}\left[\tilde{Z}_{Tt}(h'_T S_{Tt}) \mid \mathcal{F}_{T,t-1}\right] = \sum_{t=1}^{T} c_{Tt}\mathrm{E}_{\mathrm{P_T}}\left[\tilde{Z}_{Tt}\sqrt{LR}_{Tt}(h'_T S_{Tt}) \mid \mathcal{F}_{T,t-1}\right]$$

24

$$= \tilde{I}_T + \sum_{t=1}^{T}(c_{Tt} - 1)\mathrm{E}_{\mathrm{P_T}}\left[Z_{Tt}(h_T' S_{Tt}) \mid \mathcal{F}_{T,t-1}\right]$$

$$+ \sum_{t=1}^{T} c_{Tt}\mathrm{E}_{\mathrm{P_T}}\left[(\tilde{Z}_{Tt}\sqrt{LR}_{Tt} - Z_{Tt})(h_T' S_{Tt}) \mid \mathcal{F}_{T,t-1}\right].$$

Condition (f) and (C.18) imply (C.19) since $\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}}\left[(h_T' S_{Tt})^2 \mid \mathcal{F}_{T,t-1}\right] = O_P(1)$ (see (C.4)) and $\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P_T}}\left[Z_{Tt}^2 \mid \mathcal{F}_{T,t-1}\right] = O_P(1)$ (see Condition (g)).

Turning to (C.20), first note that $\sum_{t=1}^{T}(h_T' S_{Tt})^2 = O_P(1)$ and $\sum_{t=1}^{T} \tilde{Z}_{Tt}^2 = O_P(1)$ by an application of Hall and Heyde (1980, Theorem 2.23) and (C.3), (C.4), Condition (g) and Condition (h), respectively. Hence $\sum_{t=1}^{T} |\tilde{Z}_{Tt}||h_T' S_{Tt}| = O_P(1)$ and $\sum_{t=1}^{T} \mathrm{E}_{\mathrm{P'_T}}[|\tilde{Z}_{Tt}||h_T' S_{Tt}| \mid \mathcal{F}_{T,t-1}] = O_P(1)$. Let $\epsilon, \delta > 0$. In view of the previous remarks, we can find B and T_1 such that, for $T \geq T_1$, $\mathrm{P'_T}(\mathcal{A}_{\delta}^{(\mathrm{T})}) \leq \delta/6$ for

$$\mathcal{A}_{\delta}^{(T)} := \left\{ \sum_{t=1}^{T} \left| (h_T' S_{Tt})\tilde{Z}_{Tt} - \mathrm{E}_{\mathrm{P'_T}}\left[(h_T' S_{Tt})\tilde{Z}_{Tt} \mid \mathcal{F}_{T,t-1}\right] \right| > B \right\}.$$

Set $\eta = \min\{1, \sqrt{\delta}\epsilon(108(B+2))^{-1/2}\}$ and introduce the event

$$\mathcal{A}_{\eta,Tt} := \left\{|Z_{Tt}| \leq \eta\right\} \bigcap \left\{|h_T' S_{Tt}| \leq \eta\right\}.$$

Decompose

$$\sum_{t=1}^{T} \tilde{Z}_{Tt}(h_T' S_{Tt}) - \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P'_T}}\left[\tilde{Z}_{Tt}(h_T' S_{Tt}) \mid \mathcal{F}_{T,t-1}\right] = p_T^{(1)} - p_T^{(2)} + p_T^{(3)},$$

with

$$p_T^{(1)} := \sum_{t=1}^{T} \tilde{Z}_{Tt}(h_T' S_{Tt})1_{\mathcal{A}_{\eta,Tt}^c},$$

$$p_T^{(2)} := \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P'_T}}\left[\tilde{Z}_{Tt}(h_T' S_{Tt})1_{\mathcal{A}_{\eta,Tt}^c} \mid \mathcal{F}_{T,t-1}\right], \text{ and}$$

$$p_T^{(3)} := \sum_{t=1}^{T} \tilde{Z}_{Tt}(h_T' S_{Tt})1_{\mathcal{A}_{\eta,Tt}} - \sum_{t=1}^{T} \mathrm{E}_{\mathrm{P'_T}}\left[\tilde{Z}_{Tt}(h_T' S_{Tt})1_{\mathcal{A}_{\eta,Tt}} \mid \mathcal{F}_{T,t-1}\right].$$

Let us show that there exists $T^\star$ such that, for all $T \geq T^\star$, $\mathrm{P'_T}\left(|p_T^{(i)}| > \epsilon/3\right) \leq \delta/3$, which, as $\epsilon > 0$ and $\delta > 0$ can be taken arbitrarily small, yields (C.20). Applying Hall and Heyde (1980, Theorem 2.23) and (C.4), (C.3), Condition (g) and Condition (h), we obtain

$$\sum_{t=1}^{T} \tilde{Z}_{Tt}^2 1\{|\tilde{Z}_{Tt}| > \eta\} + \sum_{t=1}^{T}(h_T' S_{Tt})^2 1\{|h_T' S_{Tt}| > \eta\} = o_P(1).$$

25

This yields, using (C.4) and Condition (g) again,

$$|p_T^{(1)}| \leq \sqrt{\sum_{t=1}^T (h_T' S_{Tt})^2 1\{|h_T' S_{Tt}| > \eta\}} \sqrt{\sum_{t=1}^T \tilde{Z}_{Tt}^2} + \sqrt{\sum_{t=1}^T (h_T' S_{Tt})^2} \sqrt{\sum_{t=1}^T \tilde{Z}_{Tt}^2 1\{|\tilde{Z}_{Tt}| > \eta\}}$$

$$= o_P(1).$$

From (C.3), (C.4), Condition (g) and Condition (h), we also obtain

$$|p_T^{(2)}| \leq \sqrt{\sum_{t=1}^T c_{Tt}^2 \mathrm{E}_{\tilde{P}_T}\left[\tilde{Z}_{Tt}^2 1\{|\tilde{Z}_{Tt}| > \eta\} \mid \mathcal{F}_{T,t-1}\right]} \sqrt{\sum_{t=1}^T \mathrm{E}_{P_T}\left[(h_T' S_{Tt})^2 \mid \mathcal{F}_{T,t-1}\right]}$$

$$+ \sqrt{\sum_{t=1}^T c_{Tt}^2 \mathrm{E}_{\tilde{P}_T}\left[\tilde{Z}_{Tt}^2 \mid \mathcal{F}_{T,t-1}\right]} \sqrt{\sum_{t=1}^T \mathrm{E}_{P_T}\left[(h_T' S_{Tt})^2 1\{|h_T' S_{Tt}| > \eta\} \mid \mathcal{F}_{T,t-1}\right]} = o_P(1).$$

