T E C H N I C A L
R E P O R T

0428

OPTIMAL TESTS FOR NON-CORRELATION BETWEEN MULTIVARIATE TIME SERIES

HALLIN, M. and A. SAIDI

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Optimal Tests for Non-Correlation between Multivariate Time Series

Marc Hallin*
I.S.R.O., E.C.A.R.E.S., and Département de Mathématique
Université Libre de Bruxelles
Brussels, Belgium

and

Abdessamad Saidi†
Département de mathématiques et de statistique
Université de Montréal
Montréal, Canada

Abstract

The problem of testing non-correlation between two multivariate time series, $\{\mathbf{X}_t^{(1)},\ t \in \mathbb{Z}\}$, with values in $\mathbb{R}^{d_1}$, and $\{\mathbf{X}_t^{(2)}, t \in \mathbb{Z}\}$, with values in $\mathbb{R}^{d_2}$, is considered. Assuming that the global process $\{\mathbf{X}_t, t \in \mathbb{Z}\} := \{((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T)^T, t \in \mathbb{Z}\}$ admits a joint vector autoregressive (VAR) representation, we first show that the hypothesis of non-correlation between $\{\mathbf{X}_t^{(1)}\}$ and $\{\mathbf{X}_t^{(2)}\}$ is equivalent to the hypothesis that all off-diagonal blocks in the matrix coefficients and the innovation covariance in the joint VAR representation are zero. Then, we establish an adequate local asymptotic normality (LAN) property for this VAR model in the vicinity of the hypothesis of non-correlation. This LAN structure allows for constructing locally and asymptotically optimal pseudo-Gaussian tests for the null hypothesis of non-correlation between $\{\mathbf{X}_t^{(1)}\}$ and $\{\mathbf{X}_t^{(2)}\}$, and for comparing their local asymptotic powers with those of the various tests (Haugh-El Himdi-Roy, and Koch-Yang-Hallin-Saidi) proposed in the literature.

KEY WORDS: Non-correlation tests; Local asymptotic normality; Multivariate autoregressive moving average model; Time series; VAR.

*Supported by an A.R.C. project of the Communauté Française de Belgique, and a PAI of the Belgian federal Government.

†Supported by a P.I.P. contract of the Administration Générale de la Coopération au Développement of the Belgian federal Government. Part of this work was completed as a postdoctoral researcher at the Université de Montréal, under grants from the Natural Science and Engineering Research Council of Canada, the Network of Centres of Excellence on The Mathematics of Information Technology and Complex Systems (MITACS), and the Institut de Finance Mathématique de Montréal (IFM2).

1 Introduction.

Detecting cross-dependencies or cross-correlations between series of observations is an essential issue in practically all fields dealing with the analysis of time series data, such as econometrics, meteorology, seismology, hydrology, clinical monotoring, or environmetrics. From a methodological point of view, tests of independence or non-correlation play a fundamental role in the model-building process, certainly so in a multivariate context. The same tests also are the main tool in the analysis of such phenomenons as Granger-causality, etc.

Despite its importance, however, this fundamental problem of testing independence or non-correlation between observed series has received relatively little attention, and, when it has, the solutions proposed are essentially heuristic.

In the univariate case, Haugh (1976) developed a procedure for testing non-correlation between two ARIMA series. Non-correlation under ARIMA assumptions is equivalent to non-correlation between the respective innovation processes of the series under study. Denoting by $\hat{\eta}_t^{(1)}$ and $\hat{\eta}_t^{(2)}$ the residuals resulting from fitting ARIMA models to each of the two series separately, and by $r_{\hat{\boldsymbol{\eta}}}^{(12)}(k)$ the corresponding empirical cross-correlations at lag k, information on possible dependencies between the two series can be expected to be contained in vectors of the form

$$\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M) := \left(r_{\hat{\boldsymbol{\eta}}}^{(12)}(-M), \ldots, r_{\hat{\boldsymbol{\eta}}}^{(12)}(0), \ldots, r_{\hat{\boldsymbol{\eta}}}^{(12)}(M)\right)^T.$$

Haugh shows that, under the null hypothesis of non-correlation between the original series, $\sqrt{N}\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M)$ is asymptotically multivariate normal, with mean zero and identity covariance matrix (N denotes the series lengths). This result leads to the definition of a *portmanteau* test ϕ_H^M, say, which rejects the null hypothesis of non-correlation at significance level α whenever

$$Q_H^M := N \sum_{k=-M}^{M} \left(r_{\hat{\boldsymbol{\eta}}}^{(12)}(k)\right)^2 \tag{1.1}$$

exceeds the $(1-\alpha)$ quantile $\chi^2_{2M+1,1-\alpha}$ of the chi-square distribution with $(2M+1)$ degrees of freedom.

Haugh's test is a purely heuristic portmanteau type test, which is not directed at any specific alternative. Pierce (1977) already observed that its power against some alternatives appearing in an econometric context is extremely low. This weakness of Haugh's procedure was further substantiated in an extensive study conducted by Geweke (1981). In order to palliate this lack of power, Koch and Yang (1986) proposed another test, denoted as $\phi_{KY;i}^M$, based on statistics of the form

$$Q_{KY;i}^M := N \sum_{k=-M}^{M} \left[\sum_{l=0}^{i} r_{\hat{\boldsymbol{\eta}}}^{(12)}(k+l)\right]^2, \quad i = 1, 2, \ldots \tag{1.2}$$

On the basis of a Monte Carlo study, Koch and Yang conclude that their test is preferable to Haugh's against a wide range of alternatives. However, they do not provide any theoretical power or optimality argument.

The multivariate version of this problem was not considered until recently. El Himdi and Roy (1997) generalized the approach of Haugh (1976) to the case of two multivariate ARMA series. A multivariate version of the Koch and Yang procedure has been proposed in Hallin and Saidi (2003).

Surprisingly, little is known about the respective performances of the proposed tests. To the best of our knowledge, and apart from Monte-Carlo investigations, the only attempt to compare local powers in this context is Geweke (1981)'s derivation of Bahadur approximate slopes for Haugh's univariate test. As for optimality issues, it seems they have not been addressed so far.

The purpose of this paper is to investigate this problem via the Le Cam Local Asymptotic Normality (LAN) approach, and to propose an optimal pseudo-Gaussian solution, that is, a test which is locally asymptotically optimal, in the Le Cam sense, under Gaussian assumptions, but remains valid under a broad class of non-Gaussian innovation densities. This approach, as we shall see, does not only provide optimal tests, but also allows for deriving the asymptotic distributions of the various existing test statistics under local alternatives, hence for a computation of their respective asymptotic relative efficiencies (AREs). Now, optimality only makes sense against some specified class of alternatives: the alternatives we are considering here are of the joint vector autoregressive (VAR) type.

LAN for linear time series models was established in the univariate AR case with linear trend by Swensen (1985), in the ARMA case by Kreiss (1987); a multivariate version of these results is given by Garel and Hallin (1995). Still in the univariate case, a more general approach, allowing for nonlinearities, has been taken in Hwang and Basawa (1993), in Drost, Klaassen, and Werker (1997), and in several papers by Koul and Schick (1996, 1997); see Taniguchi and Kakizawa (2000) for a survey of LAN for time series. The LAN result we need here however is more delicate, as it combines the features of location and scale parameters, in a multivariate setting.

This LAN property is established in Section 2. In Section 3, we derive the locally asymptotically most stringent test for testing independence/non-correlation under innovation density f. The form of this test regrettably implies that its validity is limited to the innovation density f for which it is optimal. This density being unspecified in applications, such tests are of little practical interest. Fortunately, Gaussian densities are an exception, indicating that pseudo-Gaussian tests are possible. The methods by El Himdi and Roy (1997) and Hallin and Saidi (2003) are briefly presented in Section 4.1. In Section 4.2 we derive their asymptotic powers under local alternatives. The particular case of testing for non-correlation in the bivariate Gaussian VAR(1) model is described in details in Section 5, where a numerical investigation is conducted.

Boldface throughout denote vectors and matrices; the superscript ^{T} indicates transpose; vec$\mathbf{A}$ as usual stands for the vector resulting from stacking the columns of a matrix $\mathbf{A}$ on top of each other, Diag$\mathbf{A}$ for the diagonal matrix whose diagonal elements are those of $\mathbf{A}$, and $\mathbf{A} \otimes \mathbf{B}$ for the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. The vech operator stacks the elements lying on and below the main diagonal of a square matrix; an *elimination matrix* of order m is a $(m(m+1)/2 \times m^2)$ matrix $\boldsymbol{\mathcal{L}}_m$ such that, for any real $(m \times m)$ matrix $\mathbf{A}$, vech$\mathbf{A} = \boldsymbol{\mathcal{L}}_m$ vec$\mathbf{A}$; finally a *lower duplication matrix* of order m is a $(m^2 \times m(m+1)/2)$ matrix $\boldsymbol{\mathcal{D}}_m$ such that, for any lower triangular $(m \times m)$ matrix $\mathbf{A}$, vec$\mathbf{A} = \boldsymbol{\mathcal{D}}_m$ vech$\mathbf{A}$.

2 Local asymptotic normality.

2.1 Notation and main assumptions.

Let $\mathbf{X} := \{\mathbf{X}_t = ((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T)^T, t \in \mathbb{Z}\}$ denote a d-variate process partitioned into $\mathbf{X}^{(1)} := \{\mathbf{X}_t^{(1)}, t \in \mathbb{Z}\}$, with values in $\mathbb{R}^{d_1}$, and $\mathbf{X}^{(2)} := \{\mathbf{X}_t^{(2)}, t \in \mathbb{Z}\}$, with values in $\mathbb{R}^{d_2}$, $d_1 + d_2 = d$. Throughout the paper, $\mathbf{X}$ is assumed to be a centered vector autoregressive VAR(p) process, satisfying a stochastic difference equation of the form

$$\mathbf{X}_t - \sum_{j=1}^{p} \mathbf{A}_j \mathbf{X}_{t-j} = \mathbf{a}_t, \quad t \in \mathbb{Z}, \tag{2.1}$$

where $\mathbf{A}_j$, $j = 1, ..., p$ are $d \times d$ real matrices, and

(A1) $\{\mathbf{a}_t, t \in \mathbb{Z}\}$ is d-variate white noise, i.e., a process of independent, identically distributed (iid) random vectors with mean zero, positive definite covariance matrix $\mathbf{\Sigma}$, and probability density $g_{\mathbf{\Sigma}}$ such that $\boldsymbol{\varepsilon}_t := \mathbf{A}_0^{-1}\mathbf{a}_t$ has density $f := g_{\mathbf{I}}$, where $\mathbf{A}_0 := \mathbf{\Sigma}^{\frac{1}{2}}$ is the unique lower triangular root of $\mathbf{\Sigma}$, with positive diagonal entries; see, for instance, Graybill (1983, Theorem 8.6.2).

Equation (2.1) is easily rewritten as

$$\mathbf{X}_t = \sum_{j=1}^{p} \mathbf{A}_j \mathbf{X}_{t-j} + \mathbf{A}_0 \boldsymbol{\varepsilon}_t. \tag{2.2}$$

The partition of $\mathbf{X}$ into $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ induces a partition of $\mathbf{a}_t$ and $\boldsymbol{\varepsilon}_t$ into $\mathbf{a}_t = (\mathbf{a}_t^{(1)T}, \mathbf{a}_t^{(2)T})^T$ and $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_t^{(1)T}, \boldsymbol{\varepsilon}_t^{(2)T})^T$, respectively. Similarly, the matrices $\mathbf{A}_j$, $j = 0, ..., p$, are partitioned into

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{A}_0^{(11)} & \mathbf{0} \\ \mathbf{A}_0^{(21)} & \mathbf{A}_0^{(22)} \end{pmatrix}, \qquad \text{and} \qquad \mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{(11)} & \mathbf{A}_j^{(12)} \\ \mathbf{A}_j^{(21)} & \mathbf{A}_j^{(22)} \end{pmatrix}, \qquad j = 1, ..., p,$$

where $\mathbf{A}_0^{(11)}$ and $\mathbf{A}_0^{(22)}$ are lower triangular.

Denote by

$$\boldsymbol{\theta} := \left(\mathrm{vec}^T \mathbf{A}_1, ..., \mathrm{vec}^T \mathbf{A}_p, \mathrm{vech}^T \mathbf{A}_0\right)^T \tag{2.3}$$

the K-dimensional vector of parameters involved in (2.2); note that

$$K = \frac{d \times (d+1)}{2} + pd^2 = d\frac{(2p+1)d+1}{2}.$$

This new parameterization of the VAR model (2.1) will be convenient in the sequel, as non-correlation between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ reduces to the hypothesis that the parameter $\boldsymbol{\theta}$ in (2.3) lies in some linear subspace of $\mathbb{R}^K$.

Some additional assumptions are needed on the parameter values.

(A2) The roots of the determinantal equation associated with (2.1) all lie outside the unit disk, that is,

$$\left| I_d - \sum_{j=1}^{p} \mathbf{A}_j z^j \right| \neq 0, \quad \forall\, |z| \leq 1,\ z \in \mathbb{C};$$

The subset of parameter values $\boldsymbol{\theta} \in \mathbb{R}^K$ such that (A1) and (A2) hold will be denoted as $\mathbf{\Theta}$.

Proposition 2.1 *Under* (A1)*, and* (A2)*, the following three statements are equivalent:*

(i) $\mathbf{X}^{(1)}$ *and* $\mathbf{X}^{(2)}$ *are mutually orthogonal at all leads and lags;*

(ii) $\mathbf{A}_j = \begin{pmatrix} \times & \mathbf{0} \\ \mathbf{0} & \times \end{pmatrix}$, $\quad j = 1, ..., p, \quad$ *and* $\boldsymbol{\Sigma} = \begin{pmatrix} \times & \mathbf{0} \\ \mathbf{0} & \times \end{pmatrix}$;

(iii) $\mathbf{A}_j = \begin{pmatrix} \times & \mathbf{0} \\ \mathbf{0} & \times \end{pmatrix}$, $\quad j = 1, ..., p, \quad$ *and* $\mathbf{A}_0^{(21)} = \mathbf{0}$;

($\times$ stands for an unspecified submatrix of appropriate dimension, compatible with the assumption that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$). In case $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_t^{(1)T}, \boldsymbol{\varepsilon}_t^{(2)T})^T := \mathbf{A}_0^{-1}\mathbf{a}_t$ is such that $\boldsymbol{\varepsilon}_t^{(1)}$ and $\boldsymbol{\varepsilon}_t^{(2)}$ are independent, then the orthogonality property in statement (i) can be replaced by independence.

Proof. The proof is elementary, and is left to the reader. □

The null hypothesis under which (ii) (equivalently, (iii) or (i)) holds will be denoted as $\mathcal{H}_0$. It follows from Proposition 2.1 that $\mathcal{H}_0$, according to the assumption made on $\boldsymbol{\varepsilon}_t$ (that is, on $g_{\boldsymbol{\Sigma}}$), is to be interpreted as $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ being either uncorrelated or independent (at all leads and lags). Non-correlation being less restrictive than independence, we henceforth refer to $\mathcal{H}_0$ as a hypothesis of non-correlation.

It also follows from Proposition 2.1 that $\mathcal{H}_0$ takes the form of a set of $(2p+1)d_1 \times d_2$ linear restrictions on the parameter value $\boldsymbol{\theta}$. Let $\mathcal{A}$ be the set of all $(p+1)$-tuples $(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_p)$ of $d \times d$ real matrices of the block-diagonal form $\mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{(11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_j^{(22)} \end{pmatrix}$, $j = 0, ..., p$, where $\mathbf{A}_0$ is lower triangular, with positive diagonal elements. Then $\mathcal{H}_0$ holds iff $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, where

$$\boldsymbol{\Theta}_0 := \left\{ \boldsymbol{\theta} = \left(\mathrm{vec}^T\mathbf{A}_1, ..., \mathrm{vec}^T\mathbf{A}_p, \mathrm{vech}^T\mathbf{A}_0\right)^T \in \boldsymbol{\Theta} \middle| \, (\mathbf{A}_0, ..., \mathbf{A}_p) \in \mathcal{A} \right\}$$

is the intersection of $\boldsymbol{\Theta}$ with a $((1/2)(d_1 - d_2)^2 + p(d_1^2 + d_2^2))-$dimensional subspace of $\mathbb{R}^K$.

