

EFFICIENT DETECTION OF RANDOM COEFFICIENTS IN AUTOREGRESSIVE MODELS

BY ABDELHADI AKHARIF¹ AND MARC HALLIN²

Faculté des Sciences et Techniques de Tanger and Université Libre de Bruxelles

The problem of detecting randomness in the coefficients of an $AR(p)$ model, that is, the problem of testing ordinary $AR(p)$ dependence against the alternative of a random coefficient autoregressive [RCAR(p)] model is considered. A nonstandard LAN property is established for RCAR(p) models in the vicinity of $AR(p)$ ones. Two main problems arise in this context. The first problem is related to the statistical model itself: Gaussian assumptions are highly unrealistic in a nonlinear context, and innovation densities should be treated as nuisance parameters. The resulting semiparametric model however appears to be severely nonadaptive. In contrast with the linear ARMA case, pseudo-Gaussian likelihood methods here are invalid under non-Gaussian densities; even the innovation variance cannot be estimated without a strict loss of efficiency. This problem is solved using a general result by Hallin and Werker, which provides semiparametrically efficient central sequences without going through explicit tangent space calculations. The second problem is related to the fact that the testing problem under study is intrinsically one-sided, while the case of multiparameter one-sided alternatives is not covered by classical asymptotic theory under LAN. A concept of *locally asymptotically most stringent somewhere efficient* test is proposed in order to cope with this one-sided nature of the problem.

1. Introduction.

1.1. *Random coefficient time series models.* Nonlinear models, in time-series analysis as well as in many other areas, were developed as a reaction against the supremacy of linear ones—a situation inherited from strong, though often implicit, Gaussian assumptions. Much attention has been given, in this respect, to bilinear models, threshold models, and the many refinements of ARCH models; see, for example, the monographs by Granger and Andersen (1978), Priestley (1988), Tong (1990), Guégan (1994) and Taniguchi and Kakizawa (2000).

Random coefficient time series models are another tool for handling the possible nonlinear features of real-life data [see Vervaat (1979)]. A *random coefficient*

Received February 2001; revised June 2002.

¹Supported by the Fonds d'Encouragement à la Recherche de l'Université Libre de Bruxelles.

²Supported by a P.A.I. contract of the Belgian Federal Government and an Action de Recherche Concertée of the Communauté française de Belgique.

AMS 2000 subject classifications. 62G10, 62M10.

Key words and phrases. Random coefficient autoregressive models, local asymptotic normality, one-sided multiparameter tests, one-sided multiparameter alternatives.

autoregressive model of order p [in short, RCAR(p) model] is a stochastic difference equation of the form

$$(1.1) \quad X_t = \sum_{i=1}^p (a_i + u_{i;t}) X_{t-i} + \varepsilon_t, \quad t \in \mathbb{Z},$$

where $\mathbf{a} = (a_1, \dots, a_p)'$ is a vector of autoregressive parameters, $\{\mathbf{u}_t = (u_{1;t}, \dots, u_{p;t})'; t \in \mathbb{Z}\}$ an i.i.d. sequence of unobservable p -dimensional random vectors, and $\{\varepsilon_t; t \in \mathbb{Z}\}$ another unobservable i.i.d. sequence; $\{\mathbf{u}_t\}$ and $\{\varepsilon_t\}$ throughout are assumed to be mutually independent, and to have mean zero. In the same spirit as in random-effect analysis-of-variance models, the autoregressive coefficients a_i are thus randomly perturbed; when \mathbf{u}_t is almost surely zero, (1.1) reduces to an ordinary autoregressive model, of order less than or equal to p .

Early contributions to the study of RCAR models are due to Anděl (1976), Conlisk (1976), Robinson (1978), Nicholls and Quinn (1980, 1981), Feigin and Tweedie (1985), Weiss (1985) and Guyton, Zhang and Foutz (1986). A comprehensive treatment can be found in the monograph by Nicholls and Quinn (1982). A systematic investigation of asymptotic inference in a general class of nonlinear models (which includes that of random coefficient autoregressive ones) has been carried out in a series of papers by Hwang and Basawa (1993a, b, 1998) and Hwang, Basawa and Reeves (1994); Schick (1996) has proposed a \sqrt{n} -consistent estimator, and Koul and Schick (1996) an adaptive estimator, of the mean a of the random autoregressive parameter in a RCAR(1) model. The local asymptotic normality (LAN) results obtained by these authors however do not hold on the boundary between random coefficient models and the classical ones, and thus cannot be invoked in the detection problem we are considering here.

1.2. *Detection of random coefficients: nonadaptivity.* Before proceeding to identifying and estimating a RCAR model [as in Pagan (1980), Hwang and Basawa (1993a, b, 1998), Schick (1996) or Koul and Schick (1996)], one should make sure that such a sophisticated model is justified. The problem of testing traditional AR dependence against RCAR dependence is thus of basic importance in this context.

This detection problem was first considered in Nicholls and Quinn (1982), where a Gaussian Lagrange multiplier test is derived for the problem. This approach however is not totally appropriate in the context, as it is extremely inefficient for high values of p , due to the intrinsic *one-sidedness* of the alternative (see Section 2). An interesting attempt towards a non-Gaussian perspective is proposed by Ramanathan and Rajarshi (1994) for the first-order [RCAR(1)] case. Their approach is based on the ranks of squared residuals, and thus leads to tests of the *signed rank* type—since ranking squared residuals is equivalent to ranking their absolute values. Signed rank tests however require a symmetry assumption on the innovation density, which is highly unnatural in this context. Moreover, optimality

issues are not addressed (only Wilcoxon scores are considered). And, higher order extensions ($p > 1$) would run into similar one-sidedness problems as the Lagrange multiplier approach. Still for the RCAR(1) models, a locally best invariant test is studied by Lee (1998), which, unlike the preceding methods, does not require independence between the innovation process $\{\varepsilon_t\}$ and the random perturbations of the autoregressive coefficient $\{u_t\}$, but unfortunately is not valid unless the very strong assumption can be made that $\{(\varepsilon_t, u_t)\}$ are jointly Gaussian.

The same detection problem is addressed here, in the general case of order p , and in a semiparametric context where neither the density f of ε_t , nor that of \mathbf{u}_t , is specified. Except for standard regularity conditions (finiteness of moments, finiteness of Fisher information, . . .), no restrictions are placed on these densities. Emphasis is put on asymptotic optimality issues. A LAN property, relying on a nonstandard mean-square differentiability property (quadratic mean differentiability here involves second-order derivatives of innovation densities), is established in Section 2 for fixed f . Even in the first-order case, this LAN property does not follow from and does not have the same form as the one established by Koul and Schick (1996) for the “open” RCAR(1) model; just as in the traditional AR case in the vicinity of unit roots, a discontinuity in the nature of local experiments is thus observed here on the boundary of the model.

Inspection of the central sequence reveals that the model is *strongly* nonadaptive: semiparametric efficiency is thus strictly less than parametric efficiency, at all innovation densities, including the Gaussian. Even the variance of the innovation—which in classical time-series models safely can be replaced with its empirical counterpart—cannot be estimated here without a loss of efficiency. As a consequence, the traditional pseudo-Gaussian approach to the problem is not valid.

Several approaches can be considered in order to solve these problems related to unspecified innovation densities.

(ia) The “fully orthodox” semiparametric approach consists in extending the “fixed- f ” LAN result into a “joint” LAN result involving both the parameter of interest and the unknown innovation density f ; tangent space techniques, and an adequate estimation \hat{f}_n of f then in principle allow for constructing uniformly semiparametrically efficient estimation and testing techniques for the parameter of interest (“uniformly” here means “uniformly in f ”). Unfortunately, in contrast to the more classical i.i.d. situation [treated in the monograph by Bickel, Klaassen, Ritov and Wellner (1993)], no general “joint” LAN results are available in the time-series context, and the existing ones [Drost, Klaassen and Werker (1996); Koul and Schick (1996, 1997)] do not cover the present case. Moreover, experience shows that “uniform efficiency,” just as adaptivity (which is a particular case) has a high finite-sample cost (in terms of efficiency), due to the slow convergence of \hat{f}_n .

(ib) A slightly less ambitious objective is that of reaching semiparametric efficiency at some chosen (up to a scale parameter) f only, thus avoiding the

problems arising from an estimation of f . This objective, when semiparametric efficiency is sought at Gaussian densities, is (or should be) the starting point of pseudo-Gaussian methods. Its implementation requires the derivation of semiparametrically efficient Gaussian central sequences and thus, in principle, the same LAN results and tangent space techniques as in (ia); from a mathematical point of view, (ib) thus is just a variant of (ia).

(ii) A third approach, based on a classical invariance principle, leads to rank-based inference [based on ranks, not on signed ranks as in Ramanathan and Rajarshi (1994)]. This approach can be viewed [see Hallin and Werker (2003)] as a nonstandard sample-splitting method, where the sample is split into residual ranks on one side, and the order statistic of residuals on the other. It allows either for somewhere semiparametrically efficient (at chosen f) tests, or for uniformly (at all f) semiparametrically efficient permutation tests. This third approach is the subject of ongoing research, and is not touched here.

We deliberately put emphasis on the pseudo-Gaussian approach (ib) because we feel its practical implementation is easier; moreover, as we shall see, it involves autocorrelations of residuals and squared residuals, which are well-accepted tools in other time series contexts.

The method we are using is based on results by Hallin and Werker (2003) which, via invariance arguments, allow for bypassing the usual steps related to establishing “joint” LAN results and explicit tangent space calculations. These results actually provide the efficient score functions for a wide range of time series models. The same results also could be invoked for deriving the “uniformly efficient” procedures, mentioned in (ia) above, but this would overload the paper while obscuring the implementation of the Schaafsma and Smid (1966) stringency concept, which we now briefly present.

1.3. *Detection of random coefficients: one-sidedness of alternatives.* Another important question to be faced is related to the intrinsic one-sided nature of the testing problem under study. Local experiments under LAN indeed converge weakly, in the Le Cam distance, to p -dimensional Gaussian shift experiments. In these Gaussian experiments, the alternative of interest is the positive orthant in \mathbb{R}^p . For $p = 1$, this orthant reduces to a half line, and the theory of one-sided tests in one-parameter location families applies; overlooking the one-sided nature of the alternative wastes away about 1/2 of the type I error of the test. For $p = 2$, depending on the covariance matrix, the same attitude may cost about 3/4 of the type I error; in general, this wasted proportion of type I error can be as high as $1 - (1/2)^p$, which is clearly prohibitive as p increases!

This problem of testing against convex cones other than half-spaces has seldom been considered in the LAN context; see, however, Ehm, Mammen and Müller (1995). The related question of asymptotic variance bounds for estimators at

boundary points of the parameter space is briefly touched in van der Vaart (1988, 1989), and more systematically treated in Rieder (2000).

Testing problems involving one-sided multiparameter alternatives also have been treated in the more traditional setting of finite-sample Gaussian location models, where no universally satisfactory solution seems to have emerged (see Section 3.3 for a short review). We are adapting here to the LAN context a concept of *most stringent somewhere most powerful tests* introduced by Schaafsma and Smid (1966). Combined with efficiency considerations, this concept allows for defining *locally asymptotically most stringent, somewhere efficient* tests of the pseudo-Gaussian type.

1.4. *Outline of the paper.* The paper is organized as follows. Section 2 is devoted to the “fixed- f ” LAN structure of the model (Proposition 2.1). A “correlogram-based version” of the same LAN property is given in Proposition 2.2; this version allows for a better intuitive interpretation of the central sequence, and an easier diagnosis of its adaptivity/nonadaptivity features. Section 3 deals with semiparametric efficiency: the explicit form of the semiparametrically efficient Gaussian central sequence is derived in Proposition 3.1, and its asymptotic linearity property is established in Proposition 3.2. Section 4 then deals with the proposed tests and detection procedures. The proof of the nonstandard quadratic mean differentiability property which serves as a basis for the LAN results of Section 2 is given in Section 5.

2. Local asymptotic normality (LAN).

2.1. *Notation and main assumptions.* Let f and $g_{\mathbf{u}}$ denote the probability densities of ε_t and \mathbf{u}_t , respectively, in the RCAR(p) model (1.1). The null hypothesis we are interested in is that of traditional AR(p) dependence, with unspecified innovation density f , that is, the hypothesis under which $P[\mathbf{u}_t = \mathbf{0}] = 1$. The tests we are deriving should be valid (have level α), or asymptotically valid (have asymptotic level α) under arbitrary innovation density f , except perhaps for some mild technical assumptions. As usual, unless estimated-score procedures are adopted, uniformly (with respect to f) optimal tests typically do not exist, and a specific choice of the density f at which optimality is expected has to be made. The situation is somewhat better with the perturbation density $g_{\mathbf{u}}$, since only the correlation matrix $\rho_{\mathbf{u}}$ of \mathbf{u}_t happens to have an impact upon asymptotic Fisher information quantities. Deriving an optimal procedure against specific correlation structures $\rho_{\mathbf{u}}$ would make little sense, though, and we restrict our optimality considerations to diagonal structures, under which $\rho_{\mathbf{u}}$ is diagonal, with diagonal elements equal either to zero (absence of randomness in the corresponding component) or one. This can be achieved through the following parametrization.

