Methodol Comput Appl Probab (2012) 14:793–810
DOI 10.1007/s11009-011-9270-7

Weak Convergence of Marked Empirical Processes for Focused Inference on AR(p) vs AR($p+1$) Stationary Time Series

Alejandra Cabaña · Enrique M. Cabaña · Marco Scavino

Received: 13 October 2010 / Revised: 26 July 2011 /
Accepted: 18 November 2011 / Published online: 7 December 2011

Abstract The technique applied by the authors to construct consistent and focused tests of fit for i.i.d. samples and regression models is extended to AR models for stationary time series. This approach leads to construct a consistent goodness-of-fit test for the null hypothesis that a stationary series is governed by an autoregressive model of a given order p. In addition of the consistency, the test is focused to detect efficiently the alternative of an AR($p+1$) model. The basic functional statistic conveying the information provided by the series is the process of accumulated sums of the residuals computed under the model of the null hypothesis of fit, reordered as concomitants of the conveniently delayed process. This process is transformed in order to obtain a new process with the same limiting Gaussian law encountered in earlier applications of the technique. Therefore, a Watson type quadratic statistic computed from this process has the same asymptotic laws under the null hypothesis of fit, and also under the alternatives of focusing, than the test statistics used in those applications. As a consequence, the resulting test has the same desirable performance as the tests previously developed by applying the same kind of transformations of processes.

A. Cabaña was partially supported by TIN2008-06582-C03-02 and MTM2009-10893, Ministerio de Ciencia y Tecnología, Spain.

E.M. Cabaña and M. Scavino were partially supported by CSIC-Udelar, Uruguay, Centre de Recerca Matemàtica, Barcelona, Spain and Carolina Foundation, Spain.

A. Cabaña (✉)
Departament de Matemàtiques, Universitat Autònoma de Barcelona, Barcelona, Spain
e-mail: acabana@mat.uab.cat

E. M. Cabaña · M. Scavino
Departamento de Métodos Matemático-Cuantitativos, Universidad de la República, Montevideo, Uruguay

Keywords Goodness-of-fit · Autoregressive processes · Marked processes · Transformations of processes in inference

AMS 2000 Subject Classifications 62G10 · 62M10

1 Introduction

Let $(X_i)_{-p\leq i\leq n}$ denote observations of a stationary causal AR model with integral parameter i.

We wish to test the null hypothesis

$\mathcal{H}_0$ *The observed values are distributed according to the stationary Gaussian model*

$$X_i = \phi_1 X_{i-1} + \cdots + \phi_p X_{i-p} + \sigma Z_i, \quad i = 1, 2, \ldots, n \tag{1}$$

of order p with i.i.d. innovations Z_i, $\mathbf{E} Z_i = 0$, $\mathbf{E} Z_i^2 = 1$.

against any alternative model, and also wish to focus the inference on the sequence of local alternatives

$\mathcal{H}_1$ *The model*

$$X_i = \phi_1 X_{i-1} + \cdots + \phi_p X_{i-p} + \frac{\delta}{\sqrt{n}} X_{i-p-1} + \sigma Z_i, \quad i = 1, 2, \ldots, n$$

holds for $(X_i)_{-p\leq i\leq n}$ with some $\delta \neq 0$ and the same assumptions on the innovations.

Under $\mathcal{H}_0$ the linear combination $\check{Z}_i = \frac{1}{\sigma}[X_i - \sum_{j=1}^{p} \phi_j X_{i-j}]$ is equal to Z_i and therefore is independent of $\{X_k : k < i\}$. On the other hand, under the alternative $\mathcal{H}_1$, it is equal to $\frac{\delta}{\sigma\sqrt{n}} X_{i-p-1} + Z_i$ and hence, correlated with X_{i-p-1}.

This implies that for $0 < x_1 < x_2$ the expectations under $\mathcal{H}_1$ of the random variables $\mathbf{1}_{\{x_1 < X_{i-p-1} < x_2\}} \check{Z}_i$ have the same sign as δ, while for $x_1 < x_2 < 0$ they have the opposite sign. Consequently the sum

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{1}_{\{X_{i-p-1} < x\}} \check{Z}_i \tag{2}$$

is centred for all x under $\mathcal{H}_0$, and is expected to be either increasing or decreasing for $x < 0$ and to exhibit the opposite behaviour for $x > 0$, depending of the sign of δ, under the alternative. This suggests to decide rejecting $\mathcal{H}_0$ or not according to the behaviour of Eq. 2.

Since the $\check{Z}_i$ are not observable, they must be replaced by estimates, and this is the reason why we analyse the behaviour of the empirical process marked with residuals defined by Eq. 10.

Processes of this type have been extensively used in model checks for regression and time series. A broad list of references can be found in Escanciano (2007b). There is a connection between our starting point and that in Ngatchou-Wandji and Laïb (2008), though their test refers to a nonlinear AR model of order one, and ours to linear AR models of any finite order.

On the other hand, a first study of the limiting behaviour of empirical processes of the residuals of autoregressive models can be found in Kulperger (1985). See also Bai (1994), Lee and Wei (1999) and Engler and Nielsen (2009).

J. C. Escanciano has studied the weak convergence of several marked empirical process related to time series problems, including multidimensional non linear autoregressive models of order one and heteroskedastic series (Escanciano 2007a, b, 2010), and applies his results to model checks in parametric regression, and martingale hypothesis.

For our purposes of designing focused tests, we need to analyse the weak convergence of Eq. 10 not only under the null hypothesis, but also under local alternatives, and this is the purpose of Section 3.

The present article extends the methods applied in previous work on assessing goodness of fit for i.i.d. sequences and regression models (Cabaña and Cabaña 2003, 2009). These procedures share the property of being consistent, asymptotically distribution free, and with almost optimal power with respect to the alternatives of focusing. Next section presents the general ideas regarding those methods.

Finally, the test is described in Section 4, and its statistical application is illustrated with one example.

2 A Recipe for Goodness-of-Fit Testing by Transforming Processes

The empirical information applied in solving many inference problems can be described and summarized by means of certain random processes, that we will generically denote by r_n, where n is the number of observations. Important examples are the *empirical process* built from i.i.d. samples, or the processes of accumulated residuals, in regression. In both cases, r_n converges in law to a Gaussian process r: the Brownian Bridge in the i.i.d. case, and a Wiener process for the accumulated residuals in regression (see Cabaña and Cabaña 2003, 2009).

In the present context of time series, we will study the marked empirical processes $r_n = \zeta_n$ and $r_n = \hat{\zeta}_n$ defined by Eqs. 9 and 10, respectively. We state in Section 3 that these processes also have Gaussian limits.

Because in all cases the limit of the process r_n is Gaussian, it is natural to test the effect of different alternative models through the limiting distribution with the use of contiguity and Le Cam Third Lemma (Hájek and Šidák 1999, Section 7.1.4; see also Lehmann and Romano 2005, Corollary 12.3.2).

We shall briefly summarize some facts about the processes r_n which are common to all scenarios, and discuss how to make inference based on any of these processes and their Gaussian limits r in order to solve goodness-of-fit problems.

Consider r_n and r as members of the space $L^2(\mathbf{R}, P)$ of square integrable functions defined on $\mathbf{R}$ with respect to a suitably chosen probability measure P and let w_P denote a (P)-Wiener process, that is, a centred Gaussian process with independent increments and variance $\mathbf{E}(w_P(b) - w_P(a))^2 = P((a, b]) \ (a < b)$.