In order to construct locally optimal tests of $\mathcal{H}_0$, we will need uniform local asymptotic normality (ULAN) with respect to $\boldsymbol{\theta}$ in the vicinity of $\boldsymbol{\Theta}_0$. ULAN of course requires some further regularity assumptions, mainly on the innovation density f. Note that the usual assumption of quadratic mean differentiability of $\mathbf{x} \mapsto f^{\frac{1}{2}}(\mathbf{x})$, requiring the existence of a square integrable vector $\mathbf{D}f^{\frac{1}{2}}$ such that, for all $\mathbf{0} \neq \mathbf{h} \to \mathbf{0}$,

$$\left(\mathbf{h}^T\mathbf{h}\right)^{-1} \int \left(f^{\frac{1}{2}}(\mathbf{x}+\mathbf{h}) - f^{\frac{1}{2}}(\mathbf{x}) - \mathbf{h}^T\mathbf{D}f^{\frac{1}{2}}(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x} \longrightarrow 0 \qquad \text{as} \quad \mathbf{h} \to \mathbf{0}$$

(see Swensen 1985, Kreiss 1987, or Garel and Hallin 1995), is not sufficient here. Indeed, in our model, along with the $\mathbf{A}_j$ matrices $j = 1, ..., p$, which asymptotically (that is, in local limit experiments: see 3.1) plays the role of a multivariate location parameters, we also have to deal with $\mathbf{A}_0$, which plays the role of a multivariate scale. For this reason, the differentiability condition we need also includes some of the features associated with scale families.

Denoting by $\mathcal{M}_d$ the set of all lower triangular $d \times d$ matrices with positive diagonal entries, let

$$\boldsymbol{\Pi} := \left\{ \boldsymbol{\pi} \in \mathbb{R}^{\frac{d(d+3)}{2}} \middle| \boldsymbol{\pi} := \left(\boldsymbol{\mu}^T, \mathrm{vech}^T\mathbf{K}\right)^T, \ \mathbf{K} \in \mathcal{M}_d, \ \boldsymbol{\mu} \in \mathbb{R}^d \right\}$$

and $\boldsymbol{\pi}^* := \left(\mathbf{0}^T, \mathrm{vech}^T(\mathbf{I_d})\right)$, where $\mathbf{I}_d$ stands for the $d \times d$ identity matrix. Consider the parametric family of densities $\mathcal{F}_f := \{f_{\boldsymbol{\pi}}, \boldsymbol{\pi} := \left(\boldsymbol{\mu}^T, \mathrm{vech}^T\mathbf{K}\right)^T \in \boldsymbol{\Pi}\}$, where

$$f_{\boldsymbol{\pi}}(\mathbf{x}) := |\mathbf{K}|^{-1} f(\mathbf{K}^{-1}(\mathbf{x} - \boldsymbol{\mu})).$$

This family $\mathcal{F}_f$ (it will be convenient, if not totally correct, to refer to f as the *innovation density*) is a location-scale family, very much in the spirit of those considered in Drost, Klaassen, and Werker (1997) in their assumptions for ULAN. Then $\boldsymbol{\pi} \mapsto f_{\boldsymbol{\pi}}^{\frac{1}{2}}$ is differentiable in quadratic mean at $\boldsymbol{\pi}^*$ iff there exists a square integrable $\frac{d(d+3)}{2}$-dimensional vector $\boldsymbol{\rho}_{\boldsymbol{\pi}^*}(\mathbf{x})$ such that, for all $\mathbf{0} \neq \mathbf{t} = \left(\mathbf{h}^T, \text{vec}^T\boldsymbol{\delta}\right)^T$ converging to $\mathbf{0}$,

$$\left(\mathbf{t}^T\mathbf{t}\right)^{-1} \int \left(f_{\boldsymbol{\pi}^*+\mathbf{t}}^{\frac{1}{2}}(\mathbf{x}) - f_{\boldsymbol{\pi}^*}^{\frac{1}{2}}(\mathbf{x}) - \mathbf{t}^T\boldsymbol{\rho}_{\boldsymbol{\pi}^*}(\mathbf{x}) f_{\boldsymbol{\pi}^*}^{\frac{1}{2}}(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x} \longrightarrow 0. \tag{2.4}$$

The following assumptions (A3)-(A5) are shown (Lemma 7.6 of van der Vaart (1998) or Proposition 1 of Chapter 2 of Bickel, Klassen, Ritov and Wellner (1993)) to be sufficient for (2.4) to hold at $\boldsymbol{\pi}^*$.

(A3) f is a nowhere vanishing density (with respect to the Lebesgue measure on $\mathbb{R}^d$), with mean zero and identity covariance matrix .

(A4) $\mathbf{x} \mapsto f(\mathbf{x})$ is continuously differentiable, with gradient $\dfrac{\partial f}{\partial \mathbf{x}}$.

Putting $\boldsymbol{\varphi}_f := -\text{grad}\left(\log\left(f\right)\right) = -\dfrac{1}{f}\dfrac{\partial f}{\partial \mathbf{x}}$ (with values in $\mathbb{R}^d$; this gradient exists as soon as (A4) holds) and $\boldsymbol{\varphi}_{f;\boldsymbol{\pi}} := \text{grad}_{\boldsymbol{\pi}}\left(\log\left(f_{\boldsymbol{\pi}}\right)\right) := \dfrac{1}{f_{\boldsymbol{\pi}}}\dfrac{\partial f_{\boldsymbol{\pi}}}{\partial \boldsymbol{\pi}}$ (this gradient, with values in $\mathbb{R}^{\frac{d(d+3)}{2}}$, exists as soon as (A4) holds), define

$$\boldsymbol{\mathcal{I}}(f) := \int \boldsymbol{\varphi}_f(\mathbf{x})\ \boldsymbol{\varphi}_f^T(\mathbf{x})\ f(\mathbf{x})\mathrm{d}\mathbf{x},$$

and

$$\boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}} := \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}^{(11)} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}^{(12)} \\ \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}^{(21)} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}^{(22)} \end{pmatrix} := \int \boldsymbol{\varphi}_{f;\boldsymbol{\pi}}(\mathbf{x})\ \boldsymbol{\varphi}_{f;\boldsymbol{\pi}}^T(\mathbf{x})\ f_{\boldsymbol{\pi}}(\mathbf{x})\mathrm{d}\mathbf{x}.$$

The matrices $\boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}^{(11)}$ and $\boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}^{(22)}$ (with dimensions $d \times d$ and $\frac{d(d+1)}{2} \times \frac{d(d+1)}{2}$, respectively) are to be interpreted as the information matrices corresponding to the location and scale parameters, respectively.

(A5) $\boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}^*}$ is finite and nonsingular, and $\boldsymbol{\pi} \mapsto \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}$ is continuous at $\boldsymbol{\pi}^*$.

Using classical properties of matrix derivatives, one then easily obtains

$$\boldsymbol{\varphi}_{f;\boldsymbol{\pi}}(\mathbf{x}) \quad = \quad \begin{pmatrix} \left(\mathbf{K}^{-1}\right)^T \boldsymbol{\varphi}_f\left(\mathbf{K}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right) \\ \text{vech}\left(-\left(\mathbf{K}^{-1}\right)^T + \left(\mathbf{K}^{-1}\right)^T \boldsymbol{\varphi}_f\left(\mathbf{K}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right)\left(\mathbf{K}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right)^T\right) \end{pmatrix},$$

so that $\boldsymbol{\varphi}_{f;\boldsymbol{\pi}^*}(\mathbf{x}) = \begin{pmatrix} \boldsymbol{\varphi}_f\left(\mathbf{x}\right) \\ \text{vech}\left(-\mathbf{I}_d + \boldsymbol{\varphi}_f\left(\mathbf{x}\right)\mathbf{x}^T\right) \end{pmatrix}$. Hence,

$$\boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}^*} = \int \begin{pmatrix} \boldsymbol{\varphi}_f\left(\mathbf{x}\right) \\ \text{vech}\left(-\mathbf{I}_d + \boldsymbol{\varphi}_f\left(\mathbf{x}\right)\mathbf{x}^T\right) \end{pmatrix} \begin{pmatrix} \boldsymbol{\varphi}_f\left(\mathbf{x}\right) \\ \text{vech}\left(-\mathbf{I}_d + \boldsymbol{\varphi}_f\left(\mathbf{x}\right)\mathbf{x}^T\right) \end{pmatrix}^T f(\boldsymbol{x})d\mathbf{x} =: \begin{pmatrix} \boldsymbol{\mathcal{I}}(f) & \boldsymbol{\mathcal{I}}_{ls}(f) \\ \left(\boldsymbol{\mathcal{I}}_{ls}(f)\right)^T & \boldsymbol{\mathcal{I}}_s(f) \end{pmatrix}. \tag{2.5}$$

Taking this into account, and assuming that (A3) and (A4) hold, Assumption (A5) can be replaced with

(A5') $\boldsymbol{\mathcal{I}}(f)$ is finite and positive definite, and $\int \left(\mathbf{x}_j \boldsymbol{\varphi}_{f;i}(\mathbf{x})\right)^2 f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathrm{E}\left[\left(\boldsymbol{\varepsilon}_{j;t}\boldsymbol{\varphi}_{f;i}(\boldsymbol{\varepsilon}_t)\right)^2\right] < \infty$ for all i and j, where $\mathbf{x}_j$, $\boldsymbol{\varepsilon}_{j;t}$, and $\boldsymbol{\varphi}_{f;i}(\mathbf{x})$ stand for the j^{th} element of the vector $\mathbf{x}$, the j^{th} element of the random vector $\boldsymbol{\varepsilon}_t$, and the i^{th} element of the vector $\boldsymbol{\varphi}_f(\mathbf{x})$, respectively.

2.2 Local asymptotic normality.

In order to establish the required ULAN result, we shall use Theorem 2.1 of Drost, Klaassen, and Werker (1997). The likelihoods we are considering here actually are conditional likelihoods (conditional upon initial values $(\mathbf{X}_{1-p}, ..., \mathbf{X}_0)$; the influence of these initial values under assumption (A2) safely can be neglected—see, e. g., Hallin and Werker (1999)). Denote by $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}$ the distribution, under innovation density f and parameter value $\boldsymbol{\theta}$, of $\mathbf{X}^{(N)} := (\mathbf{X}_1, ..., \mathbf{X}_N)$, conditional on $(\mathbf{X}_{1-p}, ..., \mathbf{X}_0)$. Let $\boldsymbol{\tau}^{(N)} := \left(\mathrm{vec}^T\boldsymbol{\gamma}_1^{(N)}, ..., \mathrm{vec}^T\boldsymbol{\gamma}_p^{(N)}, \mathrm{vech}^T\boldsymbol{\gamma}_0^{(N)}\right)^T \in \mathbb{R}^K$ and $\tilde{\boldsymbol{\tau}}^{(N)} := \left(\mathrm{vec}^T\tilde{\boldsymbol{\gamma}}_1^{(N)}, ..., \mathrm{vec}^T\tilde{\boldsymbol{\gamma}}_p^{(N)}, \mathrm{vech}^T\tilde{\boldsymbol{\gamma}}_0^{(N)}\right)^T \in \mathbb{R}^K$, where both of $\boldsymbol{\gamma}_0^{(N)}$ and $\tilde{\boldsymbol{\gamma}}_0^{(N)}$ are lower triangular matrices, be such that $\|\boldsymbol{\tau}^{(N)}\|$ and $\|\tilde{\boldsymbol{\tau}}^{(N)}\|$ remain bounded as $N \to \infty$. Whenever $\boldsymbol{\tau}^{(N)}$ is a constant, we write $\boldsymbol{\tau} := \left(\mathrm{vec}^T\boldsymbol{\gamma}_1, ..., \mathrm{vec}^T\boldsymbol{\gamma}_p, \mathrm{vech}^T\boldsymbol{\gamma}_0\right)^T$ instead of $\boldsymbol{\tau}^{(N)}$. Defining

$$\boldsymbol{\theta}^{(N)} := \left(\mathrm{vec}^T\mathbf{A}_1^{(N)}, ..., \mathrm{vec}^T\mathbf{A}_p^{(N)}, \mathrm{vech}^T\mathbf{A}_0^{(N)}\right)^T := \boldsymbol{\theta} + N^{-1/2}\boldsymbol{\tau}^{(N)}, \tag{2.6}$$

$$\tilde{\boldsymbol{\theta}}^{(N)} := \left(\mathrm{vec}^T\tilde{\mathbf{A}}_1^{(N)}, ..., \mathrm{vec}^T\tilde{\mathbf{A}}_p^{(N)}, \mathrm{vech}^T\tilde{\mathbf{A}}_0^{(N)}\right)^T := \boldsymbol{\theta}^{(N)} + N^{-1/2}\tilde{\boldsymbol{\tau}}^{(N)}, \tag{2.7}$$

$$\mathbf{e}_t := (\mathbf{A}_0)^{-1}\left(\mathbf{X}_t - \textstyle\sum_{j=1}^p \mathbf{A}_j\mathbf{X}_{t-j}\right),$$

$$\mathbf{e}_t^{(N)} := \left(\mathbf{A}_0^{(N)}\right)^{-1}\left(\mathbf{X}_t - \sum_{j=1}^p \mathbf{A}_j^{(N)}\mathbf{X}_{t-j}\right), \text{ and } \tilde{\mathbf{e}}_t^{(N)} := \left(\tilde{\mathbf{A}}_0^{(N)}\right)^{-1}\left(\mathbf{X}_t - \sum_{j=1}^p \tilde{\mathbf{A}}_j^{(N)}\mathbf{X}_{t-j}\right), \tag{2.8}$$

the logarithm of the likelihood ratio for $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$ against $\mathrm{P}^{(N)}_{f;\tilde{\boldsymbol{\theta}}^{(N)}}$ takes the form

$$\Lambda^{(N)}_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}\left(\mathbf{X}^{(N)}\right) := \log\left(\frac{\mathrm{dP}^{(N)}_{f;\tilde{\boldsymbol{\theta}}^{(N)}}}{\mathrm{dP}^{(N)}_{f;\theta^{(N)}}}\right) = \sum_{t=1}^N \log\left(\left|\mathbf{A}_0^{(N)}\right| f(\tilde{\mathbf{e}}_t^{(N)}) \Big/ \left|\tilde{\mathbf{A}}_0^{(N)}\right| f(\mathbf{e}_t^{(N)})\right). \tag{2.9}$$

Let

$$\mathbf{W}_{N,t} := \begin{pmatrix} (\mathbf{X}_{t-1}\otimes\mathbf{I}_d)\left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T & \mathbf{0}_{d^2\times\frac{d\times(d+1)}{2}} \\ \vdots & \vdots \\ (\mathbf{X}_{t-j}\otimes\mathbf{I}_d)\left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T & \mathbf{0}_{d^2\times\frac{d\times(d+1)}{2}} \\ \vdots & \vdots \\ (\mathbf{X}_{t-p}\otimes\mathbf{I}_d)\left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T & \mathbf{0}_{d^2\times\frac{d\times(d+1)}{2}} \\ \mathbf{0}_{\frac{d(d+1)}{2}\times d} & \boldsymbol{\mathcal{D}}_d^T\left(\mathbf{I}_d\otimes\left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T\right)\boldsymbol{\mathcal{L}}_d^T \end{pmatrix}, \tag{2.10}$$

where $\boldsymbol{\mathcal{D}}_d$ is the usual $d^2 \times \frac{d(d+1)}{2}$ lower duplication matrix and $\boldsymbol{\mathcal{L}}_d$ the $\frac{d(d+1)}{2} \times d^2$ elimination matrix. We now may state the ULAN property which is the main result of this section.

