Working Paper #10-17
Economic Series (07)
Abril 2010

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

A distribution-free transform of The residuals sample Autocorrelations with application To model checking*

Miguel A. Delgado
and
Carlos Velasco
Department of Economics
Universidad Carlos III de Madrid

April 2010

Abstract

We propose an asymptotically distribution-free transform of the sample autocorrelations of residuals in general parametric time series models, possibly non-linear in variables. The residuals autocorrelation function is the basic model checking tool in time series analysis, but it is useless when its distribution is incorrectly approximated because the effects of parameter estimation or of unnoticed higher order serial dependence have not been taken into account. The limiting distribution of residuals sample autocorrelations may be difficult to derive, particularly when the underlying innovations are not independent. However, the transformation we propose is easy to implement and the resulting transformed sample autocorrelations are asymptotically distributed as independent standard normals, providing an useful and intuitive device for model checking by taking

* Miguel A. Delgado and Carlos Velasco are professors, Universidad Carlos III de Madrid, Departamento de Economía, 28903 Getafe (Madrid), Spain. (E-mails: miguelangel.delgado@uc3m.es and carlos.velasco@uc3m.es.) Research funded by Spanish Plan Nacional de I+D+i grant number SEJ2007-62908/ECON.

over the role of the standard sample autocorrelations. We also discuss in detail alternatives to the classical Box-Pierce and Bartlett's $T_p - process$ tests, showing that our transform entails no efficiency loss under Gaussianity. The finite sample performance of the procedures is examined in the context of a Monte Carlo experiment for the two goodness-of-fit tests discussed in the article. The proposed methodology is applied to modeling the autocovariance structure of the well known chemical process temperature reading data already used for the illustration of other statistical procedures

Keywords: Residuals autocorrelation function; Asymptotically pivotal statistics; Nonlinear in variables models; Long memory; Higher order serial dependence; Recursive residuals; Model checking; Local alternatives.

1. INTRODUCTION

The sample autocorrelation function of residuals is an essential tool for time series model checking. In fact, the main proposals for testing lack of autocorrelation use statistics depending on the sample autocorrelation function; e.g. the parametric pseudo Lagrange Multiplier (PLM) tests, the nonparametric Bartlett's $T_p - process$ and $U_p - process$ based tests or Portmanteau-type tests, like the popular Box and Pierce (1970) proposal. The sample autocorrelations of *iid* data are asymptotically distributed as independent standard normals, but the *iid* assumption is often of little practical relevance for specification testing. Residuals sample autocorrelations, used in model checking, are obviously no *iid*. Box and Pierce (1970) and Durbin (1970) showed that sample autocorrelations of ARMA residuals are neither independent or identically distributed, even when the underlying innovations are *iid*. Other authors have considered residuals of more general models with *iid* innovations; e.g. Li (1992) and Hwang, Basawa and Reeves (1994). Even when the putative parametric specification correctly represents the autocorrelation structure of the data, it will unlikely be able to capture other higher order serial dependence features, e.g. conditional volatility. This is why the innovations of a time series model are not expected to be independent, though they are not autocorrelated when the specification is correct. The sample autocorrelations of no independent raw data are usually neither independent or identically distributed. See e.g. Hannan and Heyde (1972) and Romano and Thombs (1996). Recently, Francq, Roy and Zakoïan (2005) have derived the asymptotic distribution of sample autocorrelations of weak ARMA residuals, where innovations are not independent. The residuals sample autocorrelations suitably scaled can be used for testing lack of autocorrelation of the innovations. However, the scale depends on the model and estimator considered, as well as on the higher order dependence of innovations.

In this article, we propose an asymptotically distribution-free transform of the sample autocorrelations of residuals of general time series models, possibly nonlinear in variables and parameters, which can be directly applied to model checking taking over the role of the standard sample autocorrelations. In particular, we consider natural alternatives to Box and Pierce (1970) and Bartlett's $T_p - process$ type tests based on these transforms.

The discussion is in terms of a strictly stationary time series process $\{X_t\}_{t\in\mathbb{Z}}$, which takes values in $\mathbb{R}^k$, and of a parametric model with residuals

$$\varepsilon_{\theta t} = \varphi_\theta(L) U_\theta(X_t), \; t \in \mathbb{Z}, \tag{1}$$

indexed by the vector of parameters $\theta \in \Theta \subset \mathbb{R}^q$, where Θ is a parameter space restricting the functions φ_θ and U_θ, such that the process $\{\varepsilon_{\theta t}\}_{t\in\mathbb{Z}}$ is strictly stationary for each $\theta \in \Theta$. The functions $\varphi_\theta : \mathbb{C} \to \mathbb{C}$ and $U_\theta : \mathbb{R}^k \to \mathbb{R}$ are known and L denotes the lag-operator. Typically $\{U_\theta(X_t)\}_{t\in\mathbb{Z}}$ are residuals of a parametric model, possibly nonlinear in variables, relating two subsets of variables in X_t, i.e. a subvector of explained variables Y_t and a subvector of explanatory variables Z_t. The leading example is the linear model with $U_\theta(X_t) = Y_t - (1, Z_t')' \theta$. However, non-linear in variables models appear naturally when variables are transformed to get more functional flexibility, e.g. Box and Cox (1964).

The transfer function φ_θ specifies the linear serial dependence behaviour of the residuals. The identifiability restriction $\varphi_\theta(0) = 1$ is usually imposed. The most popular model is the ARMA(p_1, p_2) with

$$\varphi_\theta(z) = \frac{\Phi_\delta(z)}{\Xi_\eta(z)},\ z \in \mathbb{C},$$

such that Φ_δ and Ξ_η are the autoregressive and moving average polynomials with coefficients δ and η of orders p_1 and p_2, respectively. The function U_θ is usually not indexed by the parameters (δ, η), which are restricted in such a way that Ξ_η and Φ_δ have no common roots, all lying outside the unit circle. Long memory models are also of broad applicability, such as the ARFIMA(p_1, d, p_2) specification, where $d \in (-1/2, 1/2)$ is the long memory parameter. Our assumptions do not cover such a case because φ_θ is no longer summable, cf. Assumption 3 in the Appendix. However, when X_t is a linear process, the results of Delgado, Hidalgo and Velasco (2005) can be straightforwardly applied to justify the methods proposed in this paper. In Section 4, we evaluate the finite sample performance of test statistics both for short and long memory models.

The focus of our attention is the autocorrelation function of $\{\varepsilon_{\theta t}\}_{t\in\mathbb{Z}}$,

$$\rho_\theta(j) = \frac{\gamma_\theta(j)}{\gamma_\theta(0)},\ j \in \mathbb{Z},$$

where $\gamma_\theta(j) = Cov(\varepsilon_{\theta t}, \varepsilon_{\theta t-j})$, $j \in \mathbb{Z}$, is the corresponding autocovariance function. The model (1) is correctly specified when the null hypothesis

$$H_0 : \rho_{\theta_0}(j) = 0 \text{ for all } j \in \mathbb{Z} \setminus \{0\} \text{ and some } \theta_0 \in \Theta$$

is satisfied. Given observations $\{X_t\}_{t=1}^T$, ρ_θ is estimated by the sample autocorrelation function

$$\hat\rho_{T\theta}(j) = \frac{\hat\gamma_{T\theta}(j)}{\hat\gamma_{T\theta}(0)},\ j \in \mathbb{Z},$$

where

$$\hat\gamma_{T\theta}(j) = \frac{1}{T}\sum_{t=j+1}^{T}(\varepsilon_{\theta t} - \bar\varepsilon_{\theta T})(\varepsilon_{\theta t-j} - \bar\varepsilon_{\theta T}),\ j \in \mathbb{Z},$$

is the sample autocovariance function and $\bar\varepsilon_{\theta T} = T^{-1}\sum_{t=1}^T \varepsilon_{\theta t}$ is the residuals sample mean.

When $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ are *iid* for some $\theta_0 \in \Theta_0$, it is well known that $\left\{\sqrt{T}\hat\rho_{T\theta}(j)\right\}_{j=1}^m$ are asymptotically independent distributed as standard normals. This is still the case under martingale difference restrictions on higher powers of $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$. However, there are many other serial dependence circumstances where H_0 holds while the sample autocorrelations are not asymptotically *iid*. The asymptotic distribution of the sample autocorrelations of raw data have been derived by Hannan and Heyde (1972) assuming only that $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ is a MDS, while Romano and Thombs (1996) assume general strong mixing conditions.

Define the vector containing the first m sample residuals autocorrelations, $\hat{\boldsymbol{\rho}}_{T\theta}^{(m)} = (\hat{\rho}_{\theta}(1), \ldots, \hat{\rho}_{T\theta}(m))'$. Under H_0, but with $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ exhibiting general higher order serial dependence conditions,

$$\sqrt{T}\hat{\boldsymbol{\rho}}_{T\theta_0}^{(m)} \xrightarrow{d} N\left(0, A_{\theta_0}^{(m)}\right), \quad A_{\theta}^{(m)} = \left[\frac{a_{\theta}^{(i,j)}}{\gamma_{\theta}(0)^2}\right]_{i,j=1}^{m},$$

see e.g. Romano and Thombs (1996), where

$$a_{\theta}^{(i,j)} = \sum_{\ell=-\infty}^{\infty} \mathbb{E}\left[\varepsilon_{t,\theta}\varepsilon_{t+i,\theta}\varepsilon_{t+\ell,\theta}\varepsilon_{t+\ell+j,\theta}\right], \quad i,j = 1,\ldots,m. \tag{2}$$

The asymptotic distribution of the vector $\sqrt{T}\hat{\boldsymbol{\rho}}_{T\theta_0}^{(m)}$ can be approximated with the assistance of bootstrap techniques, as Romano and Thombs (1996) suggest, or using the asymptotic approximation, after suitable scaling by a consistent estimator of $A_{\theta_0}^{(m)}$. Such estimator requires to use smoothers, unless certain restrictions on the higher serial dependence of $\{\varepsilon_{\theta t}\}_{t\in\mathbb{Z}}$ are imposed. For instance, when $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ is a MDS, $a_{\theta_0}^{(i,j)} = \mathbb{E}\left[\varepsilon_{t\theta_0}^2\varepsilon_{t+i\theta_0}\varepsilon_{t+j\theta_0}\right]$, which can be estimated by its sample analog, without need to specify any bandwidth or lag number depending on the sample size. Assuming also that $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ follows a Gaussian GARCH process, then $a_{\theta_0}^{(i,j)} = 0$, $i \neq j$, which makes the estimation easier, see Lobato, Nankervis and Savin (2002).

Consider a positive definite matrix of statistics $\hat{A}_{T\theta}^{(m)}$, such that $\hat{A}_{T\theta_0}^{(m)} = A_{\theta_0}^{(m)} + o_p(1)$ under H_0. Also, consider the vector of scaled autocorrelations,

$$\tilde{\boldsymbol{\rho}}_{T\theta}^{(m)} = \left(\tilde{\rho}_{T\theta}^{(m)}(1), \ldots, \tilde{\rho}_{T\theta}^{(m)}(m)\right)' = \hat{A}_{T\theta}^{(m)-1/2}\hat{\boldsymbol{\rho}}_{T\theta}^{(m)}.$$

Thus, under H_0 and any of the previous regularity conditions, we obtain that $T^{1/2}\tilde{\boldsymbol{\rho}}_{T\theta_0}^{(m)} \xrightarrow{d} N_m(0, I_m)$.

In practice, a preliminary estimator of θ_0 is needed. We assume that an estimator $\hat{\theta}_T$ is available, such that when $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ are not autocorrelated,

$$\hat{\theta}_T = \theta_0 + O_p\left(T^{-1/2}\right), \tag{3}$$

and

$$\hat{A}^{(m)}_{T\hat{\theta}_T} = A^{(m)}_{\theta_0} + o_p(1). \tag{4}$$

In Lemma 1 in the Appendix B we prove that this is the case for the class of estimates proposed by Lobato, Nankervis and Savin (2002) under our regularity assumptions.