Under the null hypothesis of fit, in all our previous applications, r can be written as the integral of the projection of a Gaussian white noise, that is, the formal derivative dw_P/dP, on the orthogonal complement of a finite-dimensional subspace S of $L^2(\mathbf{R}, P)$, with the scalar product given by the Wiener integral $\langle g, dw_P/dP\rangle = \int_{-\infty}^{\infty} g(s)dw_P(s)$.

For instance, in testing the fit of a random sample to a given distribution, r is a Brownian bridge and S is the one dimensional subspace generated by the constant functions. In testing the fit to a family with d parameters the dimension of S increases by d (Cabaña and Cabaña 2003, 2005). In the case of a linear model with d unknown coefficients including the intercept, S has dimension d (Cabaña and Cabaña 2009).

By selecting any orthonormal basis $(\psi_j)_{j=0,1,2,\ldots}$ such that $\psi_0,\ldots,\psi_{d-1}$ generate S, $r(t)$ can be written as the integrated Fourier expansion

$$r(t)=\sum_{j=d}^{\infty}B_j\int_{-\infty}^{t}\psi_j(s)dP(s)$$

with $B_j=\int_{-\infty}^{\infty}\psi_j(s)dw(s)$ i.i.d.standard Gaussian. In all the cases formerly developed the constant functions belong to S and therefore we choose $\psi_0=1$ with no loss of generality.

In each of those applications, a suitable family of contiguous alternatives is considered, under which the asymptotic distribution of $r_n(t)$ is $r(t)$ plus a drift $\delta\int_{-\infty}^{t}k(s)dP(s)$, where k is a function of norm one, orthogonal to S.

By further imposing $\psi_d=k$, the limiting law of $r_n(t)$ becomes

$$(B_d+\delta)\int_{-\infty}^{t}\psi_d(s)dP(s)+\sum_{j=d+1}^{\infty}B_j\int_{-\infty}^{t}\psi_j(s)dP(s). \tag{3}$$

The effect of the parameter estimation on the limiting distribution is reflected only by S, so, in order to avoid it and get a distribution free functional statistic, we apply to $r_n(t)$ the transformation that maps each expansion $\sum_{j=d}^{\infty}c_j\int_{-\infty}^{t}\psi_j(s)dP(s)$ onto $\sum_{j=0}^{\infty}c_{j+d}\int_{-\infty}^{t}\psi_j(s)dP(s)$, and hence the transformed process $w_n(t)$ has limit

$$(B_d+\delta)\int_{-\infty}^{t}dP(s)+\sum_{j=1}^{\infty}B_{j+d}\int_{-\infty}^{t}\psi_j(s)dP(s)=w(P(t))+\delta P(t)$$

with a standard Wiener process w and $P(t)=\int_{-\infty}^{t}dP(s)$.

The process w_n is obtained from r_n by means of

$$w_n(t)=\int(\mathcal{T}\mathbf{1}_t(s))dr_n(s)$$

where $\mathbf{1}_t$ is the indicator function of $(-\infty,t)$ and $\mathcal{T}$ is the inverse of the d-th power of the backshift operator associated to the given basis, and hence an isometry.

Consequently statistics of Kolmogorov–Smirnov, Cramér–von Mises and Watson type lead to consistent tests of the null hypothesis $\delta=0$, focused on the alternative $\delta\neq 0$.

In our former applications we have tested $\delta=0$ by means of the Watson type statistic

$$Q_n=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}c_{s,t}(u)dw_n(u)\right)^2dP(s)\,dP(t) \tag{4}$$

with $c_{s,t}(u)=\mathbf{1}_{\{s<u<t\}}+\mathbf{1}_{\{u<t<s\}}+\mathbf{1}_{\{t<s<u\}}$, adapted to the processes w_n that do not vanish at $+\infty$, and critical regions $Q_n>$constant.

Notice that in order to get the value of Q_n, the actual computation of the transformed process w_n is not required, because changing the order of the integrals in Eq. 4 leads to write

$$\begin{aligned} Q_n &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((|P(s)-P(t)|-1/2)^2+1/4)dw_n(s)dw_n(t) \\ &= \int_0^1\int_0^1((|u-v|-1/2)^2+1/4)dw_n(P^{-1}(u))dw_n(P^{-1}(v)) \end{aligned}$$

and hence the double Fourier expansion

$$(|P(s)-P(t)|-1/2)^2+1/4=\sum_{j,k=0}^{\infty}c_{j,k}\psi_j(s)\psi_k(t)$$

with

$$c_{j,k}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\left(|P(s)-P(t)|-\frac{1}{2}\right)^2+\frac{1}{4}\right]\psi_j(s)\psi_k(t)dP(s)dP(t)$$

reduces Q_n to the quadratic form

$$Q_n=\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}c_{j,k}\epsilon_{j+d}\epsilon_{k+d}$$

where

$$\epsilon_{j+d}=\int_{-\infty}^{\infty}\psi_j(s)dw_n(s)=\int_{-\infty}^{\infty}\psi_{j+d}(s)dr_n(s).$$

The convergence in law of r_n to r implies that

$$Q_n\xrightarrow{\mathcal{L}}Q=\frac{1}{3}(Z+\delta)^2+Q',\quad Q'=\sum_{j=1}^{\infty}\frac{C_j^2+S_j^2}{2j^2\pi^2}$$

(see Cabaña and Cabaña 2009), where $Z, C_j, S_j (j = 1, 2, \dots)$ are i.i.d. standard Gaussian.

We have preferred to use a quadratic functional of w_n and not a statistic of the Kolmogorov–Smirnov type, because the latter one requires to establish the convergence of w_n in the supremum norm, while the uniform convergence of r_n suffices to ensure the convergence in law of the quadratic statistic (see Cabaña and Cabaña 2009, Section 3.1.6).

On the other hand, the use of Watson statistic is a consequence of the discussion in Cabaña and Cabaña (2001).

In each application, we have noticed that the test with critical region $(Z+\delta)^2 >$constant is asymptotically equivalent to the two-sided test of $\delta = 0$ against $\delta \neq 0$ based on the Neyman–Pearson statistic. The addition of Q', that makes the test based on Q_n consistent against fixed alternatives, reduces its asymptotic power in relation with the power of the two-sided Neyman–Pearson test, which is not consistent. However, the loss in power is negligible as shown in Fig. 1, due to the relatively small variance of Q'.

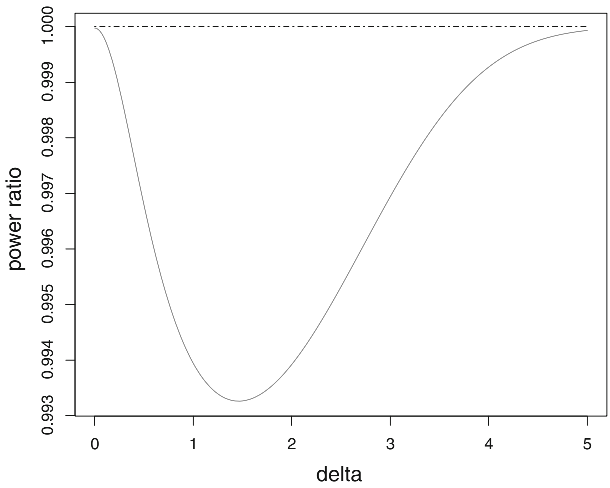

Fig. 1 Ratio of asymptotic powers of the test based on Q_n and the two-sided Neyman–Pearson test, with level 5%

In order to extend the previous methodology for testing goodness of fit in our present time series setup, we verify in next section that the limiting distribution of the empirical process marked with residuals (10) is closely related to the integral of the projection of a white noise onto a subspace of finite codimension, and replace it by a slight modification that has exactly the required asymptotic law, as shown in Corollary 1. Once this is done, the construction of the test for AR models following our general procedure is straightforward. For this present application, the space $L^2(\mathbf{R}, P)$ is replaced by $L^2([0, 1])$ with the uniform (Lebesgue) measure.