Denote by $P_{f;\mathbf{a},\sigma^2}^{(n)}$ the probability distribution of an observed series $\mathbf{X}^{(n)} = (X_1^{(n)}, \dots, X_n^{(n)})$ generated by

$$(2.1) \quad X_t = \sum_{i=1}^p (a_i + \sigma_i u_{i;t}) X_{t-i} + \varepsilon_t, \quad t \in \mathbb{Z},$$

where $\mathbf{a} := (a_1, \dots, a_p) \in \mathcal{A}$, the set of all values of the (mean value of the) autoregressive parameter \mathbf{a} satisfying the traditional *causality* assumptions, and $\sigma^2 := (\sigma_1^2, \dots, \sigma_p^2) \in \mathbb{R}_+^p$; the two noises $\{\varepsilon_t\}$ and $\{\mathbf{u}_t\} = \{(u_{1;t}, \dots, u_{p;t})'\}; t \in \mathbb{Z}\}$ are mutually independent, as in (1.1), with the additional assumption that $E[\mathbf{u}_t \mathbf{u}_t'] = \mathbf{I}_p$, the $p \times p$ unit matrix.

Writing $\mathcal{H}_f^{(n)}(a_1, \dots, a_p; \sigma_1^2, \dots, \sigma_p^2)$, or $\mathcal{H}_f^{(n)}(\mathbf{a}; \sigma^2)$ for the simple hypothesis $\{P_{f;\mathbf{a},\sigma^2}^{(n)}\}$, the null hypothesis we are interested in is

$$(2.2) \quad \mathcal{H}^{(n)} := \bigcup_f \mathcal{H}_f^{(n)} := \bigcup_f \bigcup_{\mathbf{a}} \mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0}),$$

where the union is taken over $\mathbf{a} \in \mathcal{A}$, and over all densities f satisfying assumptions A2(i) and A2(ii) below. The alternatives we would like to detect are of the general form (1.1), with completely unspecified f and nondegenerate $g_{\mathbf{u}}$. Optimality will be sought against alternatives of the form $\bigcup_{\mathbf{a}} \bigcup_{\sigma^2 \neq \mathbf{0}} \mathcal{H}_f^{(n)}(\mathbf{a}; \sigma^2)$, where the union $\bigcup_{\sigma^2 \neq \mathbf{0}}$ is taken over all vectors σ^2 such that $\sigma_i^2 > 0$ for at least one value of $i = 1, \dots, p$.

Such alternatives are typically one-sided ($\sigma_i < 0$ makes little sense); the corresponding set of parameter values is the positive orthant of \mathbb{R}^p (deprived of its apex, located at the origin). Even in simple Gaussian location models, finite-sample optimality in one-sided testing of multivariate parameters is a delicate issue, which has no clearcut solution—except for the half-space case, which essentially reduces to the one-dimensional situation. The asymptotic version of the problem apparently has not been considered so far in the LAN context—except, again, for the half-space case, for which locally asymptotically most powerful procedures are well known [see, e.g., Le Cam (1986), page 297]. Two attitudes then are possible:

(i) either a formal hypothesis testing approach is maintained; one-sided procedures then can be imported, via the weak convergence property of local experiments, from the literature on one-sample location testing against one-sided multivariate alternatives;

(ii) or, emphasis is put on heuristic detection rather than on formal hypothesis testing; for instance, p one-dimensional alternatives of the form $\sigma_i = \sigma \delta(i, i_0)$, $i = 1, \dots, p$ [$\sigma > 0$; $\delta(\cdot, \cdot)$ stands for the Kronecker symbol], one for each lag $i_0 = 1, \dots, p$, are considered separately; this latter approach is closer to the spirit and practice of correlogram inspection methods, which most time-series

practitioners would favor. Both approaches are considered, in Sections 4.1, 4.2 and 4.3, respectively.

The basic tool, as usual, is the local asymptotic normality structure of the family of likelihoods associated with (1.1), in the vicinity of $\sigma^2 = \mathbf{0}$. Establishing LAN requires some technical assumptions which, for convenient reference, we are listing here. Not all of them are required to hold throughout.

(A1) The autoregressive parameter $\mathbf{a} \in \mathbb{R}^p$ is such that the roots of $z^p - \sum_{i=1}^p a_i z^{p-i} = 0$, $z \in \mathbb{C}$ all lie inside the unit disc; the resulting parameter set is denoted by \mathcal{A} .

(A2) The innovation density f is such that:

- (i) $f(x) > 0$, $x \in \mathbb{R}$, $\int x f(x) dx = 0$ and $(m_4)_f := \int x^4 f(x) dx < \infty$;
- (iia) f is absolutely continuous, with a.e. derivative \dot{f} and finite Fisher information $\mathcal{I}_\phi(f) := \int \phi_{\dot{f}}^2(x) f(x) dx < \infty$, where $\phi_f := -\dot{f}/f$;
- (iib) \dot{f} in turn is absolutely continuous, with right-continuous a.e. derivative \ddot{f} ; letting $\psi_f := \ddot{f}/f$, assume that $\mathcal{I}_\psi(f) := \int \psi_f^2(x) f(x) dx < \infty$.

Letting $\sigma_f^2 := \int x^2 f(x) dx$ and $(m_3)_f := \int x^3 f(x) dx$, note that under (A2) $\sigma_f^2 < \infty$ and $(m_3)_f < \infty$ and

$$\begin{aligned} \phi_f(x) &= \sigma^{-1} \phi_{f_1}(x/\sigma), & \psi_f(x) &= \sigma^{-2} \psi_{f_1}(x/\sigma), \\ \mathcal{I}_\phi(f_1) &= \sigma^2 \mathcal{I}_\phi(f), & \mathcal{I}_\psi(f_1) &= \sigma^4 \mathcal{I}_\psi(f), \end{aligned}$$

where $f(x) := \sigma^{-1} f_1(x/\sigma)$. Moreover,

$$\begin{aligned} \int \phi_f(z) f(z) dz &= \int \phi_f(z) z^2 f(z) dz = \int \psi_f(z) f(z) dz \\ &= \int \psi_f(z) z f(z) dz = 0, \end{aligned}$$

whereas

$$\int \phi_f(z) z f(z) dz = \frac{1}{2} \int \psi_f(z) z^2 f(z) dz = 1.$$

For Gaussian densities,

$$\begin{aligned} \phi_{f_1}(z) &= \phi_g(z) = z, & \mathcal{I}_\phi(f_1) &= 1, \\ \psi_{f_1}(z) &= \psi_g(z) = z^2 - 1 & \text{and } \mathcal{I}_\psi(f_1) &= 2, \end{aligned}$$

while, for logistic ones, we have

$$\begin{aligned} \phi_{f_1}(z) &= \frac{1 - e^{-z}}{1 + e^{-z}}, & \mathcal{I}_\phi(f_1) &= \frac{\pi^2}{9} \approx 1.097, \\ \psi_{f_1}(z) &= \frac{e^{-2z} - 4e^{-z} + 1}{(1 + e^{-z})^2} & \text{and } \mathcal{I}_\psi(f_1) &= \frac{\pi^4}{45} \approx 2.165. \end{aligned}$$

On the density $g_{\mathbf{u}}$ of the random perturbations \mathbf{u}_t , we make the following assumption:

$$(A3) \int \mathbf{v} g_{\mathbf{u}}(\mathbf{v}) d\mathbf{v} = \mathbf{0} \text{ and } \int \mathbf{v}\mathbf{v}' g_{\mathbf{u}}(\mathbf{v}) d\mathbf{v} = \mathbf{I}.$$

Finally, the continuity of Fisher information plays an essential role in local asymptotic normality. This continuity follows without any additional assumption about the Fisher information associated with autoregressive parameters; as for the information associated with the scale parameter of \mathbf{u}_t , we will assume:

(A4) Denote by $g_i, i = 1, \dots, p$, the marginal densities of $g_{\mathbf{u}}$. Letting

$$(2.3) \quad \begin{aligned} \mathfrak{I}_{i\psi}^x(f; t) &:= \int_{z=-\infty}^{\infty} \frac{1}{t^2} \frac{[\int_{w=0}^t \int \ddot{f}(z - xvw) x^2 v^2 g_i(v) dv dw]^2}{\int f(z - xtv) g_i(v) dv} dz \\ &= \int_{z=-\infty}^{\infty} \frac{1}{t^2} \left[\frac{\int f \dot{f}(z - xtv) xv g_i(v) dv}{[\int f(z - xtv) g_i(v) dv]^{1/2}} \right]^2 dz, \end{aligned}$$

$i = 1, \dots, p, t > 0,$

and $\mathfrak{I}_{i\psi}^x(f; 0) := x^4 \mathfrak{I}_{\psi}(f)$, the functions $t \mapsto \mathfrak{I}_{i\psi}^x(f; t)$ are continuous from the right at $t = 0$, for all x .

2.2. *Local asymptotic normality (LAN).* Assuming that $X_0^{(n)}, X_{-1}^{(n)}, \dots, X_{-p+1}^{(n)}$ have been observed [starting values have no influence on asymptotics; see Hallin and Werker (1999) for a detailed discussion], let

$$(2.4) \quad Z_t := Z_t^{(n)}(\mathbf{a}) := X_t^{(n)} - \sum_{i=1}^p a_i X_{t-i}^{(n)}$$

(the notation Z_t is used whenever no confusion is possible); under $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$, Z_t clearly coincides with ε_t . Considering *local* parameter sequences of the form $(\mathbf{a}'; \mathbf{0}')' + n^{-1/2} \boldsymbol{\tau}^{(n)}$, with

$$\boldsymbol{\tau}^{(n)} := \begin{pmatrix} \boldsymbol{\alpha}^{(n)} \\ \boldsymbol{\delta}^{(n)} \end{pmatrix},$$

where $\boldsymbol{\alpha}^{(n)} := (\alpha_1^{(n)}, \dots, \alpha_p^{(n)})' \in \mathbb{R}^p$ and $\boldsymbol{\delta}^{(n)} := (\delta_1^{(n)}, \dots, \delta_p^{(n)})' \in \mathbb{R}_+^p$ are such that

$$(2.5) \quad \limsup_{n \rightarrow \infty} \sum_{i=1}^p (|\alpha_i^{(n)}| + \delta_i^{(n)}) < \infty,$$

denote by $\Lambda_f^{(n)}(\mathbf{a}; \boldsymbol{\alpha}^{(n)}, \boldsymbol{\delta}^{(n)}) = \Lambda_f^{(n)}(\mathbf{a}; \boldsymbol{\alpha}^{(n)}, \boldsymbol{\delta}^{(n)}; \mathbf{X}^{(n)}; X_0^{(n)}, \dots, X_{-p+1}^{(n)})$ the logarithm of the likelihood ratio of $\mathcal{H}_f^{(n)}(\mathbf{a} + n^{-1/2} \boldsymbol{\alpha}^{(n)}; n^{-1/2} \boldsymbol{\delta}^{(n)})$ with respect

to $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$ (conditional on $X_0^{(n)}, \dots, X_{-p+1}^{(n)}$). Then,

$$(2.6) \quad \Lambda_f^{(n)}(\mathbf{a}; \boldsymbol{\alpha}^{(n)}, \boldsymbol{\delta}^{(n)}) = \sum_{t=1}^n \log \left[\int_{\mathbb{R}^p} f \left(Z_t - \sum_{i=1}^p \left(n^{-1/2} \alpha_i^{(n)} + n^{-1/4} \sqrt{\delta_i^{(n)}} v_i \right) X_{t-i} \right) g_{\mathbf{u}}(\mathbf{v}) d\mathbf{v} \right] - \sum_{t=1}^n \log[f(Z_t)] + o_P(1);$$

the $o_P(1)$ term [under $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$, as $n \rightarrow \infty$] corresponds to the influence of the starting values X_0, \dots, X_{-p+1} .