Hence, there exists T_2 such that, for all $T \geq T_2$, $\mathrm{P}_T'\left(|p_T^{(j)}| > \epsilon/3\right) \leq \delta/3$ for $j = 1, 2$. Next, define the martingales

$$\left\{ A_{Tt} := \sum_{s=1}^t \left\{ \tilde{Z}_{Tt}(h_T' S_{Tt}) 1_{\mathcal{A}_{\eta,Tt}} - \mathrm{E}_{P_T'}[\tilde{Z}_{Tt}(h_T' S_{Tt}) 1_{\mathcal{A}_{\eta,Tt}} \mid \mathcal{F}_{T,s-1}]\right\} : 1 \leq t \leq T \right\},$$

the stopping times $\mathcal{S}^{(T)} := \inf\left\{ t \in \mathbb{N} \mid \sum_{s=1}^t |\Delta A_{Ts}| > B\right\}$, and the processes

$$\left\{ M_{Tt} := A_{T,t \wedge \mathcal{S}^{(T)}} : 1 \leq t \leq T\right\},$$

namely, the stopped versions of the martingales $\{A_{Tt} : 1 \leq t \leq T\}$—which thus also are martingales. Note that $|\Delta A_{Tt}| \leq 2\eta^2$. We obtain

$$\mathrm{E}_{P_T'} M_{TT}^2 = \sum_{t=1}^T \mathrm{E}_{P_T'}(M_{Tt} - M_{T,t-1})^2 \leq \mathrm{E}_{P_T'}\left[\sum_{t=1}^{\mathcal{S}^{(T)}} (\Delta A_{Tt})^2\right]$$

$$\leq 2\eta^2 \mathrm{E}_{P_T'}\left[\sum_{t=1}^{\mathcal{S}^{(T)}} |\Delta A_{Tt}|\right] \leq 2\eta^2 (B + 2\eta^2).$$

So, for $T \geq T_1$, we have

$$\mathrm{P}_T'\left(|p_T^{(3)}| > \epsilon/3\right) = \mathrm{P}_T'\left(|A_{TT}| > \epsilon/3\right) \leq \mathrm{P}_T'\left(M_{TT} \neq A_{TT}\right) + \mathrm{P}_T'\left(|M_{TT}| > \epsilon/3\right)$$

$$\leq \mathrm{P}_T'(\mathcal{S}^{(T)} \leq T) + \mathrm{P}_T'\left(|M_{TT}| > \epsilon/3\right)$$

$$\leq \mathrm{P}_T'(\mathcal{A}_\delta^{(T)}) + \mathrm{P}_T'\left(|M_{TT}| > \epsilon/3\right) \leq \frac{\delta}{6} + \frac{18\eta^2 (B + 2)}{\epsilon^2} \leq \frac{\delta}{3}.$$

Letting $T^\star := \max\{T_1, T_2\}$ completes the proof. $\qquad\square$

D. Additional Monte Carlo results

Table 2: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, pseudo-Gaussian test, and the rank-tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 2$, $r_0 \in \{0, 1\}$, $\phi = -0.3$, $f \in \{\mathcal{N}(0, \Sigma_2), t_3(\Sigma_2), t_{10}(\Sigma_2)\}$, $\Sigma_2 \in \{I_2, \Sigma_{2,c}\}$.

	Sample size and innovation distribution								
	$T = 100$			$T = 250$			$T = 500$		
Test	$\mathcal{N}(0, I_2)$	$t_3(I_2)$	$t_{10}(I_2)$	$\mathcal{N}(0, I_2)$	$t_3(I_2)$	$t_{10}(I_2)$	$\mathcal{N}(0, I_2)$	$t_3(I_2)$	$t_{10}(I_2)$
$r_0 = 0$									
maxeig	0.053	0.055	0.052	0.051	0.052	0.048	0.051	0.052	0.049
trace	0.055	0.057	0.054	0.053	0.056	0.049	0.052	0.052	0.050
$Q_{\dagger}^{(T)}$	0.047	0.044	0.044	0.051	0.047	0.048	0.049	0.048	0.048
$Q_{\phi}^{(T)}$	0.045	0.045	0.043	0.051	0.047	0.046	0.048	0.049	0.048
$Q_{t_3}^{(T)}$	0.050	0.048	0.048	0.051	0.048	0.047	0.048	0.048	0.047
$Q_{t_{10}}^{(T)}$	0.049	0.048	0.046	0.052	0.048	0.047	0.048	0.050	0.049
$r_0 = 1$									
maxeig	0.050	0.057	0.051	0.051	0.054	0.049	0.051	0.052	0.049
trace	0.050	0.057	0.051	0.051	0.054	0.049	0.051	0.052	0.049
$Q_{\dagger}^{(T)}$	0.041	0.037	0.039	0.047	0.048	0.046	0.048	0.049	0.048
$Q_{\phi}^{(T)}$	0.039	0.038	0.036	0.047	0.047	0.045	0.047	0.048	0.047
$Q_{t_3}^{(T)}$	0.042	0.045	0.043	0.048	0.048	0.045	0.049	0.050	0.047
$Q_{t_{10}}^{(T)}$	0.041	0.042	0.040	0.047	0.048	0.044	0.049	0.050	0.048
	$T = 100$			$T = 250$			$T = 500$		
Test	$\mathcal{N}(0, \Sigma_{2,c})$	$t_3(\Sigma_{2,c})$	$t_{10}(\Sigma_{2,c})$	$\mathcal{N}(0, \Sigma_{2,c})$	$t_3(\Sigma_{2,c})$	$t_{10}(\Sigma_{2,c})$	$\mathcal{N}(0, \Sigma_{2,c})$	$t_3(\Sigma_{2,c})$	$t_{10}(\Sigma_{2,c})$
$r_0 = 0$									
maxeig	0.052	0.055	0.052	0.050	0.051	0.049	0.050	0.053	0.050
trace	0.055	0.056	0.055	0.052	0.052	0.049	0.051	0.053	0.049
$Q_{\dagger}^{(T)}$	0.047	0.040	0.044	0.051	0.047	0.048	0.049	0.048	0.048
$Q_{\phi}^{(T)}$	0.045	0.045	0.043	0.051	0.047	0.046	0.048	0.049	0.048
$Q_{t_3}^{(T)}$	0.050	0.048	0.048	0.051	0.048	0.047	0.048	0.048	0.047
$Q_{t_{10}}^{(T)}$	0.049	0.048	0.046	0.052	0.048	0.047	0.048	0.050	0.049
$r_0 = 1$									
maxeig	0.069	0.076	0.067	0.061	0.060	0.059	0.054	0.055	0.051
trace	0.069	0.076	0.067	0.061	0.060	0.059	0.054	0.055	0.051
$Q_{\dagger}^{(T)}$	0.047	0.044	0.044	0.052	0.048	0.050	0.048	0.048	0.048
$Q_{\phi}^{(T)}$	0.042	0.045	0.041	0.050	0.049	0.048	0.048	0.048	0.047
$Q_{t_3}^{(T)}$	0.048	0.053	0.048	0.049	0.052	0.048	0.049	0.051	0.050
$Q_{t_{10}}^{(T)}$	0.047	0.049	0.045	0.051	0.052	0.048	0.049	0.049	0.048