3 The Convergence Results

Let $(X_i)_{-p\le i\le n}$ denote observations of a stationary causal AR model

$$X_i = \phi_1 X_{i-1} + \cdots + \phi_p X_{i-p} + \frac{\delta}{\sqrt{n}} X_{i-p-1} + \sigma Z_i, \quad i = 1, 2, \ldots, n \tag{5}$$

with i.i.d. centred innovations Z_i with variance equal one.

In order to simplify the assumptions required for this presentation, we assume that the Z_i and hence the X_i are Gaussian. As an anonymous referee has kindly pointed out, such strong assumption is not necessary, and can be replaced by arguments as the ones applied in Koul (2002) to analyse the weak convergence of randomly weighted empirical processes.

Let $\gamma_j = \mathbf{E}X_iX_{i-j}$, $\rho_j = \gamma_j/\gamma_0$ denote the covariances and correlation coefficients of the random variables X_i, $i = 0, 1, 2, \ldots$.

When $\delta = 0$, the vectors $\boldsymbol{\gamma}$, $\boldsymbol{\phi}$ and the matrix Γ defined by

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \cdots \\ \gamma_p \end{pmatrix}, \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \cdots \\ \phi_p \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \cdots & \gamma_{p-3} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}$$

are related by the well known Yule–Walker equations $\Gamma\boldsymbol{\phi} = \boldsymbol{\gamma}$. The equation $\boldsymbol{\gamma}^{\mathrm{tr}}\boldsymbol{\phi} = \gamma_0 - \sigma^2$ also holds (see for instance Brockwell and Davis 2002).

The unknown parameters of the model, namely, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^{\mathrm{tr}}$ and σ can be estimated by conditional maximum likelihood under the assumption $\delta = 0$ corresponding to an AR(p) model. Rewrite Eq. 5 with vector notation as

$$\boldsymbol{X} = \sum_{j=1}^{p} \phi_j \boldsymbol{X}_j + \frac{\delta}{\sqrt{n}} \boldsymbol{X}_{p+1} + \sigma \boldsymbol{Z} \tag{6}$$

with $\boldsymbol{X}_j = (X_{1-j}, X_{2-j}, \ldots, X_{n-j})^{\mathrm{tr}}$ $(j = 0, 1, 2, \ldots, p+1)$, $\boldsymbol{X} = \boldsymbol{X}_0$, and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathrm{tr}}$.

With $\mathcal{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p)$ and $\delta = 0$, Eq. 6 reduces to $\boldsymbol{X} = \mathcal{X}\boldsymbol{\phi} + \sigma\boldsymbol{Z}$, which for each n leads to the conditional maximum likelihood estimators (given $X_{-p+1}, X_{-p+2}, \ldots, X_{-1}, X_0$):

$$\hat{\boldsymbol{\phi}}_n = (\mathcal{X}^{\mathrm{tr}}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{tr}}\boldsymbol{X}, \; \hat{\sigma}_n^2 = \frac{1}{n}\|\boldsymbol{X} - \mathcal{X}\hat{\boldsymbol{\phi}}_n\|^2. \tag{7}$$

Let us introduce the residuals

$$\hat{\boldsymbol{Z}} = \frac{1}{\hat{\sigma}_n}(\boldsymbol{X} - \mathcal{X}\hat{\boldsymbol{\phi}}_n) = (\hat{Z}_{1,n}, \hat{Z}_{2,n}, \ldots, \hat{Z}_{n,n})^{\mathrm{tr}}, \tag{8}$$

the random sums

$$\eta_{j,n} = \frac{1}{\sqrt{n\gamma_0}} \sum_{i=1}^{n} X_{i-j} Z_i, \;\; j = 1, 2, \ldots,$$

the empirical process marked with the true innovations

$$\zeta_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{1}_{\{X_{i-p-1} \le \sqrt{\gamma_0}\Phi^{-1}(t)\}} Z_i, \tag{9}$$

(Φ denotes the standard normal c.d.f.) and the empirical process marked with residuals

$$\hat{\zeta}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{1}_{\{X_{i-p-1} \le \sqrt{\gamma_0}\Phi^{-1}(t)\}} \hat{Z}_{i,n}. \tag{10}$$

With these notations, the following statements hold:

Theorem 1 *For $\delta = 0$, the sequence of processes $\zeta_n(t), 0 \le t \le 1$ converges weakly in $D(0,1)$ to a standard Wiener process $\xi(t)$ on $(0,1)$.*

Theorem 2 *Let $\tilde{\boldsymbol{\rho}} = (\rho_p, \rho_{p-1} \ldots, \rho_1)$ denote the vector of correlation coefficients in reversed order and φ the standard normal p.d.f.*

(i) *For $\delta = 0$, the sequence $(\zeta_n(t), \boldsymbol{\eta}_n)$ with $\boldsymbol{\eta}_n = (\eta_{1,n}, \eta_{2,n}, \ldots, \eta_{p,n})$ converges jointly in law to $(\xi(t), \boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is a centred Gaussian vector with variance matrix $R := \Gamma/\gamma_0$, and $\mathbf{E}\xi(t)\boldsymbol{\eta} = -\tilde{\boldsymbol{\rho}}\varphi(\Phi^{-1}(t))$.*

(ii) *The pair* $(\xi(t), \boldsymbol{\eta})$ *admits the representation*

$$(\xi(t), \boldsymbol{\eta}) \sim (\xi(t), \xi_1 \tilde{\boldsymbol{\rho}} + \boldsymbol{\theta})$$

with $\xi_1 = \int_0^1 \Phi^{-1}(s) d\xi(s)$ *and* $\boldsymbol{\theta}$ *centred Gaussian with variance* $V = R - \tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{\rho}}^{tr}$, *independent of* $\xi(t), 0 \le t \le 1$.

Theorem 3

(i) *For* $\delta = 0$ *the process* $\hat{\zeta}_n(t)$ *converges weakly in* $D(0, 1)$ *to the centred Gaussian process*

$$\zeta(t) = \xi(t) - \int_0^t \Phi^{-1}(s) ds \frac{1}{\gamma_0} \left(\xi_1 (\gamma_0 - \sigma^2) + \sigma \sqrt{\gamma_0 - \sigma^2} W \right) \tag{11}$$

where $\xi_1 = \int_0^1 \Phi^{-1}(s) d\xi(s)$ *and* W *is a standard Gaussian variable independent of* ξ.

(ii) *For general* δ, *the process* $\hat{\zeta}_n(t)$ *converges weakly in* $D(0, 1)$ *to*

$$\zeta(t) + \frac{\delta\sigma}{\sqrt{\gamma_0}} \int_0^t \Phi^{-1}(s) ds. \tag{12}$$

Note The correction added to $\xi(t)$ in Eq. 11 is due to the estimation by conditional ML of the parameters of the AR(p) process. A correction would also appear for other estimators such as least squares, and more generally those admitting a Bahadur representation. The correction for $\delta \neq 0$ is the addition of a deterministic term as implied by Le Cam Third Lemma.