Next proposition provides an asymptotic expansion for $\sqrt{T}\tilde{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T}$, which implies that under H_0 and fairly general regularity conditions $\sqrt{T}\tilde{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T}$ converges to a vector of independent standard normals plus a stochastic drift, which depends on the unknown parameters θ_0, i.e. the specified model, and the particular estimation method. Define

$$\boldsymbol{\xi}^{(m)}_{\theta} = A^{(m)-1/2}_{\theta}\boldsymbol{\zeta}^{(m)}_{\theta},$$

with $\boldsymbol{\xi}^{(m)}_{\theta} = \left(\xi_\theta(1)', \ldots, \xi_\theta(m)'\right)'$ and $\boldsymbol{\zeta}^{(m)}_{\theta} = \left(\zeta_\theta(1)', \ldots, \zeta_\theta(m)'\right)'$, where ζ_θ is defined by

$$\frac{\partial}{\partial\theta'}\hat{\rho}_{T\theta}(j) \xrightarrow{p} \zeta_\theta(j) \text{ each } j \in \mathbb{Z}$$

under H_0.

Proposition 1 *Under H_0, (3), (4) and Assumptions 1-3 in the Appendix,*

$$\tilde{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T} = \tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0} + \boldsymbol{\xi}^{(m)}_{\theta_0}\left(\hat{\theta}_T - \theta_0\right) + o_p\left(T^{-1/2}\right). \tag{5}$$

The asymptotic distribution of $\sqrt{T}\tilde{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T}$ under H_0 can be derived from the asymptotic joint distribution of $\left\{\sqrt{T}\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}, \sqrt{T}\left(\hat{\theta}_T - \theta_0\right)\right\}$, as Li (1982) and Hwang, Basawa and Reeves (1994) have done for nonlinear models with *iid* innovations and by Francq, Roy and Zakoïan (2005) for weak ARMA residuals. Alternative models and estimators demand different derivations, which may be cumbersome in heavy nonlinear models, possibly exhibiting long-memory or U_θ nonlinear in variables and parameters. Rather than performing these derivations, we suggest to consider an asymptotically distribution-free transform of the residuals sample autocorrelation by means of least squares fits, which are asymptotically distributed as independent standard normals. The transformed sample autocorrelations are in fact the recursive residuals of the linear least squares projection of the sample autocorrelations against the model score that defines the estimation drift. Based on these transformed autocorrelations we propose Portmanteau and $T_p - process$ type tests with pivotal asymptotic distributions. In particular, we show that the test based on the sum of squares of the first s transformed autocorrelations is asymptotically equivalent to the LM test for AR(s) and MA(s) alternatives in a Gaussian framework.

The rest of the article is organized as follows. In Section 2, we introduce the autocorrelation transformation and discuss its asymptotic properties under general regularity conditions. The transformation is applied, in Section 3, to lack of autocorrelation testing of the underlying innovations. To this end, we introduce a class of test statistics based on weighted sums of the squared transformed sample autocorrelations. The asymptotic distribution of the tests in the direction of local alternatives, converging to the null at the $\sqrt{T}$ rate is derived. The finite sample performance of these tests is illustrated in Section 4 in the context of a Monte Carlo experiment. Section 5 presents an application to time series modeling of the well known Box and Jenkins (1976) chemical process temperature readings data (series C). Regularity conditions and mathematical proofs are contained in an Appendix, at the end of the article.

2. A DISTRIBUTION-FREE TRANSFORM OF THE SAMPLE AUTOCORRELATION FUNCTION WITH ESTIMATED PARAMETERS.

The transformation of the residuals autocorrelations proposed in this section resembles the recursive least squares residuals introduced by Brown, Durbin and Evans (1976) for CUSUM tests of parameter stability in the linear regression model with fixed regressors. Notice that the asymptotic expansion (5) can be interpreted as an (approximated) "linear regression" model with fixed regressors $\left\{\xi_{\theta_0}(j)\right\}_{j=1}^{m}$, where $\left\{\tilde{\rho}_{T\hat{\theta}_T}^{(m)}(j)\right\}_{j=1}^{m}$ are the dependent variables and $\left\{\tilde{\rho}_{T\theta_0}^{(m)}(j)\right\}_{j=1}^{m}$ the errors. The idea is to project $\left\{\tilde{\rho}^{(m)}_{T\hat{\theta}_T}(j)\right\}_{j=1}^{m}$ on $\{\xi_{\theta_0}(j)\}_{j=1}^{m}$ recursively so that the resulting residuals do not depend on $\hat{\theta}_T - \theta_0$ and, hence, are not affected by the parameter estimation effect.

Since the ξ_{θ_0} are not observable, we first discuss sample approximations to them. It can be showed under general conditions that

$$\left\| \frac{\partial}{\partial \theta'} \hat{\rho}_{T\theta_0}(j) - \frac{1}{\gamma_{\theta_0}(0)} \frac{\partial}{\partial \theta'} \hat{\gamma}_{T\theta_0}(j) \right\| \xrightarrow{p} 0,\ j \neq 0,$$

under H_0, since $\hat{\gamma}_{T\theta_0}(j) \rightarrow_p 0$ for all $j \neq 0$. So, standardization by $\hat{\gamma}_{T\theta_0}(0)$ in $\hat{\rho}_{T\theta_0}$ has no asymptotic effect on ζ_{θ_0} in the expansion (5). Then, we can compute

$$\hat{\boldsymbol{\xi}}^{(m)}_{T\hat{\theta}_T} = \hat{A}^{(m)-1/2}_{T\hat{\theta}_T} \hat{\boldsymbol{\zeta}}^{(m)}_{T\hat{\theta}_T}, \tag{6}$$

where $\hat{\boldsymbol{\xi}}^{(m)}_{T\theta} = \left(\hat{\xi}_{T\theta}(1)', \ldots, \hat{\xi}_{T\theta}(m)'\right)'$ and $\hat{\boldsymbol{\zeta}}^{(m)}_{T\theta} = \left(\hat{\zeta}_{T\theta}(1)', \ldots, \hat{\zeta}_{T\theta}(m)'\right)'$, with

$$\hat{\zeta}_{T\theta}(j) = \frac{1}{T\hat{\gamma}_{T\theta}(0)} \sum_{t=j+1}^{T} \dot{\varepsilon}_{\theta t}\left(\varepsilon_{\theta t-j} - \bar{\varepsilon}_{\theta T}\right) + \frac{1}{T\hat{\gamma}_{T\theta}(0)} \sum_{t=j+1}^{T} \dot{\varepsilon}_{\theta t-j}\left(\varepsilon_{\theta t} - \bar{\varepsilon}_{\theta T}\right)$$

and $\dot{\varepsilon}_{\theta t} = (\partial/\partial\theta')\,\varepsilon_{\theta t}$. In some circumstances, as for linear models and scalar X_t, where $\varepsilon_{\theta t} = \varphi_\theta(L)\,X_t$, it is straightforward to obtain closed, easy to compute, expressions for ζ_θ without further restrictions under H_0. Under these circumstances it is simpler to use $\zeta_{\hat{\theta}_T}$ rather than $\hat{\zeta}_{T\hat{\theta}_T}$ to compute $\hat{\boldsymbol{\zeta}}^{(m)}_{T\hat{\theta}_T}$ in (6).

Consider now the recursive least squares coefficients in the linear projection of $\{\tilde{\rho}_{T\theta}(j)\}_{j=1}^{m}$ on $\left\{\hat{\xi}_{T\theta}(j)\right\}_{j=1}^{m}$,

$$\tilde{\beta}^{(j)}_{T\theta} = \left(\sum_{\ell=j+1}^{m} \hat{\xi}_{T\theta}(\ell)'\,\hat{\xi}_{T\theta}(\ell)\right)^{-1} \sum_{\ell=j+1}^{m} \hat{\xi}_{T\theta}(\ell)'\,\tilde{\rho}^{(m)}_{T\theta}(\ell),\ j = 1, \ldots, m-q,$$

and the corresponding scaled residuals,

$$\bar{\rho}^{(m)}_{T\theta}(j) = \frac{\tilde{\rho}^{(m)}_{T\theta}(j) - \hat{\xi}_{T\theta}(j)\,\tilde{\beta}^{(j)}_{T\theta}}{\sqrt{1 + \hat{\xi}_{T\theta}(j)\left(\sum_{\ell=j+1}^{m} \hat{\xi}_{T\theta}(\ell)'\,\hat{\xi}_{T\theta}(\ell)\right)^{-1} \hat{\xi}_{T\theta}(j)'}},\ j = 1, \ldots, m-q.$$

We prove that, under H_0, $\bar{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T} = \left(\bar{\rho}^{(m)}_{T\hat{\theta}_T}(1), \ldots, \bar{\rho}^{(m)}_{T\hat{\theta}_T}(m-q)\right)'$ and $\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ are asymptotically equivalent, and $\sqrt{T}\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ is asymptotically distributed as a vector of independent standard normals, as we state in the following theorem.

Theorem 1 *Under $H_0, m > q$, Assumptions 1-4 in the Appendix and with $\hat{\theta}_T$ satisfying (3) and (4),*

$$\bar{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T} = \bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0} + o_p\left(T^{-1/2}\right) \tag{7}$$

and

$$\sqrt{T}\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0} \xrightarrow{d} N_{m-q}\left(0, I_{m-q}\right). \tag{8}$$

The theorem is proved reasoning as in the seminal paper by Brown, Durbin and Evans (1976). First, applying (5), $\tilde{\beta}^{(j)}_{T\hat{\theta}_T} = \tilde{\beta}^{(j)}_{T\theta_0} + \left(\hat{\theta}_T - \theta_0\right) + o_p\left(T^{-1/2}\right)$, which justifies (7). Second, the asymptotic independence of the $\sqrt{T}\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ components also follows applying standard arguments when dealing with recursive residuals. Notice also that $\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ can be interpreted as the martingale part of the discrete parameter empirical process $\tilde{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T}$, in the lines of the martingale transformation proposed by Khmaladze (1981) for the standard empirical process. This result forms a basis for implementing asymptotic specification tests of different nature, as is discussed in next section.

3. TESTING LACK OF AUTOCORRELATION WITH ESTIMATED PARAMETERS

We consider the class of tests for H_0 expressed as weighted sums of the squared transformed autocorrelations. That is, the test statistics have the form $\psi_{T\hat{\theta}_T}(\omega)$, with

$$\psi_{T\theta}(\omega) = T\sum_{j=1}^{m-q}\omega(j)\,\bar{\rho}^{(m)}_{T\theta}(j)^2,$$

where $\omega : \mathbb{N} \to \mathbb{R}^+$ is a squared summable weight function. It follows from Theorem 1 that, under H_0,

$$\psi_{T\hat{\theta}_T}(\omega) \xrightarrow{d} \sum_{j=1}^{m-q}\omega(j)\,Z_j^2,$$

where, henceforth, $\{Z_j\}_{j\in\mathbb{N}}$ are *iid* standard normals.

The power of the different tests indexed by alternative ω sequences can be discussed in terms of local alternatives of the form

$$H_{1T} : \rho_{\theta_0}(j) = \frac{r(j)}{\sqrt{T}} + \frac{\phi_T(j)}{T} \text{ for all } j = 0, 1, \ldots,$$

where we assume that $\hat{\theta}_T \to_p \theta_0$ under H_{1T} and r and ϕ_T are such that ρ_{θ_0} is a positive semi-definite sequence for all T. These local alternatives appear in a natural way by representing the autocorrelation structure of $\{\varepsilon_{\theta t}\}_{t\in\mathbb{Z}}$ according to the linear process

$$\varepsilon_{\theta t} = \Phi_{T\theta}(L)\, \upsilon_{\theta t}, \tag{9}$$

where $\{\upsilon_{\theta t}\}_{t\in\mathbb{Z}}$ are uncorrelated with higher order dependence characterized by $a_\theta^{(i,j)}$ defined in (2) and

$$\Phi_{T\theta}(z) = 1 + \sum_{j=1}^{\infty} \frac{\alpha_{T\theta}(j)}{\sqrt{T}} z^j,$$

with $\sum_{j=1}^{\infty} \alpha_{T\theta}(j)^2 < \infty$ for all θ and $\lim_{T\to\infty} \alpha_{T\theta_0}(j) = r(j)$. The function $\Phi_{T\theta}$ can be either parametric or nonparametric. For instance, it can be given by an ARMA model with parameters vanishing to zero at a rate $1/\sqrt{T}$ as the sample size T increases.