Table 3: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, pseudo-Gaussian test, and the rank-tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 3$, $r_0 \in \{0, 1, 2\}$, $\phi = -0.3$, $f \in \{\mathcal{N}(0, \Sigma_3), t_3(\Sigma_3), t_{10}(\Sigma_3)\}$, $\Sigma \in \{I_3, \Sigma_{3,c}\}$.

	Sample size and innovation distribution								
	$T = 100$			$T = 250$			$T = 500$		
Test	$\mathcal{N}(0, I_3)$	$t_3(I_3)$	$t_{10}(I_3)$	$\mathcal{N}(0, I_3)$	$t_3(I_3)$	$t_{10}(I_3)$	$\mathcal{N}(0, I_3)$	$t_3(I_3)$	$t_{10}(I_3)$
$r_0 = 0$									
$maxeig$	0.054	0.064	0.056	0.054	0.056	0.053	0.050	0.055	0.052
$trace$	0.057	0.064	0.058	0.054	0.056	0.053	0.050	0.054	0.051
$Q_\dagger^{(T)}$	0.046	0.041	0.044	0.048	0.043	0.051	0.048	0.048	0.050
$Q_\phi^{(T)}$	0.044	0.042	0.043	0.048	0.045	0.050	0.047	0.048	0.050
$Q_{t_3}^{(T)}$	0.047	0.045	0.046	0.049	0.046	0.051	0.050	0.047	0.050
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.046	0.052	0.048	0.048	0.050
$r_0 = 1$									
$maxeig$	0.042	0.042	0.039	0.052	0.050	0.052	0.052	0.050	0.051
$trace$	0.042	0.045	0.041	0.051	0.052	0.051	0.050	0.052	0.052
$Q_\dagger^{(T)}$	0.034	0.030	0.034	0.045	0.041	0.047	0.048	0.046	0.048
$Q_\phi^{(T)}$	0.032	0.033	0.032	0.044	0.042	0.047	0.047	0.046	0.047
$Q_{t_3}^{(T)}$	0.036	0.041	0.040	0.047	0.046	0.048	0.048	0.050	0.048
$Q_{t_{10}}^{(T)}$	0.034	0.037	0.036	0.046	0.045	0.048	0.048	0.047	0.049
$r_0 = 2$									
$maxeig$	0.045	0.050	0.047	0.049	0.048	0.049	0.049	0.050	0.049
$trace$	0.045	0.050	0.047	0.049	0.048	0.049	0.049	0.050	0.049
$Q_\dagger^{(T)}$	0.034	0.031	0.035	0.045	0.043	0.045	0.047	0.048	0.048
$Q_\phi^{(T)}$	0.031	0.032	0.033	0.044	0.044	0.043	0.046	0.049	0.046
$Q_{t_3}^{(T)}$	0.037	0.040	0.040	0.046	0.046	0.045	0.049	0.051	0.047
$Q_{t_{10}}^{(T)}$	0.033	0.036	0.035	0.045	0.046	0.044	0.047	0.050	0.047
	$T = 100$			$T = 250$			$T = 500$		
Test	$\mathcal{N}(0, \Sigma_{3,c})$	$t_3(\Sigma_{3,c})$	$t_{10}(\Sigma_{3,c})$	$\mathcal{N}(0, \Sigma_{3,c})$	$t_3(\Sigma_{3,c})$	$t_{10}(\Sigma_{3,c})$	$\mathcal{N}(0, \Sigma_{3,c})$	$t_3(\Sigma_{3,c})$	$t_{10}(\Sigma_{3,c})$
$r_0 = 0$									
$maxeig$	0.054	0.065	0.056	0.055	0.056	0.053	0.050	0.056	0.052
$trace$	0.056	0.065	0.058	0.053	0.057	0.052	0.051	0.054	0.050
$Q_\dagger^{(T)}$	0.046	0.041	0.044	0.048	0.043	0.051	0.048	0.048	0.050
$Q_\phi^{(T)}$	0.044	0.042	0.043	0.048	0.045	0.050	0.047	0.048	0.050
$Q_{t_3}^{(T)}$	0.047	0.045	0.046	0.049	0.046	0.051	0.050	0.047	0.050
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.046	0.052	0.048	0.048	0.050
$r_0 = 1$									
$maxeig$	0.066	0.067	0.067	0.058	0.057	0.058	0.054	0.052	0.054
$trace$	0.073	0.072	0.073	0.061	0.060	0.062	0.055	0.055	0.054
$Q_\dagger^{(T)}$	0.040	0.038	0.038	0.047	0.041	0.046	0.048	0.046	0.047
$Q_\phi^{(T)}$	0.039	0.044	0.037	0.045	0.044	0.045	0.047	0.047	0.047
$Q_{t_3}^{(T)}$	0.045	0.050	0.045	0.048	0.048	0.048	0.048	0.048	0.049
$Q_{t_{10}}^{(T)}$	0.042	0.048	0.041	0.047	0.047	0.047	0.048	0.049	0.049
$r_0 = 2$									
$maxeig$	0.069	0.073	0.069	0.057	0.058	0.058	0.054	0.056	0.053
$trace$	0.069	0.073	0.069	0.057	0.058	0.058	0.054	0.056	0.053
$Q_\dagger^{(T)}$	0.040	0.039	0.039	0.047	0.047	0.048	0.050	0.049	0.048
$Q_\phi^{(T)}$	0.037	0.042	0.037	0.046	0.047	0.047	0.048	0.050	0.047
$Q_{t_3}^{(T)}$	0.044	0.052	0.045	0.048	0.052	0.049	0.050	0.052	0.050
$Q_{t_{10}}^{(T)}$	0.041	0.047	0.040	0.047	0.050	0.049	0.049	0.052	0.049