Corollary 1 *Let* $(g_j(t))_{j=0,1,\dots}$ *denote an orthonormal basis of* $L^2(0, 1)$ *with the uniform measure, such that* $g_0(t) = 1$, $g_1(t) = \Phi^{-1}(t)$ *and* $(w_j)_{j=0,1,2,\dots}$ *be i.i.d. standard Gaussian.*

The sequence of processes

$$\zeta_n^*(t) = \hat{\zeta}_n(t) + \left(\sqrt{\gamma_0/\sigma^2} - 1 \right) \int_0^t g_1(s) ds \int_0^1 \Phi^{-1}(s) d\hat{\zeta}_n(s) - t \int_0^1 d\hat{\zeta}_n(s)$$

converges in distribution to

$$(w_1 + \delta) \int_0^t g_1(s) ds + \sum_{j=2}^{\infty} w_j \int_0^t g_j(s) ds = w(t) - t w(1) - \delta \varphi(\Phi^{-1}(t)) \tag{13}$$

where $w(t) = \sum_{j=0}^{\infty} w_j \int_0^t g_j(s) ds$ *is a standard Wiener process.*

Remark 1 An orthonormal basis with the properties stated in the corollary is obtained with $g_j(t) = h_j(\Phi^{-1}(t))$, where $(h_j)_{j=0,1,2,\dots}$ are the normalised Hermite polynomials defined by the expansion

$$e^{-y^2 + xy} = \sum_{j=0}^{\infty} h_j(x) y^j / \sqrt{j!}.$$

Remark 2 The limiting distribution (13) indicates that ζ_n^* provides a fair asymptotic starting point to perform inference with optimal properties about the model (5) and the null hypothesis "$\delta = 0$" in particular, as discussed in next section.

4 A Test of the Null Hypothesis $\delta = 0$

4.1 On a Test Statistic Based on the Marked Process

The limiting distributions of the processes associated to $\hat{\zeta}_n(t)$ stated in Theorem 3 indicate that those processes provide information useful in making inference about Eq. 5, but all of them contain some unknowns parameters, so that they cannot be computed exclusively from the data.

For this reason we consider the pairs $(X_{i-p-1}, \hat{Z}_i)_{i=1,2,\ldots,n}$ and denote by $\hat{Z}_{(i:n)}$ the concomitants of the order statistics $X_{i-p-1:n}$, that is, the order statistics of $(\hat{Z}_i)_{i=1,2,\ldots,n}$ with the order induced by the $(X_{i-p-1})_{i=1,2,\ldots,n}$. Now we introduce a new sequence of processes, namely, the partial sums

$$z_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[(n+1)t]} \hat{Z}_{(i:n)}$$

where $[\cdot]$ denotes the integer part.

We shall notice that $\hat{\zeta}_n$ and this functional statistic are asymptotically equivalent, so that the inference can be performed from z_n that has the same desirable properties as $\hat{\zeta}_n$.

The equivalence is due to the fact that both processes are related by

$$z_n\left(\frac{n}{n+1} F_n(t)\right) = \hat{\zeta}_n(t) \tag{14}$$

where $F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_{i-p-1} \le \sqrt{\gamma_0}\Phi^{-1}(t)\}}$, because the limit in probability $\text{plim}_{n\to\infty} \sup_{0<t<1} |\frac{n}{n+1} F_n(t) - t| = \text{plim}_{n\to\infty} \sup_{0<t<1} |F_n(t) - t| = 0$ holds.

Now Corollary 1 leads to conclude that the sequence

$$\begin{aligned} z_n^*(t) &= z_n(t) - \varphi(\Phi^{-1}(t))\left(\sqrt{\hat{\gamma}_{0,n}/\hat{\sigma}_n^2} - 1\right)\int_0^1 \Phi^{-1}(s)dz_n(s) - tz_n(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{[(n+1)t]} \hat{Z}_{(i:n)} - \varphi(\Phi^{-1}(t))\left(\sqrt{\hat{\gamma}_{0,n}/\hat{\sigma}_n^2} - 1\right) \\ &\quad \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi^{-1}\left(\frac{i}{n+1}\right)\hat{Z}_{(i:n)} - t\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Z}_i \end{aligned}$$

converges in distribution to Eq. 13, that is precisely Eq. 3 with $d = 1$, $\psi_j = g_j$ and P is the uniform distribution on (0, 1).

Consequently the test statistics becomes

$$Q_n = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} c_{j,k}\epsilon_{j+1}\epsilon_{k+1}$$

where

$$c_{j,k} = \int_0^1 \int_0^1 ((|u-v| - 1/2)^2 + 1/4) g_j(u) g_k(v) du\, dv,$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\left(|\Phi(s) - \Phi(t)| - \frac{1}{2} \right)^2 + \frac{1}{4} \right] h_j(s) h_k(t) d\Phi(s) d\Phi(t)$$

and

$$\epsilon_j = \int_0^1 g_j(s) dz_n^*(s) = \begin{cases} \sqrt{\frac{\hat{\gamma}_{0,n}}{\hat{\sigma}_n^2}} \frac{1}{\sqrt{n}} \sum_{i=1}^n g_1(i/(n+1)) \hat{Z}_{(i:n)} \text{ for } j = 1 \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n g_j(i/(n+1)) \hat{Z}_{(i:n)} \text{ for } j > 1. \end{cases}$$

The coefficients $c_{j,k}$ have already appeared in the normality tests introduced in Cabaña and Cabaña (2003), where their values for $0 \le j, k \le 5$ are tabulated. Table 1 provides their values for $0 \le j, k \le 11$. For practical computations, the quadratic form Q_n is approximated by $Q_n^{(\ell)} = \sum_{j,k=0}^{\ell} c_{j,k} \epsilon_{j+1} \epsilon_{k+1}$ with some moderate value of ℓ.

While the test based on Q_n is consistent, the one based on $Q_n^{(\ell)}$ is not. In order to maintain the consistency, it would be required to let ℓ tend to infinity, as $n \to \infty$. For practical purposes, a moderate value of ℓ makes the test sensitive to a broad family of alternatives.

Table 1 Values of the coefficients

$$c_{k,j} = c_{j,k} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [(|\Phi(y) - \Phi(x)| - .5)^2 + .25] h_j(x) h_k(y) d\Phi(x) d\Phi(y).$$

The function h_j is the j-th normalized Hermite polynomial, and Φ is the standard normal distribution function

i	j	$c_{j,k}$	j	k	$c_{j,k}$	j	k	$c_{j,k}$
0	0	0.3333333	3	3	0.0137881	6	6	0.0064284
0	≥1	0.0000000	3	5	−0.0099516	6	8	−0.0049521
1	1	0.0246214	3	7	0.0069702	6	10	0.0037782
1	3	−0.0175302	3	9	−0.0048125	7	7	0.0049669
1	5	0.0114719	3	11	0.0032963	7	9	−0.0039947
1	7	−0.0073017	4	4	0.0119114	7	11	0.0031530
1	9	0.0045937	4	6	−0.0082846	8	8	0.0041280
1	11	−0.0028759	4	8	0.0058122	8	10	−0.0033748
2	2	0.0306294	4	10	−0.0040844	9	9	0.0034264
2	4	−0.0176839	5	5	0.0078756	9	11	−0.0028673
2	6	0.0107621	5	7	−0.0060138	10	10	0.0029308
2	8	−0.0067113	5	9	0.0045022	11	11	0.0025296
2	10	0.0042450	5	11	−0.0033270	$j+k$ odd		0.0000000

4.2 Comparison of the Performances of the Classical t-Test and the Test Based on Q_n

A standard way to test $\delta = 0$ is to apply the t-test to the conditional maximum likelihood estimator $\hat{\delta}$, obtained from the last component of

$$\begin{pmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\delta}/\sqrt{n} \end{pmatrix} = \begin{pmatrix} \mathcal{X}^{\mathrm{tr}}\mathcal{X} & \mathcal{X}^{\mathrm{tr}}\boldsymbol{X}_{p+1} \\ \boldsymbol{X}_{p+1}^{\mathrm{tr}}\mathcal{X} & \boldsymbol{X}_{p+1}^{\mathrm{tr}}\boldsymbol{X}_{p+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{X}^{\mathrm{tr}} \\ \boldsymbol{X}_{p+1}^{\mathrm{tr}} \end{pmatrix} \boldsymbol{X}.$$