In order to describe the asymptotic distribution of $\bar{\boldsymbol{\rho}}_{T\hat{\theta}_T}^{(m)}$ under H_{1T} define first the projected and standardized vector of autocorrelation drifts $\bar{\mathbf{h}}_\theta^{(m)} = \left(\bar{h}_\theta^{(m)}(1), \ldots, \bar{h}_\theta^{(m)}(m-q)\right)'$ where

$$\bar{h}_\theta^{(m)}(j) = h_\theta^{(m)}(j) - \xi_\theta(j)' \left(\sum_{\ell=j+1}^{m} \xi_\theta(\ell)\, \xi_\theta(\ell)'\right)^{-1} \sum_{\ell=j+1}^{m} \xi_\theta(\ell)\, h_\theta^{(m)}(\ell), \tag{10}$$

$j = 1, 2, \ldots, m-q$ and

$$h_\theta^{(m)}(j) = \sum_{i=1}^{m} \left[A_\theta^{(m)-1/2}\right]_{(j,i)} r(i).$$

Theorem 2 *Under H_{1T}, $m > q$, Assumptions 1-4 in the Appendix and with $\hat{\theta}_T$ satisfying (3) and (4),*

$$\bar{\boldsymbol{\rho}}_{T\hat{\theta}_T}^{(m)} = \bar{\boldsymbol{\rho}}_{T\theta_0}^{(m)} + o_p\left(T^{-1/2}\right)$$

and

$$\sqrt{T}\bar{\boldsymbol{\rho}}_{T\theta_0}^{(m)} \xrightarrow{d} N_{m-q}\left(\bar{\mathbf{h}}_{\theta_0}^{(m)}, I_{m-q}\right).$$

4.1 Box-Pierce type tests

Consider the uniform weights $\omega(j) = 1_{\{j\le s\}}$, $1 \le s \le m-q$, for each $j \in \mathbb{N}$, which corresponds to the test statistic,

$$\bar{B}_{T\hat{\theta}_T}^{(m)}(s) = T \sum_{j=1}^{s} \bar{\rho}_{T\hat{\theta}_T}^{(m)}(j)^2,$$

leading to a transformed version of the popular Box and Pierce (1975) test statistic $\hat{B}_{T\hat{\theta}_T}$, with

$$\hat{B}_{T\theta}(s) = T \sum_{j=1}^{s} \hat{\rho}_{T\theta}(j)^2 .$$

Box and Pierce (1975) showed that, when $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ are *iid*, and s is increasing with T, $s = o\left(T^{-1/2}\right)$, $\hat{B}_{T\hat{\theta}_T}(s) \underset{asy}{\sim} \chi^2_{(s-q)}$. This test is unable to detect local alternatives like H_{1T}, but it can detect local alternatives of this form converging to the null at the rate $\sqrt[4]{s} \Big/ \sqrt{T}$, see Hong (1996). When s remains fixed, $\hat{B}_{T\hat{\theta}_T}(s)$ has a limiting null distribution depending on the unknown parameter vector θ_0 and other unknown features of the underlying data generating process.

On the other hand, the test statistic $\bar{B}^{(m)}_{T\hat{\theta}_T}(s)$ is asymptotically $\chi^2_{(s)}$ distributed and equivalent with increasing m to the Gaussian *LM* test statistic in the local parametric directions H_{1T} where $\Phi_{T\theta}$ in (9) is an autoregressive or moving average polynomial of order s and the innovations are *iid*, so that $A^{(m)}_{\theta_0} = I_m$. We state this result in the next Proposition. Let $\chi^2_{(n)}\left(\sum_{i=1}^n \lambda_i^2\right)$ denote a noncentered chi-squared random variable with n degrees of freedom and noncentrality parameter $\sum_{i=1}^n \lambda_i^2$; i.e. $\chi^2_{(n)}\left(\sum_{i=1}^n \lambda_i^2\right) \overset{d}{=} \sum_{j=1}^n (Z_i + \lambda_i)^2$.

Proposition 2 *Under the assumptions in Theorem 2, with $A^{(m)}_{\theta_0} = I_m$ for all m, the test based on $\bar{B}^{(m)}_{T\hat{\theta}_T}(s)$ is asymptotically equivalent to the Gaussian LM test of lack of autocorrelation up to order s, so that under H_{1T} with $r(j) = 0$ for $j > s$,*

$$\bar{B}^{(m)}_{T\hat{\theta}_T}(s) \xrightarrow{d} \chi^2_{(s)}\left(\sum_{j=1}^{s} \bar{r}^{(\infty)}_{\theta_0}(j)^2\right),$$

as $m \to \infty$, where $\bar{r}^{(\infty)}_{\theta_0}(j)$ is defined as in (10) with $h^{(m)}_{\theta_0}(j) = r(j)$ for all m.

Therefore, in the context of *iid* innovations, Box-Pierce tests based on $\bar{B}_{T\hat{\theta}_T}(s)$ for testing compound hypotheses have the same interpretation than the standard tests based on $\hat{B}_{T\theta_0}(s)$ for testing the simple hypothesis of lack of autocorrelations of the true innovations. Then, under Gaussianity, the tests $\bar{B}_{T\hat{\theta}_T}(s)$ are optimal for testing lack of serial correlation of residuals up to a finite order, without need to resort to fully efficient maximum likelihood estimates (*MLE*) of θ_0, but just using estimators satisfying (3). This also points out that the procedure applied for eliminating the estimation effect in the sample autocorrelations $\tilde{\boldsymbol{\rho}}_{T\hat{\theta}_T}^{(m)}$ does not neglect any important information asymptotically.

On the other hand, while the classical Box-Pierce test is also a LM test for simple hypotheses of order s, notice that after parameter estimation and under H_{1T} it satisfies

$$\frac{\hat{B}_{T\hat{\theta}_T}(s) - s}{\sqrt{2s}} \to_d N(0,1) \quad \text{when} \quad \frac{1}{s} + \frac{s}{T} \to 0.$$

Therefore, this test is unable to detect nonparametric local alternatives in the class H_{1T}, cf. Hong (1996).

4.2 $\mathbf{T}_p-$process type tests

The sequence of weights $\omega(j) = 1/j^2$ leads to test statistics

$$\bar{T}_{T\theta}^{(m)} = T \sum_{j=1}^{m-q} \frac{\bar{\rho}_{T\theta}^{(m)}(j)^2}{j^2},$$

which resembles the spectral representation of the classical $T_p - process$ test statistic based on the Cramér-von Mises criterion, i.e.

$$\hat{T}_{T\theta} = T \sum_{j=1}^{T-1} \frac{\hat{\rho}_{T\theta}(j)^2}{j^2},$$

see e.g. Anderson (1993).

Assuming that $\{\varepsilon_{\theta_0 t}\}_{t\in\mathbb{Z}}$ are iid, so $A_{\theta_0}^{(m)} = I_m$ is known, and allowing m to diverge to infinity with T, but at a slower rate, both $\bar{T}_{T\hat{\theta}_T}^{(m)}$ and the unfeasible $\hat{T}_{T\theta_0}$ are asymptotically distributed as $\sum_{j=1}^{\infty} Z_j^2 \big/ j^2$ under H_0. The next result describes their limiting distribution under H_{1T}.

Proposition 3 *Under the assumptions of Theorem 2, with $A_{\theta_0}^{(m)} = I_m$ for all m, and H_{1T},*

$$\bar{T}_{T\hat{\theta}_T}^{(m)} \xrightarrow{d} \sum_{j=1}^{\infty} \frac{\left(Z_j + \bar{r}_{\theta_0}^{(\infty)}(j)\right)^2}{j^2},$$

as $m \to \infty$, where $\bar{r}_{\theta_0}^{(\infty)}(j)$ is defined as in (10) with $h_{\theta_0}^{(m)}(j) = r(j)$ for all m and

$$\hat{T}_{T\theta_0} \xrightarrow{d} \sum_{j=1}^{\infty} \frac{(Z_j + r(j))^2}{j^2}.$$

However, it is not possible to perform general power comparisons among $\bar{T}^{(m)}_{T\hat{\theta}_T}$ and $\hat{T}_{T\theta_0}$ because the drifts, apart from the alternative hypothesis, depend on both the weighting function and the assumed model under H_0.

5. MONTE CARLO SIMULATIONS

In this section, we compare the percentage of rejections under H_0 and H_1 of alternative tests based on residual sample autocorrelations. The comparison is made in the context of ARFIMA designs with innovations

$$\varepsilon_{\theta_0 t} = u_t \left(1 + \alpha_1 \varepsilon^2_{\theta_0 t-1}\right)^{1/2},$$

where $\alpha_1 \in \{0, 0.4, 0.8\}$. We consider sample sizes $T = 100$ and 400 and $50,000$ replications in each experiment. Parameters are estimated using Whittle's likelihood method, see e.g. Velasco and Robinson (2000). We consider three null models: AR(1) with $\zeta_{\theta_0}(j) = -\delta_{10}^{j-1}$, MA(1) with $\zeta_{\theta_0}(j) = \eta_{10}^{j-1}$ and ARFIMA $(0, d, 0)$ with $\zeta_{\theta_0}(j) = -j^{-1}$.

The first purpose of the simulations consists in comparing the classical Box-Pierce (B-P) test, $\hat{B}_{T\hat{\theta}_T}(s)$, and our alternative test based on $\bar{B}_{T\hat{\theta}_T}(s)$, which use critical values from a chi squared distribution with $s-q$ and s degrees of freedom, respectively. Also, we compute the asymptotically pivotal $T_p - \mathit{process}$ test using the Cramer von Mises criteria (CvM) proposed by Delgado, Hidalgo and Velasco (2005), which is only valid when $\hat{\rho}_{T\theta_0}$ are asymptotically iid, and its alternative, $\bar{T}^{(m)}_{T\hat{\theta}_T}$, which is asymptotically pivotal after appropriate standardization, even when the innovations exhibit higher order serial dependence. We have computed $\bar{T}^{(m)}_{T\hat{\theta}_T}$ using large values m, $m = 20$ for $T = 100$ and $m = 40$ when $T = 400$. Notice that due to the weights $1/j^2$, the test statistic is numerically not very sensitive to the choice of fairly large m's.

Tables 1 to 3 offer the percentage of rejections under H_0. Tables 1 and 2 report results when the innovations are iid $(\alpha_1 = 0)$ and serially dependent according to an ARCH process with $\alpha_1 = 0.8$, respectively. In both cases, the estimator of $A^{(m)}_{\theta_0}$ uses information on its true structure. Thus, in Table 1, $\hat{A}^{(m)}_{T\theta_T} = I_m$, and in Table 2, $\hat{A}^{(m)}_{T\hat{\theta}_T} = diag\left\{\hat{a}^{(1,1)}_{T\hat{\theta}_T}, \ldots, \hat{a}^{(m,m)}_{T\hat{\theta}_T}\right\} / \hat{\gamma}_{T\hat{\theta}_T}(0)^2$ with $\hat{a}^{(j,j)}_{T\theta} = T^{-1} \sum_{t=1+j}^{T} \varepsilon^2_{t\theta} \varepsilon^2_{t-j\theta}$. The effect of general estimation of $A_{\theta_0}^{(m)}$ is examined in Table 3, where we also compare our new tests with Francq, Roy and Zakoïan (2005) proposal, for which we use the same unrestricted $A_{\theta_0}^{(m)}$ estimates proposed by Lobato, Nankervis and Savin (2002).