Table 4: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, pseudo-Gaussian test, and the rank-tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 5$, $r_0 \in \{0, \dots, 4\}$, $\phi = -0.3$, and $f \in \{\mathcal{N}(0, I_5), t_3(I_5), t_{10}(I_5)\}$.

| Test | Sample size and innovation distribution | | | | | | | | |
	$T = 100$			$T = 250$			$T = 500$		
	$\mathcal{N}(0, I_5)$	$t_3(I_5)$	$t_{10}(I_5)$	$\mathcal{N}(0, I_5)$	$t_3(I_5)$	$t_{10}(I_5)$	$\mathcal{N}(0, I_5)$	$t_3(I_5)$	$t_{10}(I_5)$
$r_0 = 0$									
maxeig	0.063	0.083	0.065	0.053	0.065	0.055	0.049	0.061	0.054
trace	0.070	0.082	0.072	0.058	0.063	0.058	0.054	0.059	0.055
$Q_\dagger^{(T)}$	0.044	0.038	0.043	0.047	0.047	0.047	0.048	0.047	0.048
$Q_\phi^{(T)}$	0.044	0.043	0.043	0.047	0.049	0.046	0.047	0.048	0.048
$Q_{t_3}^{(T)}$	0.045	0.042	0.046	0.050	0.047	0.047	0.048	0.048	0.049
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.048	0.047	0.047	0.047	0.049
$r_0 = 1$									
maxeig	0.017	0.029	0.016	0.048	0.058	0.049	0.052	0.058	0.052
trace	0.029	0.037	0.029	0.050	0.057	0.051	0.050	0.057	0.054
$Q_\dagger^{(T)}$	0.025	0.023	0.023	0.043	0.042	0.041	0.046	0.045	0.047
$Q_\phi^{(T)}$	0.024	0.029	0.022	0.042	0.044	0.042	0.045	0.047	0.046
$Q_{t_3}^{(T)}$	0.030	0.037	0.030	0.043	0.047	0.044	0.047	0.048	0.047
$Q_{t_{10}}^{(T)}$	0.026	0.033	0.025	0.043	0.045	0.044	0.046	0.048	0.047
$r_0 = 2$									
maxeig	0.010	0.017	0.011	0.047	0.054	0.046	0.050	0.056	0.051
trace	0.017	0.026	0.019	0.049	0.055	0.048	0.051	0.055	0.051
$Q_\dagger^{(T)}$	0.018	0.018	0.017	0.039	0.039	0.040	0.043	0.043	0.046
$Q_\phi^{(T)}$	0.016	0.021	0.018	0.039	0.042	0.040	0.043	0.048	0.045
$Q_{t_3}^{(T)}$	0.025	0.033	0.027	0.042	0.047	0.043	0.044	0.050	0.047
$Q_{t_{10}}^{(T)}$	0.019	0.027	0.020	0.039	0.045	0.042	0.044	0.047	0.045
$r_0 = 3$									
maxeig	0.020	0.024	0.019	0.048	0.054	0.047	0.050	0.053	0.049
trace	0.021	0.027	0.021	0.049	0.054	0.048	0.049	0.054	0.049
$Q_\dagger^{(T)}$	0.017	0.018	0.018	0.039	0.038	0.038	0.042	0.045	0.046
$Q_\phi^{(T)}$	0.016	0.022	0.017	0.038	0.041	0.039	0.042	0.046	0.045
$Q_{t_3}^{(T)}$	0.025	0.036	0.027	0.041	0.048	0.042	0.046	0.049	0.045
$Q_{t_{10}}^{(T)}$	0.019	0.026	0.020	0.039	0.043	0.039	0.044	0.047	0.045
$r_0 = 4$									
maxeig	0.040	0.047	0.038	0.044	0.048	0.045	0.046	0.049	0.050
trace	0.040	0.047	0.038	0.044	0.048	0.045	0.046	0.049	0.050
$Q_\dagger^{(T)}$	0.026	0.025	0.024	0.040	0.040	0.040	0.044	0.046	0.047
$Q_\phi^{(T)}$	0.024	0.026	0.023	0.039	0.042	0.039	0.043	0.047	0.047
$Q_{t_3}^{(T)}$	0.031	0.039	0.032	0.041	0.046	0.044	0.045	0.048	0.048
$Q_{t_{10}}^{(T)}$	0.026	0.031	0.026	0.038	0.044	0.042	0.045	0.048	0.047