A computation based on $\mathrm{plim}\frac{1}{n}\mathcal{X}^{\mathrm{tr}}\mathcal{X} = \Gamma$ and the Yule–Walker equations show that the asymptotic variance of $\hat{\delta}$ is one. Therefore the two-sided t-test is

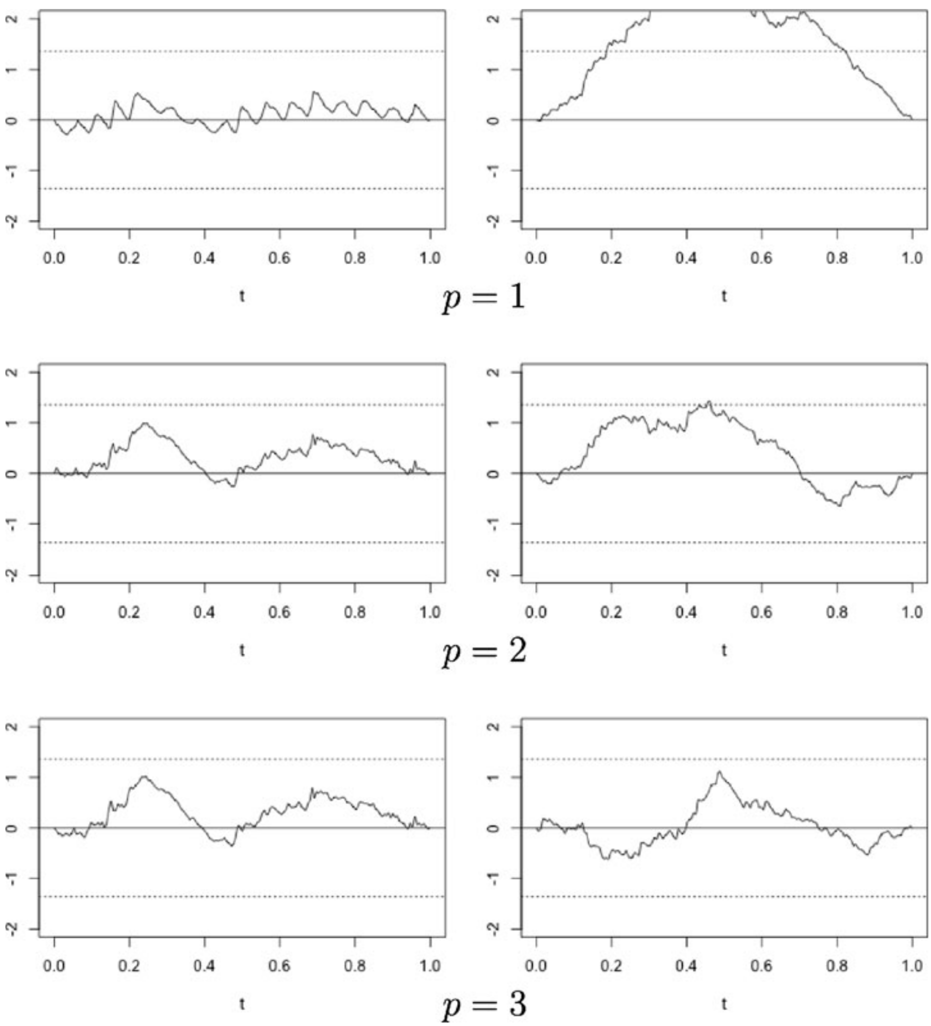

Fig. 2 Plots of the normalized accumulated sums of residuals $z_n'(t)$, $z_n^*(t)$, $0 \le t \le 1$

Table 2 Application to Wolfer's sunspot numbers: values of Q_n^ℓ ($n = 176$, $\ell = 11$, $p = 1, 2, \ldots, 16$), and approximated p-values obtained by comparing with the asymptotic law

p	1	2	3	4	5	6	7	8
Q_n^ℓ	22.88	2.14	0.20	0.04	1.60	1.77	2.74	1.12
p-value	0.0000	0.015	0.713	0.997	0.038	0.028	0.005	0.091

p	9	10	11	12	13	14	15	16
Q_n^ℓ	1.29	1.14	0.13	0.13	0.17	0.41	0.87	0.31
p-value	0.066	0.087	0.848	0.852	0.769	0.408	0.148	0.526

asymptotically equivalent to a test with rejection region $(Z + \delta)^2 >$constant, with Z standard Gaussian, and hence the comments at the end of Section 2 apply.

4.3 An Example

In order to illustrate the performance of the test based on the transformed process, we apply it to the series of 176 sunspot numbers analysed by Anderson (1971).

Figure 2 compares the plots of accumulated sums of the residuals $z_n'(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[(n+1)t]} \hat{Z}_i$ and the functional statistic $z_n^*(t)$ introduced in Section 4.1. In the diagrams at the left-hand side the polygonal lines join the points $(i/n, z_n'(i/n))$, $i = 0, 1, \ldots, n$, and apparently do not behave significatively different from a Brownian bridge. On the right-hand side, the polygonals join the points $(i/n, z_n^*(i/n))$, $i = 0, 1, \ldots, n$, and show how the accumulated reordered residuals corrected for the mean exhibit a correlation between $\hat{Z}_i$ and X_{i-p-1}, in the case $p = 1$. The graphics show clearly that the AR(1) model should be rejected.

Table 2 shows the values computed for Q_{176}^{11} and $p = 1, \ldots, 16$ and the p-values corresponding to the asymptotic distribution of Q_n under "$\delta = 0$". This corroborates the conclusion obtained from the graphs, and furthermore, suggests to adopt at least the AR(3) model to fit the data.

Remark In order to perform the tests, it is advisable to obtain the actual distribution for finite samples by Monte Carlo simulation.

5 Comments on the Statements in Section 3 and Sketch of Proof of Theorem 1

Theorems 2 and 3 help us to establish the notation. They contain and apply results not essentially different of other found in the statistical literature (see for instance Brockwell and Davis 2002; Koul 2002). Then Corollary 1 can be proved by establishing that a copy of $\hat{\zeta}_n$ converging uniformly to a copy of ζ (that is, a Skorokhod representation) leads to copies of the ζ_n^* that converge uniformly to a copy of the stated limit. These considerations motivate no further comments on their proofs.

Theorem 1 is the main convergence result, supporting the application of the general procedure described in Section 2. There is a much more general result in Koul and Stute (1999) that can be applied to provide a similar convergence result for nonlinear AR models of order one under suitable assumptions on the model and its innovations. Notwithstanding, we develop here a sketch of the proof of Theorem 1 in our restricted context of a stationary model with Gaussian innovations, because in this particular case no further assumptions are needed.

We split the proof in three steps.