Also, let $h_u = h_u(\mathbf{a})$, $u \in \mathbb{Z}$, denote the Green’s functions associated with the difference operator $\mathbf{a}(L) = 1 - \sum_{i=1}^p a_i L^i$, where L stands for the lag operator [$h_u(\mathbf{a})$ thus is characterized by $(\mathbf{a}(L))^{-1} := \sum_{u=0}^{\infty} h_u L^u$]. Put

$$(2.7) \quad \Gamma_f(\mathbf{a}) = \begin{pmatrix} \mathfrak{I}_{\phi}(f_1) \Gamma_{\mathbf{X}}(\mathbf{a}) & \frac{1}{2} \mathfrak{I}_{\phi\psi}(f_1)(m_3)_{f_1} \Gamma_{\mathbf{X}\mathbf{X}^2}(\mathbf{a}) \\ \frac{1}{2} \mathfrak{I}_{\phi\psi}(f_1)(m_3)_{f_1} \Gamma'_{\mathbf{X}\mathbf{X}^2}(\mathbf{a}) & \frac{1}{4} \mathfrak{I}_{\psi}(f_1) \Gamma_{f; \mathbf{X}^2}(\mathbf{a}) \end{pmatrix},$$

where

$$\mathfrak{I}_{\phi\psi}(f) := \int \phi_{f_1}(x) \psi_{f_1}(x) f_1(x) dx,$$

$\Gamma_{\mathbf{X}}(\mathbf{a})$ and $\Gamma_{f_1; \mathbf{X}^2}(\mathbf{a})$ stand for the covariance matrices, under $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$, of $\sigma^{-1}(X_1, \dots, X_p)'$ and $\sigma^{-2}(X_1^2, \dots, X_p^2)'$, respectively, and $(m_3)_{f_1} \Gamma_{\mathbf{X}\mathbf{X}^2}(\mathbf{a})$ for their cross-covariance matrix. Thus,

$$(2.8) \quad (\Gamma_{\mathbf{X}}(\mathbf{a}))_{ij} = \sum_{u=|i-j|}^{\infty} h_u h_{u-|i-j|},$$

$$(2.9) \quad \begin{aligned} (\Gamma_{f; \mathbf{X}^2}(\mathbf{a}))_{ij} &= (m_4)_{f_1} \sum_{u=0}^{\infty} h_u^2 h_{u-|i-j|}^2 + \sum_{0 \leq u \neq v < \infty} h_u^2 h_v^2 h_{u-|i-j|} h_{v-|i-j|} \\ &+ 4 \sum_{u=0}^{\infty} \sum_{v=u+1}^{\infty} h_u h_{u-|i-j|} h_v h_{v-|i-j|} \end{aligned}$$

and

$$(2.10) \quad (\Gamma_{\mathbf{X}\mathbf{X}^2}(\mathbf{a}))_{ij} = \begin{cases} \sum_{u=0}^{\infty} h_u h_{u-|i-j|}^2, & i \leq j, \\ \sum_{u=0}^{\infty} h_u^2 h_{u-|i-j|}, & i > j. \end{cases}$$

Neither $\Gamma_{\mathbf{X}}$ nor $\Gamma_{\mathbf{X}\mathbf{X}^2}$ depends on f ; the dependence on f lies in the multiplicative effects of the information factors $\mathfrak{I}_{\phi\psi}$ and $\mathfrak{I}_{\phi}(f_1)$, and of the skewness parameter $(m_3)_{f_1}$, whereas $\Gamma_{f;\mathbf{X}^2}$ depends on f through the kurtosis parameter $(m_4)_{f_1}$, in a slightly more complicated way. We then have the following LAN result.

PROPOSITION 2.1. *Let (A1) and (A2) hold. Then, under $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$, as $n \rightarrow \infty$, for all sequences $\boldsymbol{\tau}^{(n)} = (\boldsymbol{\alpha}^{(n)'}, \boldsymbol{\delta}^{(n)'})'$ satisfying (2.5):*

(i) $\Lambda_f^{(n)}(\mathbf{a}; \boldsymbol{\alpha}^{(n)}, \boldsymbol{\delta}^{(n)}) = \boldsymbol{\tau}^{(n)'} \boldsymbol{\Delta}_f^{(n)}(\mathbf{a}) - \frac{1}{2} \boldsymbol{\tau}^{(n)'} \boldsymbol{\Gamma}_f(\mathbf{a}) \boldsymbol{\tau}^{(n)} + o_P(1)$ with the $2p$ -dimensional random vector (the central sequence)

$$(2.11) \quad \boldsymbol{\Delta}_f^{(n)}(\mathbf{a}) := \begin{pmatrix} \boldsymbol{\Delta}_{f;I}^{(n)}(\mathbf{a}) \\ \boldsymbol{\Delta}_{f;II}^{(n)}(\mathbf{a}) \end{pmatrix} := n^{-1/2} \begin{pmatrix} \sum_{t=2}^n \phi_f(Z_t) X_{t-1} \\ \vdots \\ \sum_{t=p+1}^n \phi_f(Z_t) X_{t-p} \\ \frac{1}{2} \sum_{t=2}^n \psi_f(Z_t) X_{t-1}^2 \\ \vdots \\ \frac{1}{2} \sum_{t=p+1}^n \psi_f(Z_t) X_{t-p}^2 \end{pmatrix};$$

(ii) $\boldsymbol{\Delta}_f^{(n)}(\mathbf{a})$ is asymptotically multinormal, with mean $\mathbf{0}$ and covariance $\boldsymbol{\Gamma}_f(\mathbf{a})$.

PROOF. The proof consists in checking that Swensen’s (1985) conditions (1.2)–(1.7), which are sufficient for LAN, hold. The quantities $X_{n,t}$ and $Z_{n,t}$ appearing in Swensen’s paper here take the form (in Swensen’s notation)

$$X_{n,t} := f^{-1/2}(Z_t)$$

$$\times \left[\int_{\mathbb{R}^p} f \left(Z_t - \sum_{i=1}^p (n^{-1/2} \alpha_i^{(n)} + n^{-1/4} \sqrt{\delta_i^{(n)}} v_i) X_{t-i} \right) g_{\mathbf{u}}(\mathbf{v}) d\mathbf{v} \right]^{1/2} - 1$$

and

$$Z_{n,t} := \frac{1}{2} n^{-1/2} \left(\phi_f(Z_t) \sum_{i=1}^p \alpha_i^{(n)} X_{t-i} + \frac{1}{2} \psi_f(Z_t) \sum_{i=1}^p \sqrt{\delta_i^{(n)}} X_{t-i}^2 \right),$$

respectively. Conditions (1.3)–(1.7) follow more or less routinely from the assumptions made; the only delicate one actually is the first one [(1.2) in Swensen’s notation], which requires that

$$(2.12) \quad \lim_{n \rightarrow \infty} \sum_{t=1}^n E[(X_{n,t} - Z_{n,t})^2] = 0.$$

This however is a direct consequence (see Swensen’s Lemma 2) of the quadratic mean differentiability, at $\mathbf{t} = \mathbf{0}$ and for all $x \in \mathbb{R}$, of $\mathbf{t} \mapsto [\int f(z - x\mathbf{t}'\mathbf{v})g(\mathbf{v}) d\mathbf{v}]^{1/2}$. This quadratic mean differentiability property is established in detail in the Appendix (Lemma 5.1). \square

2.3. *A correlogram-based version of LAN.* The LAN result of Proposition 2.1 is unsatisfactory on two accounts. First, it is hardly interpretable, as the definition (2.11) of the central sequences $\Delta_f^{(n)}(\mathbf{a})$ conveys little intuitive content. Second, the behavior of $\Delta_f^{(n)}(\mathbf{a})$ under innovation densities $g \neq f$ is hard to tell from (2.11). For instance, it is all but clear whether the empirical innovation variance $\hat{\sigma}^2 = (\hat{\sigma}^{(n)})^2 := n^{-1} \sum_{t=1}^n Z_t^2$ can be substituted, in (2.11), for the exact one σ^2 (i.e., whether the scores associated with $f_{\hat{\sigma}}$ can be used in case σ^2 is unspecified). Proposition 2.2 provides an alternative formulation of LAN, based on a central sequence $\underline{\Delta}_f^{(n)}(\mathbf{a})$ which, contrary to $\Delta_f^{(n)}(\mathbf{a})$, is measurable with respect to residuals, and allows for a better interpretation and for studying more easily the impact of a variation of the underlying densities.

More precisely, $\underline{\Delta}_f^{(n)}(\mathbf{a})$ relies on a generalized concept of residual autocorrelations and higher order residual moments. Generalized residual autocorrelations are the non-Gaussian counterparts of the traditional autocorrelation coefficients; they play the same central role in all inference problems for linear ARMA processes with innovation density f as the traditional residual correlogram does in classical time series; see Hallin and Werker (1999) or Hallin and Puri (1994). The *lag k residual autocorrelation coefficient associated with density f* (and computed from the residual series Z_1, \dots, Z_n) is defined as

$$(2.13) \quad r_{f;k}^{(n)} := (n - k)^{-1} \sum_{t=k+1}^n \phi_{f_1} \left(\frac{Z_t}{\sigma} \right) \frac{Z_{t-k}}{\sigma} (\mathfrak{J}_\phi(f_1))^{-1/2}.$$

It is well known, however, that autocorrelation coefficients are not sufficient in the detection of nonlinear serial features. Bilinear models, for example, typically require a third-order generalization of classical correlograms, the so-called *cubic autocorrelations*, associated with the concept of *bispectrum* [see Benghabrit and Hallin (1996, 1998)]. In the present case of random coefficient models, another class of serial statistics quite naturally enters into the picture. Their structure looks quite similar to that of cubic autocorrelations, but with score function ψ_f instead of ϕ_f :

$$(2.14) \quad q_{f;k\ell}^{(n)} := \begin{cases} (n - k)^{-1} \sum_{t=k+1}^n \psi_{f_1} \left(\frac{Z_t}{\sigma} \right) \left(\frac{Z_{t-k}}{\sigma} \right)^2 [\mathfrak{J}_\psi(f_1)(m_4)_{f_1}]^{-1/2}, & k = \ell, \\ (n - \ell)^{-1} \sum_{t=\ell+1}^n \psi_{f_1} \left(\frac{Z_t}{\sigma} \right) \frac{Z_{t-k} Z_{t-\ell}}{\sigma^2} [\mathfrak{J}_\psi(f_1)]^{-1/2}, & k < \ell. \end{cases}$$

We then have the following result.

PROPOSITION 2.2. *The LAN result of Proposition 2.1 still holds, with*

$$\underline{\Delta}_f^{(n)}(\mathbf{a}) = \begin{pmatrix} \underline{\Delta}_{f;I}^{(n)}(\mathbf{a}) \\ \underline{\Delta}_{f;II}^{(n)}(\mathbf{a}) \end{pmatrix}$$

substituted for $\Delta_f^{(n)}(\mathbf{a})$, where [writing h_u for $h_u(\mathbf{a})$]

$$(2.15) \quad (\underline{\Delta}_{f;I}^{(n)}(\mathbf{a}))_i := \sum_{k=i}^{n-1} h_{k-i} (n-k)^{1/2} r_{f;k}^{(n)} [\mathfrak{L}_\phi(f_1)]^{1/2}, \quad i = 1, \dots, p,$$

and

$$(2.16) \quad \begin{aligned} (\underline{\Delta}_{f;II}^{(n)}(\mathbf{a}))_i &:= \frac{1}{2} \sum_{k=i}^{n-1} h_{k-i}^2 (n-k)^{1/2} q_{f;kk}^{(n)} [\mathfrak{L}_\psi(f_1)(m_4)_{f_1}]^{1/2} \\ &+ \sum_{i \leq k < \ell \leq n-1} h_{k-i} h_{\ell-i} (n-\ell)^{1/2} q_{f;k\ell}^{(n)} [\mathfrak{L}_\psi(f_1)]^{1/2}, \end{aligned}$$

$i = 1, \dots, p.$

PROOF. Central sequences always are defined up to $o_P(1)$ terms [as $n \rightarrow \infty$, under $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$]. Thus, it is sufficient to show that (2.15) and (2.16) are asymptotically equivalent to (2.11). This easily follows from noting that

$$(2.17) \quad X_t = \sum_{k=0}^{t-1} h_k(\mathbf{a}) Z_{t-k}(\mathbf{a}) + o_P(|\lambda_M|^{-t}) \quad \text{as } t \rightarrow \infty,$$

uniformly in n , where λ_M is the inverse of the root of the characteristic polynomial (see A1) which lies closest to the unit circle; the result follows from inserting (2.17) into (2.11), and rearranging the sums. \square

In case of a Gaussian f (denote by $\Delta_g^{(n)}$ the corresponding central sequence), (2.13) and (2.14) reduce to

$$(2.18) \quad r_k^{(n)} := (n-k)^{-1} \sigma^{-2} \sum_{t=k+1}^n Z_t Z_{t-k}$$

(a classical residual autocorrelation coefficient) and

$$(2.19) \quad q_{k\ell}^{(n)} := \begin{cases} (n-k)^{-1} \sum_{t=k+1}^n \left(\frac{Z_t^2}{\sigma^2} - 1 \right) \left(\frac{Z_{t-k}}{\sigma} \right)^2 [(m_4)_{f_1} ((m_4)_{f_1} - 1)]^{-1/2}, & k = \ell, \\ (n-\ell)^{-1} \sum_{t=\ell+1}^n \left(\frac{Z_t^2}{\sigma^2} - 1 \right) \frac{Z_{t-k} Z_{t-\ell}}{\sigma^2} [(m_4)_{f_1} - 1]^{-1/2}, & k < \ell, \end{cases}$$

where, moreover, $[(m_4)_{f_1} - 1] = 2$ and $[(m_4)_{f_1} ((m_4)_{f_1} - 1)] = 6$.

2.4. *Nonadaptivity.* The central sequence $\Delta_f^{(n)}(\mathbf{a})$ [equivalently, $\underline{\Delta}_f^{(n)}(\mathbf{a})$] allows for asymptotically optimal inference under innovation density f , hence for efficient detection of random components at f . This efficiency is entirely based on the fact that a “small” random component in (1.1) induces a shift in the distribution of $\Delta_f^{(n)}(\mathbf{a})$, which has expectation zero under the null hypothesis $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$. Now, if $\Delta_f^{(n)}(\mathbf{a})$ is to be considered for testing the broader semiparametric hypothesis $\mathcal{H}^{(n)}$, it is essential that a change in the underlying innovation density does not induce the same shift as the alternative to be detected: $\Delta_f^{(n)}(\mathbf{a})$ [equivalently, $\underline{\Delta}_f^{(n)}(\mathbf{a})$] thus should remain centered under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$, $g \neq f$. Unfortunately, this is not the case.