We observe in Table 1, under *iid* innovations, that the classical B-P test shows size distortions when s is either too small or too large, but the size accuracy is fine when $s = T^{1/2}$. As expected, the Type I error of the classical B-P test is out of control in Table 2, with ARCH innovations and $A_{\theta_0}^{(m)}$ diagonal, but not equal to the identity matrix. However, the new B-P test exhibits a remarkable size accuracy in the two tables for all s considered. Obviously, it performs better in Table 1 where full information on $A_{\theta_0}^{(m)}$ is used and, hence, it is not estimated.

Interestingly, the CvM test of Delgado, Hidalgo and Velasco (2005) is outperformed in Table 1 by the new alternative based on $\bar{T}_{T\hat{\theta}_T}^{(m)}$. As expected, the CvM test of DHV has the Type I error uncontrolled in Table 2 with ARCH innovations.

Notice that Assumption 4 is not satisfied for the AR(1) and MA(1) models when the parameters are set to zero. However, the performance of the new test statistics in these cases is very good. In fact, the percentage of rejections of the new tests under the null is very similar for all parameter values.

In Table 3 we examine the effect of using unrestricted estimators of $A_{\theta_0}^{(m)}$ under *iid* and ARCH innovations on the size accuracy of the new tests. As expected, the simulated size is worse than in Tables 1 and 2 due to the unnecessary randomness introduced. We also report in this table results for the $\tilde{Q}_s$ test proposed by Francq, Roy and Zakoïan (2005) for ARMA models, which is also a portmanteau test where least squares residual sample autocorrelations are scaled by a consistent variance and covariance matrix estimate derived from the joint distribution of the least squares parameter estimator and the sample autocorrelations of the innovations. Notice that in these simulations we use the same design as Francq, Roy and Zakoïan (2005) ($\alpha_1 = 0.4$) and the same choices of s. The asymptotic variance and covariance matrix of residual autocorrelations is singular when the parameter is set to zero in the MA(1) and AR(1) specifications, see Francq, Roy and Zakoïan (2005), Remark 2.

$\boxed{\text{TABLES 1, 2 \& 3 ABOUT HERE}}$

In Figures 1 and 2 we show graphically the effect of the choice of s in the B-P tests for both types of innovations in an AR(1) model with $\delta_0 = 0.8$, $\alpha_1 = \{0.0, 0.8\}$ and $T = 100$. For *iid* innovations ($\alpha_1 = 0$) the simulations confirm that the size accuracy of the test based on projected autocorrelations with $\hat{A}^{(m)}_{T\hat{\theta}_T} = I_m$ (B-P-new) is very good for small and moderate values of s, while the proportion of rejections of the classical B-P test increases monotonically with s. In the conditional heteroskedastic situation ($\alpha_1 = 0.8$), both standardizations exploiting the MDS restriction (B-P-new MD and B-P-new diag, the last one imposing also diagonality of $A^{(m)}_{\theta_0}$) perform in a similar fashion, whereas no standardized statistics cannot account for the higher order dependence in the data.

$\boxed{\text{FIGURES 1 \& 2 ABOUT HERE}}$

Table 4 reports the percentage of rejections under the alternative hypothesis for the following specifications of the null and alternative models:

a) $H_0 : AR\,(1)$ vs $H_1 : ARMA\,(1,1)\,.$

b) $H_0 : MA\,(1)$ vs $H_1 : ARMA\,(1,1)\,.$

c) $H_0 : ARFIMA\,(0,d,0)$ vs $H_1 : ARFIMA\,(1,d,0)\,.$

d) $H_0 : AR\,(1)$ vs $H_1 : ARFIMA\,(1,d,0)\,.$

Innovations are *iid* (Table 4) and ARCH (Table 5), and in the first case we impose $\hat{A}^{(m)}_{T\hat{\theta}_T} = I_m$ and compare our tests to the classical B-P test, while in the second case we leave $\hat{A}^{(m)}_{T\hat{\theta}_T}$ completely unrestricted and compare to the $\tilde{Q}_s$ test. It is confirmed that the classical B-P test detects better the alternatives the smaller s is. There is a clear trade off between size accuracy and power for the B-P and $\tilde{Q}_s$ tests. Our new tests exhibit good power performance for all the s considered. This good performance is due to the ability of the new tests of considering small s values with the Type I error under control. The $\bar{T}^{(m)}_{T\hat{\theta}_T}$ test reports better power than Delgado, Hidalgo and Velasco (2005) test and seems well indicated for detecting long memory alternatives.

TABLES 4 & 5 ABOUT HERE

6. A REAL DATA EXAMPLE

In this section we analyze the specification of the well known chemical process temperature readings (series C) from Box and Jenkins (1976), see also Beran (1995), using the transformed residuals autocorrelations proposed. Beran (1995) and Velasco and Robinson (2000) estimate a fractional integration parameter d, rather than fitting an ARIMA model with a unit root as Box and Jenkins suggested. We also work with the increments of the series, but allow for fractional integration in some specifications, all fitted using Whittle estimation. For checking the fit of every model we use the Box-Pierce based on transformed residuals autocorrelations, $\bar{B}^{(m)}_{T\hat{\theta}_T}(s)$, for $s = 1, 2, 3, 5$ and for the original Box and Pierce (1970) test, $\hat{B}_{T\hat{\theta}_T}(s)$, for $s = 5, 10, 20, 30$, which includes all the usual choices of the range of lags in similar applications given that $T = 226$. We also report the Cramér-von Mises (CvM) test $\bar{T}^{(m)}_{T\theta}$ based on transformed residuals autocorrelations and the asymptotically distribution-free CvM test proposed by Delgado, Hidalgo and Velasco (2005) based on a martingale transformation of the $T_p - process$. Both have similar asymptotic (pivotal) distributions, but the latter is based on Brownian motion rather than a Brownian bridge. The value of m is fixed to $\lfloor T/10 \rfloor + q$, the results not being very sensitive to this choice. We only report the analysis with $A^{(m)}_{\theta_0} = I_m$ for easier comparison with non transformed autocorrelations. We finally provide BIC values for the models considered and the estimate of d with its standard error for $ARFIMA$ models.

We report results for all models with up to two short memory (AR or MA) parameters, see Table 6. All models with only one short run parameter (apart from the memory parameter d) are strongly rejected by the the CvM type tests and by the Portmanteau $\bar{B}^{(m)}_{T\hat{\theta}_T}(s)$ test for all lags $s = 1, \ldots, 5$. However, the Box-Pierce test can only reject the too simplistic pure fractional specification for the smallest $s = 5$, but not for the customary $s = 10, 20$. In order to test Box and Jenkins' specification of an exact difference, we fit ARIMA models with one and two parameters. Despite having favorable BIC values compared with long memory alternative specifications, all ARIMA models are clearly rejected by tests based on transformed residuals autocorrelations, but the usual Box-Pierce test only provides strong evidence against the ARIMA$(0,1,1)$ and $(2,1,0)$ models.

TABLE 6 ABOUT HERE

We now consider the analysis of individual residuals autocorrelations for lags up to 20. Recall that transformed autocorrelations can be compared with usual $\pm 2/\sqrt{T}$ confidence bands, as when working with raw data, but recall that these confidence bands are inconsistent when parameters are estimated. In Figures 3 and 4, we have plotted the autocorrelograms of residuals, both original and transformed ones, for ARFIMA$(1,d,0)$ and ARFIMA$(0,d,1)$ models, respectively. Again, these specifications were rejected clearly by tests based on transformed autocorrelations, $\bar{\rho}^{(m)}_{T\hat{\theta}_T}$, but diagnosis based on the untransformed autocorrelations, $\hat{\rho}_{T\hat{\theta}_T}$, using an incorrect asymptotic approximation, are unable to reject these specifications. In these plots we can easily identify the source of these rejections, since the transformed autocorrelations provide evidence on serial correlation of the underlying innovations from the very first lag onwards, and can be compared to a uniform benchmark based on their asymptotic *iid* standard normal distribution.

FIGURES 3 AND 4 ABOUT HERE

APPENDIX A: PROOFS AND AUXILIARY RESULTS

In this Appendix we present the assumptions sufficient for the proofs of our results and some auxiliary results that can be of independent interest. First we introduce some notation. Given the model $\varepsilon_{\theta t} = \varphi_\theta (L) U_\theta (\mathbf{X}_t)$ so $\varepsilon_t = \varepsilon_{\theta_0 t}$ we set

$$\dot{\varepsilon}_{\theta t} = \frac{\partial}{\partial \theta'} \varepsilon_{\theta t} = \Big(\varphi_\theta (L) \dot{U}_\theta (\mathbf{X}_t) + \dot{\varphi}_\theta (L) U_\theta (\mathbf{X}_t) \Big)'$$

where $\dot{U}_\theta (x) = (\partial / \partial \theta) U_\theta (x)$ and $\dot{\varphi}_\theta (z) = (\partial / \partial \theta) \varphi_\theta (z)$, and

$$\ddot{\varepsilon}_{\theta t} = \frac{\partial^2}{\partial \theta \partial \theta'} \varepsilon_{\theta t} = \Big(\varphi_\theta (L) \ddot{U}_\theta (\mathbf{X}_t) + \dot{U}_\theta (\mathbf{X}_t) \dot{\varphi}_\theta (L)' + \dot{\varphi}_\theta (L) \dot{U}_\theta (\mathbf{X}_t)' + \ddot{\varphi}_\theta (L) U_\theta (\mathbf{X}_t) \Big),$$

where $\ddot{U}_\theta(x) = (\partial^2/\partial\theta\partial\theta') U_\theta(x)$ and $\ddot{\varphi}_\theta(z) = (\partial^2/\partial\theta\partial\theta') \varphi_\theta(z)$. Similar definitions apply for $\dot{\hat{\boldsymbol{\rho}}}^{(m)}_{T\theta} = (\partial/\partial\theta') \hat{\boldsymbol{\rho}}^{(m)}_{T\theta}$ and $\ddot{\rho}^{(m)}_{T\theta}(j) = (\partial^2/\partial\theta\partial\theta') \rho^{(m)}_{T\theta}(j)$.

Assumption 1 *$(\mathbf{X}'_t, \varepsilon_t)'$ is strictly stationary, ε_t is zero mean, $E\left[\varepsilon_t^{4+2\delta}\right] < \infty$ for some $\delta > 0$ and $(\mathbf{X}'_t, \varepsilon_t)'$ is strong mixing with coefficients α_j satisfying $\sum_{j=1}^{\infty} \alpha_j^{\delta/(2+\delta)} < \infty$, where*

$$\alpha_j = \sup_{A,B} |\Pr(AB) - \Pr(A)\Pr(B)|$$

and A and B vary over events in the σ fields generated by $\left\{(\mathbf{X}'_t, \varepsilon_t)',\ t \leq 0\right\}$ and $\left\{(\mathbf{X}'_t, \varepsilon_t)',\ t \geq j\right\}$.

Assumption 2 *$U_\theta(x)$ is twice differentiable in θ for each x and $|U_\theta(x)| + \left\|\dot{U}_\theta(x)\right\| + \left\|\ddot{U}_\theta(x)\right\| \leq U_*(x)$, where $E|U_*(\mathbf{X}_t)|^{4+2\delta} < \infty$ for some $\delta > 0$.*

Assumption 3 *$\varphi_\theta(z)$ is twice differentiable in $\theta, \varphi_\theta(0) = 1$ and the coefficients in the expansions*

$$\varphi_\theta(z) = \sum_{j=0}^{\infty} \varphi_{\theta,j} z^j,\ \ \dot{\varphi}_\theta(z) = \sum_{j=1}^{\infty} \dot{\varphi}_{\theta,j} z^j \ \text{ and } \ddot{\varphi}_\theta(z) = \sum_{j=1}^{\infty} \ddot{\varphi}_{\theta,j} z^j$$

satisfy $|\varphi_{\theta,j}| + \left\|\dot{\varphi}_{\theta,j}\right\| + \left\|\ddot{\varphi}_{\theta,j}\right\| \leq \phi_j$, uniformly for $\theta \in \Theta$, with $\sum_{j=0}^{\infty} \phi_j < \infty$.