Step 1 The first- and second-order moments of $\zeta_n(t)$ and $\xi(t)$ are the same. Introduce the indicator process $\boldsymbol{U}(t)$, a n-vector with components $U_i(t) = (\mathbf{1}_{\{X_{i-p-1}\le\sqrt{\gamma_0}\Phi^{-1}(t)\}})_{i=1,2,\ldots,n}$, and write $\zeta_n(t) = \frac{1}{\sqrt{n}}\mathbf{Z}^{\mathrm{tr}}\boldsymbol{U}(t)$. The expectation of ζ_n is zero, because each Z_i is independent of $U_i(t)$, and the covariance $\mathbf{E}\zeta_n(s)\zeta_n(t)$ is the common value for all i of

$$\mathbf{E}Z_i^2U_i(s)U_i(t) = \mathbf{P}\{X_{i-p-1}\le\sqrt{\gamma_0}(\Phi^{-1}(s)\wedge\Phi^{-1}(t))\} = s\wedge t.$$

Step 2 The sequence $\zeta_n(t)$ converges fi. di. to the Wiener process $\xi(t)$. In view of Step 1, it suffices to verify that the limiting distribution of finite dimensional vectors $(\zeta_n(t_1),\ldots,\zeta_n(t_k))$ with $t_1,\ldots,t_k$ arbitrarily chosen in $(0,1)$ is Gaussian, and this can be done by applying Theorem 5.1 in Serfling (1968).

Step 3 The sequence $\zeta_n(t)$ is tight. The tightness of the sequence $\zeta_n(t)$ can be verified by applying the moment inequality criterion in Billingsley (1999, Theorem 13.5). We shall show that for $0\le t_1<t<t_2\le 1$, the expectation

$$\mathbf{E}(\zeta_n(t)-\zeta_n(t_1))^2(\zeta_n(t_2)-\zeta_n(t))^2 \tag{15}$$

is bounded by $(G(t_2)-G(t_1))^2$ for a bounded and increasing function G.

We introduce the notation

$$U_i(t_1,t_2) = \mathbf{1}_{\{\sqrt{\gamma_0}\Phi^{-1}(t_1)<X_{i-p-1}\le\sqrt{\gamma_0}\Phi^{-1}(t_2)\}}$$

write the expectation (15) as a sum in four indices

$$\frac{1}{n^2}\sum_{h,i,j,k}\mathbf{E}Z_hU_h(t_1,t)Z_iU_i(t_1,t)Z_jU_j(t,t_2)Z_kU_k(t,t_2).$$

and notice that the terms with one of the indices greater than the other three have expectation zero because the corresponding Z is independent of the other factors in that term. The expression reduces then to a sum in three indices, by equating two of them, and stating that the remaining two are not larger.

Since the terms with $i=j, i=k, h=j, h=k$ are zero, the nonvanishing terms correspond to $i=h>j,k$ and $i,h<j=k$, so that Eq. 15 can be written, with a change of indices, and replacing the independent factor Z_i^2 by its expectation, as the sum of

$$\frac{1}{n^2}\sum_i\mathbf{E}U_i(t_1,t)\left(\sum_{j<i}Z_jU_j(t,t_2)\right)^2 \tag{16}$$

and the analogous expression obtained by interchanging the intervals (t_1,t) and (t,t_2).

In order to establish the required bound, it will be enough to verify that for any i, the inequality

$$\mathbf{E}U_i(t,t_2)\left(\frac{1}{\sqrt{n}}\sum_{j<i}Z_jU_j(t_1,t)\right)^2 < (F(t_2)-F(t_1))^{1+\alpha} \tag{17}$$

holds for some bounded increasing function F and a positive α, or, since the left hand side of Eq. 17 is the sum of

$$\frac{1}{n}\mathbf{E}U_i(t,t_2)\sum_{j<i}Z_j^2U_j(t_1,t) \tag{18}$$

and

$$2\mathbf{E}U_i(t,t_2)\frac{1}{n}\sum_{j<i}Z_jU_j(t_1,t)\sum_{k<j}Z_kU_k(t_1,t), \tag{19}$$

it suffices with establishing bounds as the one in Eq. 17 for Eqs. 18 and 19.

Let us split the remainder of the proof in several lemmas.

Lemma 1 *If* $\begin{pmatrix}X\\Y\end{pmatrix}\sim Normal\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&r\\r&1\end{pmatrix}\right)$ *and* $0<\rho<1$, *then there exist increasing and bounded functions* $F_{h,\rho}(x)$, $h=0,1,2,\ldots$ *such that the inequalities*

$$\mathbf{E}X^{2h}\mathbf{1}_{\{x_1<X<x_2\}}\mathbf{1}_{\{x_1<Y<x_2\}} < (F_{h,\rho}(x_2)-F_{h,\rho}(x_1))^2 \tag{20}$$

hold uniformly on $|r|\le\rho<1$.

Remark Let $\Phi_{2h}(x)=\int_{-\infty}^{x}x^{2h}\varphi(x)dx$ (and in particular $\Phi_0=\Phi$). The functions defined by

$$F_{h,\rho}(x)=\frac{(1+\rho)^{h+1}}{\sqrt[4]{1-\rho^2}(1-\rho)^{\frac{h+1}{2}}}\left(\Phi_{2h}\left(\frac{x}{\sqrt{1+\rho}}\right)+\Phi\left(\frac{x}{\sqrt{1+\rho}}\right)\right) \tag{21}$$

satisfy Eq. 20.

Lemma 2 *There exists a constant* C *such that the expression* (18) *is bounded by* $C(F_{2,\rho}(\Phi^{-1}(t_2))-F_{2,\rho}(\Phi^{-1}(t_1)))^2$, *where* $F_{2,\rho}$ *is the bounded increasing function defined by Eq. 21 and* $\rho=\max_{i\ne j}|\mathbf{E}X_iX_j/\gamma_0|$.

Lemma 3 *There exist a constant* C_1 *and* ρ_1, $0<\rho_1<1$, *such that the expression* (19) *is bounded by* $C_1[(F_{2,\rho_1}(t_2)-F_{2\rho_1}(t_1))^2+(F_{0,\rho_1}(t_2)-F_{0\rho_1}(t_1))^2]$, *where* F_{2,ρ_1}, F_{0,ρ_1} *are defined by Eq. 21.*

These three lemmas imply that Eq. 15 is bounded by a sum of three terms, each of which is the squared increment of an increasing and bounded function. It is plain that the sum of three (or any finite number of) such terms can be expressed as the squared increment of a new increasing and bounded function, so that this ends the verification of the tightness.

Sketch of the proof of Lemma 1 The joint density

$$\varphi_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)\right)$$

of the random variables X, Y is bounded by $\frac{1}{\sqrt{1-r^2}}\varphi(x/\sqrt{1+|r|})\varphi(y/\sqrt{1+|r|})$ and therefore

$$\begin{aligned}
\mathbf{E}X^{2h}\mathbf{1}_{\{x_1<X<x<Y<x_2\}} &\leq \frac{1}{\sqrt{1-r^2}} \int_{x_1}^{x} (x')^{2h}\varphi(x'/\sqrt{1+|r|})dx' \int_{x}^{x_2} \varphi(y/\sqrt{1+|r|})dy \\
&\leq \frac{(1+|r|)^{h+1}}{\sqrt{1-r^2}} \int_{x_1/\sqrt{1+|r|}}^{x/\sqrt{1+|r|}} (x')^{2h}\varphi(x')dx' \int_{x/\sqrt{1+|r|}}^{x_2/\sqrt{1+|r|}} \varphi(y)dy \\
&= \frac{(1+|r|)^{h+1}}{\sqrt{1-r^2}} \left(\Phi_{2h}\left(\frac{x_2}{\sqrt{1+|r|}}\right) - \Phi_{2h}\left(\frac{x_1}{\sqrt{1+|r|}}\right)\right) \\
&\quad \times \left(\Phi\left(\frac{x_2}{\sqrt{1+|r|}}\right) - \Phi\left(\frac{x_1}{\sqrt{1+|r|}}\right)\right). \qquad (22)
\end{aligned}$$