The problem, as we shall see, is caused by the presence in $\underline{\Delta}_{f;II}^{(n)}$ of the statistic $q_{f;kk}^{(n)}$. In general, when Z_t has density g , where $g \neq f$, even assuming that $\sigma_g = \sigma_f = \sigma$, we have that

$$E\left[\psi_{f_1}\left(\frac{Z_t}{\sigma}\right)\frac{Z_{t-k}^2}{\sigma^2}\right] \neq 0.$$

A change in the shape of the innovation density thus has the same impact on $\underline{\Delta}_{f;II}^{(n)}$ as the alternative we are trying to detect. The same problem does not occur with $r_{f;k}^{(n)}$, since

$$E\left[\phi_f\left(\frac{Z_t}{\sigma}\right)\left(\frac{Z_{t-k}}{\sigma}\right)\right] = 0,$$

irrespective of the (centered) density g of Z_t .

Gaussian densities at first sight are an exception. Indeed, for the central sequence associated with the $\mathcal{N}(0, \sigma^2)$ density,

$$(2.20) \quad E\left[\psi_g\left(\frac{Z_t}{\sigma}\right)\frac{Z_{t-k}^2}{\sigma^2}\right] = E\left[\left(\frac{Z_t^2}{\sigma^2} - 1\right)\frac{Z_{t-k}^2}{\sigma^2}\right] = 0,$$

under any g . This holds true, however, only if $\sigma_g = \sigma$: even under Gaussian g , (2.20) does not hold unless $\sigma_g^2 = \sigma_f^2$. A natural idea for overcoming this problem consists in substituting $(\hat{\sigma}^{(n)})^2 := n^{-1} \sum_{t=1}^n Z_t^2$ for σ^2 and $\hat{m}_4^{(n)} := \sum_{t=1}^n (Z_t^4 / \hat{\sigma}^{(n)4})$ for $(m_4)_{f_1}$ (this latter substitution clearly has no asymptotic impact); denote by $\hat{q}_{f;kk}^{(n)}$ the resulting statistic. Then, a simple computation shows that

$$\begin{aligned} \hat{q}_{f;kk}^{(n)} - q_{f;kk}^{(n)} &= \frac{1}{(n-k)[\hat{m}_4^{(n)}(\hat{m}_4^{(n)} - 1)]^{1/2}} \\ &\quad \times \left\{ \sum_{t=k+1}^n \left(\frac{Z_t^2}{\hat{\sigma}^2} - 1\right)\left(\frac{Z_{t-k}^2}{\hat{\sigma}^2}\right) - \sum_{t=k+1}^n \left(\frac{Z_t^2}{\sigma^2} - 1\right)\left(\frac{Z_{t-k}^2}{\sigma^2}\right) \right\} \end{aligned}$$

$$\begin{aligned}
 &+ o_P(n^{-1/2}) \\
 &= \frac{-1}{[\hat{m}_4^{(n)}(\hat{m}_4^{(n)} - 1)]^{1/2}} \\
 &\quad \times \left\{ \hat{\sigma}^2 \left(\frac{1}{\hat{\sigma}^2} - \frac{1}{\sigma^2} \right) - \frac{1}{n-k} \sum_{t=k+1}^n Z_t^2 Z_{t-k}^2 \left(\frac{1}{\hat{\sigma}^4} - \frac{1}{\sigma^4} \right) \right\} \\
 &+ o_P(n^{-1/2}) \\
 &= O_P(n^{-1/2}) \neq o_P(n^{-1/2}).
 \end{aligned}$$

Even under Gaussian densities, the estimated $\hat{\sigma}^2$ thus cannot be substituted for σ^2 into $q_{kk}^{(n)}$, hence into the Gaussian central sequence, without affecting its qualification as a central sequence.

Summing up, the impact on $\underline{\Delta}_{f;II}^{(n)}$ of a variation of either the scale or the shape of the innovation density in general is the same as that of a *small* random component in the autoregression coefficient. This rules out the construction of pseudo-Gaussian procedures for the semiparametric problem considered here. Efficient procedures for the parametric Gaussian model are impossible as well, as soon as the variance of the Gaussian innovation remains unspecified. The RCAR model (1.1) thus is *strongly nonadaptive*.

To conclude this section, note that part (ii) of assumption (A2) requiring the innovation ε_t to have mean zero, remains quite innocuous. If indeed a specified location parameter μ is added into the model, the obvious modification of central sequences consists in replacing Z_t with $Z_t - \mu$. Elementary calculations show that, in such a case, μ in turn can be replaced with $\bar{Z}_n := n^{-1} \sum_{t=1}^n Z_t$ in $r_{f;k}^{(n)}$, $q_{f;k\ell}^{(n)}$ and $q_{f;kk}^{(n)}$ provided that ϕ_f and ψ_f satisfy some mild regularity assumptions, which are trivially satisfied by the Gaussian score functions ϕ_g and ψ_g .

3. Semiparametrically efficient pseudo-Gaussian methods.

3.1. *Semiparametrically efficient Gaussian central sequence.* Since the problem under study is not adaptive, parametric efficiency at given f , as measured by the Fisher information matrix $\Gamma_f(\mathbf{a})$, cannot be reached when underlying innovation densities remain unspecified. The best that can be hoped for is *semiparametric efficiency*—at any density f if “estimated-score” methods (requiring an estimation of the actual innovation density) are considered; at some chosen f if “fixed-score” methods are used.

The results of Section 2, as well as those of Hallin and Werker (2003) which are used in the sequel, allow for both types of methods. Instead of developing uniformly efficient procedures [in the spirit of Choi, Hall and Schick (1996)], we rather concentrate on a semiparametrically efficient pseudo-Gaussian method, that

is, a “fixed-score” method which is valid under unspecified innovation density but reaches semiparametric efficiency under Gaussian conditions.

Explicit derivation of semiparametrically efficient methods usually requires tedious projections along tangent spaces. In the type of time series models under study, a general result of Hallin and Werker (2003) allows for avoiding such calculations. For Gaussian f , this latter result yields the following proposition. Define

$$\hat{r}_k^{(n)} := \frac{\sum_{t=k+1}^n Z_t Z_{t-k}}{(n-k)\hat{\sigma}^2} = r_k^{(n)} + o_P(n^{-1/2}),$$

$$\hat{q}_{k\ell}^{(n)} := \frac{\sum_{t=\ell+1}^n (Z_t^2 - \hat{\sigma}^2) Z_{t-k} Z_{t-\ell}}{(n-\ell)\hat{\sigma}^4[\hat{m}_4^{(n)} - 1]^{1/2}} = q_{k\ell}^{(n)} + o_P(n^{-1/2}),$$

[where o_P 's are taken under any $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$ such that g satisfies Assumption (A2)] and

$$q_{kk*}^{(n)} := \frac{\sum_{t=k+1}^n (Z_t^2 - \sigma^2)(Z_{t-k}^2 - \sigma^2)}{(n-k)\sigma^4[(m_4)_{f_1} - 1]},$$

$$\hat{q}_{kk*}^{(n)} := \frac{\sum_{t=k+1}^n (Z_t^2 - \hat{\sigma}^2)(Z_{t-k}^2 - \hat{\sigma}^2)}{(n-k)\hat{\sigma}^4[\hat{m}_4^{(n)} - 1]};$$

note that $\hat{q}_{kk*}^{(n)}$ coincides with the autocorrelation (at lag k) of squared residuals—a traditional heuristic tool in the analysis of nonlinear time series models.

PROPOSITION 3.1. *The semiparametrically efficient central sequence for Gaussian innovation density f has the form*

$$\Delta_*^{(n)}(\mathbf{a}) := \begin{pmatrix} \Delta_I^{(n)}(\mathbf{a}) \\ \Delta_{*II}^{(n)}(\mathbf{a}) \end{pmatrix},$$

with

$$(\Delta_I^{(n)}(\mathbf{a}))_i := \sum_{k=i}^{n-1} h_{k-i} (n-k)^{1/2} \hat{r}_k^{(n)}, \quad i = 1, \dots, p,$$

and

$$(3.1) \quad (\Delta_{*II}^{(n)}(\mathbf{a}))_i := \frac{1}{2} \sum_{k=i}^{n-1} h_{k-i}^2 (n-k)^{1/2} \hat{q}_{kk*}^{(n)} [\hat{m}_4^{(n)} - 1]$$

$$+ \sum_{i \leq k < \ell \leq n-1} h_{k-i} h_{\ell-i} (n-\ell)^{1/2} \hat{q}_{k\ell}^{(n)} [\hat{m}_4^{(n)} - 1]^{1/2},$$

$i = 1, \dots, p.$

Under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$, $\Delta_*^{(n)}(\mathbf{a})$ is asymptotically multinormal as $n \rightarrow \infty$, with mean $\mathbf{0}$ and covariance matrix

$$(3.2) \quad \mathbf{W}_{*g}(\mathbf{a}) := \begin{pmatrix} \Gamma_{\mathbf{X}}(\mathbf{a}) & \frac{1}{2}(m_3)_{g_1}^2 \Gamma_{\mathbf{X}\mathbf{X}^2}(\mathbf{a}) \\ \frac{1}{2}(m_3)_{g_1}^2 \Gamma'_{\mathbf{X}\mathbf{X}^2}(\mathbf{a}) & \frac{1}{4}[(m_4)_{g_1} - 1] \Gamma_{*g; \mathbf{X}^2}(\mathbf{a}) \end{pmatrix},$$

where

$$(3.3) \quad (\Gamma_{*g; \mathbf{X}^2}(\mathbf{a}))_{ij} := [(m_4)_{g_1} - 1] \sum_{u=0}^{\infty} h_u^2 h_{u-|i-j|}^2 + \sum_{0 \leq u \neq v < \infty} \sum h_u^2 h_{v-|i-j|}^2 + 4 \sum_{u=0}^{\infty} \sum_{v=u+1}^{\infty} h_u h_{u-|i-j|} h_v h_{v-|i-j|}.$$

Note that, contrary to the Gaussian central sequence $\Delta_g^{(n)}$, the efficient central sequence $\Delta_*^{(n)}$ does not involve any unspecified quantity any more. Moreover, asymptotic normality here is proved under $\mathcal{H}_g^{(n)}$, whereas Propositions 2.1 and 2.2 provide information on the behavior of $\Delta_g^{(n)}$ under Gaussian density f only.

COROLLARY 3.1. *The efficient information matrix (for Gaussian innovation density) is*

$$(3.4) \quad \Gamma_{*g}(\mathbf{a}) := \begin{pmatrix} \Gamma_{\mathbf{X}}(\mathbf{a}) & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \Gamma_{*g; \mathbf{X}^2}(\mathbf{a}) \end{pmatrix},$$

with

$$(\Gamma_{*g; \mathbf{X}^2}(\mathbf{a}))_{ij} := 2 \sum_{u=0}^{\infty} h_u^2 h_{u-|i-j|}^2 + \sum_{0 \leq u \neq v < \infty} \sum h_u^2 h_{v-|i-j|}^2 + 4 \sum_{u=0}^{\infty} \sum_{v=u+1}^{\infty} h_u h_{u-|i-j|} h_v h_{v-|i-j|}.$$

PROOF OF PROPOSITION 3.1. The proof consists in showing that the Assumptions (I) and (J) required for Propositions 3.1 and 3.3 in Section 3 of Hallin and Werker (2003) to hold are satisfied. It follows from Proposition 2.2 that $\Delta_g^{(n)}$ is a linear combination (with coefficients depending on θ but not on f) of three types of statistics: $(n - k)^{1/2} r_k^{(n)}$ defined in (2.18),

$$(n - \ell)^{1/2} [(m_4)_{f_1} - 1]^{1/2} q_{k\ell}^{(n)} = (n - \ell)^{-1/2} \sigma^{-4} \sum_{t=\ell+1}^n (Z_t^2 - \sigma^2) Z_{t-k} Z_{t-\ell}$$

and

$$(n - k)^{1/2}[(m_4)_{f_1}((m_4)_{f_1} - 1)]^{1/2} q_{kk}^{(n)} = (n - k)^{-1/2} \sigma^{-4} \sum_{t=\ell+1}^n (Z_t^2 - \sigma^2) Z_{t-k}^2.$$

Each of these three statistics has the form $(n - p)^{-1/2} \sum_{t=p+1}^n J(Z_t, \dots, Z_{t-p})$ considered in Assumptions (I) of Hallin and Werker (2003). Let us show that Assumptions (J(i))–(J(iii)) (same reference) are also satisfied.

For $(n - k)^{1/2} r_k^{(n)}$, the score function J is (up to a factor σ^{-2}) $J_{r_k} : (z_0, z_1, \dots, z_p) \mapsto J_{r_k}(z_0, z_1, \dots, z_p) := z_0 z_k$. This function trivially satisfies Assumptions (J(i)) and (J(ii)). The fact that, denoting by $(\varepsilon_0, \dots, \varepsilon_p)$ an i.i.d. $\mathcal{N}(0, \sigma^2)$ p -tuple, we have $E[J_{r_k}(\varepsilon_0, \dots, \varepsilon_p) | \varepsilon_0] = \varepsilon_0 E[\varepsilon_k] = 0$ implies that $(n - k)^{1/2} r_k^{(n)}$ enters without modification the efficient central sequence: in the notation of Hallin and Werker (2003), $J_{r_k} = J_{r_k}^*$. Assumption (J(iii)) thus reduces to asymptotic linearity of $r_k^{(n)}$ with respect to \mathbf{a} , which is an immediate consequence of the ULAN property of $AR(p)$ models.