Theorem 2.1 *Suppose that Assumptions* (A1)-(A5) *are satisfied. Let* $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, $\boldsymbol{\theta}^{(N)}$ *and* $\tilde{\boldsymbol{\theta}}^{(N)}$ *as defined in* (2.6) *and* (2.7), *respectively. Then,*

$$\Lambda^{(N)}_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}\left(\mathbf{X}^{(N)}\right) = (\tilde{\boldsymbol{\tau}}^{(N)})^T \boldsymbol{\Delta}^{(N)}_{f;\boldsymbol{\theta}^{(N)}} - \frac{1}{2}(\tilde{\boldsymbol{\tau}}^{(N)})^T \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}}(\boldsymbol{\theta})\tilde{\boldsymbol{\tau}}^{(N)} + o_{\mathrm{P}}(1),$$

under $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$, *as* $N \to \infty$, *with (the central sequence)*

$$\boldsymbol{\Delta}^{(N)}_{f;\boldsymbol{\theta}^{(N)}} := N^{-\frac{1}{2}} \sum_{t=1}^{N} \begin{pmatrix} \left(\mathbf{X}_{t-1} \otimes \left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T\right) \boldsymbol{\varphi}_f(\mathbf{e}_t^{(N)}) \\ \vdots \\ \left(\mathbf{X}_{t-j} \otimes \left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T\right) \boldsymbol{\varphi}_f(\mathbf{e}_t^{(N)}) \\ \vdots \\ \left(\mathbf{X}_{t-p} \otimes \left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T\right) \boldsymbol{\varphi}_f(\mathbf{e}_t^{(N)}) \\ \boldsymbol{\mathcal{D}}_d^T \left(\mathbf{I}_d \otimes \left(\left(\mathbf{A}_0^{(N)}\right)^{-1}\right)^T\right) \boldsymbol{\mathcal{L}}_d^T \mathrm{vech}\left[-\mathbf{I}_d + \boldsymbol{\varphi}_f(\mathbf{e}_t^{(N)})(\mathbf{e}_t^{(N)})^T\right] \end{pmatrix}, \tag{2.11}$$

and (the information matrix)

$$\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}} := \begin{pmatrix} \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{11} & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{1i} & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{1j} & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{1p} & \mathbf{0} \\ \vdots & \ddots & \vdots & & \vdots & & \vdots & \vdots \\ \left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{1i}\right)^T & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{ii} & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{ij} & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{ip} & \mathbf{0} \\ \vdots & & \vdots & \ddots & \vdots & & \vdots & \vdots \\ \left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{1j}\right)^T & \cdots & \left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{ij}\right)^T & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{jj} & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{jp} & \mathbf{0} \\ \vdots & & \vdots & & \vdots & \ddots & \vdots & \vdots \\ \left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{1p}\right)^T & \cdots & \left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{ip}\right)^T & \cdots & \left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{jp}\right)^T & \cdots & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{pp} & \mathbf{0} \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \cdots & \mathbf{0}^T & \cdots & \mathbf{0}^T & \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{00} \end{pmatrix}, \tag{2.12}$$

where

$$\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{00} := \boldsymbol{\mathcal{D}}_d^T \left(\mathbf{I}_d \otimes \left(\mathbf{A}_0^{-1}\right)^T\right) \boldsymbol{\mathcal{L}}_d^T \boldsymbol{\mathcal{I}}_s(f) \boldsymbol{\mathcal{L}}_d \left(\mathbf{I}_d \otimes \left(\mathbf{A}_0^{-1}\right)\right) \boldsymbol{\mathcal{D}}_d,$$

and

$$\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{ij} := \left(\begin{pmatrix} \boldsymbol{\Gamma}^{(11)}_{\boldsymbol{\theta}}(j-i) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}^{(22)}_{\boldsymbol{\theta}}(j-i) \end{pmatrix} \otimes \left(\left(\mathbf{A}_0^{-1}\right)^T \boldsymbol{\mathcal{I}}(f) \mathbf{A}_0^{-1}\right)\right),$$

with

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{(ii)}(k) := \mathrm{E}_{\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}}\left[\mathbf{X}_t^{(i)}\left(\mathbf{X}_{t-k}^{(i)}\right)^T\right], \quad i = 1, 2.$$

Moreover, $\boldsymbol{\Delta}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$ *(still under* $\mathrm{P}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$*, as* $N \to \infty$*) is asymptotically normal, with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}$*.*

Proof. By Theorem 2.1 of Drost, Klaassen, and Werker (1997), ULAN holds if the log-likelihood ratio $\Lambda_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}^{(N)}\left(\mathbf{X}^{(N)}\right)$ can be written as

$$\Lambda_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}^{(N)}\left(\mathbf{X}^{(N)}\right) = \sum_{t=1}^{N}\left\{\log\left(f_{\boldsymbol{\pi}_t^N}(\mathbf{e}_t^{(N)})\right) - \log\left(f_{\boldsymbol{\pi}^*}(\mathbf{e}_t^{(N)}))\right)\right\}, \tag{2.13}$$

where $\boldsymbol{\pi}_t^N = \boldsymbol{\pi}^* + (\mathbf{W}_{N,t})^T\left(\tilde{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}^{(N)}\right)$ with $\mathbf{W}_{N,t}$ defined in (2.10), and Assumptions A-D (same reference) are satisfied. By definition (2.9) we have that

$$\begin{aligned}\Lambda_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}^{(N)}\left(\mathbf{X}^{(N)}\right) &= \sum_{t=1}^{N}\log\left(\left|\mathbf{A}_0^{(N)}\right| f(\tilde{\mathbf{e}}_t^{(N)})\Big/\left|\tilde{\mathbf{A}}_0^{(N)}\right| f(\mathbf{e}_t^{(N)})\right)\\ &= \sum_{t=1}^{N}\log\left(\left|\left(\mathbf{A}_0^{(N)}\right)^{-1}\tilde{\mathbf{A}}_0^{(N)}\right|^{-1} f(\tilde{\mathbf{e}}_t^{(N)})\right) - \log\left(f(\mathbf{e}_t^{(N)})\right).\end{aligned}$$

Thus, in order to show (2.13) we only need to verify that

$$f_{\boldsymbol{\pi}_t^N}(\mathbf{e}_t^{(N)}) = \left|\left(\mathbf{A}_0^{(N)}\right)^{-1}\tilde{\mathbf{A}}_0^{(N)}\right|^{-1} f(\tilde{\mathbf{e}}_t^{(N)})$$

Indeed, from the definition of $\mathbf{W}_{N,t}$ and the properties of the Kronecker product, the vec and the vech operators, we have that

$$\boldsymbol{\pi}_t^N = \begin{pmatrix}\mathbf{0}\\ \mathrm{vech}^T(\mathbf{I}_d)\end{pmatrix} + \begin{pmatrix}\left(\mathbf{A}_0^{(N)}\right)^{-1}\sum_{j=1}^{p}\left(\tilde{\mathbf{A}}_j^{(N)} - \mathbf{A}_j^{(N)}\right)\mathbf{X}_{t-j}\\ \left(\mathbf{A}_0^{(N)}\right)^{-1}\left(\tilde{\mathbf{A}}_0^{(N)} - \mathbf{A}_0^{(N)}\right)\end{pmatrix}.$$

It follows that

$$\begin{aligned}f_{\boldsymbol{\pi}_t^N}(\mathbf{e}_t^{(N)}) &= \left|\mathbf{A}_0^{(N)}\right|\left|\tilde{\mathbf{A}}_0^{(N)}\right|^{-1} f\left(\left(\tilde{\mathbf{A}}_0^{(N)}\right)^{-1}\left(\left(\mathbf{X}_t - \sum_{j=1}^{p}\mathbf{A}_j^{(N)}\mathbf{X}_{t-j}\right) - \sum_{j=1}^{p}\left(\tilde{\mathbf{A}}_j^{(N)} - \mathbf{A}_j^{(N)}\right)\mathbf{X}_{t-j}\right)\right)\\ &= \left|\left(\mathbf{A}_0^{(N)}\right)^{-1}\tilde{\mathbf{A}}_0^{(N)}\right|^{-1} f(\tilde{\mathbf{e}}_t^{(N)}).\end{aligned}$$

Hence, (2.1) in Drost, Klaassen, and Werker (1997) is satisfied.

Turning to their Assumptions A-D, Assumption A follows from the fact that the influence of initial values under (A1) and (A2) is asymptotically negligible, and Assumption B is a direct consequence of (A3)-(A5). That Assumption C holds is a consequence of the fact that the random $pd^2 + \frac{d(d+1)}{2} \times d + \frac{d(d+1)}{2}$ matrix process $\mathbf{W}_{N,t}$, which depends on $\boldsymbol{\theta}^{(N)}$ and is measurable with respect to the past, is square-integrable. As for Assumption D, by Lemma A.1 (same reference), it is sufficient to check that there exists a square-integrable random $\left(pd^2 + \frac{d(d+1)}{2}\right) \times \left(d + \frac{d(d+1)}{2}\right)$ matrix $\mathbf{W}_t(\boldsymbol{\theta}^{(N)})$, measurable with respect to the past, such that (all convergences are under $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}$, as $N \to \infty$)

(i) $N^{-1} \sum_{t=1}^{N} \mathbf{W}_t(\boldsymbol{\theta}) \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}^*} \left(\mathbf{W}_t(\boldsymbol{\theta})\right)^T \xrightarrow{\mathrm{P}} \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}}$,

(ii) for all $\delta > 0$, $N^{-1} \sum_{t=1}^{N} \|\mathbf{W}_t(\boldsymbol{\theta})\|^2 I \left(N^{-\frac{1}{2}} \|\mathbf{W}_t(\boldsymbol{\theta})\| > \delta\right) \xrightarrow{\mathrm{P}} 0$, and

(iii) $\mathbf{W}_{N,t}$ converges to $\mathbf{W}_t(\boldsymbol{\theta}^{(N)})$, in the sense that

$$\sum_{t=1}^{N} \left| \left(\mathbf{W}_{N,t} - \mathbf{W}_t(\boldsymbol{\theta}^{(N)})\right)^T \left(\tilde{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}^{(N)}\right) \right|^2 \xrightarrow{\mathrm{P}} 0$$

An appropriate choice of $\mathbf{W}_t(\boldsymbol{\theta}^{(N)})$ here is $\mathbf{W}_{N,t}$; by assumption (A2) and an ergodicity argument, one easily checks that conditions (i)-(iii) are satisfied.

This completes the proof of the ULAN property. □

Remark 2.1 The construction of locally asymptotically optimal tests will require the computation of the covariance matrix $\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}}$, hence that of the autocovariance matrices $\boldsymbol{\Gamma}^{(ii)}_{\boldsymbol{\theta}}(l)$ $(0 \leq l \leq p)$. Since $\mathbf{X}^{(1)}_t$ and $\mathbf{X}^{(2)}_t$ under the null are stationary AR(p) processes, the Yule-Walker relations

$$\boldsymbol{\Gamma}^{(ii)}_{\boldsymbol{\theta}}(l) := \mathrm{E}_{\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}} \left[\mathbf{X}^{(i)}_t \left(\mathbf{X}^{(i)}_{t-l}\right)^T\right] = \sum_{j=1}^{p} \mathbf{A}^{(ii)}_j \boldsymbol{\Gamma}^{(ii)}_{\boldsymbol{\theta}}(l-j), \qquad l = 1, ..., p,$$

and

$$\boldsymbol{\Gamma}^{(ii)}_{\boldsymbol{\theta}}(0) := \mathrm{E}_{\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}} \left[\mathbf{X}^{(i)}_t \left(\mathbf{X}^{(i)}_t\right)^T\right] = \mathbf{A}^{(ii)}_0 \left(\mathbf{A}^{(ii)}_0\right)^T + \sum_{j=1}^{p} \mathbf{A}^{(ii)}_j \boldsymbol{\Gamma}^{(ii)}_{\boldsymbol{\theta}}(-j),$$

allow for computing $\boldsymbol{\Gamma}^{(ii)}_{\boldsymbol{\theta}}(l)$ for $i = 1, 2$ and $l = 0, ..., p$—see the subroutine COVARS of Shea (1989). The same subroutine also allows, via the recursions

$$\boldsymbol{\delta}^{(ii)}_{\boldsymbol{\theta}}(0) = \mathbf{A}^{(ii)}_0 \qquad\qquad \boldsymbol{\delta}^{(ii)}_{\boldsymbol{\theta}}(l) = \sum_{j=1}^{p} \mathbf{A}^{(ii)}_j \boldsymbol{\delta}^{(ii)}_{\boldsymbol{\theta}}(l-j), \tag{2.14}$$

for a computation of the matrices $\boldsymbol{\delta}^{(ii)}_{\boldsymbol{\theta}}(l) := \mathrm{E}_{\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}} \left[\mathbf{X}^{(i)}_t \left(\boldsymbol{\varepsilon}^{(i)}_{t-l}\right)^T\right]$ entering the expression of local asymptotic powers (see Proposition 4.1).

3 Locally asymptotically most stringent test.

3.1 Weak convergence of statistical experiments.

Local asymptotic normality (LAN) at $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ implies the weak convergence of the sequence of local experiments (localized at $\boldsymbol{\theta}$) $\mathcal{E}^{(N)}_f(\boldsymbol{\theta}) := \{\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau}}, \boldsymbol{\tau} \in \mathbb{R}^K\}$ to the K-dimensional Gaussian shift experiment

$$\mathcal{E}_f(\boldsymbol{\theta}) := \left\{\mathcal{N}\left(\boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}} \boldsymbol{\tau}, \boldsymbol{\Gamma}^{\boldsymbol{\Delta}}_{f;\boldsymbol{\theta}}\right) | \boldsymbol{\tau} \in \mathbb{R}^K\right\}.$$

This convergence implies that all power functions that are implementable from the sequence $\mathcal{E}_f^{(N)}(\boldsymbol{\theta})$ converge, as $N \to \infty$, pointwise in $\boldsymbol{\tau}$ but uniformly with respect to the set of all possible testing procedures, to the power functions that are implementable in the limit Gaussian experiment $\mathcal{E}_f(\boldsymbol{\theta})$. Conversely, all risk functions associated with $\mathcal{E}_f(\boldsymbol{\theta})$ can be obtained as limits of sequences of risk functions associated with $\mathcal{E}_f^{(N)}(\boldsymbol{\theta})$. Denoting by $\boldsymbol{\Delta}$ the (k-dimensional) observation in $\mathcal{E}_f(\boldsymbol{\theta})$, it follows that, if a test $\phi(\boldsymbol{\Delta})$ enjoys some exact optimality property in the Gaussian experiment $\mathcal{E}_f(\boldsymbol{\theta})$, then the sequence $\phi\left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right)$ inherits, under local and asymptotic form, the same optimality properties in the sequence of experiments $\mathcal{E}_f^{(N)}(\boldsymbol{\theta})$—see, e.g., Section 11.9 of Le Cam (1986).

3.2 Locally asymptotically most stringent test.

Denote by $\mathbf{Q}$ a $K \times (K-r)$ matrix of maximal rank $K-r$, and by $\mathcal{M}(\mathbf{Q})$ the linear subspace of $\mathbb{R}^K$ spanned by the columns of $\mathbf{Q}$. The null hypothesis $\mathcal{H}_0 : \boldsymbol{\tau} \in \mathcal{M}(\mathbf{Q})$ is equivalent to $\mathcal{H}_0 : \boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}} \boldsymbol{\tau} \in \mathcal{M}(\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}\, \mathbf{Q})$, a set of linear constraints on the location parameter of the Gaussian shift experiment $\mathcal{E}(\boldsymbol{\theta})$. The most stringent α-level test for this problem, consists in rejecting $\mathcal{H}_0$ whenever

$$\boldsymbol{\Delta}^T \left[(\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}})^{-1} - \mathbf{Q} \left(\mathbf{Q}^T \boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}} \mathbf{Q} \right)^{-1} \mathbf{Q}^T \right] \boldsymbol{\Delta} > \chi^2_{r,1-\alpha}, \tag{3.1}$$

where $\chi^2_{r,1-\alpha}$ denotes the $(1-\alpha)$-quantile of a chi-square variable with r degrees of freedom. A locally asymptotically most stringent (at $\boldsymbol{\theta}$) test thus is obtained by substituting $\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}$ for $\boldsymbol{\Delta}$ in (3.1).