The derivative $\frac{\partial}{\partial x}\Phi_{2h}(\frac{x}{\sqrt{1+|r|}}) = (\frac{x}{\sqrt{1+|r|}})^{2h}\varphi(\frac{x}{\sqrt{1+|r|}})\frac{1}{\sqrt{1+|r|}}$ is uniformly bounded by $(\frac{x}{\sqrt{1-\rho}})^{2h}\varphi(\frac{x}{\sqrt{1+\rho}})/\sqrt{1-\rho}$ in $|r| \leq \rho$, so that

$$\begin{aligned}
\Phi_{2h}\left(\frac{x_2}{\sqrt{1+|r|}}\right) - \Phi_{2h}\left(\frac{x_1}{\sqrt{1+|r|}}\right) &\leq \int_{x_1}^{x_2} \left(\frac{x}{\sqrt{1-\rho}}\right)^{2h} \varphi\left(\frac{x}{\sqrt{1+\rho}}\right)\frac{dx}{\sqrt{1-\rho}} \\
&= \left(\frac{\sqrt{1+\rho}}{\sqrt{1-\rho}}\right)^{2h+1} \\
&\quad \times \left(\Phi_{2h}\left(\frac{x_2}{\sqrt{1+\rho}}\right) - \Phi_{2h}\left(\frac{x_1}{\sqrt{1+\rho}}\right)\right)
\end{aligned}$$

holds for all $|r| \leq \rho$ and in particular, with $h = 0$,

$$\Phi\left(\frac{x_2}{\sqrt{1+|r|}}\right) - \Phi\left(\frac{x_1}{\sqrt{1+|r|}}\right) \leq \frac{\sqrt{1+\rho}}{\sqrt{1-\rho}}\left(\Phi\left(\frac{x_2}{\sqrt{1+\rho}}\right) - \Phi\left(\frac{x_1}{\sqrt{1+\rho}}\right)\right).$$

From Eq. 22 we conclude that $\mathbf{E}X^{2h}\mathbf{1}_{\{x_1<X<x<Y<x_2\}}$ is bounded by

$$\begin{aligned}
&\frac{(1+\rho)^{2h+2}}{\sqrt{1-\rho^2}(1-\rho)^{h+1}}\left(\Phi_{2h}\left(\frac{x_2}{\sqrt{1+\rho}}\right) - \Phi_{2h}\left(\frac{x_1}{\sqrt{1+\rho}}\right)\right) \\
&\quad\times\left(\Phi\left(\frac{x_2}{\sqrt{1+\rho}}\right) - \Phi\left(\frac{x_1}{\sqrt{1+\rho}}\right)\right)
\end{aligned}$$

and hence by $(F_{h,\rho}(x_2) - F_{h,\rho}(x_1))^2$ with $F_{h,\rho}(x)$ given by Eq. 21. □

Sketch of the proof of Lemma 2 The sum (18) has less than n terms, hence it suffices to verify that the stated inequality is satisfied by each term

$$\mathbf{E}U_i(t,t_2)Z_j^2U_j(t_1,t), \quad j<i. \qquad (23)$$

By introducing the standard Gaussian variables $X = X_{i-p-1}/\sqrt{\gamma_0}$, $Y = X_{j-p-1}/\sqrt{\gamma_0}$, and $Z = Z_j$ with covariances $\mathbf{E}XY = r$, $\mathbf{E}XZ = s$, $\mathbf{E}YZ = 0$, Eq. 23

can be written as $\mathbf{E}\mathbf{1}_{\{x_1<X<x\}}\mathbf{1}_{\{x<Y<x_2\}}Z^2$ with $x_1 = \Phi^{-1}(t_1)$, $x = \Phi^{-1}(t)$, $x_2 = \Phi^{-1}(t_2)$. The covariance r is bounded in absolute value by $\rho = \max_{i\neq j}|\mathbf{E}X_iX_j/\gamma_0| < 1$.

Now we express Z as a linear combination $aX + bY + cW$ of X, Y and a new standard Gaussian variable W independent of X, Y and apply Lemma 1 to each term in the right-hand side of the inequality

$$\begin{aligned}\mathbf{E}\mathbf{1}_{\{x_1<X<x\}}\mathbf{1}_{\{x<Y<x_2\}}Z^2 \leq\ & 3a^2\mathbf{E}\mathbf{1}_{\{x_1<X<x_2\}}\mathbf{1}_{\{x_1<Y<x_2\}}X^2\\ &+3b^2\mathbf{E}\mathbf{1}_{\{x_1<X<x_2\}}\mathbf{1}_{\{x_1<Y<x_2\}}Y^2\\ &+3c^2\mathbf{E}\mathbf{1}_{\{x_1<X<x_2\}}\mathbf{1}_{\{x_1<Y<x_2\}}W^2\end{aligned}$$

thus providing a bound for Eq. 23 with the required format. □

Sketch of the proof of Lemma 3 The expression (19) is the sum of less than n terms $2\mathbf{E}U_i(t,t_2)Z_jU_j(t_1,t)\sum_{k<j}Z_kU_k(t_1,t)$, $j<i\leq n$ divided by n, and each term can be written as

$$\sum_{k<j} 2\mathbf{E}Z_jZ_k\mathbf{1}_{\{x\leq Y_i\leq x_2\}}\mathbf{1}_{\{x_1\leq Y_j\leq x\}}\mathbf{1}_{\{x_1\leq Y_k\leq x\}}, \tag{24}$$

where $Y_h = X_{h-p-1}/\sqrt{\gamma_0}$, $x_1 = \Phi^{-1}(t_1)$, $x = \Phi^{-1}(t)$ and $x_2 = \Phi^{-1}(t_2)$. The statement will be proved by showing that Eq. 24 admits a bound as the required one, uniformly in i, j. □

Let us introduce the notations $s_h = \mathbf{E}Y_hZ_0$ and use $\rho_h = \mathbf{E}Y_hY_0$. Because the model is causal, $s_h = 0$ for $h \leq p$. Moreover, the stationarity implies

$$\sum_{h=-\infty}^{\infty}|\rho_h| < \infty, \quad \sum_{h=-\infty}^{\infty}|s_h| = \sum_{h=p+1}^{\infty}|s_h| < \infty. \tag{25}$$

The centred vector $\begin{pmatrix}Y_i\\Y_j\\Y_k\\Z_j\\Z_k\end{pmatrix}$ has variance $\begin{pmatrix}1 & \rho_{i-j} & \rho_{i-k} & s_{i-j} & s_{i-k}\\ \rho_{i-j} & 1 & \rho_{j-k} & 0 & s_{j-k}\\ \rho_{i-k} & \rho_{j-k} & 1 & 0 & 0\\ s_{i-j} & 0 & 0 & 1 & 0\\ s_{i-k} & s_{j-k} & 0 & 0 & 1\end{pmatrix}$.