Assumption 4 *For some $m > q$,*

$$\sum_{j=m-q+1}^{m} \boldsymbol{\xi}_{\theta_0}(j) \boldsymbol{\xi}_{\theta_0}(j)' > 0.$$

Remark. This type of assumption must always be satisfied when using recursive residuals in different contexts and it is more restrictive than the absence of multicollineality assumption when applying ordinary least squares. See e.g. Brown, Durbin and Evans (1976), Khamaladze (1981) or Delgado, Hidalgo and Velasco (2005) in different contexts. The assumption is not satisfied in some situations where the asymptotic variance and covariance matrix of residuals sample autocorrelations is singular. It may happen, for instance, where fitting an AR(1) to a strong white noise, as Franq, Roy and Zakoïan (2005) point out in their Remark 2. We have considered this situation in our simulations, when in AR(1) and MA(1) models the true parameters are set to zero, not satisfying Assumption 4. However, $\bar{B}_{T\hat{\theta}_T}(s)$ exhibits an excellent level accuracy in small samples in all occasions for the two sample sizes considered. The assumption could be relaxed by using generalized inverses when computing the recursive residuals, as proposed by Tsigroshvili (1998) in the related problem of constructing chi-squared tests using innovation martingales in the classical goodness-of-fit problem. Duchesne and Franq (2008) suggested also to construct Portmanteau-tests using generalized inverses of the asymptotic variance and covariance matrix of the residuals sample autocorrelations. This extension to our case is beyond the scope of this article.

Proof of Proposition 1. The statement follows from

$$\hat{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T} = \hat{\boldsymbol{\rho}}^{(m)}_{T\theta_0} + \boldsymbol{\zeta}^{(m)}_{\theta_0}\left(\hat{\theta}_T - \theta_0\right) + o_p\left(T^{-1/2}\right), \tag{11}$$

where $\boldsymbol{\zeta}^{(m)}_{\theta} = p\lim_{T\to\infty}\left(\partial/\partial\theta'\right)\hat{\boldsymbol{\rho}}^{(m)}_{T\theta}$, and (4) because $\hat{\boldsymbol{\rho}}^{(m)}_{T\theta_0} = O_p\left(T^{-1/2}\right)$ under H_0 or H_{1T}. We assume without loss of generality that $E\left[\varepsilon_t^2\right] = 1$ to prove (11). Now write $\hat{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T} - \hat{\boldsymbol{\rho}}^{(m)}_{T\theta_0} = \dot{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\left(\hat{\theta}_T - \theta_0\right) + D_T$, where each element of the vector D_T is

$$D_T(j) = \left(\hat{\theta}_T - \theta_0\right)' \ddot{\rho}^{(m)}_{T\theta^*_{T,j}}(j)\left(\hat{\theta}_T - \theta_0\right)$$

and $\theta^*_{T,j}$ are such that $\left\|\theta^*_{T,j} - \theta_0\right\| \leq \left\|\hat{\theta}_T - \theta_0\right\|$. Then for $j = 1, \ldots, m$,

$$\frac{\partial}{\partial\theta'}\hat{\rho}_{T\theta}(j) = \frac{\frac{\partial}{\partial\theta'}\hat{\gamma}_{T\theta}(j)}{\hat{\gamma}_{T\theta}(0)} - \frac{\hat{\gamma}_{T\theta}(j)}{\hat{\gamma}_{T\theta}(0)}\frac{\frac{\partial}{\partial\theta'}\hat{\gamma}_{T\theta}(0)}{\hat{\gamma}_{T\theta}(0)}.$$

The mean correction in $\hat{\gamma}_{T\theta_0}(j)$ has no asymptotic effect, since $\hat{\gamma}_{T\theta_0}(j) = \gamma_{T\theta_0}(j) + O_p\left(T^{-1}\right)$, where $\gamma_{T\theta}(j) = T^{-1}\sum_{t=j+1}^{T}\varepsilon_{\theta t}\varepsilon_{\theta t-j}$, $j \in \mathbb{Z}$, because $\bar{\varepsilon}_{\theta_0 T} = O_p\left(T^{-1/2}\right)$ under Assumption 1. Next, using that $\hat{\gamma}_{T\theta_0}(j) = \gamma_{\theta_0}(j) + o_p(1)$ (in particular $\gamma_{\theta_0}(0) = 1$ and $\gamma_{\theta_0}(j) = 0$ for $j \neq 0$ under H_0) and that $\frac{\partial}{\partial\theta'}\gamma_{T\theta_0}(0) = O_p(1)$ under Assumptions 1-3, as we now show, we conclude that the normalization of $\hat{\boldsymbol{\rho}}^{(m)}_{T\hat{\theta}_T}$ has no asymptotic effect under H_0, so that

$$\frac{\partial}{\partial\theta'}\hat{\rho}_{T\theta_0}(j) = \frac{\partial}{\partial\theta'}\gamma_{T\theta_0}(j) + o_p(1).$$

Write now

$$\frac{\partial}{\partial\theta'}\gamma_{T\theta_0}(j) = \frac{1}{T}\sum_{t=j+1}^{T}\dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j} + \frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{\theta_0 t}\dot{\varepsilon}_{\theta_0 t-j} := A_{T,1}(j) + A_{T,2}(j).$$

Setting $\boldsymbol{\zeta}_{\theta_0}(j) = \boldsymbol{\zeta}^{(1)}_{\theta_0}(j) + \boldsymbol{\zeta}^{(2)}_{\theta_0}(j)$ where $\boldsymbol{\zeta}^{(i)}_{\theta_0}(j) := \lim_{T\to\infty} E\left[A_{T,i}(j)\right],$ we wish to show that $A_{T,i}(j) = \boldsymbol{\zeta}^{(i)}_{\theta_0}(j) + o_p(1)$, $i = 1, 2,$ $j = 1, 2, \ldots$. We first show that

$$E\left\|A_{T,1}(j) - E\left[A_{T,1}(j)\right]\right\|^2 = \frac{1}{T^2}\sum_{t=1+j}^{T}\sum_{r=1+j}^{T} E\left[e(t,t-j)' e(r,r-j)\right],$$

is $o(1)$, where $e(t,t-j)' = \varepsilon_{\theta_0 t}\dot{\varepsilon}_{\theta_0 t-j} - E\left[\varepsilon_{\theta_0 t}\dot{\varepsilon}_{\theta_0 t-j}\right]$ and we omit dependence on θ_0 in the notation. Then, for some $n > 0$ fixed with T, $E\left\|A_{T,1}(j) - E\left[A_{T,1}(j)\right]\right\|^2$ is

$$\frac{1}{T^2}\sum_{t=1+j}^{T} E\left[e(t,t-j)' e(t,t-j)\right] + \frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{t-n-j\le r<t}^{T} E\left[e(t,t-j)' e(r,r-j)\right] \tag{12}$$
$$+\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{1+j\le r<t-n-j}^{T} E\left[e(t,t-j)' e(r,r-j)\right].$$

The first two terms of (12) are $O\left(T^{-1}\right) = o(1)$ since involve at most $T+n$ elements with bounded absolute expectation because by Assumptions 1-3 and Minkowski and Hölder inequalities,

$$\begin{aligned} E\left\|\dot{\varepsilon}_{\theta_0 t}\right\|^4 &\le E\left\|\dot{\varphi}_\theta(L)\, U_\theta(\mathbf{X}_t)\right\|^4 + E\left\|\varphi_\theta(L)\,\dot{U}_\theta(\mathbf{X}_t)\right\|^4 \\ &\le 2\left(\sum_{j=1}^{\infty}|\phi_j|\right)^4 E|U_*(\mathbf{X}_t)|^4 < \infty. \end{aligned} \tag{13}$$

Now write $\dot{\varepsilon}_{\theta t} = \dot{\varepsilon}^{(0,n)}_{\theta t} + \dot{\varepsilon}^{(n+1,\infty)}_{\theta t}$, $\dot{\varepsilon}^{(r,s)}_{\theta t} = \sum_{j=r}^{s}\left(\varphi_{\theta,j}\dot{U}_\theta(\mathbf{X}_{t-j}) + \dot{\varphi}_{\theta,j}U_\theta(\mathbf{X}_{t-j})\right)$ and $e^s_r(t,t-j) = \varepsilon_{\theta_0 t}\dot{\varepsilon}^{(r,s)}_{\theta_0 t-j}$. Then $e^n_0(t,t-j)$ is mixing with mixing coefficients $\beta_k \le \alpha_{k+j+n}$. The third term in (12) is then equal to

$$\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{r<t-n-j}^{T} E\left[e^n_0(t,t-j)' e(r,r-j)\right] \tag{14}$$
$$+\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{r<t-n-j}^{T} E\left[e^\infty_{n+1}(t,t-j)' e_j(r,r-j)\right].$$

Using Assumptions 2 and 3, the first term in (14) is bounded in absolute value by

$$\frac{C}{T^2}\left(E\left\|e^n_0(t,t-j)\right\|^{2+\delta} E\left\|e(r,r-j)\right\|^{2+\delta}\right)^{\frac{1}{2+\delta}}\sum_{t=1+j}^{T}\sum_{r<t-n-j}^{T}\alpha^{\delta/(2+\delta)}_{t-n-j-r} = O\left(T^{-1}\right) = o(1)$$

by Roussas and Ioannidies (1987) and Cauchy-Schwarz inequality.

Using again Assumptions 2 and 3, $\left|E\left[e_{n+1}^{\infty}(t,t-j)' e(r,r-j)\right]\right|$ can be made arbitrarily small choosing n large enough since

$$E\left|\sum_{j=n+1}^{\infty}\left\{\varphi_{\theta_0,j}\dot{U}_{\theta_0}(\mathbf{X}_{t-j})+\dot{\varphi}_{\theta_0,j}U_{\theta_0}(\mathbf{X}_{t-j})\right\}'\varepsilon_{\theta_0 t}e(r,r-j)\right|=O\left(\sum_{j=n+1}^{\infty}|\phi_j|\right),$$

and

$$E\left\|\sum_{j=n+1}^{\infty}\left\{\varphi_{\theta_0,j}\dot{U}_{\theta_0}(\mathbf{X}_{t-j})+\dot{\varphi}_{\theta_0,j}U_{\theta_0}(\mathbf{X}_{t-j})\right\}\varepsilon_{\theta_0 t}\right\|=O\left(\sum_{j=n+1}^{\infty}|\phi_j|\right),$$

because of the same reasoning as for (13). Then we conclude that the second term in (14) and the third term in (12) are $o_p(1)$.