29

Figure 8: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 0 \text{ versus } H' : r = 1}$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_2), t_3(I_2), t_{10}(I_2)\}$.
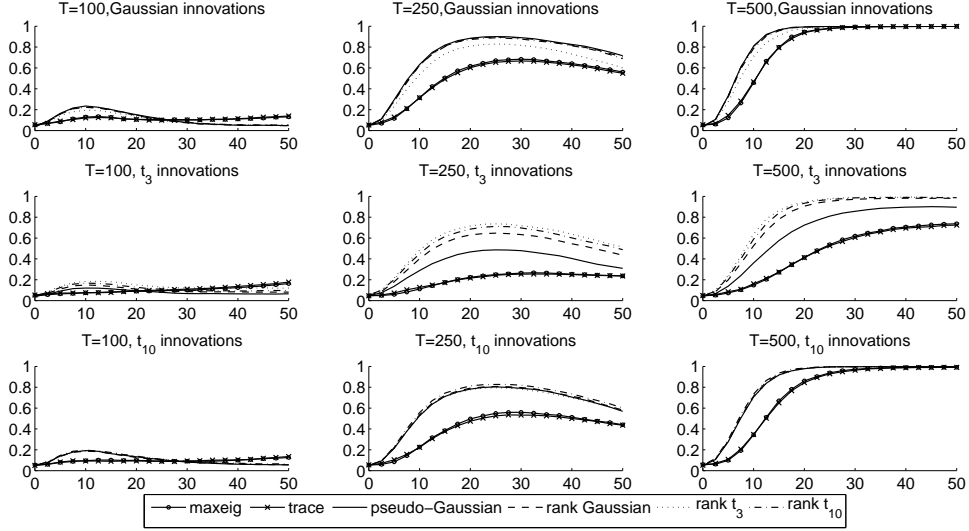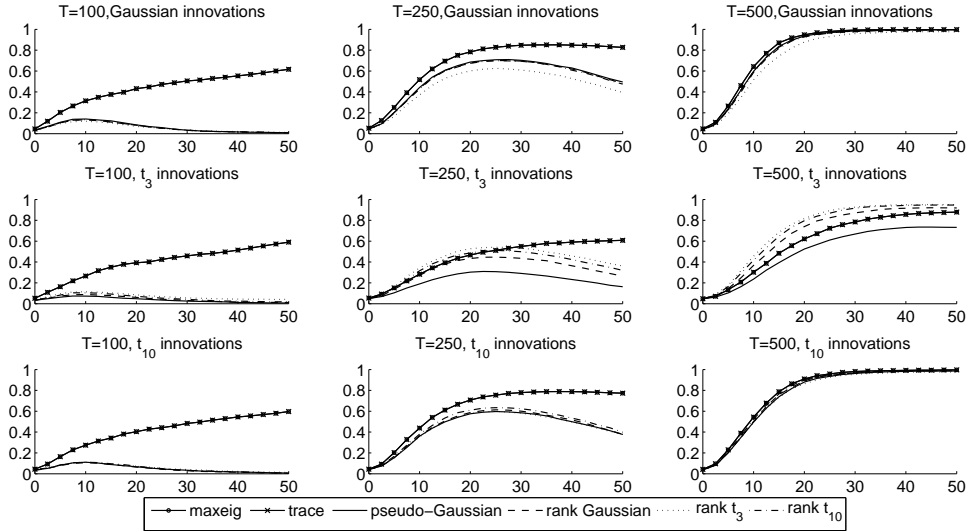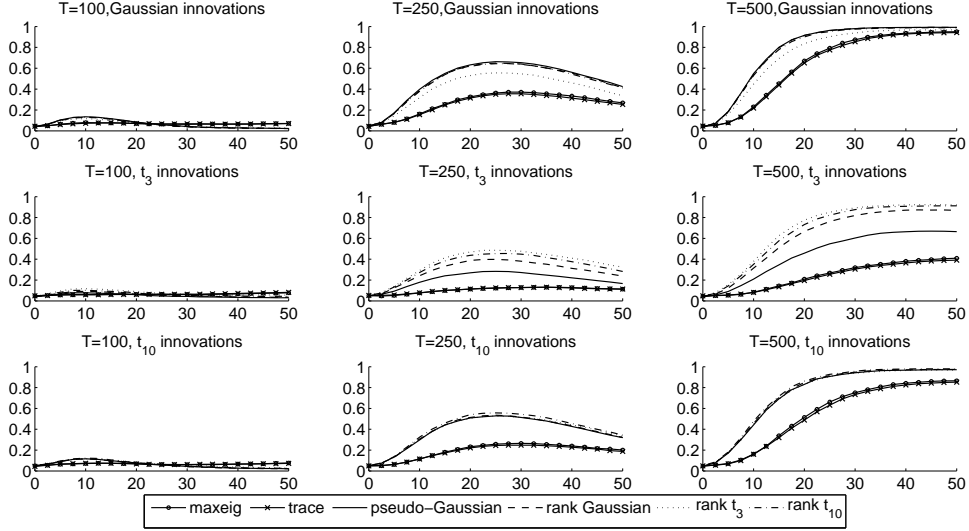

Figure 9: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 1 \text{ versus } H' : r = 2}$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_2), t_3(I_2), t_{10}(I_2)\}$.