In view of Proposition 2.1, the null hypothesis $\mathcal{H}_0$ of non-correlation between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ takes the form $\mathcal{H}_0 : \mathbf{Q}_\perp^T \boldsymbol{\theta} = \mathbf{0}$, with

$$\mathbf{Q}_\perp^T := \begin{pmatrix}
\mathbf{L}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times d^2} & & \cdots & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times \frac{d(d+1)}{2}} \\
\mathbf{S}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times d^2} & & \cdots & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times \frac{d(d+1)}{2}} \\
\mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{L}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times d^2} & \cdots & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times \frac{d(d+1)}{2}} \\
\mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{S}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times d^2} & \cdots & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times \frac{d(d+1)}{2}} \\
\vdots & & & & & \vdots \\
\mathbf{0}_{d_1 d_2 \times d^2} & \cdots & & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{L}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times \frac{d(d+1)}{2}} \\
\mathbf{0}_{d_1 d_2 \times d^2} & \cdots & & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{S}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times \frac{d(d+1)}{2}} \\
\mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{0}_{d_1 d_2 \times d^2} & & \cdots & \mathbf{0}_{d_1 d_2 \times d^2} & \mathbf{L}\boldsymbol{\mathcal{D}}_d
\end{pmatrix}.$$

where

$$\mathbf{L} = \left(\begin{pmatrix} \mathbf{I}_{d_1} \\ \mathbf{0}_{d_2 \times d_1} \end{pmatrix}^T \otimes \begin{pmatrix} \mathbf{0}_{d_2 \times d_1} & \mathbf{I}_{d_2} \end{pmatrix} \right), \qquad \mathbf{S} = \left(\begin{pmatrix} \mathbf{0}_{d_1 \times d_2} \\ \mathbf{I}_{d_2} \end{pmatrix}^T \otimes \begin{pmatrix} \mathbf{I}_{d_1} & \mathbf{0}_{d_1 \times d_2} \end{pmatrix} \right),$$

and $\boldsymbol{\mathcal{D}}_d$ is the $\left(d^2 \times \frac{d(d+1)}{2}\right)$ lower duplication matrix. An alternative form for $\mathcal{H}_0$ is $\mathcal{H}_0 : \boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$, with

$$\mathbf{Q} := \begin{pmatrix} \mathbf{U}_{d^2\times d_1d_1} & \mathbf{V}_{d^2\times d_2d_2} & \mathbf{0}_{d^2\times d_1d_1} & \mathbf{0}_{d^2\times d_2d_2} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}_{d^2\times d_1d_1} & \mathbf{0}_{d^2\times d_2d_2} & \mathbf{U}_{d^2\times d_1d_1} & \mathbf{V}_{d^2\times d_2d_2} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & & \ddots & & & & \\ \mathbf{0}_{d^2\times d_1d_1} & \mathbf{0}_{d^2\times d_2d_2} & \mathbf{0}_{d^2\times d_1d_1} & \mathbf{0}_{d^2\times d_2d_2} & \cdots & \mathbf{U} & \mathbf{V} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}_{\frac{d(d+1)}{2}\times d_1d_1} & \mathbf{0}_{\frac{d(d+1)}{2}\times d_2d_2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{D}}_d^T\mathbf{U}\boldsymbol{\mathcal{L}}_{d_1}^T & \boldsymbol{\mathcal{D}}_d^T\mathbf{V}\boldsymbol{\mathcal{L}}_{d_2}^T \end{pmatrix},$$

where

$$\mathbf{U} = \left(\begin{pmatrix} \mathbf{I}_{d_1} \\ \mathbf{0}_{d_2\times d_1} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{I}_{d_1} \\ \mathbf{0}_{d_2\times d_1} \end{pmatrix}\right), \qquad \mathbf{V} = \left(\begin{pmatrix} \mathbf{0}_{d_1\times d_2} \\ \mathbf{I}_{d_2} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{0}_{d_1\times d_2} \\ \mathbf{I}_{d_2} \end{pmatrix}\right),$$

and $\boldsymbol{\mathcal{L}}_m$ is the $\frac{m(m+1)}{2} \times m^2$ elimination matrix. Referring to (3.1), a sequence of locally (at $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$) asymptotically most stringent α-level tests for the null hypothesis $\mathcal{H}_0$ of non-correlation is $\phi_f^* := \phi_f^*\left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right) := I\left[Q_f^* > \chi^2_{r,1-\alpha}\right]$, where

$$Q_f^* := Q_f^*\left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right) := \left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right)^T \left[{\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)} \tag{3.2}$$

with $r = (2p+1)d_1d_2$. The power of this test against $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau}}$ satisfies

$$\lim_{N\to+\infty} \mathrm{E}_{\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau}}}\left[\phi_f^*\left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right)\right] = 1 - F^r_{\chi^2}\left(\chi^2_{r,1-\alpha}; \psi_f^2(\boldsymbol{\tau},\boldsymbol{\theta})\right), \tag{3.3}$$

where $\psi_f^2(\boldsymbol{\tau},\boldsymbol{\theta}) := \boldsymbol{\tau}^T\left[\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}} - \boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}\mathbf{Q}\left(\mathbf{Q}^T\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\boldsymbol{\Gamma}_{f;\boldsymbol{\theta}}^{\boldsymbol{\Delta}}\right]\boldsymbol{\tau}$ and $F^r_{\chi^2}(.;\psi^2)$ denotes the distribution function of the non central chi-square variable with r degrees of freedom and noncentrality parameter ψ^2.

This test $\phi_f^*\left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right)$ however is of little practical use as long as it explicitly depends on an unspecified parameter value $\boldsymbol{\theta}$. In order to construct a version which is locally and asymptotically optimal at any $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$, let us assume that a sequence of estimators $\hat{\boldsymbol{\theta}}^{(N)}$ is available, with the following properties:

(A6) (i) $\hat{\boldsymbol{\theta}}^{(N)} \in \mathcal{M}(\mathbf{Q})$;

(ii) (root-N consistency) for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and $\epsilon > 0$, there exist $b(\boldsymbol{\theta},\epsilon)$ and $N(\boldsymbol{\theta},\epsilon)$ such that

$$\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}\left[\left\|N^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}\right)\right\| > b(\boldsymbol{\theta},\epsilon)\right] < \epsilon \quad \text{for all} \quad N \geq N(\boldsymbol{\theta},\epsilon);$$

(iii) $\hat{\boldsymbol{\theta}}^{(N)}$ is *locally asymptotically discrete*, that is, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and $c > 0$, there exists $J = J(\boldsymbol{\theta};c)$ such that the number of possible values of $\hat{\boldsymbol{\theta}}^{(N)}$ in balls of the form $\left\{\mathbf{t} \in \mathbb{R}^K : \left\|\sqrt{N}(\mathbf{t}-\boldsymbol{\theta})\right\| \leq c\right\}$ is bounded by J, uniformly as N tends to infinity.

It is a classical result (see, e.g., Chapter 11 of Le Cam (1986)) that, under ULAN (which entails the asymptotic linearity of $\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}$) and Assumptions (A6), substituting $\hat{\boldsymbol{\theta}}^{(N)}$ for $\boldsymbol{\theta}$ has no influence on the asymptotic behavior of $\phi_f^*\left(\boldsymbol{\Delta}_{f;\boldsymbol{\theta}}^{(N)}\right)$, hence on its local asymptotic optimality.

3.3 Pseudo-Gaussian tests.

A fatal shortcoming of the optimal test described in Section 3.2 is that its validity, in general, is limited to innovation densities f. Indeed, the last block of the central sequence has expectation zero under density f only, which means that a perturbation of this density has the same effect (a nonzero noncentrality parameter), asymptotically, as the alternative to be detected.

In practice, f is never specified. Therefore, this optimal test is of little practical value. Fortunately, the Gaussian case is a remarkable exception, as the expectation of $\boldsymbol{\varphi}_f(\mathbf{e}_t^{(N)})(\mathbf{e}_t^{(N)})^T - \mathbf{I}_d$, which, for Gaussian f, reduces to $\mathbf{e}_t^{(N)}(\mathbf{e}_t^{(N)})^T - \mathbf{I}_d$, is $\mathbf{0}$ irrespective of the density f, provided that second order moments are finite. Therefore, in the sequel, we will concentrate on the Gaussian version Q^* of the test statistic Q_f^* described in Section 3.2. The results of that section however will allow for a computation of the asymptotic relative efficiencies, under f, of the resulting pseudo-Gaussian ϕ^* test with respect to the various tests existing in the literature.

4 Asymptotic performance.

4.1 The Haugh-El Himdi-Roy and Koch-Yang-Hallin-Saidi tests.

Though their starting point is slightly different from the one adopted here (they assume that $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ admit individual VAR(p) representations, whereas we start from the joint VAR(p) model (2.1)), Haugh, El Himdi and Roy, Koch and Yang, and Hallin and Saidi all end up testing the same null hypothesis $\mathcal{H}_0$ as we do, namely the block-diagonal form of $\mathbf{A}_0, \ldots \mathbf{A}_p$ in the parameterization (2.2) of (2.1).

The spirit of their approach is essentially Gaussian, as it is entirely based on second-order moments. Denote by $\hat{\mathbf{A}}_j^{(N)}$, $j = 0, \ldots, p$ the estimated values of $\mathbf{A}_j$, $j = 0, \ldots, p$ associated with $\hat{\boldsymbol{\theta}}^{(N)}$, and by $\hat{\mathbf{e}}_t$ the corresponding estimated residuals (obtained from substituting $\hat{\mathbf{A}}_j^{(N)}$ for $\mathbf{A}_j^{(N)}$ in (2.8)). Put $\boldsymbol{\eta}_t := (\mathbf{A}_0)\mathbf{e}_t$, $\hat{\boldsymbol{\eta}}_t := (\hat{\mathbf{A}}_0^{(N)})\hat{\mathbf{e}}_t$,

$$\mathbf{C}_{\hat{\boldsymbol{\eta}}}(k) = \begin{pmatrix} \mathbf{C}_{\hat{\boldsymbol{\eta}}}^{(11)}(k) & \mathbf{C}_{\hat{\boldsymbol{\eta}}}^{(12)}(k) \\ \mathbf{C}_{\hat{\boldsymbol{\eta}}}^{(21)}(k) & \mathbf{C}_{\hat{\boldsymbol{\eta}}}^{(22)}(k) \end{pmatrix} := \begin{cases} N^{-1} \sum_{t=k+1}^{N} \hat{\boldsymbol{\eta}}_t(\hat{\boldsymbol{\eta}}_{t-k})^T & 0 \le k \le N-1 \\ N^{-1} \sum_{t=1}^{N+k} \hat{\boldsymbol{\eta}}_t(\hat{\boldsymbol{\eta}}_{t-k})^T & 1-N \le k \le 0, \end{cases}$$

and

$$\mathbf{R}_{\hat{\boldsymbol{\eta}}}(k) = \begin{pmatrix} \mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(11)}(k) & \mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(12)}(k) \\ \mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(21)}(k) & \mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(22)}(k) \end{pmatrix} := (\mathrm{Diag}\mathbf{C}_{\hat{\boldsymbol{\eta}}}(0))^{-\frac{1}{2}}\mathbf{C}_{\hat{\boldsymbol{\eta}}}(k)(\mathrm{Diag}\mathbf{C}_{\hat{\boldsymbol{\eta}}}(0))^{-\frac{1}{2}}.$$

The El Himdi and Roy test then takes the form $\phi_{HR}^M := I[Q_{HR}^M > \chi^2_{(2M+1)d_1d_2,1-\alpha}]$, where

$$Q_{HR}^M := N \sum_{k=-M}^{M} \left(\mathrm{vec}\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(12)}(k)\right)^T \left(\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(22)}(0) \otimes \mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(11)}(0)\right)^{-1} \left(\mathrm{vec}\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(12)}(k)\right), \tag{4.1}$$

is asymptotically chi-square, with $(2M+1)d_1d_2$ degrees of freedom, under $\mathcal{H}_0$; for $d_1 = d_2 = 1$, it reduces to Haugh's statistic Q_H^M given in (1.1). Moreover, the statistic (4.1) is asymptotically unaffected if we replace $\hat{\boldsymbol{\eta}}_t$ by $\hat{\mathbf{e}}_t$.

Using the same notation, the tests $\phi^M_{HS;i}$ considered in Hallin and Saidi (2003) are based on the statistics

$$Q^M_{HS;i} := \sum_{k=1}^{(2M+1)d_1d_2-i} \left[\sum_{l=0}^{i} \boldsymbol{\nu}_M\,(k+l)\right]^2,\ i = 0, 1, ..., Md_1d_2 - 1, \tag{4.2}$$

where $\boldsymbol{\nu}_M := \sqrt{N}\mathbf{I}_{2M+1} \otimes \left(\mathbf{R}^{(22)}_{\hat{\boldsymbol{\eta}}}(0) \otimes \mathbf{R}^{(11)}_{\hat{\boldsymbol{\eta}}}(0)\right)^{-\frac{1}{2}} \mathbf{r}^{(12)}_{\hat{\boldsymbol{\eta}}}(M)$, with

$$\mathbf{r}^{(12)}_{\hat{\boldsymbol{\eta}}}(M) = \left(\left(\mathrm{vec}\mathbf{R}^{(12)}_{\hat{\boldsymbol{\eta}}}(-M)\right)^T, \ldots, \left(\mathrm{vec}\mathbf{R}^{(12)}_{\hat{\boldsymbol{\eta}}}(M)\right)^T\right)^T. \tag{4.3}$$

For $i = 0$, (4.2) reduces to El Himdi and Roy's test statistic (4.1), whereas, for ($d_1 = d_2 = 1$), it coincides with the Koch and Yang univariate statistic $Q^M_{KY;i}$ given in (1.2). Under the null hypothesis, $Q^M_{HS;i}$ has asymptotically the same distribution as the random variable

$$Q\left(\Lambda^M_i\right) := \sum_{j=1}^{(2M+1)d_1d_2} \lambda^{(i)}_M(j) W^2_{i,j},$$

where the $W_{i,j}$'s are independent standard normal and the $\lambda^{(i)}_M(j)$'s are the eigenvalues of $\mathbf{A}^{(i)}_M = \left(\mathbf{C}^{(i)}_M\right)^T \mathbf{C}^{(i)}_M$, where $\mathbf{C}^{(i)}_M = \left[\mathbf{L}^{(i)}_1, ..., \mathbf{L}^{(i)}_{(2M+1)d_1d_2-i}\right]$ with, writing $\mathbf{1}_{(i+1)\times 1} = (1, \ldots, 1)^T \in \mathbb{R}^{i+1}$ and $\mathbf{0}_{l\times 1} = (0, \ldots, 0)^T \in \mathbb{R}^l$, for $l \in \mathbb{N}^*$, $\mathbf{L}^{(i)}_k = \left(\mathbf{0}^T_{(k-1)\times 1}, \mathbf{1}^T_{i+1\times 1}, \mathbf{0}^T_{(((2M+1)d_1d_2)-(k+i))\times 1}\right)^T \in \mathbb{R}^{(2M+1)d_1d_2}$.

4.2 Local asymptotic powers.

The local asymptotic normality results of Theorem 2.1 also allow, via Le Cam's third lemma (Le Cam 1986), for computing local asymptotic powers, under innovation density f, for the various procedures described in Section 4. In order to derive local powers, we need the joint asymptotic distribution, under $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}$, of $\left(\left(\sqrt{N}\mathbf{r}^{(12)}_{\boldsymbol{\eta}}(M)\right)^T, \Lambda^{(N)}_{\boldsymbol{\theta}^{(N)}/\boldsymbol{\theta}}(\mathbf{X}^{(N)})\right)^T$.