On the other hand $\boldsymbol{\zeta}_{\theta_0}^{(2)}(j)$ is $\lim_{T\to\infty}E\left[A_{T,2}(j)\right]=E\left[\dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j}\right]$, which is different from zero if $\varphi_\theta(L)$ contains lags and/or if $U_\theta(\mathbf{X}_t)$ contains lagged non strictly exogenous explanatory variables. Then,

$$E\left\|A_{T,2}(j)-E\left[A_{T,2}(j)\right]\right\|^2=\frac{1}{T^2}\sum_{t=1+j}^{T}\sum_{r=1+j}^{T}E\left[e(t-j,t)' e(r-j,r)\right],$$

and for some $n>m$ fixed with T, this is

$$\frac{1}{T^2}\sum_{t=1+j}^{T}E\left[e(t-j,t)' e(t-j,t)\right]+\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{t-n\le r<t}^{T}E\left[e(t-j,t)' e(r-j,r)\right]$$
$$+\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{r<t-n}^{T}E\left[e(t-j,t)' e(r-j,r)\right]. \tag{15}$$

The first two terms are $O\left(T^{-1}\right)$ since involve at most $T+n$ elements with bounded absolute expectation by Assumptions 1-3. Writing $e(t-j,t)=e_0^n(t-j,t)+e_{n+1}^{\infty}(t-j,t)$, the third term of (15) is equal to

$$\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{r<t-n}^{T}E\left[e_0^n(t-j,t)' e(r-j,r)\right]+\frac{2}{T^2}\sum_{t=1+j}^{T}\sum_{r<t-n}^{T}E\left[e_{n+1}^{\infty}(t-j,t)' e(r-j,r)\right], \tag{16}$$

so that $e^{(0,n)}(t-j,t)$ is mixing, with mixing coefficients $\beta_k\le\alpha_{k-\max\{j,n-j\}}$. The first term in (16) is $o(1)$ because it is bounded in absolute value by

$$\frac{C}{T^2}\left(E\left\|e_0^n(t-j,t)\right\|^{2+\delta}E\left\|e(t-j,t)\right\|^{2+\delta}\right)^{\frac{1}{2+\delta}}\sum_{t=1+j}^{T}\sum_{1+j\le r<t-n-j}^{T}\alpha_{t-n-r}^{\delta/(2+\delta)}=O\left(T^{-1}\right)$$

by Roussas and Ioannidies (1987) and Assumption 1.

Using Assumption 3, $\left|E\left[e_{n+1}^{\infty}(t-j,t)' e(r-j,r)\right]\right|$ in (16) can be made arbitrarily small choosing $n>m$ large enough since by Minkowski inequality and Assumptions 1 and 3

$$E\left\|\sum_{j=n+1}^{\infty}\left\{\dot{\varphi}_{\theta_0,j}U_{\theta_0}\left(\mathbf{X}_{t-j}\right)+\varphi_{\theta_0,j}\dot{U}_{\theta_0}\left(\mathbf{X}_{t-j}\right)\right\}'\varepsilon_{\theta_0 t-j}e\left(r-j,r\right)\right\|=O\left(\sum_{j=n+1}^{\infty}\left|\phi_j\right|\right),$$

and

$$E\left\|\sum_{j=n+1}^{\infty}\left\{\dot{\varphi}_{\theta_0,j}U_{\theta_0}\left(\mathbf{X}_{t-j}\right)+\varphi_{\theta_0,j}\dot{U}_{\theta_0}\left(\mathbf{X}_{t-j}\right)\right\}\varepsilon_{\theta_0 t-j}\right\|=O\left(\sum_{j=n+1}^{\infty}\left|\phi_j\right|\right),$$

so that (16) is $o(1)$ and we conclude that $A_{T,2}(j)=\zeta_{\theta_0}^{(2)}(j)+o_p(1)$.

Finally for $j=1,\ldots,m$ we have that

$$\ddot{\rho}_{T\theta^*}^{(m)}(j)=\frac{1}{T}\sum_{t=1+j}^{T}\left\{\dot{\varepsilon}_{\theta^*t-j}\dot{\varepsilon}_{\theta^*t}'+\ddot{\varepsilon}_{\theta^*t-j}\varepsilon_{\theta^*t}+\ddot{\varepsilon}_{\theta^*t}\varepsilon_{\theta^*t-j}+\dot{\varepsilon}_{\theta^*t}\dot{\varepsilon}_{\theta^*t-j}'\right\},$$

and we can show that $\ddot{\rho}_{T\theta^*}^{(m)}(j)=O_p(1)$, $j=1,\ldots,m$, since $E\left[\varepsilon_{\theta^*t}^2\right]+E\left\|\dot{\varepsilon}_{\theta^*t}\right\|^2+E\left\|\ddot{\varepsilon}_{\theta^*t}\right\|^2<\infty$ using Assumptions 2, 3 and similar techniques. □

Proof of Theorem 1. If the projection of $\tilde{\rho}_{T\hat{\theta}_T}^{(m)}(j)$ is calculated with the true $\xi_{\theta_0}(j)$, so that we set

$$P_{\theta_0}[\boldsymbol{\rho}](j)=\rho(j)-\xi_{\theta_0}(j)\,\beta_{\theta_0}^{(j)}[\boldsymbol{\rho}],$$

with $\beta_{\theta_0}^{(j)}[\boldsymbol{\rho}]=\left(\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\xi_{\theta_0}(\ell)\right)^{-1}\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\rho(\ell)$, $j=1,\ldots,m-q$, by standard algebra using Assumption 4 and Proposition 1, up to $o_p\left(T^{-1/2}\right)$ terms,

$$\begin{aligned}P_{\theta_0}\left[\tilde{\boldsymbol{\rho}}_{T\hat{\theta}_T}^{(m)}\right](j) &= \tilde{\rho}_{T\theta_0}^{(m)}(j)+\xi_{\theta_0}(j)\left(\hat{\theta}_T-\theta_0\right)-\xi_{\theta_0}(j)'\left(\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\xi_{\theta_0}(\ell)\right)^{-1}\\ &\quad\times\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\left\{\tilde{\rho}_{T\theta_0}^{(m)}(\ell)+\xi_{\theta_0}(\ell)\left(\hat{\theta}_T-\theta_0\right)\right\}\\ &= P_{\theta_0}\left[\tilde{\boldsymbol{\rho}}_{T\theta_0}^{(m)}\right](j).\end{aligned}$$

Then, when using $\hat{\xi}_{T\hat{\theta}_T}(j)$, $\bar{\boldsymbol{\rho}}_{T\hat{\theta}_T}^{(m)}=\bar{\boldsymbol{\rho}}_{T\theta_0}^{(m)}+o_p\left(T^{-1/2}\right)$ follows, because of (4) and $\hat{\xi}_{T\hat{\theta}_T}(j)\to_p\xi_{\theta_0}(j)$, which can be proved with the methods of Proposition 1 noting that replacing θ_0 by $\hat{\theta}_T$ in the definition of $A_{T,1}(j)$ and $A_{T,2}(j)$ has no asymptotic effect because of Assumptions 2, 3 and (3).

Finally, the CLT for $\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ follows from the CLT for $\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ under Assumptions 1, (4), H_0, and from the fact that the projections $\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ are standardized by construction if $\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ is already standardized, i.e $\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ has asymptotic variance $AVar\left(T^{1/2}\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right) = I_m$, as can be showed by immediate calculations. Thus $AVar\left(T^{1/2}\bar{\rho}^{(m)}_{T\theta_0}(j)\right)$ is equal to

$$\begin{aligned} AVar\left(T^{1/2}\left(\tilde{\rho}^{(m)}_{T\theta_0}(j) - \hat{\xi}_{T\hat{\theta}_T}(j)\,\tilde{\beta}^{(j)}_{T\hat{\theta}_T}\right)\right) &= AVar\left(T^{1/2}\left(\tilde{\rho}^{(m)}_{T\theta_0}(j) - \xi_{\theta_0}(j)\,\beta^{(j)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right]\right)\right) \\ &= 1 + \xi_{\theta_0}(j)\left(\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\,\xi_{\theta_0}(\ell)\right)^{-1}\xi_{\theta_0}(j)', \end{aligned}$$

while for $1 \leq j < k \leq m$, $ACov\left(T^{1/2}\bar{\rho}^{(m)}_{T\theta_0}(j), T^{1/2}\bar{\rho}^{(m)}_{T\theta_0}(k)\right)$ is given by

$$\begin{aligned} & ACov\left(T^{1/2}\left(\tilde{\rho}^{(m)}_{T\theta_0}(j) - \hat{\xi}_{T\hat{\theta}_T}(j)\,\tilde{\beta}^{(j)}_{T\hat{\theta}_T}\right), T^{1/2}\left(\tilde{\rho}^{(m)}_{T\theta_0}(k) - \hat{\xi}_{T\hat{\theta}_T}(k)\,\tilde{\beta}^{(k)}_{T\hat{\theta}_T}\right)\right) \\ = \; & ACov\left(T^{1/2}\left(\tilde{\rho}^{(m)}_{T\theta_0}(j) - \xi_{\theta_0}(j)\,\beta^{(j)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right]\right), T^{1/2}\left(\tilde{\rho}^{(m)}_{T\theta_0}(k) - \xi_{\theta_0}(k)\,\beta^{(k)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right]\right)\right) \\ = \; & ACov\left(T^{1/2}\tilde{\rho}^{(m)}_{T\theta_0}(j), T^{1/2}\tilde{\rho}^{(m)}_{T\theta_0}(k)\right) - ACov\left(T^{1/2}\tilde{\rho}^{(m)}_{T\theta_0}(j), T^{1/2}\xi_{\theta_0}(k)\,\tilde{\beta}^{(k)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right]\right) \\ & -ACov\left(T^{1/2}\xi_{\theta_0}(j)\,\tilde{\beta}^{(j)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right], T^{1/2}\tilde{\rho}^{(m)}_{T\theta_0}(k)\right) \\ & +ACov\left(T^{1/2}\xi_{\theta_0}(j)\,\tilde{\beta}^{(j)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right], T^{1/2}\xi_{\theta_0}(k)\,\tilde{\beta}^{(k)}_{\theta_0}\left[\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}\right]\right), \end{aligned}$$

where these terms are, 0, 0 (because $j < k$), $-\xi_{\theta_0}(j)\left(\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\,\xi_{\theta_0}(\ell)\right)^{-1}\xi_{\theta_0}(k)'$, and $\xi_{\theta_0}(j)\left(\sum_{j+1}^{m}\xi_{\theta_0}(\ell)'\,\xi_{\theta_0}(\ell)\right)^{-1}\sum_{j\vee k+1}^{m}\xi_{\theta_0}(\ell)'\,\xi_{\theta_0}(\ell)\left(\sum_{k+1}^{m}\xi_{\theta_0}(\ell)'\,\xi_{\theta_0}(\ell)\right)^{-1}\xi_{\theta_0}(k)'$ $= \xi_{\theta_0}(j)\left(\sum_{\ell=j+1}^{m}\xi_{\theta_0}(\ell)'\,\xi_{\theta_0}(\ell)\right)^{-1}\xi_{\theta_0}(k)'$, respectively, and the asymptotic covariance of the projections is 0. □

Proof of Theorem 2. It follows as Theorem 1 using for the CLT of $\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ Assumption 1 under H_{1T}, which only affects the drift of the limiting normal distribution, $\mathbf{h}^{(m)}_{\theta_0} = \left(h^{(m)}_{\theta_0}(1), \ldots, h^{(m)}_{\theta_0}(1)\right)'$. Then the drift $\bar{\mathbf{h}}^{(m)}_{\theta_0}$ of the asymptotic distribution of $\bar{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$ is equal to that of $\tilde{\boldsymbol{\rho}}^{(m)}_{T\theta_0}$, given by H_{1T}, after standardization by $A^{(m)-1/2}_{\theta_0}$ and linear projection of $\mathbf{h}^{(m)}_{\theta_0}$. □

Proof of Proposition 2. We do the proof in two steps. First we find a suitable representation of the LM tests in terms of $\hat{\rho}_{T\theta_0}(j)$. Then we show that this representation can be calculated as $\bar{B}_{T\hat{\theta}_T}(s)$ where s depends on the alternative against the LM test is directed to.