The same properties, and the same conclusion, hold for $(n - k)^{1/2} \times [(m_4)_{f_1}((m_4)_{f_1} - 1)]^{1/2} q_{k\ell}^{(n)}$, $k \neq \ell$, with (up to a factor σ^{-4}) the score function

$$J_{q_{k\ell}}(Z_t, \dots, Z_{t-p}) := (Z_t^2 - \sigma^2) Z_{t-k} Z_{t-\ell} = J_{q_{k\ell}}^*(Z_t, \dots, Z_{t-p}).$$

As expected, the case of $(n - k)^{1/2} [(m_4)_{f_1}((m_4)_{f_1} - 1)]^{1/2} q_{kk}^{(n)}$, with scores (up to a factor σ^{-4}) $J_{q_{kk}}(Z_t, \dots, Z_{t-p}) := (Z_t^2 - \sigma^2) Z_{t-k}^2$, is different. Conditions (J(i))–(J(iii)) are still satisfied, but $E[J_{q_{kk}}(\varepsilon_0, \dots, \varepsilon_p) | \varepsilon_0] = (\varepsilon_0^2 - \sigma^2) E[\varepsilon_k^2] = (\varepsilon_0^2 - \sigma^2) \sigma^2 \neq 0$ implies that $J_{q_{kk}}^*(\varepsilon_0, \dots, \varepsilon_p) = J_{q_{kk}}(\varepsilon_0, \dots, \varepsilon_p) - (\varepsilon_0^2 - \sigma^2) \sigma^2$, so that

$$\begin{aligned} & (n - k)^{-1/2} \sigma^{-4} \sum_{t=k+1}^n \{ (Z_t^2 - \sigma^2) Z_{t-k}^2 - (Z_t^2 - \sigma^2) \sigma^2 \} \\ &= (n - k)^{-1/2} \sigma^{-4} \sum_{t=k+1}^n (Z_t^2 - \sigma^2) (Z_{t-k}^2 - \sigma^2) \end{aligned}$$

has to be substituted for $(n - k)^{1/2} [(m_4)_{f_1} - 1]^{1/2} q_{kk}^{(n)}$ in order to obtain the efficient central sequence.

In order to complete the proof, recall that $r_k^{(n)} - \hat{r}_k^{(n)}$ and $q_{k\ell}^{(n)} - \hat{q}_{k\ell}^{(n)}$ are $o_P(n^{-1/2})$; so is $q_{kk*}^{(n)} - \hat{q}_{kk*}^{(n)}$ since, under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$,

$$\begin{aligned} & \sum_{t=k+1}^n (Z_t^2 - \sigma^2) (Z_{t-k}^2 - \sigma^2) - \sum_{t=k+1}^n (Z_t^2 - \hat{\sigma}^2) (Z_{t-k}^2 - \hat{\sigma}^2) \\ &= -(\hat{\sigma}^2 - \sigma^2) \sum_{t=k+1}^n (Z_t^2 + Z_{t-k}^2) + n(\hat{\sigma}^4 - \sigma^4) \end{aligned}$$

$$\begin{aligned}
 &= n(2\hat{\sigma}^2\sigma^2 - \hat{\sigma}^4 - \sigma^4) + O_P(1) \\
 &= n(2(\hat{\sigma}^2 - \sigma^2)\sigma^2 - (\hat{\sigma}^4 - \sigma^4)) + O_P(1) \\
 &= n(\hat{\sigma}^2 - \sigma^2)(\sigma^2 - \hat{\sigma}^2) + O_P(1) \\
 &= -n(\hat{\sigma}^2 - \sigma^2)^2 + O_P(1) = O_P(1) = o_P(n^{1/2}).
 \end{aligned}$$

The asymptotic normality of $\Delta_*^{(n)}(\mathbf{a})$ under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$ follows along the same lines as that of $\Delta_f^{(n)}(\mathbf{a})$ under $\mathcal{H}_f^{(n)}(\mathbf{a}; \mathbf{0})$. In the derivation of the off-diagonal blocks of the covariance matrix, the $(m_3)_{g_1}^2$ factor comes from the fact that

$$\begin{aligned}
 &E_{g_1}[(\Delta_I^{(n)}(\mathbf{a}))_i(\Delta_{*II}^{(n)}(\mathbf{a}))_j] \\
 &= \frac{1}{2} \sum_{k=i}^{n-1} h_{k-i} h_{k-j}^2 (n-k)^{-1} \sum_{t=k+1}^n E_{g_1}[Z_t^3] E_{g_1}[Z_{t-k}^3] \\
 &= \frac{1}{2} (m_3)_{g_1}^2 \sum_{k=i}^{n-1} h_{k-i} h_{k-j}^2. \quad \square
 \end{aligned}$$

3.2. *Local asymptotic linearity.* The nonrandom part \mathbf{a} of the autoregression coefficient throughout plays the role of a nuisance, and thus has to be estimated under the null hypothesis $\mathcal{H}^{(n)}$. In order to control for the effect of substituting an estimated value $\hat{\mathbf{a}}^{(n)}$ for the exact one in the efficient central sequence $\Delta_*^{(n)}(\mathbf{a})$, we need the following asymptotic linearity result.

PROPOSITION 3.2. *Assume that (A1) and (A3) hold, and that the density g satisfies (A2). Then, for all sequences $\boldsymbol{\tau}^{(n)} := (\boldsymbol{\alpha}^{(n)'} \boldsymbol{\delta}^{(n)'})'$ satisfying (2.5):*

(i)

$$\Delta_*^{(n)}(\mathbf{a}) - \begin{pmatrix} \boldsymbol{\Gamma}_X(\mathbf{a})\boldsymbol{\alpha}^{(n)} \\ \frac{1}{2}\boldsymbol{\Gamma}_{*g;X^2}(\mathbf{a})\boldsymbol{\delta}^{(n)} \end{pmatrix}$$

is asymptotically multinormal as $n \rightarrow \infty$, under $\mathcal{H}_g^{(n)}(\mathbf{a} + n^{-1/2}\boldsymbol{\alpha}^{(n)}; n^{-1/2}\boldsymbol{\delta}^{(n)})$, with mean $\mathbf{0}$ and covariance matrix $\mathbf{W}_{*g}(\mathbf{a})$;

(ii)

$$\Delta_*^{(n)}(\mathbf{a} + n^{-1/2}\boldsymbol{\alpha}^{(n)}) - \Delta_*^{(n)}(\mathbf{a}) = - \begin{pmatrix} \boldsymbol{\Gamma}_X(\mathbf{a})\boldsymbol{\alpha}^{(n)} \\ \mathbf{0}_{p \times 1} \end{pmatrix} + o_P(1)$$

as $n \rightarrow \infty$, under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$, hence also under $\mathcal{H}_g^{(n)}(\mathbf{a} + n^{-1/2}\boldsymbol{\alpha}^{(n)}; n^{-1/2}\boldsymbol{\delta}^{(n)})$.

PROOF. Part (i) of the proposition follows as an application of Le Cam’s *Third Lemma*. Consider indeed the arbitrary linear combination (with $\beta \in \mathbb{R}$, $\boldsymbol{\gamma}_I \in \mathbb{R}^p$, $\boldsymbol{\gamma}_{II} \in \mathbb{R}^p$)

$$\begin{aligned} \eta^{(n)} &:= \beta \Lambda_g^{(n)}(\mathbf{a}; \boldsymbol{\alpha}^{(n)}, \boldsymbol{\delta}^{(n)}) + (\boldsymbol{\gamma}'_I, \boldsymbol{\gamma}'_{II}) \boldsymbol{\Delta}_*^{(n)}(\mathbf{a}) \\ &= n^{-1/2} \sum_{t=1}^n \sum_{i=1}^p \{ [\beta \alpha_i \phi_g(Z_t) + \gamma_{Ii} Z_t] X_{t-i} \\ &\quad + \frac{1}{2} [\beta \delta_i \psi_g(Z_t) + \gamma_{IIi} (Z_t^2 - 1)] X_{t-i}^2 \} \\ &\quad - (\beta/2) (\boldsymbol{\tau}^{(n)'} \boldsymbol{\Gamma}_g(\mathbf{a}) \boldsymbol{\tau}^{(n)}) + o_P(1). \end{aligned}$$

The first term is a sum of p -dependent variables with finite variance. From classical central limit results, $\eta^{(n)}$ is asymptotically normal, with mean $-\frac{1}{2} d_g^2 := -\frac{\beta}{2} (\boldsymbol{\tau}^{(n)'} \boldsymbol{\Gamma}_g(\mathbf{a}) \boldsymbol{\tau}^{(n)})$ and asymptotic variance

$$(\beta, \boldsymbol{\gamma}'_I, \boldsymbol{\gamma}'_{II}) \begin{pmatrix} d_g^2 & (\boldsymbol{\alpha}^{(n)'} \boldsymbol{\Gamma}_X(\mathbf{a}), \frac{1}{2} \boldsymbol{\delta}^{(n)'} \boldsymbol{\Gamma}_{*g; X^2}(\mathbf{a})) \\ \left(\begin{matrix} \boldsymbol{\Gamma}_X(\mathbf{a}) \boldsymbol{\alpha}^{(n)} \\ \frac{1}{2} \boldsymbol{\Gamma}_{*g; X^2}(\mathbf{a}) \boldsymbol{\delta}^{(n)} \end{matrix} \right) & \mathbf{W}_{*g}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} \beta \\ \boldsymbol{\gamma}_I \\ \boldsymbol{\gamma}_{II} \end{pmatrix}.$$

The result follows from the usual Cramér–Wold argument and Le Cam’s Third Lemma. Due to the extremely simple form of the score functions J^* entering the Gaussian efficient central sequence $\boldsymbol{\Delta}_*^{(n)}$, the asymptotic linearity property in part (ii) of the proposition follows from straightforward direct computation. \square

Next, consider an estimate $\hat{\mathbf{a}}^{(n)}$ of \mathbf{a} enjoying the following root- n consistency and asymptotic discreteness properties.

- (A5) (i) For all g satisfying (A2), all \mathbf{a} satisfying (A1) and all $\varepsilon > 0$, there exist a positive constant $B = B(g, \mathbf{a}, \varepsilon) \in \mathbb{R}_+$ and an integer $N = N(g, \mathbf{a}, \varepsilon) \in \mathbb{N}$ such that, under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$, $P[n^{1/2} \|(\hat{\mathbf{a}}^{(n)} - \mathbf{a})\| > B] < \varepsilon$ for all $n \geq N$ ($\|\cdot\|$ denotes the Euclidean norm).
- (ii) For all \mathbf{a} satisfying (A1) and all $c > 0$, there exists a positive number $K = K(\mathbf{a}, c)$ such that the number of possible values of $\hat{\mathbf{a}}^{(n)}$ in balls of the form $\{\mathbf{t} \in \mathbb{R}^p : \|\mathbf{t} - \mathbf{a}\| \leq c\}$ is bounded by K as $n \rightarrow \infty$.

Part (i) of assumption (A5) is satisfied by all classical estimates of \mathbf{a} under the null hypothesis of AR dependence (Yule–Walker, exact or approximate MLEs, robust estimates, adaptive estimates, R -estimates, . . .), but also by the estimates of Schick (1996), which take into account the possible RCAR nature of the model. Part (ii) is the traditional requirement of asymptotic discreteness. It can be obtained through classical discretization methods, but has little practical implications if any. The following proposition is an immediate corollary of Proposition 3.2.

PROPOSITION 3.3. *Let $\hat{\mathbf{a}}^{(n)}$ satisfy Assumption (A5). Then, under the same conditions as in Proposition 3.2, $\Delta_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)}) - \Delta_{*II}^{(n)}(\mathbf{a}) = o_P(1)$ as $n \rightarrow \infty$, under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$, hence also under $\mathcal{H}_g^{(n)}(\mathbf{a} + n^{-1/2}\boldsymbol{\alpha}^{(n)}; n^{-1/2}\boldsymbol{\delta}^{(n)})$.*

The proposition is restricted to the asymptotic behavior of $\Delta_{*II}^{(n)}$, which is needed in the sequel; of course, it also holds that $\Delta_{*I}^{(n)}(\hat{\mathbf{a}}^{(n)}) - \Delta_{*I}^{(n)}(\mathbf{a}) = -n^{1/2} \times \Gamma_{\mathbf{X}}(\mathbf{a})(\hat{\mathbf{a}}^{(n)} - \mathbf{a}) + o_P(1)$.

4. Optimal pseudo-Gaussian detection methods.

4.1. *Locally asymptotically most stringent somewhere efficient tests.* The problem of testing linear restrictions on the mean of a multivariate Gaussian location model with known covariance structure against one-sided multivariate alternatives has a long story; see Bartholomew (1961), Chacko (1963), Kudô (1963), Nüesch (1966), Shorack (1967), Perlman (1969) and many others, who all consider variants of Gaussian likelihood-ratio tests. A different approach is developed in Schaafsma and Smid (1966), where the concept of *most stringent somewhere most powerful tests*, extending Abelson and Tukey’s (1963) idea of a *maximin contrast*, is introduced.