Proposition 4.1 *Assume that the process* (2.1) *satisfies* (A1)-(A5) *and that all fourth-order cumulants of the process* $\boldsymbol{\varepsilon}$ *are zero. Then, under* $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}}$,

$$\begin{pmatrix} \sqrt{N}\mathbf{r}^{(12)}_{\boldsymbol{\eta}}(M) \\ \Lambda^{(N)}_{\boldsymbol{\theta}^{(N)}/\boldsymbol{\theta}}(\mathbf{X}^{(N)}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ -\frac{\nu^2}{2} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right) & \mathbf{V} \\ \mathbf{V}^T & \nu^2 \end{pmatrix}\right),$$

where $\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(i)}}(0) = \left(\mathrm{Diag}\left(\mathbf{A}^{(ii)}_0 \left(\mathbf{A}^{(ii)}_0\right)^T\right)\right)^{-\frac{1}{2}} \left(\mathbf{A}^{(ii)}_0 \left(\mathbf{A}^{(ii)}_0\right)^T\right) \left(\mathrm{Diag}\left(\mathbf{A}^{(ii)}_0 \left(\mathbf{A}^{(ii)}_0\right)^T\right)\right)^{-\frac{1}{2}}$,

$$\begin{aligned} \nu^2 &:= -\left(\mathrm{vec}^T \boldsymbol{\gamma}^T_0 \mathrm{vec}\mathbf{A}^{-1}_0\right)^2 + \mathrm{E}\left[\left(\mathbf{A}^{-1}_0 \boldsymbol{\gamma}_0 \boldsymbol{\varepsilon}_t\right)^T \boldsymbol{\varphi}(\boldsymbol{\varepsilon}_t)\right]^2 \\ &\qquad + \mathrm{tr}\left[\boldsymbol{\mathcal{I}}(f) \sum_{i=1}^{p}\sum_{j=1}^{p} \mathbf{A}^{-1}_0 \boldsymbol{\gamma}_i \begin{pmatrix} \boldsymbol{\Gamma}^{(11)}_{\boldsymbol{\theta}}(j-i) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}^{(22)}_{\boldsymbol{\theta}}(j-i) \end{pmatrix} \boldsymbol{\gamma}^T_j \left(\mathbf{A}^{-1}_0\right)^T\right], \end{aligned}$$

$$\mathbf{V} := (\mathbf{L}\ \mathrm{vec}\mathbf{V}(-M), \ldots, \mathbf{L}\ \mathrm{vec}\mathbf{V}(-1), \mathbf{L}\ \mathrm{vec}\mathbf{V}(0), \mathbf{L}\ \mathrm{vec}\mathbf{V}(1), \ldots, \mathbf{L}\ \mathrm{vec}\mathbf{V}(M))^T, \tag{4.4}$$

with

$$\mathbf{L} := \left(\left(\mathrm{Diag}\left(\mathbf{A}_0^{(22)}\left(\mathbf{A}_0^{(22)}\right)^T\right)\right)^{-\frac{1}{2}}\left(\mathbf{A}_0^{(22)}\right)\right) \otimes \left(\left(\mathrm{Diag}\left(\mathbf{A}_0^{(11)}\left(\mathbf{A}_0^{(11)}\right)^T\right)\right)^{-\frac{1}{2}}\left(\mathbf{A}_0^{(11)}\right)\right),$$

$$\mathbf{V}(0) = \mathrm{E}\left[\boldsymbol{\varepsilon}_t^{(1)}\left((\mathbf{A}_0)^{-1}\boldsymbol{\gamma}_0\boldsymbol{\varepsilon}_t\right)^T\boldsymbol{\varphi}(\boldsymbol{\varepsilon}_t)\left(\boldsymbol{\varepsilon}_t^{(2)}\right)^T\right],$$

and, for $1 \leq k \leq M$, *letting* $\boldsymbol{\delta}_{\boldsymbol{\theta}}^{(ii)}(k-j) := \mathrm{E}_{\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}}\left[\mathbf{X}_{t-j}^{(i)}\left(\boldsymbol{\varepsilon}_{t-k}^{(i)}\right)^T\right]$ *(see (2.14)),*

$$\mathbf{V}(-k) := \sum_{j=1}^{\min(p,k)}\left(\boldsymbol{\delta}_{\boldsymbol{\theta}}^{(11)}(k-j)\right)^T\left(\boldsymbol{\gamma}_j^{(21)}\right)^T\left(\left(\mathbf{A}_0^{(22)}\right)^{-1}\right)^T,$$

$$\mathbf{V}(k) := \sum_{j=1}^{\min(p,k)}\left(\mathbf{A}_0^{(11)}\right)^{-1}\left(\boldsymbol{\gamma}_j^{(12)}\right)\left(\boldsymbol{\delta}_{\boldsymbol{\theta}}^{(22)}(k-j)\right).$$

The assumption that all fourth-order cumulants of $\boldsymbol{\varepsilon}$ are zero—an assumption which is of couse satisfied under Gaussian densities—is required for the validity of the El Himdi-Roy and Hallin-Saidi tests; our optimal tests ϕ^* do not require this assumption, and a more general result on asymptotic powers could be stated here, with much heavier expressions involving cumulants of all orders (see Roy 1989). For the sake of simplicity, and since the purpose of Proposition 4.1 is to allow for power comparisons between ϕ^*, ϕ_{HR}^M, and $\phi_{HS;i}^M$, we restrict ourselves to densities under which this cumulant assumption is satisfied.

Proof of Proposition 4.1. In order to prove this proposition, we use a multivariate version of Theorem 2.23 of Hall and Heyde (1980). For instance, note that, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$,

$$\sqrt{N}\mathrm{vec}\mathbf{R}_{\boldsymbol{\eta}}^{(12)}(k) - \left(\left(\mathrm{Diag}\left(\mathbf{A}_0^{(22)}\left(\mathbf{A}_0^{(22)}\right)^T\right)\right)^{-\frac{1}{2}} \otimes \left(\mathrm{Diag}\left(\mathbf{A}_0^{(11)}\left(\mathbf{A}_0^{(11)}\right)^T\right)\right)^{-\frac{1}{2}}\right)\sqrt{N}\mathrm{vec}\mathbf{C}_{\boldsymbol{\eta}}^{(12)}(k)$$

is $o_{\mathrm{P}}(1)$. Since, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$, $\mathbf{e}_t = \boldsymbol{\varepsilon}_t$, we have, for $k \geq 0$,

$$\mathbf{C}_{\boldsymbol{\eta}}^{(12)}(k) = N^{-1}\sum_t \boldsymbol{\eta}_t^{(1)}\left(\boldsymbol{\eta}_{t-k}^{(2)}\right)^T = N^{-1}\sum_t \mathbf{A}_0^{(11)}\boldsymbol{\varepsilon}_t^{(1)}\left(\boldsymbol{\varepsilon}_{t-k}^{(2)}\right)^T\left(\mathbf{A}_0^{(22)}\right)^T,$$

we obtain $\mathrm{vec}\mathbf{C}_{\boldsymbol{\eta}}^{(12)}(k) = N^{-1}\sum_t\left(\mathbf{A}_0^{(22)} \otimes \mathbf{A}_0^{(11)}\right)\mathrm{vec}\left(\boldsymbol{\varepsilon}_t^{(1)}\left(\boldsymbol{\varepsilon}_{t-k}^{(2)}\right)^T\right)$. Similar results hold for $k < 0$. Thus, $\sqrt{N}\mathrm{vec}\mathbf{R}_{\boldsymbol{\eta}}^{(12)}(k) = \mathbf{L}\sqrt{N}\mathrm{vec}\mathbf{C}_{\boldsymbol{\varepsilon}}^{(12)}(k)$, where

$$\mathbf{L} := \left(\left(\mathrm{Diag}\left(\mathbf{A}_0^{(22)}\left(\mathbf{A}_0^{(22)}\right)^T\right)\right)^{-\frac{1}{2}} \otimes \left(\mathrm{Diag}\left(\mathbf{A}_0^{(11)}\left(\mathbf{A}_0^{(11)}\right)^T\right)\right)^{-\frac{1}{2}}\right)\left(\mathbf{A}_0^{(22)} \otimes \mathbf{A}_0^{(11)}\right).$$

Applying ULAN with $\boldsymbol{\theta}^{(N)} = \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}^{(N)} = \boldsymbol{\theta}^{(N)}$ we obtain that, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$, as $N \to \infty$,

$$\Lambda_{\boldsymbol{\theta}^{(N)}/\boldsymbol{\theta}}^{(N)}\left(\mathbf{X}^{(N)}\right) := \log\left(\frac{d\mathrm{P}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}}{d\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}}\right) = \sum_{t=1}^{N}\zeta_{N,t} - \frac{1}{2}\nu^2 + o_{\mathrm{P}}(1),$$

where $\zeta_{N,t} := N^{-\frac{1}{2}}\{-\mathrm{vec}^T\boldsymbol{\gamma}_0^T\mathrm{vec}\mathbf{A}_0^{-1} + \left(\mathbf{A}_0^{-1}\boldsymbol{\gamma}_0\boldsymbol{\varepsilon}_t\right)^T\boldsymbol{\varphi}(\boldsymbol{\varepsilon}_t) + (\mathbf{A}_0^{-1}\sum_{j=1}^{p}\boldsymbol{\gamma}_j\mathbf{X}_{t-j})^T\boldsymbol{\varphi}(\boldsymbol{\varepsilon}_t)\}$.

Now, define

$$\mathbf{Y}_{N,t} =: \begin{pmatrix} \mathbf{L}\,\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t-M}^{(1)}\left(\boldsymbol{\varepsilon}_{t}^{(2)}\right)^{T}\right) \\ \vdots \\ \mathbf{L}\,\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t}^{(1)}\left(\boldsymbol{\varepsilon}_{t}^{(2)}\right)^{T}\right) \\ \mathbf{L}\,\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t}^{(1)}\left(\boldsymbol{\varepsilon}_{t-1}^{(2)}\right)^{T}\right) \\ \vdots \\ \mathbf{L}\,\mathrm{vec}\left(\boldsymbol{\varepsilon}_{t}^{(1)}\left(\boldsymbol{\varepsilon}_{t-M}^{(2)}\right)^{T}\right) \\ -\mathrm{vec}^{T}\boldsymbol{\gamma}_0^{T}\,\mathrm{vec}\mathbf{A}_0^{-1} + \left(\mathbf{A}_0^{-1}\boldsymbol{\gamma}_0\boldsymbol{\varepsilon}_t\right)^{T}\boldsymbol{\varphi}(\boldsymbol{\varepsilon}_t) + \left(\mathbf{A}_0^{-1}\sum_{j=1}^{p}\boldsymbol{\gamma}_j\mathbf{X}_{t-j}\right)^{T}\boldsymbol{\varphi}(\boldsymbol{\varepsilon}_t) \end{pmatrix}.$$

From Theorem 2.1, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$, we have $\begin{pmatrix} \sqrt{N}\mathbf{r}_{\boldsymbol{\eta}}^{(12)}(M) \\ \sum_t \zeta_{N,t} \end{pmatrix} = N^{-\frac{1}{2}}\sum_t \mathbf{Y}_{N,t} + o_{\mathrm{P}}(1)$. It is easy to check that $\mathbf{Y}_{N,t}$ defines a square-integrable martingale difference. In order to apply the classical central limit theory for martingale differences, we must check that, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$,

$$\frac{1}{N}\sum_t \mathrm{E}\left[\mathbf{Y}_{N,t}\left(\mathbf{Y}_{N,t}\right)^{T}|\mathcal{A}_{t-1}^{(N)}\right] \xrightarrow{\mathrm{P}} \boldsymbol{\Upsilon}, \tag{4.5}$$

for some non-random matrix $\boldsymbol{\Upsilon}$, and that, for all $\epsilon > 0$,

$$\frac{1}{N}\sum_t \mathrm{E}\left[\|\mathbf{Y}_{N,t}\|^2 I\left(\|\mathbf{Y}_{N,t}\| > \sqrt{N}\epsilon\right)|\mathcal{A}_{t-1}^{(N)}\right] \xrightarrow{\mathrm{P}} 0. \tag{4.6}$$

One easily can verify that under the assumption that the fourth-order cumulants of $\boldsymbol{\varepsilon}$ are zero, condition (4.5) is satisfied, with

$$\boldsymbol{\Upsilon} := \begin{pmatrix} \mathbf{I}_{2M+1}\otimes\left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0)\otimes\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right) & \mathbf{V} \\ \mathbf{V}^{T} & \nu^2 \end{pmatrix},$$

where $\mathbf{V}$ is defined in (4.4). In order to prove that (4.6) holds, it is sufficient to show that $\sum_t \mathrm{E}\left[\left\|\frac{\mathbf{Y}_{N,t}}{\sqrt{N}}\right\|^2 I\left(\|\mathbf{Y}_{N,t}\| > \sqrt{N}\epsilon\right)\right] \longrightarrow 0$. Remark that

$$\sum_t \mathrm{E}\left[\left\|\frac{\mathbf{Y}_{N,t}}{\sqrt{N}}\right\|^2 I\left(\|\mathbf{Y}_{N,t}\| > \sqrt{N}\epsilon\right)\right] \leq \mathrm{E}\left[\sum_t\left\|\frac{\mathbf{Y}_{N,t}}{\sqrt{N}}\right\|^2 I\left(\max_t\|\mathbf{Y}_{N,t}\| > \sqrt{N}\epsilon\right)\right];$$

hence, we just have to show that $\sum_t\left\|\frac{\mathbf{Y}_{N,t}}{\sqrt{N}}\right\|^2$ is uniformly integrable (Hall and Heyde 1980, p. 53). Uniform integrability of $\sum_t\left\|\frac{\mathbf{Y}_{N,t}}{\sqrt{N}}\right\|^2$ readily follows from the fact that $\sum_t\zeta_{N,t}^2$ is bounded. $\square$

Le Cam's third lemma provides the distribution under $\mathrm{P}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$ of $\sqrt{N}\mathbf{r}_{\boldsymbol{\eta}}^{(12)}(M)$. This is not sufficient for obtaining the local asymptotic powers of the test statistics (4.1) and (4.2), which require the asymptotic distribution of $\sqrt{N} r_{\hat{\boldsymbol{\eta}}}^{(12)}(M)$ defined in (4.3). However, under non-correlation between $\mathbf{X}_t^{(1)}$ and $\mathbf{X}_t^{(2)}$, we have (see (3.6) in El Himdi and Roy 1997)

$$\sqrt{N}\mathbf{r}_{\boldsymbol{\eta}}^{(12)}(M) - \sqrt{N}\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M) = o_{\mathrm{P}}(1). \tag{4.7}$$

We then have the following corollary to Proposition 4.1.

Corollary 4.1 . *Suppose that assumptions of Proposition 4.1 hold. Then, under* $P_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$,

(i) $Q_{HR}^M = N \sum_{k=-M}^{M} \left(\mathrm{vec}\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(12)}(k)\right)^T \left(\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(22)}(0) \otimes \mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(11)}(0)\right)^{-1} \left(\mathrm{vec}\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(12)}(k)\right) \xrightarrow{\mathcal{L}} \chi^2_{(2M+1)d_1d_2,\delta^2}$, *where,* $\chi^2_{(2M+1)d_1d_2,\delta^2}$ *stand for a non central chi-square variable with* $(2M+1)d_1d_2$ *degrees of freedom and non-centrality parameter* $\delta^2 = \left\| \left(\mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right)^{-\frac{1}{2}}\right) \mathbf{V} \right\|^2$;

$Q_{HS;i}^M = \sum_{k=1}^{(2M+1)d_1d_2-i} \left[\sum_{l=0}^{i} \boldsymbol{\nu}_M(k+l)\right]^2 \xrightarrow{\mathcal{L}} \sum_{j=1}^{(2M+1)d_2d_1} \lambda_M^{(i)}(j) W_{i,j}^2$, *where the coefficients* $\lambda_M^{(i)}(j)$ *are the eigenvalues of*

$$\mathbf{A}_M^{(i)} = \sum_{k=1}^{(2M+1)d_1d_2-i} \mathbf{L}_k^{(i)} \mathbf{L}_k^{(i)T}, \quad \textit{with} \quad \mathbf{L}_k^{(i)} := \left(\mathbf{0}^T_{(k-1)\times 1}, \mathbf{1}^T_{(i+1)\times 1}, \mathbf{0}^T_{((2M+1)d_1d_2-(k+i))\times 1}\right)^T,$$

and, denoting by $\mathbf{P}_M^{(i)}$ *the orthogonal matrices whose columns are the eigenvectors of* $\mathbf{A}_M^{(i)}$ *corresponding to the eigenvalues* $\lambda_M^{(i)}(j)$, *the variables* $W_{i,j}$, $j = 1, \ldots, (2M+1)\, d_1 d_2$, *are independent normal variables with mean*

$$\mathrm{E}\left[W_{i,j}\right] = \left(\left(\mathbf{P}_M^{(i)}\right)^T \left(\mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right)^{-\frac{1}{2}}\right) \mathbf{V}\right)_j$$

and variance $\mathrm{E}\left[W_{i,j} - \mathrm{E}\left[W_{i,j}\right]\right]^2 = 1$.

Proof of Corollary 4.1. By Proposition 4.1, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$,

$$\begin{pmatrix} \sqrt{N}\mathbf{r}_{\boldsymbol{\eta}}^{(12)}(M) \\ \Lambda_{\boldsymbol{\theta}^{(N)}/\boldsymbol{\theta}}^{(N)}(\mathbf{X}^{(N)}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ -\frac{\nu^2}{2} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right) & \mathbf{V} \\ \mathbf{V}^T & \nu^2 \end{pmatrix}\right),$$

Hence by (4.7), if follows that, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(N)}$,

$$\begin{pmatrix} \sqrt{N}\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M) \\ \Lambda_{\boldsymbol{\theta}^{(N)}/\boldsymbol{\theta}}^{(N)}(\mathbf{X}^{(N)}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ -\frac{\nu^2}{2} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right) & \mathbf{V} \\ \mathbf{V}^T & \nu^2 \end{pmatrix}\right),$$

Le Cam's third lemma thus implies that, under $P_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$,

$$\sqrt{N}\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{V}, \mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right)\right). \tag{4.8}$$

The proof of (i) follows from (4.8) and the fact that $\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(22)}(0)$ and $\mathbf{R}_{\hat{\boldsymbol{\eta}}}^{(11)}(0)$ are consistent estimators of $\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0)$ and $\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)$ respectively.