Set the sequence of $1 \times s$ row vectors $d_s(j) = \left(1_{\{j=1\}}, \ldots, 1_{\{j=s\}}\right)$ for $j = 1, 2, \ldots, s$ and $d_s(j) = \mathbf{0}$, for $j > s$, where $\mathbf{0}$ denotes a conformable matrix of zeros. An LM test statistic against MA(s) or AR(s) alternatives (not nested in the model specified by H_0) has the form

$$
\begin{aligned}
LM_T(s) &= TS_{T,1}\left(\tilde{\theta}_T\right)' H_T^{11}\left(\tilde{\theta}_T\right) S_{T,1}\left(\tilde{\theta}_T\right) \\
&= TS_T\left(\tilde{\theta}_T\right)' A_T^{-1}\left(\tilde{\theta}_T\right) S_T\left(\tilde{\theta}_T\right),
\end{aligned}
$$

where $S_{T,1}(\theta) = \sum_{j=1}^{T-1} d_s(j)' \hat{\rho}_{T\theta}(j) = \left(\hat{\rho}_{T,\theta}(1), \ldots, \hat{\rho}_{T,\theta}(s)\right)'$ and $H_T^{11}(\theta) = \left\{A_T^{-1}(\theta)\right\}_{11}$, with $S_T(\theta)$ and $A_T(\theta) = \sum_{j=1}^{T-1} \delta_\theta(j)' \delta_\theta(j)$ for $\delta_\theta(j) = (d_s(j), \ \zeta_\theta(j))$ being first order approximations to the corresponding score and Hessian of the objective function $Q_T(\theta) = \sum_{k=1}^{T} \varepsilon_{\theta k}^2$ for estimation of the complete model, cf. Theorem 1 in Hosking (1980). $\{A\}_{r,s}$ and S_r denote the corresponding blocks of A and S accordingly to the definition of δ_θ, while $\tilde{\theta}_T$ is any restricted estimate of θ_0 that asymptotically behaves as the *MLE*, i.e. admits this stochastic expansion under H_{1T},

$$
T^{1/2}\left(\tilde{\theta}_T - \theta_0\right) = -T^{1/2} A_{T,22}(\theta_0)^{-1} S_{T,2}(\theta_0) + o_p(1), \tag{17}
$$

where $A_{T,22}(\theta) = \sum_{j=1}^{T-1} \zeta_\theta(j)' \zeta_\theta(j)$ and $S_{T,2}(\theta) = \sum_{j=1}^{T-1} \zeta_\theta(j)' \hat{\rho}_{T\theta}(j)$ and Assumption 4 guarantees now that $\lim_{T\to\infty} A_T(\theta_0) > 0$.

Next, we first define the class of statistics

$$
\Psi_{T,\theta}^{(m)}(\omega) := T \sum_{j=1}^{m} \omega(j) \hat{\rho}_{T\theta}(j) \left(\sum_{j=1}^{m} \omega(j)' \omega(j)\right)^{-1} \sum_{j=1}^{m} \omega(j)' \hat{\rho}_{T\theta}(j)
$$

for any sequence of row vectors $\omega(j)$, and the residuals of the linear projection of $d_s(j)$ on $\mathbb{X}_1^m$, $m \geq q$, where $\mathbb{X}_j^k = \left(\zeta_\theta(j)', \ldots, \zeta_\theta(k)'\right)'$, $k \geq j$,

$$
\hat{d}_{s,\theta}^{(m)}(j) = d_s(j) - \zeta_\theta(j) \left(\sum_{k=1}^{m} \zeta_\theta(k)' \zeta_\theta(k)\right)^{-1} \sum_{k=1}^{m} \zeta_\theta(k)' d_s(k).
$$

Then it is easy to generalize (5) in Proposition 1 and Theorem 1 exploiting the orthogonality of $\hat{d}^{(m)}_{s,\theta}(j)$ and $\zeta_\theta(j)$ and show that under H_{1T} and Assumptions 1-3,

$$\Psi^{(T-1)}_{T,\hat{\theta}_T}\left(\hat{d}^{(T-1)}_{s,\hat{\theta}_T}\right) = \Psi^{(T-1)}_{T,\theta_0}\left(\hat{d}^{(T-1)}_{s,\theta_0}\right) + o_p(1) = LM_T(s) + o_p(1)$$

for any $\sqrt{T}$-consistent estimator $\hat{\theta}_T$ of θ_0, while the second equality follows because of (17) and noting that $H^{11}_T(\theta)^{-1} = \sum_{j=1}^{T-1}\hat{d}^{(T-1)}_{s,\theta}(j)'\,\hat{d}^{(T-1)}_{s,\theta}(j) = I_s - \mathbb{X}^s_1(\mathbb{X}^{s\prime}_1\mathbb{X}^s_1)\mathbb{X}^{s\prime}_1 = I_s - \left(\zeta_\theta(1)', \ldots, \zeta_\theta(s)'\right)'\left(\sum_{k=1}^{T-1}\zeta_\theta(k)'\zeta_\theta(k)\right)^{-1}\left(\zeta_\theta(1)', \ldots, \zeta_\theta(s)'\right)$.

Second. We now show that the Box-Pierce statistic $\bar{B}^{(m)}_{T\hat{\theta}_T}(s)$ provides an alternative way of computing $\Psi^{(m)}_{T,\theta}\left(\hat{d}^{(m)}_{s,\theta}\right)$ for any θ and $m \geq s+q$ under Assumption 4, i.e.

$$T\sum_{j=1}^{s}\bar{\rho}^{(m)}_{T,\theta}(j)^2 = \Psi^{(m)}_{T,\theta}\left(\hat{d}^{(m)}_{s,\theta}\right), \quad s = 1, \ldots, m-q. \tag{18}$$

For that we note that $\sum_{j=1}^{s}\bar{\rho}^{(m)}_{T,\theta}(j)^2 = S^{(m)}_m - S^{(m)}_{m-s}$ using equation (5) in Brown et al. (1975), where

$$S^{(m)}_{m-s} = \hat{\boldsymbol{\rho}}^{(m)\prime}_{T,\theta}\left(\begin{pmatrix}\mathbf{0}_s & \mathbf{0}\\ \mathbf{0} & I_{m-s}\end{pmatrix} - \begin{pmatrix}\mathbf{0}\\ \mathbb{X}^m_{s+1}\end{pmatrix}\left(\mathbb{X}^{m\prime}_{s+1}\mathbb{X}^m_{s+1}\right)^{-1}\left(\mathbf{0}\ \ \mathbb{X}^{m\prime}_{s+1}\right)\right)\hat{\boldsymbol{\rho}}^{(m)}_{T,\theta}$$

is the sum of least squares residuals in the linear projection of $\left\{\hat{\rho}_{T,\theta}(j)\right\}_{j=s+1}^{m}$ on $\mathbb{X}^m_{s+1}$ and $\hat{\boldsymbol{\rho}}^{(m)}_{T,\theta} = \tilde{\boldsymbol{\rho}}^{(m)}_{T,\theta} = \left(\hat{\rho}_{T,\theta}(1), \ldots, \hat{\rho}_{T,\theta}(m)\right)'$ since $A^{(m)}_{\theta_0} = I_m$.

Thus, it suffices to show that $\Psi^{(m)}_{T,\theta}\left(\hat{d}^{(m)}_{s,\theta}\right) = T\left(S^{(m)}_m - S^{(m)}_{m-s}\right)$. To this end, write exploiting the definition of $d_s(j)$

$$\Psi^{(m)}_{T,\theta}\left(\hat{d}^{(m)}_{s}\right) = T\hat{\boldsymbol{\rho}}^{(m)\prime}_{T,\theta}G^{(m)}_s\hat{\boldsymbol{\rho}}^{(m)}_{T,\theta},$$

where $G^{(m)}_s = P^{(m)}V^{(m)\prime}_s H^{(m)}_s V^{(m)}_s P^{(m)}$, with $V^{(m)}_s = \left(d_s(1)', \ldots, d_s(m)'\right) = (I_s\ \mathbf{0})$, $H^{(m)}_s = (I_s - \mathbb{X}^s_1(\mathbb{X}^{m\prime}_1\mathbb{X}^m_1)^{-1}\mathbb{X}^{s\prime}_1)^{-1}$ and $P^{(m)} = I_m - \mathbb{X}^m_1(\mathbb{X}^{m\prime}_1\mathbb{X}^m_1)^{-1}\mathbb{X}^{m\prime}_1$. Then we can use the facts that $H^{(m)}_s = I_s + \mathbb{X}^s_1\left(\mathbb{X}^{m\prime}_{s+1}\mathbb{X}^m_{s+1}\right)^{-1}\mathbb{X}^{s\prime}_1$ and that $\mathbb{X}^{m\prime}_1\mathbb{X}^m_1 = \mathbb{X}^{s\prime}_1\mathbb{X}^s_1 + \mathbb{X}^{m\prime}_{s+1}\mathbb{X}^m_{s+1}$ to show that (18) follows after standard algebraic manipulations.

In particular we show that $G^{(m)}_s$ is equal to the difference in the weight matrices of $S^{(m)}_m$ and $S^{(m)}_{m-s}$, i.e.

$$\begin{pmatrix}I_s & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} - \mathbb{X}^m_1(\mathbb{X}^{m\prime}_1\mathbb{X}^m_1)^{-1}\mathbb{X}^{m\prime}_1 + \left(\mathbf{0}\ \mathbb{X}^{m\prime}_{s+1}\right)'\left(\mathbb{X}^{m\prime}_{s+1}\mathbb{X}^m_{s+1}\right)^{-1}\left(\mathbf{0}\ \mathbb{X}^{m\prime}_{s+1}\right). \tag{19}$$

For that, and using that $P^{(m)}V_s^{(m)\prime} = V_s^{(m)\prime} - \mathbb{X}_1^m \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{s\prime}$, note that $G_s^{(m)}$ is equal to the first matrix in (19) plus

$$
\begin{aligned}
&\left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right)' \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right) - \mathbb{X}_1^m \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right) \\
&-\mathbb{X}_1^m \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{s\prime}\mathbb{X}_1^s \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right) \\
&-\left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right)' \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{m\prime} - \left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right)' \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \mathbb{X}_1^{s\prime}\mathbb{X}_1^s \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{m\prime} \\
&+\mathbb{X}_1^m \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{s\prime}\mathbb{X}_1^s \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{m\prime} \\
&+\mathbb{X}_1^m \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{s\prime}\mathbb{X}_1^s \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \mathbb{X}_1^{s\prime}\mathbb{X}_1^s \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right)
\end{aligned}
$$

and write this as $\sum_{j=1}^{7} G_j$, say. Next, using $\mathbb{X}_1^{s\prime}\mathbb{X}_1^s = \mathbb{X}_1^{m\prime}\mathbb{X}_1^m - \mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m$, we have that $\sum_{j=1}^{3} B_j$ is equal to

$$
-\left(\mathbf{0}\ \mathbb{X}_{s+1}^{m\prime}\right)' \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right), \tag{20}
$$

while $\sum_{j=4}^{7} B_j$ is $-\mathbb{X}_1^m \left(\mathbb{X}_1^{m\prime}\mathbb{X}_1^m\right)^{-1} \mathbb{X}_1^{m\prime}$ plus

$$
-\left(\mathbb{X}_1^{s\prime}\ \mathbf{0}\right)' \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \mathbb{X}_1^{m\prime} + \mathbb{X}_1^m \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \mathbb{X}_1^{m\prime}. \tag{21}
$$

Then it is easy to check after straightforward calculation that (20) plus (21) is equal to $\left(\mathbf{0}\ \mathbb{X}_{s+1}^{m\prime}\right)' \left(\mathbb{X}_{s+1}^{m\prime}\mathbb{X}_{s+1}^m\right)^{-1} \left(\mathbf{0}\ \mathbb{X}_{s+1}^{m\prime}\right)$, concluding the proof of (18) in the light of (19).