30

Figure 10: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 0 \text{ versus } H' : r = 1}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{2,c}), t_3(\Sigma_{2,c}), t_{10}(\Sigma_{2,c})\}$.
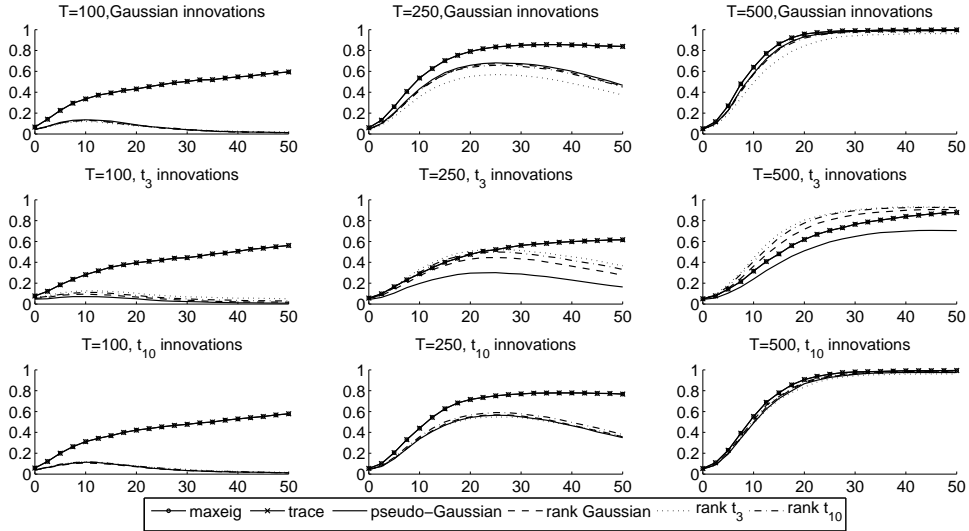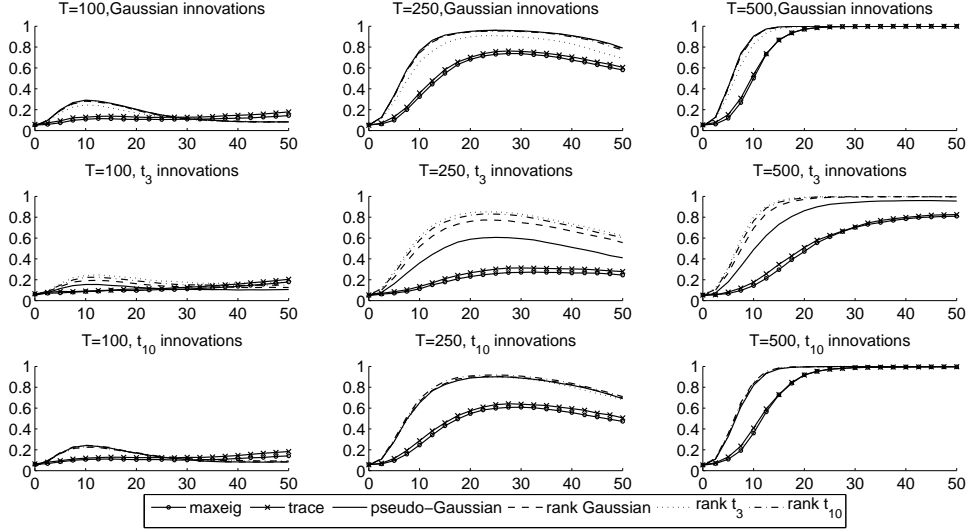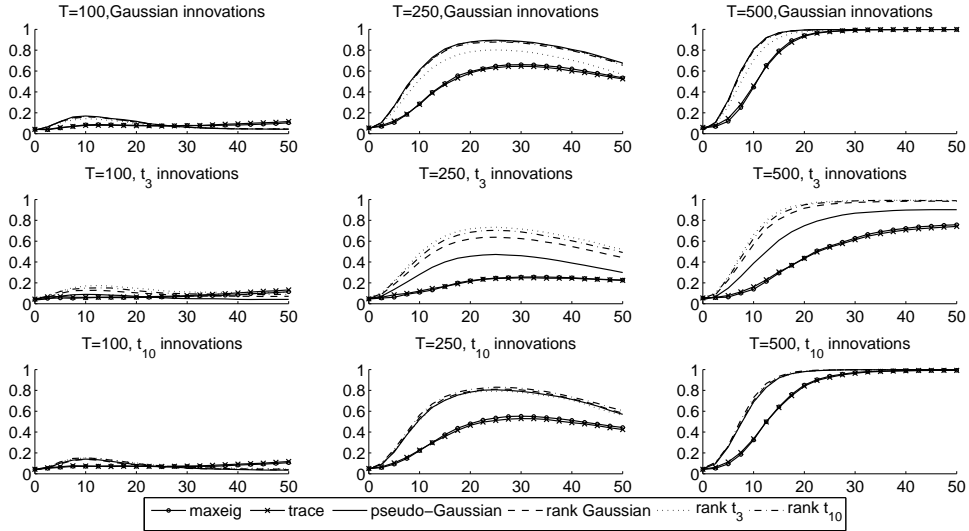

Figure 11: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 1 \text{ versus } H' : r = 2}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{2,c}), t_3(\Sigma_{2,c}), t_{10}(\Sigma_{2,c})\}$.
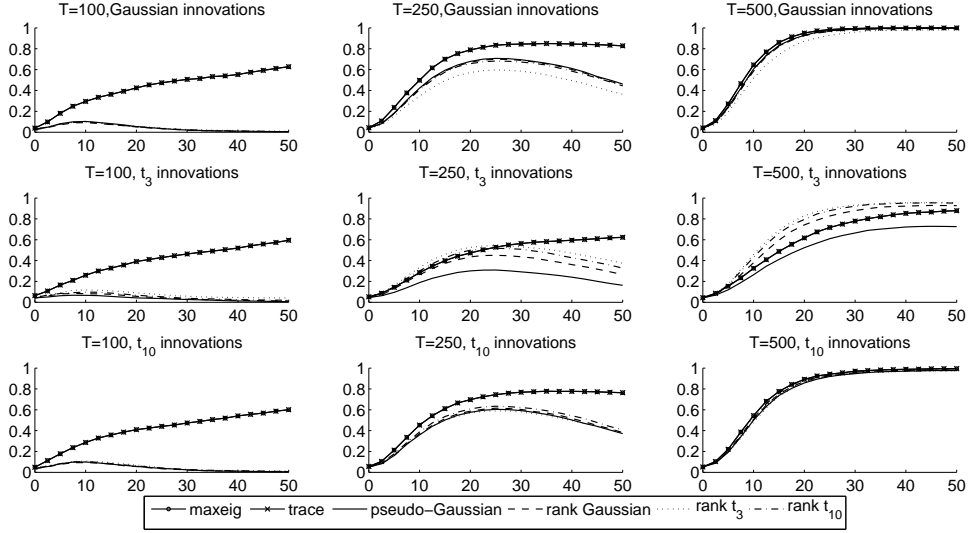

31

Figure 12: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 0 \text{ versus } H' : r = 1}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_3), t_3(I_3), t_{10}(I_3)\}$.