Though it has been shown that neither of the two methods uniformly dominates the other, the advantage over likelihood-ratio tests of the Schaafsma and Smid approach is that it relies on a clear finite-sample optimality concept, which makes it particularly well suited in the LAN context. It readily extends, via the limit Gaussian shift experiments, into a local asymptotic version. On the contrary, the justification of likelihood-ratio tests lies in their asymptotic properties, which make little sense when applied to a limit Gaussian shift experiment in which the observation is unique in essence. Moreover, numerous anomalies of Gaussian likelihood-ratio methods have been reported in the context of location problems with known covariance structure: see Gutmann (1987), Berger (1989), Menéndez and Salvador (1991), Menéndez, Rueda and Salvador (1992), among several others. We thus propose to develop an asymptotic procedure along the lines of the Schaafsma and Smid approach, which we now describe.

Denote by $\boldsymbol{\xi}$ a q -dimensional observation described by the full-rank Gaussian location model $(\mathbb{R}^q, \mathcal{B}^q, \mathcal{P} = \{\mathcal{N}(\boldsymbol{\mu}, \mathbf{W}) \mid \boldsymbol{\mu} \in \mathbb{R}^q\})$. Consider the null hypothesis \mathcal{H} under which $\boldsymbol{\mu}$ lies in the $(q - p)$ -dimensional linear subspace \mathcal{M} of \mathbb{R}^q characterized by the $1 \leq p < q$ independent linear constraints $\mathbf{H}'\boldsymbol{\mu} = \mathbf{0}$, and the one-sided alternative \mathcal{K} under which the p components of $\mathbf{H}'\boldsymbol{\mu}$ are nonnegative, with at least one inequality strict. Thus, under \mathcal{K} , $\boldsymbol{\mu}$ belongs to the half-cone $\mathcal{C} := \{\mathbf{x} \in \mathbb{R}^q \mid \mathbf{H}'\mathbf{x} > \mathbf{0}\}$. Denote by \mathcal{M}^\perp the p -dimensional orthogonal complement of \mathcal{M} , where orthogonality is defined in the metric associated with the covariance structure \mathbf{W} [namely, $\mathbf{x} \perp \mathbf{y}$ iff $\mathbf{x}'\mathbf{W}^{-1}\mathbf{y} = 0$, so that $\mathcal{M}^\perp = \mathcal{M}(\mathbf{W}\mathbf{H})$, the linear subspace of \mathbb{R}^q spanned by the columns of $\mathbf{W}\mathbf{H}$]. For each $\ell \in \mathcal{C}$, define \mathcal{K}_ℓ

as the one-dimensional, one-sided sub-alternative under which $\mu \in \{\kappa \ell \mid \kappa > 0\}$; let $\ell^\perp = \mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}'\ell$, and $\mathcal{C}^\perp = \mathbf{W}\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}'\mathcal{C}$ denote the projections of ℓ and \mathcal{C} onto \mathcal{M}^\perp (still in the metric associated with \mathbf{W}). A uniformly most powerful size- α test for \mathcal{H} against \mathcal{K}_ℓ is $\phi^{\ell^\perp}(\xi)$, which rejects \mathcal{H} whenever the test statistic

$$(4.1) \quad T^{\ell^\perp}(\xi) := \frac{\xi' \mathbf{W}^{-1} \ell^\perp}{[\ell'^\perp \mathbf{W}^{-1} \ell^\perp]^{1/2}} = \frac{\xi' \mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1} \mathbf{H}' \ell}{[\ell' \mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1} \mathbf{H}' \ell]^{1/2}}$$

exceeds the $(1 - \alpha)$ quantile z_α of the standard normal distribution.

Write $\mathcal{T} := \{\phi^{\ell^\perp}(\xi) \mid \ell \in \mathcal{C}\} = \{\phi^\ell(\xi) \mid \ell \in \mathcal{C}^\perp\}$ for this class of *somewhere most powerful* tests: Schaafsma and Smid (1966) show that \mathcal{T} contains a *most stringent* element, that is, a test ϕ^{ℓ^*} whose maximum shortcoming, with respect to \mathcal{T} , over the alternative $\mu \in \mathcal{C}$, is minimal. The corresponding test statistic is T^{ℓ^*} , where $\ell^* \in \mathcal{C}^\perp$ is such that, after orthogonalizing the experiment by means of the linear transformation $\mathbf{W}^{-1/2}$, $\mathbf{W}^{-1/2}\ell^*$ lies along

- (i) the axis of the semicone of revolution circumscribing, in the q -dimensional linear space $\mathbf{W}^{-1/2}\mathcal{M}^\perp$ now equipped with the Euclidean norm, the polyhedral angle $\mathcal{C}' := \mathbf{W}^{-1/2}\mathcal{C}^\perp$, in case this axis itself belongs to \mathcal{C}' , or, in case it does not,
- (ii) the axis of the semicone of revolution circumscribing the sub-polyhedral angle \mathcal{C}'_j formed by a subset of j vertices of \mathcal{C}' such that (a) this axis belongs to \mathcal{C}'_j and, (b) the angle of this axis with any of the vertices of \mathcal{C}' which do not belong to \mathcal{C}'_j is no greater than the j equal angles of this axis with the vertices characterizing \mathcal{C}'_j .

Abelson and Tukey [(1963), Sections 8 and 17–20] establish the existence and uniqueness of this solution ℓ^* , and show (Sections 10–12) how to compute it.

The LAN structure (Proposition 2.1) of the model under study entails the weak convergence (in the Le Cam sense), as $n \rightarrow \infty$, of the sequences of local experiments, of the form

$$\mathcal{E}_f^{(n)}(\mathbf{a}) := \left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \{P_{f;\mathbf{a}+n^{-1/2}\alpha, n^{-1/2}\delta}^{(n)} \mid \alpha \in \mathbb{R}^p, \delta \in \mathbb{R}_+^p\} \right)$$

to the Gaussian experiment

$$\mathcal{E}_f(\mathbf{a}) := \left(\mathbb{R}^{2p}, \mathcal{B}^{2p}, \mathcal{P} = \left\{ \mathcal{N} \left(\mu = \Gamma_f(\mathbf{a}) \begin{pmatrix} \alpha \\ \delta \end{pmatrix}, \Gamma_f(\mathbf{a}) \right) \mid \alpha \in \mathbb{R}^p, \delta \in \mathbb{R}_+^p \right\} \right).$$

Recall that this convergence entails convergence of the risk functions associated with $\mathcal{E}_f^{(n)}(\mathbf{a})$ to the corresponding ones in $\mathcal{E}_f(\mathbf{a})$ —see Le Cam (1986) or Le Cam and Yang (1990) for details. Denote by $\mathbf{\Delta}$ an observation described by the limit experiment $\mathcal{E}_f(\mathbf{a})$. If $\phi^*(\mathbf{\Delta})$ is a most stringent somewhere most powerful size- α test in the experiment $\mathcal{E}_f(\mathbf{a})$, for some linear space \mathcal{M} and some half-cone \mathcal{C} of \mathbb{R}^{2p} , then the sequences of tests $\phi^*(\mathbf{\Delta}_f^{(n)}(\mathbf{a}))$ inherit, in the original LAN

model, and under local and asymptotic form, the same optimality properties as $\phi^*(\mathbf{\Delta})$ in $\mathcal{E}_f(\mathbf{a})$. This can be taken as the definition of a *locally asymptotically most stringent somewhere most powerful test*, or, when applied to an efficient sequence, a *locally asymptotically most stringent somewhere efficient test* (in short, LAMSSE) for the problem under study.

Applying the general method of construction just described in the efficient limit experiment $\mathcal{E}_g^*(\mathbf{a})$ associated with Gaussian innovation densities, with $q = 2p$, $\mathcal{M} = \mathcal{M}((\mathbf{I}_p, \mathbf{0})')$ [the linear subspace of \mathbb{R}^{2p} spanned by the columns of $(\mathbf{I}_p, \mathbf{0})'$], $\mathcal{M}^\perp = \mathcal{M}(\mathbf{W}(\mathbf{0}, \mathbf{I}_p)')$, with the Gaussian information matrix (3.4), yields $\mathcal{M}^\perp = \mathcal{M}((\mathbf{0}, \mathbf{I}_p)')$, $\mathcal{C} = \{(\mathbf{x}', \mathbf{y}')' \in \mathbb{R}^{2p} \mid \mathbf{\Gamma}_{*g; \mathbf{X}^2}^{-1}(\mathbf{a})\mathbf{y} \in \mathbb{R}_+^p\}$ and $\mathcal{C}^\perp = \{(\mathbf{0}', \mathbf{y}')' \in \mathbb{R}^{2p} \mid \mathbf{\Gamma}_{*g; \mathbf{X}^2}^{-1}(\mathbf{a})\mathbf{y} \in \mathbb{R}_+^p\}$. Writing $\mathbf{\Delta}_* = (\mathbf{\Delta}'_I, \mathbf{\Delta}'_{*II})'$ for $\mathbf{\Delta}_*^{(n)}(\mathbf{a})$ and $\boldsymbol{\ell}^* = (\boldsymbol{\ell}'_I, \boldsymbol{\ell}'_{II})'$ for $(\mathbf{0}', \boldsymbol{\ell}'_{II})'$, denote by

$$\begin{aligned}
 T_{\mathbf{\Gamma}_*^{(n)}; \mathbf{a}}^{\boldsymbol{\ell}^*}(\mathbf{\Delta}_*) &= \frac{\mathbf{\Delta}'_* \mathbf{\Gamma}_*^{-1}(\mathbf{a}) \boldsymbol{\ell}^*}{[\boldsymbol{\ell}'_* \mathbf{\Gamma}_*^{-1}(\mathbf{a}) \boldsymbol{\ell}^*]^{1/2}} \\
 (4.2) \qquad &= \frac{\sqrt{2} \mathbf{\Delta}'_{*II} (\mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}))^{-1} \boldsymbol{\ell}^*_{II}}{[\boldsymbol{\ell}'_{II} (\mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}))^{-1} \boldsymbol{\ell}^*_{II}]^{1/2}} := T_{\mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}); \mathbf{a}}^{\boldsymbol{\ell}^*}(\mathbf{\Delta}_{*II})
 \end{aligned}$$

the resulting test statistic: the corresponding tests are LAMSSE for the problem of detecting randomness in the coefficients of the autoregressive model (2.1) if the innovation density is assumed to be Gaussian under the alternative. Since $\boldsymbol{\ell}^*_{II} = \boldsymbol{\ell}^*_{II}(\mathbf{a}, \mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}))$, they are $\mathbf{\Delta}_{*II}^{(n)}$ -measurable and depend on $\mathbf{\Gamma}_*(\mathbf{a})$ only through $\mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a})$.

These tests however remain unsatisfactory on two accounts: they explicitly involve the parameter \mathbf{a} which is unspecified under the hypothesis $\mathcal{H}^{(n)}$ to be tested; and, their asymptotic standard normal distribution relies on a Gaussian assumption which should be dropped.

In order to palliate these two problems, we first show that the unspecified \mathbf{a} can be replaced by any estimate $\hat{\mathbf{a}}^{(n)}$ of \mathbf{a} satisfying (A7). The matrix $\mathbf{\Gamma}_{*g; \mathbf{X}^2}$ and the cone \mathcal{C} indeed are continuous functions of \mathbf{a} . Hence, letting $\hat{\mathbf{\Gamma}}_{*g} := \mathbf{\Gamma}_{*g; \mathbf{X}^2}(\hat{\mathbf{a}}^{(n)})$ and $\hat{\boldsymbol{\ell}}^* := \boldsymbol{\ell}^*_{II}(\hat{\mathbf{a}}^{(n)}, \hat{\mathbf{\Gamma}}_{*g})$, we have $T_{\hat{\mathbf{\Gamma}}_{*g}; \hat{\mathbf{a}}^{(n)}}^{\hat{\boldsymbol{\ell}}^*}(\mathbf{\Delta}_{*II}^{(n)}(\mathbf{a})) = T_{\mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}); \mathbf{a}}^{\boldsymbol{\ell}^*}(\mathbf{\Delta}_{*II}^{(n)}(\mathbf{a})) + o_P(1)$, under $\mathcal{H}^{(n)}$. Now, in view of Proposition 3.3, the effect of substituting $\hat{\mathbf{a}}^{(n)}$ for \mathbf{a} in $\mathbf{\Delta}_*^{(n)}(\mathbf{a})$ [namely, the difference $\mathbf{\Delta}_*^{(n)}(\hat{\mathbf{a}}^{(n)}) - \mathbf{\Delta}_*^{(n)}(\mathbf{a})$] asymptotically belongs to \mathcal{M} , and hence is orthogonal to $\boldsymbol{\ell}^* \in \mathcal{M}^\perp$. It follows that the difference $T_{\hat{\mathbf{\Gamma}}_{*g}; \hat{\mathbf{a}}^{(n)}}^{\hat{\boldsymbol{\ell}}^*}(\mathbf{\Delta}_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)})) - T_{\mathbf{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}); \mathbf{a}}^{\boldsymbol{\ell}^*}(\mathbf{\Delta}_{*II}^{(n)}(\mathbf{a}))$ is $o_P(1)$ under $\mathcal{H}^{(n)}$ and under contiguous alternatives, as $n \rightarrow \infty$.