Turning to (ii), write $Q_{HS;i}^{M}$ as a quadratic form in the vector $\boldsymbol{\nu}_M$:

$$Q_{HS;i}^{M} = \sum_{k=1}^{(2M+1)d_1d_2-i} \left[\sum_{l=k}^{k+i} \boldsymbol{\nu}_M(l)\right]^2 = (\boldsymbol{\nu}_M)^T \sum_{k=1}^{(2M+1)d_1d_2-i} \mathbf{L}_k^{(i)} \left(\mathbf{L}_k^{(i)}\right)^T \boldsymbol{\nu}_M.$$

Thus, $Q_{HS;i}^{M}$ takes the form $(\boldsymbol{\nu}_M)^T \mathbf{C}_M^{(i)}(\mathbf{C}_M^{(i)})^T\boldsymbol{\nu}_M$, with $\mathbf{C}_M^{(i)} = \left[\mathbf{L}_1^{(i)}, \ldots, \mathbf{L}_{(2M+1)d_1d_2-i}^{(i)}\right]$, which factorizes into $(\boldsymbol{\nu}_M)^T \mathbf{P}_M^{(i)}\boldsymbol{\Lambda}_M^{(i)}\left(\mathbf{P}_M^{(i)}\right)^T\boldsymbol{\nu}_M$, where $\mathbf{P}_M^{(i)}$ is orthogonal and $\boldsymbol{\Lambda}_M^{(i)}$ is diagonal with positive elements $\lambda_M^{(i)}(j)$. Therefore,

$$Q_{HS;i}^{M} = \sum_{j=1}^{(2M+1)d_2d_1} \lambda_M^{(i)}(j) \left(\left(\mathbf{P}_M^{(i)}\right)^T \boldsymbol{\nu}_M\right)_j^2.$$

Hence (ii) follows by the fact that under $\mathrm{P}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$,

$$\left(\mathbf{P}_M^{(i)}\right)^T \boldsymbol{\nu}_M \xrightarrow{\mathcal{L}} \mathcal{N}\left(\left(\mathbf{P}_M^{(i)}\right)^T \left(\mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right)^{-\frac{1}{2}}\right)\mathbf{V}, \mathbf{I}_{(2M+1)d_2d_1}\right).$$

This completes the proof of the corollary. □

Remark 4.1

(1) The asymptotic powers of the Haugh and Koch and Yang tests ϕ_H^M and $\phi_{KY;i}^M$ under local alternatives of course readily follow from the more general results on ϕ_{HR}^M and $\phi_{HS;i}^M$ given in Corollary 4.1.

(2) It follows from (4.8) that the asymptotic distribution of $\sqrt{N}\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M)$ under $\mathrm{P}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$ does not depend on $\boldsymbol{\gamma}_j^{(11)}$ and $\boldsymbol{\gamma}_j^{(22)}$ for $j = 1, ..., p$. Further, if we perturb only $\mathbf{A}_0^{(11)}$ and $\mathbf{A}_0^{(22)}$ in the matrix $\mathbf{A}_0$, i. e., if $\boldsymbol{\gamma}_0^{(21)} = 0$, we can check that $\mathbf{V}(0) = \mathbf{0}$ (a somewhat expected finding). This conclusion confirms the general fact that, under the null hypothesis,

$$\sqrt{N}\mathbf{r}_{\hat{\boldsymbol{\eta}}}^{(12)}(M) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{2M+1} \otimes \left(\boldsymbol{\varphi}_{\boldsymbol{\eta}^{(2)}}(0) \otimes \boldsymbol{\varphi}_{\boldsymbol{\eta}^{(1)}}(0)\right)\right).$$

Indeed, under $\mathrm{P}_{f;\boldsymbol{\theta}^{(N)}}^{(N)}$, with $\boldsymbol{\gamma}_j^{(12)} = \mathbf{0}$ and $\boldsymbol{\gamma}_j^{(21)} = \mathbf{0}$ for $j = 1, ..., p$, and $\boldsymbol{\gamma}_0^{(21)} = \mathbf{0}$, we are still under the null hypothesis of non-correlation, so that we should have $\mathbf{V} = \mathbf{0}$.

5 The bivariate VAR(1) case, and some Monte Carlo results.

As an illustration, and in order to investigate the finite sample performance of our tests, we consider in some detail the bivariate autoregressive VAR model of order 1 ($p = 1$ and $d = 2$). For simplicity, we focus on the Gaussian case, i.e., we assume that the density f is $\mathcal{N}(\mathbf{0}, \mathbf{I})$. We first give an explicit forms for the central sequence and the related information matrix. Then we give the explicit form of the locally asymptotically most stringent test for non-correlation between the two components of this VAR(1) model.

The notation will be adapted to this bivariate context as follows. We will denote by $\left\{(X_t, Y_t)^T, t \in \mathbb{Z}\right\}$ the observed bivariate process, and write $\left\{(u_t, v_t)^T, t \in \mathbb{Z}\right\}$ for the innovation $\{\mathbf{e}_t, t \in \mathbb{Z}\}$. Equation (2.2) then takes the form

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} A_1^{(11)} & A_1^{(12)} \\ A_1^{(21)} & A_1^{(22)} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} A_0^{(11)} & 0 \\ A_0^{(21)} & A_0^{(22)} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix} = \begin{pmatrix} a_t^{(1)} \\ a_t^{(2)} \end{pmatrix};$$

the vector of parameters is $\boldsymbol{\theta} = \left(A_1^{(11)}, A_1^{(21)}, A_1^{(12)}, A_1^{(22)}, A_0^{(11)}, A_0^{(21)}, A_0^{(22)}\right)^T$.

Let $\boldsymbol{\theta} = (\phi, 0, 0, \theta, \sigma_1, 0, \sigma_2)$ be an arbitrary parameter value satisfying the null hypothesis of non-correlation between X_t and Y_t. Define

$$\left\{\begin{array}{rcl} u_t & = & \sigma_1^{-1}(X_t - \phi X_{t-1}) \\ v_t & = & \sigma_2^{-1}(Y_t - \theta Y_{t-1}) \end{array}\right., \qquad \text{and} \qquad \left\{\begin{array}{rcl} \eta_t^{(1)} & = & (X_t - \phi X_{t-1}) \\ \eta_t^{(2)} & = & (Y_t - \theta Y_{t-1}) \end{array}\right. .$$

The central sequence (2.11) then takes the simple form

$$\boldsymbol{\Delta}_{\mathcal{N};\boldsymbol{\theta}}^{(N)} = N^{-\frac{1}{2}} \sum_{t=1}^{N} \begin{pmatrix} \sigma_1^{-1} X_{t-1} u_t \\ \sigma_2^{-1} X_{t-1} v_t \\ \sigma_1^{-1} Y_{t-1} u_t \\ \sigma_2^{-1} Y_{t-1} v_t \\ \sigma_1^{-1}\left(-1 + u_t^2\right) \\ \sigma_2^{-1} u_t v_t \\ \sigma_2^{-1}\left(-1 + v_t^2\right) \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^{+\infty} \phi^k N^{-\frac{1}{2}} \sum_{t=1}^{N} u_{t-1-k} u_t \\ \frac{\sigma_1}{\sigma_2} \sum_{k=0}^{+\infty} \phi^k N^{-\frac{1}{2}} \sum_{t=1}^{N} u_{t-1-k} v_t \\ \frac{\sigma_2}{\sigma_1} \sum_{k=0}^{+\infty} \theta^k N^{-\frac{1}{2}} \sum_{t=1}^{N} v_{t-1-k} u_t \\ \sum_{k=0}^{+\infty} \theta^k N^{-\frac{1}{2}} \sum_{t=1}^{N} v_{t-1-k} v_t \\ N^{-\frac{1}{2}} \sum_{t=1}^{N} \sigma_1^{-1}\left(-1 + u_t^2\right) \\ N^{-\frac{1}{2}} \sum_{t=1}^{N} \sigma_2^{-1} u_t v_t \\ N^{-\frac{1}{2}} \sum_{t=1}^{N} \sigma_2^{-1}\left(-1 + v_t^2\right) \end{pmatrix}. \tag{5.1}$$

The corresponding information matrix is diagonal. Indeed, in the Gaussian case, $\boldsymbol{\mathcal{I}}(f)$ is the identity matrix, and (cf. equation (2.5))

$$\boldsymbol{\mathcal{I}}_s(f) = \mathrm{E}\left[\mathrm{vech}\left(-\mathbf{I}_2 + \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T\right) \mathrm{vech}^T\left(-\mathbf{I}_2 + \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T\right)\right] = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

It follows that the information matrix (2.12) is

$$\Gamma^{\boldsymbol{\Delta}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{1-\phi^2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{1-\phi^2}\frac{\sigma_1^2}{\sigma_2^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1-\theta^2}\frac{\sigma_2^2}{\sigma_1^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{1-\theta^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{\sigma_1^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sigma_2^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{\sigma_2^2} \end{pmatrix}. \tag{5.2}$$

5.1 Optimal pseudo-Gaussian test.

Let $\hat{\boldsymbol{\theta}} = \left(\hat{\phi}, 0, 0, \hat{\theta}, \hat{\sigma}_1, 0, \hat{\sigma}_2\right)^T$ be any estimator that satisfies assumption (A6). Define

$$\hat{\eta}_t^{(1)} = \left(X_t - \hat{\phi}X_{t-1}\right) \qquad \hat{\eta}_t^{(2)} = \left(Y_t - \hat{\theta}Y_{t-1}\right).$$

Then, from (3.2), (5.1) and (5.2), the sequence ϕ^* of locally asymptotically most stringent α-level pseudo-Gaussian tests for the null hypothesis of non-correlation or independence between X_t and Y_t rejects the null hypothesis at asymptotic significance level α whenever the test statistic $Q^* = Q^{(N)*}$, with

$$Q^{(N)*} := \frac{1-\hat{\phi}^2}{\hat{\sigma}_1^2\hat{\sigma}_2^2}\left(N^{-\frac{1}{2}}\sum_{t=1}^{N} X_{t-1}\hat{\eta}_t^{(2)}\right)^2 + \frac{1-\hat{\theta}^2}{\hat{\sigma}_1^2\hat{\sigma}_2^2}\left(N^{-\frac{1}{2}}\sum_{t=1}^{N} Y_{t-1}\hat{\eta}_t^{(1)}\right)^2 + \frac{1}{\hat{\sigma}_1^2\hat{\sigma}_2^2}\left(N^{-\frac{1}{2}}\sum_{t=1}^{N}\hat{\eta}_t^{(1)}\hat{\eta}_t^{(2)}\right)^2, \tag{5.3}$$

exceeds the (1-α) quantile of the chi-square distribution with 3 degrees of freedom. Note that this test statistic can also be expressed in term of residual cross-correlations. Indeed, when the sample size is large, using the fact that both X_t and Y_t admit an infinite moving average representation, the same test statistic takes the form

$$Q^{(N)*} = N\left[\left(1-\hat{\phi}^2\right)\left(\sum_{k=0}^{N-2}\hat{\phi}^k r_{\hat{\boldsymbol{\eta}}}^{(12)}(-1-k)\right)^2 + \left(1-\hat{\theta}^2\right)\left(\sum_{k=0}^{N-2}\hat{\theta}^k r_{\hat{\boldsymbol{\eta}}}^{(12)}(1+k)\right)^2 + \left(r_{\hat{\boldsymbol{\eta}}}^{(12)}(0)\right)^2\right]. \tag{5.4}$$

Remark 5.1 The test statistic (5.4) is a sum of three terms. The third term coincides with Haugh (1976)'s statistic, with $M = 0$. The first and the second ones are exploiting relevant information contained in the off-diagonal perturbations of the parameters under the alternative. The relative weakness of Haugh's procedure (1 degree of freedom) is entirely due to the fact that such perturbations (2 degrees of freedom) are entirely neglected in his test statistic.

5.2 Computation of local asymptotic powers.

In this Section, we use Corollary 4.1 to compute local asymptotic powers of the Haugh-El Himdi-Roy test ϕ_H^M (1.1), the Koch-Yang-Hallin-Saidi test $\phi_{KY;i}^M$ (1.2), and our optimal test ϕ^* (5.3) in the particular context of Gaussian bivariate VAR(1) (so that El Himdi-Roy reduces to Haugh, and Hallin-Saidi to Koch-Yang).

Let $\boldsymbol{\tau} := \left(\mathrm{vec}^T\boldsymbol{\gamma}_1, \mathrm{vech}^T\boldsymbol{\gamma}_0\right)^T$, with $\boldsymbol{\gamma}_0 = \begin{pmatrix} \gamma_0^{(11)} & 0 \\ \gamma_0^{(21)} & \gamma_0^{(22)} \end{pmatrix}$ and $\boldsymbol{\gamma}_1 = \begin{pmatrix} \gamma_1^{(11)} & \gamma_1^{(12)} \\ \gamma_1^{(21)} & \gamma_1^{(22)} \end{pmatrix}$, and consider the local alternatives

$$\begin{aligned}\boldsymbol{\theta}^{(N)} &= \boldsymbol{\theta} + N^{-1/2}\boldsymbol{\tau} \\ &= \Big(\phi + N^{-1/2}\gamma_1^{(11)}, N^{-1/2}\gamma_1^{(21)}, N^{-1/2}\gamma_1^{(12)}, \theta + N^{-1/2}\gamma_1^{(22)}, \\ &\qquad \sigma_1 + N^{-1/2}\gamma_0^{(11)}, N^{-1/2}\gamma_0^{(21)}, \sigma_2 + N^{-1/2}\gamma_0^{(22)}\Big)^T.\end{aligned}$$

Under $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$, the data generating process is given by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \left\{\begin{pmatrix} \phi & 0 \\ 0 & \theta \end{pmatrix} + N^{-1/2}\boldsymbol{\gamma}_1\right\}\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \left\{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} + N^{-1/2}\boldsymbol{\gamma}_0\right\}\begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix}.$$

From (3.3), the asymptotic power of the optimal test against $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$ is

$$\Pi^{\boldsymbol{\theta}^{(N)}}_{\phi^*} = 1 - F^3_{\chi^2}\left(\chi^2_{3,1-\alpha}; \psi^2\right), \tag{5.5}$$

with

$$\psi^2 := \frac{\left(\gamma_0^{(21)}\right)^2}{\sigma_2^2} + \frac{\left(\gamma_1^{(21)}\right)^2}{1-\phi^2}\,\frac{\sigma_1^2}{\sigma_2^2} + \frac{\left(\gamma_1^{(12)}\right)^2}{1-\theta^2}\,\frac{\sigma_2^2}{\sigma_1^2}. \tag{5.6}$$

Now, in order to compute local asymptotic powers for Haugh's test, we need the explicit form of the vector $\mathbf{V}$ in (4.4); one easily obtains

$$\mathbf{V} = \left(\gamma_1^{(21)}\,\phi^{M-1}\,\frac{\sigma_1}{\sigma_2}, ..., \gamma_1^{(21)}\,\frac{\sigma_1}{\sigma_2}, \frac{\left(\gamma_0^{(21)}\right)}{\sigma_2}, \gamma_1^{(12)}\,\frac{\sigma_2}{\sigma_1}, ..., \gamma_1^{(12)}\,\theta^{M-1}\,\frac{\sigma_2}{\sigma_1}\right)^T.$$

Using *(i)* in Corollary 4.1, the asymptotic power of Haugh's test ϕ_H^M against $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$ is thus

$$\Pi^{\boldsymbol{\theta}^{(N)}}_{\phi_H^M} = 1 - F^{2M+1}_{\chi^2}\left(\chi^2_{2M+1,1-\alpha}; \delta^2\right), \tag{5.7}$$

where

$$\delta^2 = \|\mathbf{V}\|^2 = \frac{\left(\gamma_0^{(21)}\right)^2}{\sigma_2^2} + \left(\gamma_1^{(21)}\right)^2\,\frac{\sigma_1^2}{\sigma_2^2}\,\frac{1-\phi^{2M}}{1-\phi^2} + \left(\gamma_1^{(12)}\right)^2\,\frac{\sigma_2^2}{\sigma_1^2}\,\frac{1-\theta^{2M}}{1-\theta^2}. \tag{5.8}$$