Figure 13: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 1 \text{ versus } H' : r = 2}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_3), t_3(I_3), t_{10}(I_3)\}$.
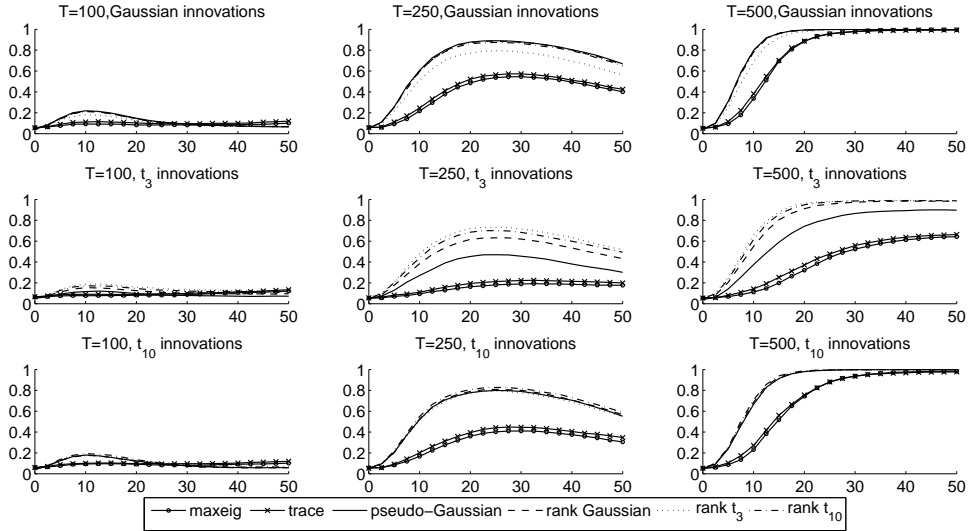

32

Figure 14: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 2 \text{ versus } H' : r = 3}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_3), t_3(I_3), t_{10}(I_3)\}$.
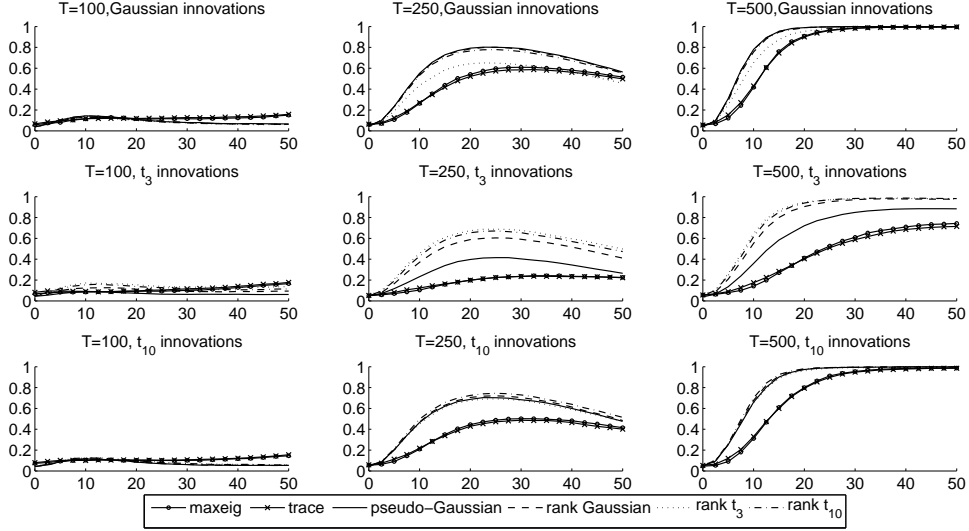

Figure 15: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 0 \text{ versus } H' : r = 1}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{3,c}), t_3(\Sigma_{3,c}), t_{10}(\Sigma_{3,c})\}$.
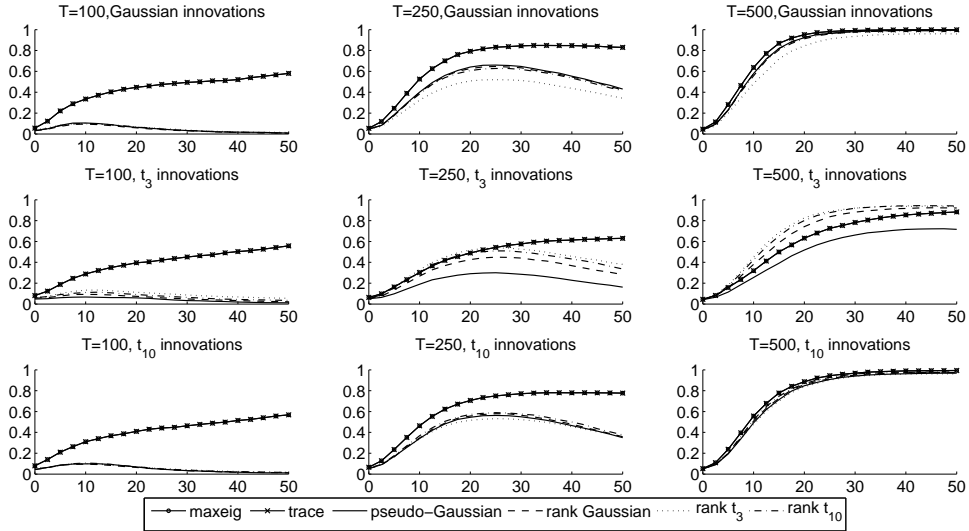

33

Figure 16: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 1 \text{ versus } H' : r = 2}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{3,c}), t_3(\Sigma_{3,c}), t_{10}(\Sigma_{3,c})\}$.

Figure 17: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 2 \text{ versus } H' : r = 3}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, \Sigma_{3,c}), t_3(\Sigma_{3,c}), t_{10}(\Sigma_{3,c})\}$.
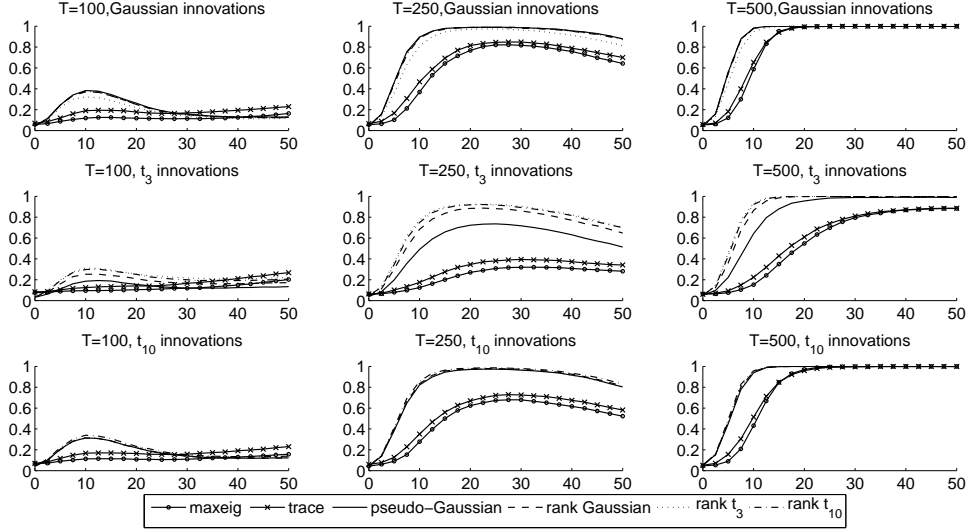