Second, the fact that $T_{\hat{\mathbf{\Gamma}}_{*g}; \hat{\mathbf{a}}^{(n)}}^{\hat{\boldsymbol{\ell}}^*}(\mathbf{\Delta}_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)}))$ is asymptotically standard normal under Gaussian densities only is due to the Gaussian form of the standardizing covariance matrix $\hat{\mathbf{\Gamma}}_{*g; \hat{\mathbf{a}}^{(n)}}$. In order to get rid of this Gaussian assumption, note that, under density g , the asymptotic covariance matrix (3.3) of $\mathbf{\Delta}_{*II}^{(n)}(\mathbf{a})$ [namely,

$\Gamma_{*g; \mathbf{X}^2}(\mathbf{a})$ which, under Gaussian g , coincides with $\frac{1}{2}\Gamma_{*g; \mathbf{X}^2}(\mathbf{a})]$ depends on g through $[(m_4)_{g_1} - 1]$ only. This quantity is estimated consistently by $(\hat{m}_4^{(n)} - 1)$. Denote by $\hat{\Gamma}(\mathbf{a})$ the matrix resulting from plugging this estimate into $\Gamma_{*g; \mathbf{X}^2}(\mathbf{a})$: the sequence $\hat{\Gamma}(\hat{\mathbf{a}}^{(n)})$ is asymptotically equivalent to $\frac{1}{4}[(m_4)_{g_1} - 1]\Gamma_{*g; \mathbf{X}^2}(\mathbf{a})$ under $\mathcal{H}_g^{(n)}(\mathbf{a}; \mathbf{0})$, for any density g satisfying Assumption (A2). Part (i) of Proposition 3.1 then entails that the statistic

$$(4.3) \quad T^* := T_{\hat{\Gamma}(\hat{\mathbf{a}}^{(n)}); \hat{\mathbf{a}}^{(n)}}^{\hat{\boldsymbol{\ell}}^*}(\boldsymbol{\Delta}_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)})),$$

in the definition of which the matrix $\hat{\Gamma}(\hat{\mathbf{a}}^{(n)})$ is used throughout (this includes the construction of $\hat{\boldsymbol{\ell}}^*$), is asymptotically standard normal under any $\mathcal{H}_g^{(n)}$ such that g satisfies Assumption (A2), and asymptotically equivalent, under Gaussian g , to the Gaussian LAMSSE test statistic (4.2). Summing up, we have established the following proposition [the local power, in part (iv), readily follows from the distribution, under local alternatives, of $\boldsymbol{\Delta}_*^{(n)}$; see Propositions 3.1 and 3.2].

PROPOSITION 4.1. *The test rejecting $\mathcal{H}^{(n)}$ whenever the statistic T^* given in (4.3) exceeds the $(1 - \alpha)$ quantile of the standard normal distribution:*

- (i) *has asymptotic size α ;*
- (ii) *is locally and asymptotically most stringent somewhere efficient at the Gaussian experiment within the class of tests having asymptotic size α , at the Gaussian experiment;*
- (iii) *in the one-dimensional AR(1) case ($p = 1$), is locally asymptotically most powerful, at asymptotic size α , against Gaussian alternatives [of the form $\bigcup_{-1 < a < 1} \bigcup_{\sigma > 0} \mathcal{H}_f^{(n)}(a; \sigma^2)$, where f is Gaussian];*
- (iv) *has asymptotic power*

$$(4.4) \quad 1 - \Phi\left(z_\alpha - \boldsymbol{\delta}' \boldsymbol{\ell}_g^* \left[[(m_4)_{g_1} - 1] \boldsymbol{\ell}_g^{*'} \boldsymbol{\Gamma}_{*g; \mathbf{X}^2}^{-1}(\mathbf{a}) \boldsymbol{\ell}_g^* \right]^{-1/2}\right),$$

where $\boldsymbol{\ell}_g^* := \boldsymbol{\ell}_{II}^*(\mathbf{a}, \boldsymbol{\Gamma}_{*g; \mathbf{X}^2}(\mathbf{a}))$ is obtained in the same way as $\boldsymbol{\ell}_{II}^*(\mathbf{a}, \boldsymbol{\Gamma}_{g; \mathbf{X}^2}(\mathbf{a}))$ in (4.2), but with $\frac{1}{4}[(m_4)_{g_1} - 1]\boldsymbol{\Gamma}_{g; \mathbf{X}^2}(\mathbf{a})$ substituted for $\frac{1}{2}\boldsymbol{\Gamma}_{g; \mathbf{X}^2}$, against alternatives of the form $\mathcal{H}_g^{(n)}(\mathbf{a}; n^{-1/2}\boldsymbol{\delta})$, for which (A3) holds and g satisfies (A2).

4.2. *A simple heuristic test.* Obviously, the main difficulty with the method just described lies in the derivation of the *most stringent direction* $\boldsymbol{\ell}^*$. This derivation is relatively simple for small values of p , but its complexity is likely to increase with p . A simpler method, based on a heuristic choice of a direction $\boldsymbol{\ell}$ then may be preferable. Basically, instead of selecting the axis $\boldsymbol{\ell}^*$ of a cone of revolution in the metric determined by the (estimated) efficient information matrix, one may choose $\boldsymbol{\ell}_1 = (0, \dots, 0, 1, \dots, 1)'$, which is the axis of the cone of revolution circumscribing the positive orthant in the parameter space equipped with the

Euclidean norm—with intuitive justification that such direction corresponds to alternatives under which all components of the vector σ^2 are given equal weights. A locally asymptotically most powerful test then can be derived, for the one-dimensional alternative associated with ℓ_1 , along the same lines as in the previous section. Denote by ℓ_1^\perp the projection of ℓ_1 onto $\mathcal{M}^\perp = \mathcal{M}((\mathbf{0}, \mathbf{I}_p)')$. After due substitution of $\hat{\mathbf{a}}^{(n)}$ and $\hat{m}_4^{(n)}$ for \mathbf{a} and $(m_4)_{g_1}$, respectively, the resulting test statistic takes the very simple form

$$(4.5) \quad T^{\ell_1}(\Delta_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)})) = \frac{2(\Delta_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)}))' \hat{\Gamma}_{\mathbf{X}^2}^{-1}(\hat{\mathbf{a}}^{(n)}) \mathbf{1}_p}{[[\hat{m}_4^{(n)} - 1] \mathbf{1}'_p \hat{\Gamma}_{\mathbf{X}^2}^{-1}(\hat{\mathbf{a}}^{(n)}) \mathbf{1}_p]^{1/2}},$$

still to be compared with the $(1 - \alpha)$ standard normal quantile z_α . We thus have the following results (local power readily follows from Propositions 3.1 and 3.2).

PROPOSITION 4.2. *The test rejecting $\mathcal{H}^{(n)}$ whenever the statistic (4.5) exceeds the $(1 - \alpha)$ quantile z_α of the standard normal distribution:*

- (i) *has asymptotic size α ;*
- (ii) *has asymptotic power*

$$(4.6) \quad 1 - \Phi\left(z_\alpha - \delta' \mathbf{1}_p \left[[(m_4)_{g_1} - 1] \mathbf{1}'_p \Gamma_{*g; \mathbf{X}^2}^{-1}(\mathbf{a}) \mathbf{1}_p\right]^{-1/2}\right)$$

against alternatives of the form $\mathcal{H}_g^{(n)}(\mathbf{a}; n^{-1/2} \delta)$, where (A3) holds and g satisfies (A2).

PROOF. The result follows from Proposition 3.1 and, after some algebra, from a routine application of Le Cam’s Third Lemma. \square

The main advantage of (4.5) over (4.2) is that it can be expressed in closed form, and is easily computed. It would be interesting to compare the local powers (4.4) and (4.6), but, unfortunately, ℓ_g^* being a complicated function of \mathbf{a} , this comparison in general is not possible. Note however that, in the first-order case ($p = 1$), the two tests coincide.

4.3. An order-identification/detection method. Instead of a formal hypothesis testing approach, one may prefer a more flexible procedure, based on the inspection of a series of statistics, one for each component of the parameter σ^2 under study. Denote by T_j , $j = 1, \dots, p$, the asymptotically most powerful test statistics of the form (4.1) associated with the unit vectors $\mathbf{e}_j = (e_{jk})$, with $e_{jk} := \delta_j, p+k, k = 1, \dots, p$:

$$(4.7) \quad T_j(\Delta_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)})) = \frac{2(\Delta_{*II}^{(n)}(\hat{\mathbf{a}}^{(n)}))' \hat{\Gamma}_{\mathbf{X}^2}^{-1}(\hat{\mathbf{a}}^{(n)}) \mathbf{e}_j}{[[\hat{m}_4^{(n)} - 1] \mathbf{e}'_j \hat{\Gamma}_{\mathbf{X}^2}^{-1}(\hat{\mathbf{a}}^{(n)}) \mathbf{e}_j]^{1/2}}.$$

An inspection of the p values T_j , $j = 1, \dots, p$, and their comparison with the standard normal quantile z_α allows for the identification of the lags at which random effects may be present, whereas the tests described in the previous sections only can detect the global presence of such effects. Such inspection of course should be performed in the same spirit as correlogram inspection, and the probability level α cannot be interpreted as a global type I risk any more.

APPENDIX:
A NONSTANDARD Q.M. DIFFERENTIABILITY PROPERTY

In this section, we establish the quadratic mean (q.m.) differentiability property on which the proof of Proposition 2.1 relies. This q.m. differentiability property is somewhat nonstandard, as it involves the second-order derivatives \ddot{f} of the density f .

Letting

$$h_{z;\mathbf{x}}(\mathbf{t}^a, \mathbf{t}^\delta) := \int f \left(z - \sum_{i=1}^p t_i^a x_i - \sum_{i=1}^p t_i^\delta v_i x_i \right) g_{\mathbf{u}}(\mathbf{v}) d\mathbf{v}, \quad \mathbf{x}, \mathbf{t}^a \in \mathbb{R}^p, \mathbf{t}^\delta \in \mathbb{R}_+^p,$$

the log-likelihood (2.6) takes the form

$$\sum_{t=1}^n [\log h_{Z_t; \mathbf{X}_t}(\mathbf{t}^a, \mathbf{t}^\delta) - \log h_{Z_t; \mathbf{X}_t}(\mathbf{0}, \mathbf{0})] + o_P(1),$$

with $\mathbf{X}_t = (X_{t-1}, \dots, X_{t-p})'$, $\mathbf{t}^a = n^{-1/2} \boldsymbol{\alpha}^{(n)}$ and $\mathbf{t}^\delta = n^{-1/4} (\sqrt{\delta_1^{(n)}}, \dots, \sqrt{\delta_p^{(n)}})'$.

LEMMA A.0.1. *The function $(t_1^a, \dots, t_p^a, (t_1^\delta)^2, \dots, (t_p^\delta)^2) \mapsto [h_{z;\mathbf{x}}(\mathbf{t}^a, \mathbf{t}^\delta)]^{1/2}$ is differentiable in quadratic mean at $\mathbf{t}^a = \mathbf{0} = \mathbf{t}^\delta$ (“from the right,” as far as \mathbf{t}^δ is concerned), with q.m. gradient*

$$-\frac{1}{2} \left(\frac{\dot{f}(z)}{f^{1/2}(z)} x_1, \dots, \frac{\dot{f}(z)}{f^{1/2}(z)} x_p, -\frac{1}{2} \frac{\ddot{f}(z)}{f^{1/2}(z)} x_1^2, \dots, -\frac{1}{2} \frac{\ddot{f}(z)}{f^{1/2}(z)} x_p^2 \right)',$$

that is,

$$(A.1) \quad \lim_{(\mathbf{t}^a, \mathbf{t}^\delta) \rightarrow \mathbf{0}} \left[\sum_{i=1}^p (t_i^a)^2 + \sum_{i=1}^p (t_i^\delta)^4 \right]^{-1} \times \int \left\{ h_{z;\mathbf{x}}^{1/2}(\mathbf{t}^a, \mathbf{t}^\delta) - h_{z;\mathbf{x}}^{1/2}(\mathbf{0}, \mathbf{0}) + \frac{1}{2} \frac{\dot{f}(z)}{f^{1/2}(z)} \sum_{i=1}^p t_i^a x_i - \frac{1}{4} \frac{\ddot{f}(z)}{f^{1/2}(z)} \sum_{i=1}^p (t_i^\delta)^2 x_i^2 \right\}^2 dz = 0.$$

PROOF. In view of Lemma 2.1 in Garel and Hallin (1995), it is sufficient to establish partial q.m. differentiability (at $\mathbf{t}^a = \mathbf{0} = \mathbf{t}^\delta$) of $h^{1/2}$ with respect to each component of \mathbf{t}^a and $((t_1^\delta)^2, \dots, (t_p^\delta)^2)'$. The problem for \mathbf{t}^a reduces (at $\mathbf{t}^\delta = \mathbf{0}$, of course) to the classical AR(p) case, treated by Swensen (1985) and Kreiss (1987). We thus concentrate on q.m. differentiability with respect to $(t_i^\delta)^2$. Without loss of generality, we assume $p = 1$, writing $t, x, h_{z;x}(t)$, and v instead of $\mathbf{t}, \mathbf{x}, h_{z;\mathbf{x}}(\mathbf{t}, \mathbf{0})$, and \mathbf{v} , and g for $g_{\mathbf{u}}$. The proof is inspired by Hájek (1972) [Assumption (A2) of a nowhere vanishing density allows for some simplification], and decomposes into three parts.