The asymptotic power of Koch and Yang's test $\phi^M_{KY;i}$ under $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$ follows from *(ii)* in Corollary 4.1. Denote by $F^n_{\chi^2}\left(.;\boldsymbol{\lambda};\boldsymbol{\mu}\right) = F^n_{\chi^2}\left(.;\lambda_1,...,\lambda_n:\mu_1,...,\mu_n\right)$ the distribution function of $\sum_{j=1}^n \lambda_j\mathcal{N}(\mu_j), 1^2$, which is a linear combination of independent non-central χ^2 with one degree of freedom and non-centrality parameter μ_j^2, and by $\chi^2_{n,1-\alpha}\left(\boldsymbol{\lambda}\right)$ the $(1-\alpha)$ quantile of $\sum_{j=1}^n \lambda_j\chi_1^2$, which is a linear combination of independent central χ^2 with one degree of freedom. The asymptotic power of Koch and Yang's test (1.2) under $\mathrm{P}^{(N)}_{f;\boldsymbol{\theta}^{(N)}}$ then is given by

$$\Pi^{\boldsymbol{\theta}^{(N)}}_{\phi^M_{KY;i}} = 1 - F^{2M+1}_{\chi^2}\left(\chi^2_{2M+1,1-\alpha}\left(\boldsymbol{\lambda}_M^{(i)}\right); \boldsymbol{\lambda}_M^{(i)}; \left(\mathbf{P}_M^{(i)}\right)^T\mathbf{V}\right), \tag{5.9}$$

where the coefficients $\boldsymbol{\lambda}_M^{(i)} := \left(\lambda_M^{(i)}(1), ..., \lambda_M^{(i)}(2M+1)\right)^T$ and the columns of $\mathbf{P}_M^{(i)}$ are the eigenvalues and eigenvectors of

$$\mathbf{A}_M^{(i)} = \sum_{k=1}^{(2M+1)-i} \mathbf{L}_k^{(i)} \mathbf{L}_k^{(i)T}, \quad \text{with} \quad \mathbf{L}_k^{(i)} := \left(\mathbf{0}_{(k-1)\times 1}^T, \mathbf{1}_{(i+1)\times 1}^T, \mathbf{0}_{((2M+1)-(k+i))\times 1}^T\right)^T.$$

From (5.5), (5.7), and (5.9), it is clear that these powers do not depend on $\gamma_0^{(11)}$, $\gamma_0^{(22)}$, $\gamma_1^{(11)}$, and $\gamma_1^{(22)}$. Indeed, the diagonal elements of $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ do not affect the null hypothesis, and, in the investigation of local powers, they safely can be put to zero.

Even with these explicit forms, comparisons between (1.1), (1.2), and (5.3) still are difficult, due to the fact that the asymptotic distributions, under local alternatives, of the Haugh and Koch and Yang statistics are not of the traditional noncentral chi-square type, and depend, in a complicated manner, on ϕ, θ, σ_1, σ_2, $\gamma_0^{(21)}$, $\gamma_1^{(12)}$, $\gamma_1^{(21)}$, as well as, for Haugh's test, on M (for Koch and Yang's, on M and i). General conclusions are hardly possible, and comparisons do not reduce to the computation of a few asymptotic relative efficiencies.

We start with some qualitative comparisons; then, for a few typical alternatives, we provide some numerical measures of performance. Note that our optimal tests are (locally and asymptotically) most stringent, not most powerful—so that they can be dominated, at particular alternatives, by their Haugh or Koch and Yang competitors.

We will distinguish three classes of local alternatives:

(i) *Type 1:* $\gamma_1^{(12)} = 0 = \gamma_1^{(21)}$; $\gamma_0^{(21)} \neq 0$;

(ii) *Type 2:* $\gamma_0^{(21)} = 0$; $\gamma_1^{(12)} \neq 0$ and/or $\gamma_1^{(21)} \neq 0$;

(iii) *Type 3:* $\gamma_0^{(21)} \neq 0$, $\gamma_1^{(12)} \neq 0$, and/or $\gamma_1^{(21)} \neq 0$.

Under *Type 1* alternatives, local asymptotic powers do not depend on the values of ϕ, θ and σ_1. For fixed $\gamma_0^{(21)}$ and σ_2, Haugh's power function (5.7) is a decreasing function of M; in fact, one can check numerically that

$$1 - F_{\chi^2}^{m}\left(\chi^2_{m,1-\alpha}; \omega^2\right) \geq 1 - F_{\chi^2}^{m+1}\left(\chi^2_{m+1,1-\alpha}; \omega^2\right), \tag{5.10}$$

for all α, ω^2 and $M \geq 0$. Further, for all α and $M \geq 0$, $1 - F_{\chi^2}^{2M+1}\left(\chi^2_{2M+1,1-\alpha}; \omega^2\right)$ is increasing with ω^2. It follows that the best Haugh test is ϕ_H^0, wich dominates $\phi_H^1 = \phi^*$. Under this type of alternatives, Koch and Yang's test will be worst. Intuitively, this is not a surprising conclusion, because, in this very particular case, only instantaneous correlation exists, which is perfectly captured by ϕ_H^0, while larger values of M clearly induce a loss of power.

Under *Type 2* alternatives, using (5.6), (5.8) and (5.10) the optimal test ϕ^* is better than ϕ_H^M, irrespective of M. The maximum power of ϕ_H^M will be achieved for some $M \geq 1$ that depends on the perturbations and the parameters ϕ, θ, σ_1, and σ_2. For such alternatives, $\phi_{KY;i}^M$ could do better than ϕ_H^M because under such alternatives the series X_t and Y_t are related over a long distributed lag. Further, it even may happen that $\phi_{KY;i}^M$, for an adequate choice of M and i, beat the optimal test. However, the optimal test ϕ^* in general is sizeably better than ϕ_H^M and $\phi_{KY;i}^M$, irrespective of M and i.

For the *Type 3* alternatives, the conclusions are intermediate between the previous two. Indeed, if $\gamma_0^{(21)}$ is large compared to $\gamma_1^{(21)}$ and $\gamma_1^{(21)}$, Haugh with $M = 0$ will be best; if not, depending on the perturbations and the parameter values ϕ, θ, σ_1, and σ_2, Koch and Yang or the optimal test will prevail.

For more quantitative conclusion, we now focus on the following three special cases:

(B) *Alternative B: Type 1* alternatives, with $\sigma_2 = 1$. The observed process is generated by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} \phi & 0 \\ 0 & \theta \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ N^{-1/2}\gamma_0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix},$$

with $\gamma_0 \neq 0$. Note that, under such alternatives, local asymptotic powers do not depend on ϕ, θ, nor σ_1.

(C) *Alternatives C: Type 2* alternatives, with $\phi = \theta = 0.5$, $\sigma_1 = \sigma_2 = 1.00$ and $\gamma_1^{(12)} = \gamma_1^{(21)} = \gamma_1 > 0$. The observed process is generated by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} 0.5 & N^{-1/2}\gamma_1 \\ N^{-1/2}\gamma_1 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix}.$$

(D) *Alternatives D: Type 3* alternatives, with $\phi = \theta = 0.5$, $\sigma_1 = \sigma_2 = 1.00$, $\gamma_0^{(21)} = 0.5$, and $\gamma_1^{(12)} = \gamma_1^{(21)} = \gamma_1 > 0$. The observed process is generated by

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} 0.5 & N^{-1/2}\gamma_1 \\ N^{-1/2}\gamma_1 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ N^{-1/2}/2 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix}.$$

Thus, under alternative B, the only perturbation is γ_0 whereas, under alternatives C and D, the only perturbation is γ_1.

Tables 1, 2, and 3 report, for various values of the perturbations, the local asymptotic powers at significance level $\alpha = 0.05$, under alternatives B, C, and D, respectively, and for several values of M and i. Inspection of the three tables confirms the previous qualitative conclusions. The case of ϕ_H^0 is somewhat special, since it is aimed, exclusively, at detecting positive values of γ_0 under Alternative B. In Table 1, ϕ_H^0 thus is clearly best. On the other hand, its performance under all other types of alternatives only can be pretty poor: under Alternatives C and D, for instance, ϕ_H^0 is not even consistent, its power being uniformly equal to 0.0500 (under C) and 0.0791 (under D), irrespective of γ_1. Therefore, it cannot be considered a serious competitor to the other procedures described, and we do not include it in Tables 2 and 3.

In Table 1, leaving aside the special case of ϕ_H^0, the optimal test ϕ^* appears sizeably more powerful than both Haugh and Koch and Yang's. In Table 2, Koch and Yang with smallest M and $i \neq 0$ slightly dominates the optimal test for small values of the perturbation ($\gamma_1^2 \leq 2$); the optimal test however prevails for $\gamma_1^2 \geq 3$; both are uniformly beating Haugh. The same phenomenon is observed in Table 3, with a more significant dominance of Koch and Yang for small γ_1 values.

			γ_0^2								
		Test	0.5	1	2	3	4	5	10	15	20
		$\phi^{(N)*}$	**.0811**	**.1157**	**.1922**	**.2746**	**.3585**	**.4405**	**.7611**	**.9170**	**.9751**
M=0	i=0	ϕ_H^0	.1090	.1701	.2930	.4100	.5160	.6088	.8854	.9721	.9940
M=5	i=0	ϕ_H^5	.0643	.0802	.1168	.1590	.2058	.2562	.5228	.7416	.8781
	i=4	$\phi_{KY;4}^5$	.0618	.0741	.0995	.1260	.1534	.1815	.3269	.4687	.5957
M=10	i=0	ϕ_H^{10}	.0597	.0704	.0948	.1228	.1544	.1891	.3925	.5995	.7648
	i=4	$\phi_{KY;4}^{10}$	.0562	.0628	.0766	.0914	.1069	.1233	.2151	.3174	.4227
	i=8	$\phi_{KY;8}^{10}$	.0563	.0629	.0763	.0901	.1047	.1188	.1951	.2752	.3561
M=20	i=0	ϕ_H^{20}	.0566	.0638	.0797	.0977	.1179	.1402	.2781	.4418	.6024
	i=4	$\phi_{KY;4}^{20}$	.0537	.0577	.0662	.0752	.0847	.0947	.1523	.2202	.2954
	i=8	$\phi_{KY;8}^{20}$	.0532	.0566	.0636	.0709	.0785	.0863	.1292	.1774	.2299
	i=12	$\phi_{KY;12}^{20}$	.0531	.0563	.0629	.0697	.0767	.0838	.1217	.1631	.2071
	i=16	$\phi_{KY;16}^{20}$	.0533	.0567	.0635	.0705	.0776	.0848	.1223	.1618	.2029
M=30	i=0	ϕ_H^{30}	.0533	.0610	.0734	.0873	.1027	.1196	.2252	.3574	.4991
	i=4	$\phi_{KY;4}^{30}$	.0529	.0560	.0625	.0694	.0767	.0843	.1281	.1804	.2396
	i=8	$\phi_{KY;8}^{30}$	.0523	.0548	.0600	.0654	.0710	.0768	.1087	.1449	.1851
	i=12	$\phi_{KY;12}^{30}$	.0521	.0544	.0591	.0639	.0688	.0739	.1012	.1313	.1640
	i=16	$\phi_{KY;16}^{30}$	.0521	.0542	.0587	.0633	.0679	.0727	.0978	.1250	.1540
	i=20	$\phi_{KY;20}^{30}$	.0521	.0543	.0587	.0632	.0677	.0724	.0966	.1219	.1494

Table 1. Asymptotic powers, under Alternative B, of the optimal test ϕ^*, Haugh's test ϕ_H^M, and Koch and Yang's test $\phi_{KY;i}^M$, for $i = 4$, 8, 12, 16, 20, and $M = 0, 5, 10, 20, 30$, at significance level $\alpha = 0.05$. Boldface indicate the winner (leaving aside ϕ_H^0) in each column.

			γ_1^2								
		Test	0.5	1	2	3	4	5	10	15	20
		$\phi^{(N)*}$	.1402	.2468	.4670	**.6541**	**.7904**	**.8797**	**.9957**	**.9999**	**1.0000**
M=5	i=0	ϕ_H^5	.0917	.1442	.2733	.4169	.5558	.6770	.9622	.9977	.9999
	i=4	$\phi_{KY;4}^5$	**.1568**	**.2689**	**.4802**	.6513	.7765	.8619	.9914	.9996	.9999
M=10	i=0	ϕ_H^{10}	.0781	.1131	.2012	.3072	.4212	.5336	.9009	.9886	.9991
	i=4	$\phi_{KY;4}^{10}$	.1158	.1912	.3530	.5082	.6418	.7484	.9710	.9978	.9998
	i=8	$\phi_{KY;8}^{10}$	.1341	.2233	.3995	.5555	.6824	.7796	.9741	.9956	.9998
M=20	i=0	ϕ_H^{20}	.0688	.0915	.1480	.2183	.2991	.3862	.7759	.9506	.9927
	i=4	$\phi_{KY;4}^{20}$	.0901	.1382	.2510	.3739	.4948	.6052	.9211	.9900	.9990
	i=8	$\phi_{KY;8}^{20}$	.0954	.1479	.2661	.3901	.5090	.6157	.9187	.9884	.9987
	i=12	$\phi_{KY;12}^{20}$	.0972	.1497	.2640	.3813	.4932	.5942	.8981	.9815	.9973
	i=16	$\phi_{KY;16}^{20}$	.1009	.1551	.2683	.3806	.4859	.5808	.8772	.9721	.9947
M=30	i=0	ϕ_H^{30}	.0649	.0825	.1255	.1790	.2416	.3114	.6737	.8984	.9779
	i=4	$\phi_{KY;4}^{30}$	.0806	.1174	.2059	.3074	.4135	.5168	.8702	.9775	.9971
	i=8	$\phi_{KY;8}^{30}$	.0835	.1228	.2148	.3172	.4221	.5228	.8652	.9741	.9962
	i=12	$\phi_{KY;12}^{30}$	.0835	.1216	.2084	.3035	.4000	.4943	.8318	.9593	.9921
	i=16	$\phi_{KY;16}^{30}$	.0832	.1203	.2026	.2913	.3814	.4689	.7985	.9413	.9859
	i=20	$\phi_{KY;20}^{30}$	.0837	.1230	.1998	.2839	.3687	.4511	.7708	.9237	.9786

Table 2. Asymptotic powers, under Alternative C, of the optimal test ϕ^*, Haugh's test ϕ_H^M, and Koch and Yang's test $\phi_{KY;i}^M$, for $i = 4$, 8, 12, 16, 20, and $M = 5$, 10, 20, 30, at significance level $\alpha = 0.05$. Boldface indicate the winner in each column.

		Test	γ_1^2 0.5	1	2	3	4	5	10	15	20
		$\phi^{(N)*}$	.1593	.2677	.4864	.6691	.8006	.8861	**.9960**	**.9999**	**1.000**
M=5	i=0	ϕ_H^5	.1007	.1551	.2865	.4304	.5680	.6873	.9639	.9978	.9999
	i=4	$\phi_{KY;4}^5$	**.2147**	**.3463**	**.5679**	**.7287**	**.8363**	**.9043**	.9953	.9998	.9999
M=10	i=0	ϕ_H^{10}	.0842	.1203	.2106	.3177	.4320	.5438	.9044	.9891	.9992
	i=4	$\phi_{KY;4}^{10}$	.1509	.2435	.4254	.5848	.7122	.8072	.9820	.9988	.9999
	i=8	$\phi_{KY;8}^{10}$	.1724	.2770	.4674	.6233	.7421	.8282	.9831	.9988	.9999
M=20	i=0	ϕ_H^{20}	.0728	.0961	.1541	.2255	.3071	.3945	.7812	.9522	.9930
	i=4	$\phi_{KY;4}^{20}$	.1121	.1730	.3060	.4410	.5658	.6733	.9456	.9941	.9995
	i=8	$\phi_{KY;8}^{20}$	.1171	.1814	.3171	.4508	.5722	.6760	.9414	.9927	.9993
	i=12	$\phi_{KY;12}^{20}$	.1182	.1812	.3105	.4362	.5507	.6500	.9228	.9875	.9983
	i=16	$\phi_{KY;16}^{20}$	.1224	.1862	.3123	.4312	.5389	.6324	.9033	.9801	.9965
M=30	i=0	ϕ_H^{30}	.0680	.0860	.1301	.1845	.2479	.3182	.6796	.9009	.9786
	i=4	$\phi_{KY;4}^{30}$	.0974	.1443	.2506	.3653	.4790	.5843	.9050	.9859	.9984
	i=8	$\phi_{KY;8}^{30}$	.0996	.1483	.2560	.3697	.4808	.5830	.8974	.9826	.9977
	i=12	$\phi_{KY;12}^{30}$	.0985	.1449	.2453	.3501	.4529	.5488	.8660	.9708	.9949
	i=16	$\phi_{KY;16}^{30}$	.0977	.1422	.2365	.3338	.4294	.5193	.8340	.9559	.9903
	i=20	$\phi_{KY;20}^{30}$	.0979	.1414	.2316	.3235	.4133	.4982	.8070	.9406	.9845

Table 3. Asymptotic powers under Alternative D, of the optimal test ϕ^*, Haugh's test ϕ_H^M, and Koch and Yang's test $\phi_{KY;i}^M$, for $i = 4, 8, 12, 16, 20$ and $M = 5, 10, 15, 20, 30$, at significance level $\alpha = 0.05$. Boldface indicate the winner in each column.