34

Figure 18: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 0 \text{ versus } H' : r = 1}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_5), t_3(I_5), t_{10}(I_5)\}$.
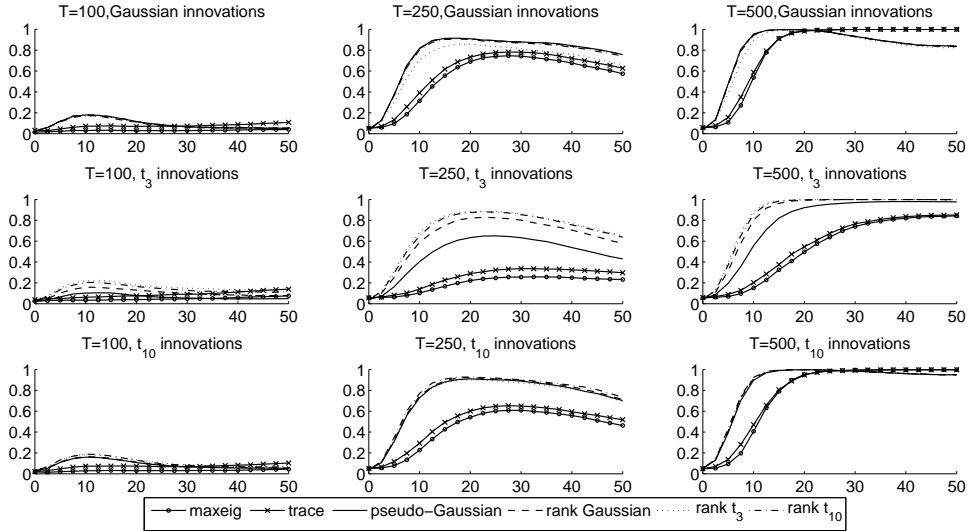

Figure 19: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 1 \text{ versus } H' : r = 2}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_5), t_3(I_5), t_{10}(I_5)\}$.
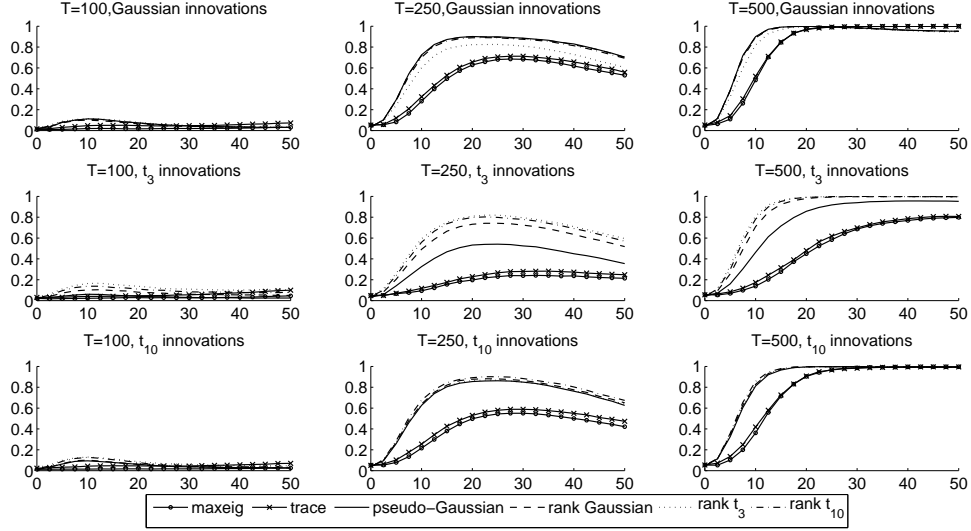

35

Figure 20: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 2 \text{ versus } H' : r = 3}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_5), t_3(I_5), t_{10}(I_5)\}$.

Figure 21: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 3 \text{ versus } H' : r = 4}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_5), t_3(I_5), t_{10}(I_5)\}$.
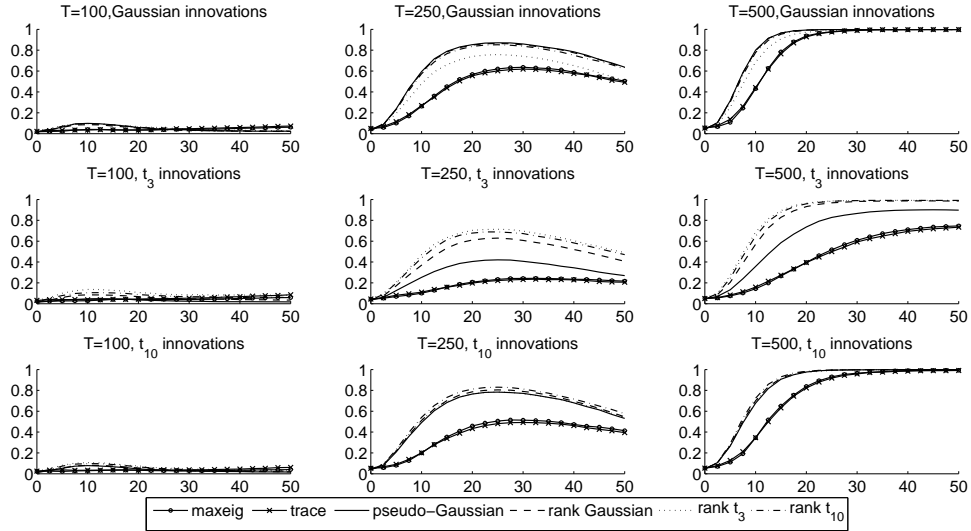

36

Figure 22: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.3), $g \in \{\phi, t_3, t_{10}\}$, for testing $\underline{H : r = 4 \text{ versus } H' : r = 5}$ under (5.2), for $h \in \{0, 2.5, 5, \ldots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, and $f \in \{\mathcal{N}(0, I_5), t_3(I_5), t_{10}(I_5)\}$.

37