(i) With the above notation, $t^2 \mapsto h_{z;x}(t) = \int_{v=-\infty}^{\infty} f(z - xtv)g(v) dv$ is absolutely continuous in a right-neighborhood of $t = 0$, with a.e. derivative

$$(A.2) \quad \Upsilon_{z;x}(t) := \frac{1}{2t} \int_{w=0}^t \int_{v=-\infty}^{\infty} \ddot{f}(z - xwv)x^2v^2g(v) dv dw.$$

Indeed, from the absolute continuity of f and \dot{f} and Fubini's theorem, we obtain

$$(A.3) \quad \begin{aligned} h(t) - h(0) &= \int_{v=-\infty}^{\infty} [f(z - xtv) - f(z)]g(v) dv \\ &= - \int_{v=-\infty}^{\infty} \int_{a=0}^t \dot{f}(z - xav)xv da g(v) dv \\ &= \int_{v=-\infty}^{\infty} \int_{a=0}^t \int_{w=0}^a \ddot{f}(z - xwv)x^2v^2 dw da g(v) dv \\ &= \frac{1}{2} \int_{b=0}^{t^2} b^{-1/2} \int_{w=0}^{b^{1/2}} \int_{v=-\infty}^{\infty} \ddot{f}(z - xwv)x^2v^2g(v) dv dw db. \end{aligned}$$

The value (A.2) of the a.e. derivative follows for each $t > 0$. At $t = 0$, the right derivative is defined as the limit, as $t \downarrow 0$, of $[h(t) - h(0)]/t^2$, for which (A.3) yields $\frac{0}{0}$. L'Hospital's rule applies, however, leading to

$$(A.4) \quad \begin{aligned} &\lim_{t \downarrow 0} [h(t) - h(0)]/t^2 \\ &= \left[\lim_{t \downarrow 0} \Upsilon_{z;x}(t) / \lim_{t \downarrow 0} 1 \right] \\ &= \frac{1}{2} \ddot{f}(z)x^2 \int_{v=-\infty}^{\infty} v^2g(v) dv = \frac{1}{2} \ddot{f}(z)x^2 \end{aligned}$$

since, from Assumption (A2), $z \mapsto \ddot{f}(z)$ is right-continuous. Thus, for all z , this right derivative at $t = 0$ exists, and is equal to $\frac{1}{2} \ddot{f}(z)x^2$.

(ii) It follows that $t^2 \mapsto s_{z;x}(t) := [h_{z;x}(t)]^{1/2}$ is also absolutely continuous in a neighborhood of $t = 0$, with a.e. derivative

$$(A.5) \quad \dot{s}_{z;x}(t) = \frac{1}{4t} \int_{w=0}^t \frac{\int_{v=-\infty}^{\infty} \ddot{f}(z - xwv)x^2v^2g(v) dv}{[\int_{v=-\infty}^{\infty} f(z - xtv)g(v) dv]^{1/2}} dw.$$

L'Hospital's rule (same reasoning as above) at $t = 0$ yields

$$\dot{s}_{z;x}(0) = \frac{1}{4} \frac{\ddot{f}(z)x^2}{f^{1/2}(z)}.$$

It follows that, for all z ,

$$(A.6) \quad \lim_{t \rightarrow 0} \left\{ \frac{1}{t^2} [s_{z;x}(t) - s_{z;x}(0)] \right\} = \dot{s}_{z;x}(0) = \frac{1}{4} \frac{\ddot{f}(z)x^2}{f^{1/2}(z)}.$$

(iii) The partial quadratic mean differentiability property to be proved takes the form

$$(A.7) \quad \lim_{t \rightarrow \infty} \int \left\{ \frac{1}{t^2} [s_{z;x}(t) - s_{z;x}(0)] - \frac{1}{4} \dot{s}_{z;x}(0) \right\}^2 dz = 0.$$

From (ii) above,

$$\begin{aligned} \left\{ \frac{1}{t^2} [s_{z;x}(t) - s_{z;x}(0)] \right\}^2 &= \left(\frac{1}{t^2} \right)^2 \left(\int_{\lambda=0}^{t^2} \dot{s}_{z;x}(\sqrt{\lambda}) d\lambda \right)^2 \\ &\leq \frac{1}{t^2} \int_{\lambda=0}^{t^2} [\dot{s}_{z;x}(\sqrt{\lambda})]^2 d\lambda, \end{aligned}$$

for all z ; hence, from Fubini's theorem and (A.5),

$$(A.8) \quad \begin{aligned} &\int_{z=-\infty}^{\infty} \left\{ \frac{1}{t^2} [s_{z;x}(t) - s_{z;x}(0)] \right\}^2 dz \\ &\leq \frac{1}{t^2} \int_{\lambda=0}^{t^2} \int_{z=-\infty}^{\infty} [\dot{s}_{z;x}(\sqrt{\lambda})]^2 dz d\lambda \\ &= \frac{1}{16t^2} \int_{\lambda=0}^{t^2} \mathcal{J}_{\psi}^x(f; \sqrt{\lambda}) d\lambda, \end{aligned}$$

with $\mathcal{J}_{\psi}^x(f; t)$ defined in (2.3). The continuity assumption in (A4) implies that this latter quantity converges, as $t \rightarrow 0$, to

$$\frac{1}{16} \mathcal{J}_{\psi}^x(f; 0) = \int_{z=-\infty}^{\infty} [\dot{s}_{z;x}(0)]^2 dz,$$

which, together with (A.8), entails that

$$(A.9) \quad \limsup_{t \rightarrow 0} \int_{z=-\infty}^{\infty} \left\{ \frac{1}{t^2} [s_{z;x}(t) - s_{z;x}(0)] \right\}^2 dz \leq \int_{z=-\infty}^{\infty} [\dot{s}_{z;x}(0)]^2 dz.$$

In view of Theorem V.I.3 of Hájek and Šidák (1967) [also in Hájek, Šidák and Sen (1999)], (A.6) and (A.9) jointly imply (A.7). This completes the proof. \square

Acknowledgment. The authors express their gratitude to two anonymous referees for a very careful reading and insightful discussion of the initial version of this paper. Their comments allowed for several improvements.

REFERENCES

- ABELSON, R. P. and TUKEY, J. W. (1963). Efficient utilization of non-numerical information in qualitative analysis: General theory and the case of simple order. *Ann. Math. Statist.* **34** 1347–1369.
- ANDĚL, J. (1976). Autoregressive series with random parameters. *Math. Oper. Statist.* **7** 735–741.
- BARTHOLOMEW, D. J. (1961). A test of homogeneity of means under restricted alternatives. *J. Roy. Statist. Soc. Ser. B* **23** 239–281.
- BENGHABRIT, Y. and HALLIN, M. (1996). Locally asymptotically optimal tests for autoregressive against bilinear serial dependence. *Statist. Sinica* **6** 147–169.
- BENGHABRIT, Y. and HALLIN, M. (1998). Locally asymptotically optimal tests for AR(p) against diagonal bilinear dependence. *J. Statist. Plann. Inference* **68** 47–63.
- BERGER, R. L. (1989). Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. *J. Amer. Statist. Assoc.* **84** 192–199.
- BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore, MD.
- CHACKO, V. J. (1963). Testing homogeneity against ordered alternatives. *Ann. Math. Statist.* **34** 945–956.
- CHOI, S., HALL, W. J. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861.
- CONLISK, J. (1976). A further note on stability in a random coefficient model. *Internat. Econom. Rev.* **17** 759–764.
- DROST, F. C., KLAASSEN, C. A. J. and WERKER, B. J. M. (1997). Adaptive estimation in time-series models. *Ann. Statist.* **25** 786–817.
- EHM, W., MAMMEN, E. and MÜLLER, D. W. (1995). Power robustification of approximately linear tests. *J. Amer. Statist. Assoc.* **90** 1025–1033.
- FEIGIN, P. D. and TWEEDIE, R. L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6** 1–14.
- GAREL, B. and HALLIN, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Statist. Math.* **47** 551–579.
- GRANGER, C. W. J. and ANDERSEN, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Göttingen.
- GUÉGAN, D. (1994). *Séries chronologiques non linéaires à temps discret*. Economica, Paris.
- GUTMANN, S. (1987). Tests uniformly more powerful than uniformly most powerful monotone tests. *J. Statist. Plann. Inference* **17** 279–292.
- GUYTON, D. A., ZHANG, N. F. and FOUTZ, R. V. (1986). A random parameter process for modeling and forecasting time series. *J. Time Ser. Anal.* **7** 105–115.
- HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.
- HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
- HÁJEK, J., ŠIDÁK, Z. and SEN, P. K. (1999). *Theory of Rank Tests*, 2nd ed. Academic Press, New York.
- HALLIN, M. and PURI, M. L. (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.* **50** 175–237.

- HALLIN, M. and WERKER, B. (1999). Optimal testing for semiparametric AR models: From Gaussian Lagrange multipliers to autoregression rank scores and adaptive tests. In *Asymptotics, Nonparametrics and Time Series* (S. Ghosh, ed.) 295–350. Dekker, New York.
- HALLIN, M. and WERKER, B. (2003). Semi-parametric efficiency, distribution-freeness, and invariance. *Bernoulli* **9** 137–165.
- HWANG, S. Y. and BASAWA, I. V. (1993a). Asymptotic optimal inference for a class of nonlinear time series models. *Stochastic Process. Appl.* **46** 91–113.
- HWANG, S. Y. and BASAWA, I. V. (1993b). Parameter estimation in a regression model with random coefficient autoregressive errors. *J. Statist. Plann. Inference* **36** 57–67.
- HWANG, S. Y., BASAWA, I. V. and REEVES, J. (1994). The asymptotic distribution of residual autocorrelations and related tests of fit for a class of nonlinear time series models. *Statist. Sinica* **4** 107–125.
- HWANG, S. Y. and BASAWA, I. V. (1998). Parameter estimation for generalized random coefficient autoregressive processes. *J. Statist. Plann. Inference* **68** 323–337.
- KOUL, H. L. and SCHICK, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.* **24** 1025–1052.
- KOUL, H. L. and SCHICK, A. (1997). Efficient estimation in nonlinear autoregressive time series models. *Bernoulli* **3** 247–277.
- KREISS, J. P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.
- KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50** 403–418.
- LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
- LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York.
- LEE, S. (1998). Coefficient constancy test in a random coefficient autoregressive model. *J. Statist. Plann. Inference* **74** 93–101.
- MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Dominance of likelihood ratio tests under cone constraints. *Ann. Statist.* **20** 2087–2099.
- MENÉNDEZ, J. A. and SALVADOR, B. (1991). Anomalies of the likelihood ratio test for testing restricted hypotheses. *Ann. Statist.* **19** 889–898.
- NICHOLLS, D. F. and QUINN, B. G. (1980). The estimation of random coefficient autoregressive models. I. *J. Time Ser. Anal.* **1** 37–46.
- NICHOLLS, D. F. and QUINN, B. G. (1981). The estimation of random coefficient autoregressive models. II. *J. Time Ser. Anal.* **2** 185–203.
- NICHOLLS, D. F. and QUINN, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction. Lecture Notes in Statist.* **11**. Springer, New York.
- NÜESCH, P. E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *Ann. Math. Statist.* **37** 113–119.
- PAGAN, A. R. (1980). Some identification and estimation results for regression models with stochastically varying coefficients. *J. Econometrics* **13** 341–363.
- PERLMAN, M. D. (1969). One-sided testing problems in multivariate analysis. *Ann. Math. Statist.* **40** 549–567. [Correction (1971) **42** 1777.]
- PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, London.
- RAMANATHAN, T. V. and RAJARSHI, M. B. (1994). Rank tests for testing the randomness of autoregressive coefficients. *Statist. Probab. Lett.* **21** 115–120.
- RIEDER, H. (2000). One-sided confidence about functionals over tangent cones. Dept. Mathematics, Univ. Bayreuth. Available at www.uni-bayreuth.de/departments/math/org/mathe7/rieder/publications.html.

- ROBINSON, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scand. J. Statist.* **5** 163–168.
- SCHAAFSMA, W. and SMID, L. J. (1966). Most stringent somewhere most powerful tests against alternatives restricted by a number of linear alternatives. *Ann. Math. Statist.* **37** 1161–1172.
- SCHICK, A. (1996). \sqrt{n} -consistent estimation in a random coefficient autoregressive model. *Austral. J. Statist.* **38** 155–160.
- SHORACK, G. R. (1967). Testing against ordered alternatives in model I analysis of variance: Normal theory and nonparametric. *Ann. Math. Statist.* **38** 1740–1752.
- SWENSEN, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16** 54–70.
- TANIGUCHI, M. and KAKIZAWA, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.
- TONG, H. (1990). *Nonlinear Time Series: A Dynamical Approach*. Oxford Univ. Press.
- VAN DER VAART, A. (1988). *Statistical Estimation in Large Parameter Spaces*. CWI, Amsterdam.
- VAN DER VAART, A. (1989). On the asymptotic information bound. *Ann. Statist.* **17** 1487–1500.
- VERVAAT, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. in Appl. Probab.* **11** 750–783.
- WEISS, A. A. (1985). The stability of the AR(1) process with an AR(1) coefficient. *J. Time Ser. Anal.* **6** 181–186.

DÉPARTEMENT DE MATHÉMATIQUES
FACULTÉ DES SCIENCES ET TECHNIQUES
TANGIERS
MOROCCO

I.S.R.O., E.C.A.R.E.S. AND
DÉPARTEMENT DE MATHÉMATIQUE
UNIVERSITÉ LIBRE DE BRUXELLES
BRUSSELS
BELGIUM
E-MAIL: mhallin@ulb.ac.be