5.3 A Monte Carlo study.

We conclude this Section with a Monte Carlo investigation of the finite sample behaviors of the statistics discussed in the previous sections. In order to do so, we consider four particular cases of the data generating equation

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} 0.5 & N^{-1/2}\gamma_1 \\ N^{-1/2}\gamma_1 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ N^{-1/2}\gamma_0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix}. \tag{5.11}$$

- *Experiment A:* $\gamma_0 = \gamma_1 = 0$. Under this experiment, X_t and Y_t are independent; this experiment allows for checking the validity of asymptotic distributions under the null.
- *Experiment B:* $\gamma_0^2 = 10$. Alternative B; X_t and Y_t are dependent at lag zero only.
- *Experiment C:* $\gamma_1^2 = 5$. Alternative C; X_t and Y_t are dependent over all lags, except for lag zero.
- *Experiment D:* $\gamma_0^2 = 0.25$ and $\gamma_1^2 = 5$. Alternative D; X_t and Y_t are dependent over all lags.

For each of these four experiments, 10000 replications of two independent standard Gaussian white noises $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ of length 500 were generated from the G05EAF subroutine of the NAG library. These sequences were plugged into the various models, yielding 10000 replications, of length 500, of the processes X_t and Y_t, respectively. Initial values X_0 and Y_0 were put to zero. In order to prevent starting values to affect the stationarity of the generated series, only the subseries of length $N = 100$ (respectively, $N = 200$) resulting from dropping the 400 (respectively, 300) first observations were considered for the analysis.

For each of the replications thus obtained, under experiments A through D, using subroutine G13DCF of the NAG library, an AR(1) model was fitted to each component series, yielding two vectors of estimated residuals, $\hat{\eta}^{(1)}(t)$, and $\hat{\eta}^{(2)}(t)$. Note that (4.7) remains valid when an

ARMA(p,q) such that $p \geq 1$ and $q \geq 0$ is adjusted to each series; the asymptotic local powers of Haugh's and Koch and Yang's tests thus remain unaffected if ARMA orders are oversestimated. From these residuals, we computed

- the optimal test statistic Q^* given in (5.3);
- the residual cross-correlations (via subroutine G13DMF), and the statistics Q_H^M and $Q_{KY;i}^M$ given in (1.1) and (1.2), for various values of M ($M = 0, 5, 10, 15, 20, 30$) and i ($i = 4, 8, 12, 16$);
- the modified versions $\tilde{\phi}_{HR}^M$ (here reducing to $\tilde{\phi}_H^M$) of the Haugh-El Himdi-Roy tests proposed in El Himdi and Roy (1997), and the modified versions $\tilde{\phi}_{HS;i}^M$ (here reducing to $\tilde{\phi}_{KY;i}^M$) of the Koch-Yang-Hallin-Saidi tests proposed in Hallin and Saidi (2003), based on
$$\tilde{Q}_{KY;i}^M := \sum_{k=1}^{(2M+1)-i} \left[\sum_{l=0}^{i} \tilde{\boldsymbol{\nu}}_M (k+l) \right]^2, \quad i = 0, 1, \ldots, M-1, \tag{5.12}$$
with $\tilde{\boldsymbol{\nu}}_M := \left(\frac{N}{\sqrt{N-M}} r_{\hat{\boldsymbol{\eta}}}^{(12)}(-M), \ldots, \frac{N}{\sqrt{N-|k|}} r_{\hat{\boldsymbol{\eta}}}^{(12)}(k), \ldots, \frac{N}{\sqrt{N-M}} r_{\hat{\boldsymbol{\eta}}}^{(12)}(M) \right)^T$ (for $\tilde{\phi}_{HR}^M$ being obtained for $i = 0$); these modified versions are supposed to allow for a better control of asymptotic significance levels.

For each replication, these statistics were compared with their exact asymptotic critical values; these values for the Koch and Yang statistics were computed using the Imhof (1961) algorithm (in Koch and Yang (1986) and Hallin and Saidi (2003), only approximate asymptotic critical values, based on Satterthwaite (1941, 1946)'s approximation were used). Rejection frequencies are reported for two series length ($N = 100$ and $N = 200$), in Tables 4 and 5, at nominal α-values 0.05 and 0.01, and for various values of M and i.

Rejection frequencies for Experiment A are reported in Table 4. For all series lengths and significance levels, the rejection frequencies for ϕ^* are significantly closer to nominal α-values than those for $\phi_{KY;i}^M$ and $\tilde{\phi}_{KY;i}^M$; the latter two tests, by the way, appear to be seriously biased (the corresponding rejection frequencies are significantly less than α). The corrected versions $\tilde{\phi}_{KY;i}^M$ improve over the original one, but still yield a significant bias.

Table 5 reports rejection frequencies, under Experiments B, C, and D, at probability level $\alpha = 0.05$. In view of the severe bias of the Haugh and Koch and Yang tests, the critical values we consider here are based on both the asymptotic distributions and the observed rejection frequencies reported in Table 4. The figures in the table very clearly indicate that the Koch and Yang procedure, in Experiment B, is uniformly, and quite significantly weakest (significance of power differences can be tested by means of a Mc Nemar test—see, e.g., Armitage and Colton 1998; recall indeed that the various simulations are generated from the same pseudo-white noise). For instance, for a series length of $N = 100$, the Haugh procedure based on asymptotic critical values yields an empirical power of .4331 with $M = 5$, .2785 with $M = 10$, whereas our procedure for the same sample size reaches .7125. The same conclusions still hold when empirical critical values are used. Under Experiments C and D, where dependencies are distributed over a long period, the optimal test appears to perform best. Even with empirical critical values, ϕ^* performs best (Koch and Yang with $M = 5$ and $i = 4$ does slightly better than ϕ^* in Experiment D, for $N = 200$, but this advantage appears to be non significant). An other interesting conclusion from Table 5 is that the empirical local powers, under Experiments B, C, and D, are close to the theoretical figures reported in Tables 1, 2, and 3, which confirms the relevance of the asymptotic theory developed in this paper.

α	N	ϕ^*	M	ϕ_H^M	$\phi_{KY;4}^M$	$\phi_{KY;8}^M$	$\phi_{KY;12}^M$	$\tilde{\phi}_H^M$	$\tilde{\phi}_{KY;4}^M$	$\tilde{\phi}_{KY;8}^M$	$\tilde{\phi}_{KY;12}^M$
	100	.0084	5	.0065	.0038			.0075	.0043		
			10	.0041	.0027	.0028		.0084	.0040	.0033	
			15	.0056	.0048	.0055	.0061	.0104	.0041	.0037	.0031
			20	.0028	.0018	.0022	.0026	.0105	.0037	.0038	.0040
			25	.0015	.0012	.0017	.0025	.0120	.0048	.0037	.0041
			30	.0014	.0012	.0020	.0021	.0137	.0055	.0052	.0046
0.01											
	200	.0095	5	.0067	.0070			.0075	.0075		
			10	.0065	.0055	.0053		.0091	.0069	.0057	
			15	.0056	.0048	.0055	.0061	.0101	.0066	.0061	.0069
			20	.0043	.0046	.0059	.0063	.0110	.0064	.0073	.0070
			25	.0024	.0035	.0045	.0058	.0103	.0680	.0068	.0074
			30	.0022	.0037	.0047	.0053	.0096	.0072	.0079	.0082
	100	.0439	5	.0372	.0343			.0474	.0358		
			10	.0300	.0218	.0236		.0466	.0293	.0266	
			15	.0220	.0185	.0180	.0187	.0478	.0286	.0254	.0237
			20	.0136	.0136	.0153	.0162	.0518	.0274	.0232	.0222
			25	.0096	.0099	.0133	.0142	.0505	.0297	.0247	.0234
			30	.0063	.0091	.0117	.0133	.0525	.0301	.0283	.0262
0.05											
	200	.0505	5	.0433	.0432			.0470	.0445		
			10	.0384	.0362	.0357		.0475	.0407	.0377	
			15	.0340	.0345	.0333	.0332	.0503	.0410	.0367	.0357
			20	.0283	.0289	.0307	.0323	.0490	.0398	.0375	.0360
			25	.0242	.0270	.0299	.0308	.0531	.0419	.0392	.0377
			30	.0165	.0234	.0278	.0312	.0525	.0395	.0412	.0409

Table 4. Rejection frequencies in 10000 replications of Experiment A, for the optimal test ϕ^*, the Haugh and modified Haugh tests ϕ_H^M and $\tilde{\phi}_H^M$, the Koch and Yang and modified Koch and Yang tests $\phi_{KY;i}^M$ and $\tilde{\phi}_{KY;i}^M$, for $M = 5, 10, 15, 20, 25, 30$, and for various values of $i = 0, 4, 8, 12$, at significance levels $\alpha = 0.05$ and 0.01, for series lengths $N = 100$ and 200, respectively.

			ASYMPTOTIC CRITICAL VALUE					EMPIRICAL CRITICAL VALUE				
EXP	N	M	ϕ_H^M	$\phi_{KY;4}^M$	$\phi_{KY;8}^M$	$\phi_{KY;12}^M$	ϕ^*	$\tilde{\phi}_H^M$	$\tilde{\phi}_{KY;4}^M$	$\tilde{\phi}_{KY;8}^M$	$\tilde{\phi}_{KY;12}^M$	ϕ^*
	100	5	.4338	.2471			**.7125**	.4750	.2949			**.7310**
		10	.2785	.1207	.1020			.3605	.2022	.1682		
		15	.1794	.0751	.0608	.0615		.2959	.1731	.1266	.1134	
		20	.1169	.0528	.0450	.0443		.2563	.1360	.1093	.1009	
		25	.0707	.0371	.0373	.0364		.2374	.1290	.1027	.0924	
		30	.0443	.0292	.0290	.0319		.2195	.1192	.0937	.0887	
B												
	200	5	.4670	.2922			**.7362**	.4929	.3131			**.7354**
		10	.3277	.1639	.1427			.3683	.2007	.1809		
		15	.2451	.1197	.0946	.0955		.3001	.1637	.1352	.1359	
		20	.1899	.0957	.0781	.0757		.2681	.1384	.1176	.1131	
		25	.1467	.0777	.0655	.0633		.2283	.1251	.0983	.0989	
		30	.1088	.0640	.0588	.0574		.2200	.1188	.0952	.0913	
	100	5	.4910	.7537			**.8545**	.5340	.7934			**.8677**
		10	.3078	.5371	.6089			.3943	.6540	.7154		
		15	.2034	.3829	.4461	.4357		.3288	.5695	.6070	.5755	
		20	.1351	.2698	.3391	.3329		.2872	.4749	.5311	.5059	
		25	.0833	.1965	.2771	.2699		.2647	.4294	.4809	.4469	
		30	.0527	.1453	.2244	.2285		.2404	.3858	.4344	.4087	
C												
	200	5	.5970	.8253			**.8745**	.6146	.8383			**.8739**
		10	.4240	.6708	.7150			.4724	.7113	.7585		
		15	.3139	.5596	.5918	.5710		.3777	.6317	.6614	.6395	
		20	.2404	.4683	.4999	.4380		.3347	.5590	.5862	.5595	
		25	.1890	.4009	.4380	.4134		.2861	.5021	.5290	.4977	
		30	.1446	.3389	.3829	.3623		.2712	.4701	.4850	.4511	
	100	5	.4577	.8029			**.8549**	.5011	.8382			**.8689**
		10	.2816	.5938	.6486			.3631	.7050	.7513		
		15	.1792	.4297	.4821	.4627		.3028	.6148	.6417	.6066	
		20	.1163	.3048	.3689	.3515		.2630	.5198	.5638	.5304	
		25	.0717	.2194	.2960	.2851		.2427	.4684	.5092	.4673	
		30	.0439	.1616	.2375	.2394		.2203	.4182	.5479	.4258	
D												
	200	5	.5853	.8700			**.8776**	.6085	**.8823**			.8772
		10	.4126	.7299	.7612			.4610	.7707	.8048		
		15	.3034	.6250	.6413	.6148		.3651	.6884	.7076	.6839	
		20	.2330	.5327	.5486	.5179		.3252	.6193	.6347	.6005	
		25	.1811	.4520	.4822	.4501		.2783	.5604	.5737	.5355	
		30	.1365	.3870	.4222	.3960		.2630	.5228	.5260	.4857	

Table 5. Rejection frequencies in 10000 replications of Experiments B, C, and D, for the optimal test ϕ^*, the Haugh and modified Haugh tests ϕ_H^M and $\tilde{\phi}_H^M$, the Koch and Yang and modified Koch and Yang tests $\phi_{KY;i}^M$ and $\tilde{\phi}_{KY;i}^M$, for various values of M and i, at significance level $\alpha = 0.05$, for series lengths $N = 100$ and 200, respectively. Rejection is based on asymptotic and empirical critical values, respectively. Boldface indicate the winner.

References

[1] Armitage, P., and Colton, T. (1998). McNemar test, in Lachenbruch, P. A., Ed., *Encyclopedia of Biostatistics*, Vol. 3. New York: John Wiley, pp. 2486-2487.

[2] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and adaptative Estimation for Semiparametric Models*. New York: Springer.

[3] Drost, F. C., Klaassen, C. A. J., and Werker, B. J. M. (1997). Adaptive estimation in time-series models, *Annals of Statistics* **25**, 786-818.

[4] El Himdi, K., and Roy, R. (1997). Tests for noncorrelation of two multivariate ARMA time series. *The Canadian Journal of Statistics* **25**, 233-256.

[5] Garel, B., and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics* **47**, 551-579.

[6] Geweke, J. (1981). A comparison of tests of independence of two covariance-stationary time series. *Journal of the American Statistical Association* **47**, 551-579.

[7] Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Second Edition. Belmont: Wadsworth.

[8] Hall, P., and Heyde, C. C. (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.

[9] Hallin, M., and Werker, B. J. M. (1999). Optimal testing for semi-parametric autoregressive models: from Gaussian Lagrange multipliers to regression rank scores and adaptive tests. In *Asymptotics, Nonparametrics, and Time Series* (S. Ghosh, Ed.), 295-358. New York: M. Dekker.

[10] Hallin, M., and Saidi, A. (2003). Testing noncorrelation of VARMA series. *Journal of Time Series Analysis*, to appear.

[11] Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association* **71**, 378-385.

[12] Hwang, S. Y., and Basawa, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series. *Stochastic Processes and their Applications* **46**, 91-114.

[13] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419-426.

[14] Koch, P. D., and Yang, S. S. (1986). A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. *Journal of the American Statistical Association* **81**, 533-544.

[15] Koul, H. L., and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model, *Annals of Statistics* **24**, 1025-1052.

[16] Koul, H. L., and Schick, A. (1997). Efficient estimation in nonlinear autoregressive time series models, *Bernoulli* **3**, 247-277.

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

[18] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.

[19] Le Cam, L., and Yang, G. L. (1990). *Asymptotics in Statistics*. New York: Springer-Verlag.

[20] Pierce, D. A. (1977). Relationships—and the lack thereof—between economic time series, with special reference to money and interest rates. *Journal of the American Statistical Association* **72**, 11-22.

[21] Satterthwaite, F.E. (1941). Synthesis of variance. *Psychometrica* **6**, 309-316.

[22] Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110-114.

[23] Shea, B. L. (1989). The exact likelihood of a vector autoregressive moving average model. *Journal of the Royal Statistical Society*, **C**, **38**, 161-184.

[24] Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* **16**, 54-70.

[25] Taniguchi, M., and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer.

[26] van der Vaart A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.