We write now the last two components of that vector as linear combinations

$$Z_j = a_jY_i + b_jY_j + c_jY_k + d_jW + e_jW_j \tag{26}$$

$$Z_k = a_kY_i + b_kY_j + c_kY_k + d_kW + e_kW_k \tag{27}$$

of the standard Gaussian variables Y_i, Y_j, Y_k, W, W_j, W_k with the last three mutually independent and independent of the former ones, and compute the expectation

$$\begin{aligned}&2\mathbf{E}Z_jZ_k\mathbf{1}_{\{x\leq Y_i\leq x_2\}}\mathbf{1}_{\{x_1\leq Y_j\leq x\}}\mathbf{1}_{\{x_1\leq Y_k\leq x\}}\\ &= 2\mathbf{E}(a_jY_i+b_jY_j+c_jY_k)(a_kY_i+b_kY_j+c_kY_k)\\ &\quad\times\mathbf{1}_{\{x\leq Y_i\leq x_2\}}\mathbf{1}_{\{x_1\leq Y_j\leq x\}}\mathbf{1}_{\{x_1\leq Y_k\leq x\}}\\ &\quad+d_jd_k\mathbf{E}\mathbf{1}_{\{x\leq Y_i\leq x_2\}}\mathbf{1}_{\{x_1\leq Y_j\leq x\}}\mathbf{1}_{\{x_1\leq Y_k\leq x\}}.\end{aligned}$$

By using repeatedly the estimate $2|Y_h Y_l| \leq Y_h^2 + Y_l^2$, and introducing $\sigma_j = |a_j| + |b_j| + |c_j|, \sigma_k = |a_k| + |b_k| + |c_k|$, this leads to the estimate

$$
\begin{aligned}
&2\mathbf{E}|Z_j Z_k|\mathbf{1}_{\{x \leq Y_i \leq x_2\}}\mathbf{1}_{\{x_1 \leq Y_j \leq x\}}\mathbf{1}_{\{x_1 \leq Y_k \leq x\}} \\
&\leq [|a_j|\sigma_k + |a_k|\sigma_j]\mathbf{E}Y_i^2\mathbf{1}_{\{x_1 \leq Y_i \leq x_2\}}\mathbf{1}_{\{x_1 \leq Y_k \leq x_2\}} \\
&\quad + [|b_j|\sigma_k + |b_k|\sigma_j]\mathbf{E}Y_j^2\mathbf{1}_{\{x_1 \leq Y_j \leq x_2\}}\mathbf{1}_{\{x_1 \leq Y_k \leq x_2\}} \\
&\quad + [|c_j|\sigma_k + |c_k|\sigma_j]\mathbf{E}Y_k^2\mathbf{1}_{\{x_1 \leq Y_i \leq x_2\}}\mathbf{1}_{\{x_1 \leq Y_k \leq x_2\}} \\
&\quad + |d_j d_k|\mathbf{E}\mathbf{1}_{\{x_1 \leq Y_i \leq x_2\}}\mathbf{1}_{\{x_1 \leq Y_k \leq x_2\}}.
\end{aligned}
$$

Lemma 1 and the fact that $|\rho_h| \leq \rho = \max_{-\infty<h<\infty} |\rho_h| < 1$ imply that each of the four expectations is smaller that the increment of a bounded function uniformly in i, j, k and hence the conclusion of the lemma is obtained by verifying that the sums of coefficients

$$
\sum_{k<j}[|a_j|\sigma_k + |a_k|\sigma_j], \sum_{k<j}[|b_j|\sigma_k + |b_k|\sigma_j], \sum_{k<j}[|c_j|\sigma_k + |c_k|\sigma_j], \sum_{k<j} |d_j d_k|
$$

are uniformly bounded.

Some cumbersome but straightforward computations show that $\Delta_{i,j,k} := \det \mathbf{Var}(Y_i, Y_j, Y_k) = 1 - \rho_{i-j}^2 - \rho_{i-k}^2 - \rho_{j-k}^2 + 2\rho_{i-j}\rho_{i-k}\rho_{j-k}$, and the coefficients in Eqs. 26 and 27 are

$$
\begin{aligned}
a_j &= s_{i-j}(1 - \rho_{j-k}^2)/\Delta_{i,j,k} \\
b_j &= s_{i-j}(\rho_{i-k}\rho_{j-k} - \rho_{i-j})/\Delta_{i,j,k} \\
c_j &= s_{i-j}(\rho_{i-j}\rho_{j-k} - \rho_{i-k})/\Delta_{i,j,k} \\
a_k &= s_{i-k}(1 - \rho_{j-k}^2)/\Delta_{i,j,k} + s_{j-k}(\rho_{i-k}\rho_{j-k} - \rho_{i-j})/\Delta_{i,j,k} \\
b_k &= s_{i-k}(\rho_{i-k}\rho_{j-k} - \rho_{i-j})/\Delta_{i,j,k} + s_{j-k}(1 - \rho_{i-k}^2)/\Delta_{i,j,k} \\
c_k &= s_{i-k}(\rho_{i-j}\rho_{j-k} - \rho_{i-k})/\Delta_{i,j,k} + s_{j-k}(\rho_{i-j}\rho_{i-k} - \rho_{j-k})/\Delta_{i,j,k} \\
d_j d_k &= s_{i-k}(-s_{i-j}(1 - \rho_{j-k}^2))/\Delta_{i,j,k}^2 + s_{j-k}(-s_{i-j}(\rho_{i-k}\rho_{j-k} - \rho_{i-j}))/\Delta_{i,j,k}^2
\end{aligned}
$$

The expressions involving the correlations ρ_h are all absolutely bounded by 2, the absolute values of the correlations s_{i-j} between the X's and the Z's are trivially bounded by 1, and the determinant of the variance of any three different standardized Y's is uniformly bounded off zero, then each term of the sums in Eq. 25 is absolutely bounded by some constant K, the same for all i, j, k, times $|s_{i-k}| + |s_{j-k}|$, so that the sums are uniformly bounded.

Acknowledgements The helpful comments of the referees that lead to an improved version of the manuscript are gratefully acknowledged.

References

Anderson TW (1971) The statistical analysis of time series. Wiley, New York

Bai J (1994) Weak convergence of the sequential empirical processes of residuals in ARMA models. Ann Stat 22:2051–2061
Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley, New York
Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting, 2nd edn. Springer, Berlin
Cabaña A, Cabaña EM (2001) Goodness-of-fit tests based on quadratic functionals of transformed empirical processes. Statistics 35:171–189
Cabaña A, Cabaña EM (2003) Tests of normality based on transformed empirical processes. Methodol Comput Appl Probab 5:309–335
Cabaña A, Cabaña EM (2005) Goodness-of-fit to the exponential distribution, focused on Weibull alternatives. Commun Stat Simul Comput 34:711–724
Cabaña A, Cabaña EM (2009) Goodness-of-fit tests for continuous regression. Methodol Comput Appl Probab 11:119–144
Escanciano JC (2007) Model checks using residual marked empirical processes. Stat Sin 17:115–138
Escanciano JC (2007) Weak convergence of non-stationary multivariate marked processes with applications to martingale testing. J Multivar Anal 98:1321–1336
Escanciano JC (2010) Asymptotic distribution-free diagnostic tests for heteroskedastic time series models. Econom Theory 26:744–773
Hájek J, Šidák Z (1999) Theory of rank tests, 2nd edn. Academic, New York
Koul HL (2002) Weighted empirical processes in dynamic nonlinear models, 2nd edn. Lecture notes in statistics, vol 1666. Springer, New York
Koul HL, Stute W (1999) Nonparametric model checks for time series. Ann Stat 27:204–236
Lee S, Wei C-Z (1999) On residual empirical processes of stochastic regression models with applications to time series. Ann Stat 27:237–261
Lehmann EL, Romano JP (2005) Testing statistical hypotheses, 3rd edn. Springer texts in statistics. Springer, New York
Kulperger RJ (1985) On the residuals of autoregressive processes and polynomial regression. Stoch Process Their Appl 21:107–118
Engler E, Nielsen B (2009) The empirical process of autoregressive residuals. Econom J 12:367–381
Ngatchou-Wandji J, Laïb N (2008) Local power of a Crámer–von Mises type test for parametric autoregressive models of order one. Comput Math Appl 56:918–929
Serfling RJ (1968) Contributions to central limit theory for dependent variables. Ann Math Stat 39:1158–1175