Then the proof of the Proposition is completed letting m to increase with T. $\square$

Proof of Proposition 3. We set $h_{\theta_0}^{(m)}(j) = r(j)$ for all m, and then $\bar{r}_{\theta_0}^{(m)}(j) = \bar{h}_{\theta_0}^{(m)}(j)$. Next, we note that for m fixed with T, $\bar{T}_{T\hat{\theta}_T}^{(m)} \xrightarrow{d} \sum_{j=1}^{m} \left(Z_j + \bar{r}_{\theta_0}^{(m)}(j)\right) \Big/ j^2$ as $T \to \infty$ by Theorem 2. Finally, using Theorem 3.2 in Billingsley (1999), we only need to show that

$$
\lim_{m\to\infty} \limsup_{T\to\infty} \Pr\left(\left|\bar{T}_{T\hat{\theta}_T}^{(m)} - \bar{T}_{T\hat{\theta}_T}^{(\infty)}\right| > \epsilon\right) = 0
$$

for any $\epsilon > 0$, but this follows by the proof of Proposition 1 and Markov's inequality. $\square$

APPENDIX B: COVARIANCE MATRIX ESTIMATES

We consider Lobato, Nankervis and Savin (2002)'s version of Newey and West (1987) estimate,

$$\hat{A}_{T\theta}^{(m)} = g_{\theta}^{(m)}(0) + \sum_{j} k\left(\frac{j}{\ell}\right)\left\{g_{\theta}^{(m)}(j) + g_{\theta}^{(m)}(j)'\right\},$$

where $w_{\theta t}^{(m)} = (w_{\theta,1t}, \ldots, w_{\theta,mt})'$ and $w_{\theta,kt} = \varepsilon_{\theta t}\varepsilon_{\theta t-k}$, $g_{\theta}^{(m)}(j) = T^{-1}\sum_{t=1+j}^{T} w_{\theta t}^{(m)} w_{\theta t-j}^{(m)\prime}$, ℓ is the bandwidth parameter and k is the kernel or lag window, for which we assume the following.

Assumption 5 *The kernel k belongs K where K is the class of functions $K = \{k : \mathbb{R} \to [-1,1]\}$ that is symmetric around zero, continuous at zero at all but a finite number of points, and satisfies*

$$k(0) = 1, \ \int_{-\infty}^{\infty} |k(x)|\,dx < \infty, \quad \int_{-\infty}^{\infty} |\psi(\xi)|\,d\xi < \infty$$

where $\psi(\xi) = (2\pi)^{-1}\int_{-\infty}^{\infty} k(x)\, e^{i\xi x} dx$.

Then we obtain the following result, which is valid under both H_0 and H_{1T}.

Lemma 1 *Under Assumptions 1-3 and 5, (3) and $1/\ell + \ell/T^{1/2} \to 0$,*

$$\hat{A}_{T\hat{\theta}_T}^{(m)} = A_{\theta_0}^{(m)} + o_p(1).$$

Proof of Lemma 1. By Lemma 1 of Lobato, Nankervis and Savin (2002), see also Theorem 2.1 in Davidson and De Jong (2000), it follows that $\hat{A}_{T\theta_0}^{(m)} = A_{\theta_0}^{(m)} + o_p(1)$ by Assumptions 1 and 5. Now we provide the proof of $\hat{A}_{T\hat{\theta}_T}^{(m)} = \hat{A}_{T\theta_0}^{(m)} + o_p(1)$. The basic argument as in Newey and West (2002, Lemma A.3) is to bound uniformly the expected value of the derivatives of $\hat{A}_{T\theta}^{(m)}$. For our case is enough to consider the first derivative of each element (r,s) of $\hat{A}_{T\theta}^{(m)}$, so we can show for any θ^* so that $\theta^* \to_p \theta_0$,

$$E\left\|\frac{\partial}{\partial\theta}\hat{A}_{T\theta}^{(m)}(r,s)\right\|_{\theta=\theta^*} \le 2\sum_{j}\left|k\left(\frac{j}{\ell}\right)\right| E\left\|\frac{\partial}{\partial\theta}g_{\theta,r,s}^{(m)}(j)\right\|_{\theta=\theta^*} = O(\ell),$$

because $E\left\|(\partial/\partial\theta)\, g_{\theta,r,s}^{(m)}(j)\right\| < \infty$ uniformly in θ and j under Assumptions 1-3, cf. equation (13). Then it follows that $\left\|(\partial/\partial\theta)\,\hat{A}_{T\theta}^{(m)}(r,s)\right\|_{\theta=\theta^*} = O_p(\ell)$ and that $\hat{A}_{T\hat{\theta}_T}^{(m)} = A_{\theta_0}^{(m)} = O_p\left(\ell T^{-1/2}\right) + o_p(1) = o_p(1)$ using (3), and the results follows. $\square$

REFERENCES

[1] Anderson, T.W. (1993), "Goodness of fit tests for spectral distributions", *Annals of Statistics,* **21**, 830-847.

[2] Beran, J. (1995), "Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models," *Journal of the Royal Statistical Society, Series B*, **57**, 659-672.

[3] Billingsley, P. (1999), *Convergence of Probability Measures,* Second Edition, New York: Wiley.

[4] Box, G.E.P. and Cox, D. (1964), "An analysis of transformations", *Journal of the Royal Statistical Society*, *Series* B, **26**, 211–52.

[5] Box, G.E.P. and Pierce, D.A. (1970), "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models". *Journal of the American Statistical Association,* **65,** 1509-1526.

[6] Box, G.P. and M.J. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, California.

[7] Brown, R.L., Durbin, J. and Evans, J.M. (1975), "Techniques for testing constancy of regression relationships over time" (with discussion). *Journal of the Royal Statistical Society, Series B,* **37,** 149-192.

[8] Delgado, M.A., Hidalgo, J. and Velasco, C. (2005), "Distribution Free Goodness-of-fit Tests for Linear Processes", *Annals of Statistics*, **33**, 2568-2609.

[9] Duchesne, P. and C. Franq (2008), "On diagnostic checking time series models with Pormanteau test statistics based on generalized inverses and $\{2\}-$Inverses", COMSTAT 2008, Proceedings in Computational Statistics, P. Brito Ed., pp. 143-154, Physica-Verlag.

[10] Durbin, J. (1970), "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables", *Econometrica,* **38**, 410-421.

[11] Francq, C., R. Roy and J-M. Zakoïan (2005), "Diagnostic checking in ARMA models with uncorrelated errors", *Journal of the American Statistical Association,* **100**, 532-544.

[12] Hannan, E.J. and C.C. Heyde (1972), "On Limit Theorems for Quadratic Functions of Discrete Time Series", *Annals of Mathematical Statistics,* **43**, 2058-2066.

[13] Hwang, S.Y., I.V. Basawa and J. Reeves (1994), "The asymptotic distributions of residual autocorrelations and related tests of fit for a class of nonlinear time series models", *Statistica Sinica,* **4**, 107-125.

[14] Hong, Y. (1996), "Consistent testing for serial correlation of unknown form", *Econometrica,* **64**, 837-864.

[15] Hosking, J.R.M. (1980), "Lagrange-multiplier tests of time-series models", *Journal of the Royal Statistical Society, Series B,* **42,** 170-181.

[16] Imhof, J.P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, **48**, 419-426.

[17] Khmaladze, E.V. (1981), "Martingale approach to the goodness of fit tests", *Probability Theory and its Applications,* **26**, 246-265.

[18] Li, W.K. (1992), "On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling", *Biometrika*, **79**, 435-437.

[19] Lobato, I., J.C. Nankervis and N.E. Savin (2002), "Testing for zero autocorrelation in the presence of statistical dependence", *Econometric Theory*, **18**, 730-743.

[20] Romano, J.P. and L.A. Thombs (1996), "Inference for Autocorrelations under Weak Assumptions," *Journal of the American Statistical Association,* **91,** 590-600.

[21] Roussas, G.G. and D.A. Ioannidies (1987), "Moment inequalities for mixing sequences of random variables," *Stochastic Analysis and Applications*, **5**, 61–120.

[22] Tsigroshvili, Z. (1998), "Some notes on goodness-of-fit tests and innovation martingales", *Proceedings of A. Razmadze Mathematical Institute,* **117,** 89-102.

[23] Velasco, C. and P.M. Robinson (2000), "Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series", *Journal of the American Statistical Association,* **95**, 229-1243.

Table 1. Empirical size of CvM and Portmanteau tests at 5% of significance. IID Innovations: $A_T = I_m$.

	$T=100$										$T=400$									
	CvM	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\hat{B}_{T\hat{\theta}_T}(s)$				CvM	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\hat{B}_{T\hat{\theta}_T}(s)$			
$s:$			1	2	3	5	5	10	15	20			1	2	3	5	5	10	20	40
	H$_0$: AR(1)										H$_0$: AR(1)									
δ_{10}																				
-0.8	8.6	4.5	4.8	4.5	4.3	4.3	4.3	5.3	6.0	6.7	6.5	4.8	5.0	4.8	4.8	4.8	4.6	4.8	5.6	6.1
-0.5	3.8	4.5	4.9	4.5	4.2	4.1	4.8	5.5	6.1	6.8	6.3	4.8	5.0	4.9	4.8	4.7	4.3	4.7	5.4	5.9
0.0	3.3	4.6	5.0	4.5	4.2	4.0	4.6	5.1	5.8	6.6	6.0	4.7	4.9	4.9	4.8	4.8	4.3	4.6	5.3	5.9
0.5	2.5	4.6	4.9	4.6	4.4	4.2	4.7	5.3	6.0	6.6	5.8	4.9	5.0	4.9	4.9	4.9	4.3	4.6	5.3	6.0
0.8	1.9	4.7	4.9	4.8	4.7	4.6	4.9	5.5	6.1	6.7	5.5	4.9	5.0	4.9	4.9	5.0	4.4	4.8	5.4	6.1
	H$_0$: MA(1)										H$_0$: MA(1)									
η_{10}																				
-0.8	8.1	4.4	4.5	4.3	4.2	4.4	6.7	7.3	7.8	8.3	6.3	4.7	4.9	4.6	4.6	4.7	5.0	5.1	5.8	6.3
-0.5	3.5	4.1	4.4	4.3	4.0	4.0	4.8	5.5	6.2	6.8	6.2	4.7	4.8	4.8	4.7	4.7	4.3	4.7	5.4	6.0
0.0	3.2	4.2	4.5	4.3	4.0	3.9	4.5	5.1	5.8	6.5	6.0	4.6	4.7	4.8	4.8	4.8	4.3	4.6	5.3	6.0
0.5	2.5	4.2	4.5	4.2	4.1	4.0	4.6	5.2	5.9	6.4	5.7	4.8	4.9	4.9	4.8	4.8	4.3	4.6	5.3	6.0
0.8	1.8	8.6	7.9	7.9	8.6	9.8	11.5	11.2	11.3	11.6	5.5	4.8	5.0	4.8	4.8	5.0	4.8	5.1	5.6	6.2
	H$_0$: I(d)										H$_0$: I(d)									
d_0																				
0.0	5.6	3.8	4.3	3.5	3.0	3.2	3.8	4.8	5.5	6.2	5.7	4.5	4.7	4.3	4.1	4.1	4.4	4.7	5.3	6.0
0.2	2.5	3.8	4.2	3.4	3.0	3.2	4.8	5.5	6.1	6.7	5.7	4.4	4.7	4.2	4.1	4.1	4.4	4.7	5.3	6.1
0.4	2.6	3.7	4.1	3.3	2.9	3.1	4.8	5.4	6.0	6.6	5.8	4.2	4.4	4.1	4.0	4.0	4.4	4.8	5.5	6.0

Note: CvM denotes the Cramer-von Mises test of Delgado, Hidalgo and Velasco (2005) and $\bar{T}^{(s)}_{T\hat{\theta}_T}$ denotes the new truncated $T_p - process$ test discussed in Section 4.2. $\bar{B}_{T\hat{\theta}_T}(s)$ and $\hat{B}_{T\hat{\theta}_T}(s)$ denote Box-Pierce tests with projected and usual residuals autocorrelations, respectively.

Table 2. Empirical size of CvM and Portmanteau tests at 5% of significance. ARCH Innovations: Diagonal $\hat{A}_T$.

$T=100$

	CvM	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\hat{B}_{T\hat{\theta}_T}(s)$			
$s:$			1	2	3	5	5	10	15	20
			H$_0$: AR(1)							
δ_{10}										
-0.8	19.8	4.3	4.2	4.6	4.9	5.5	19.0	14.8	12.4	10.7
-0.5	15.1	4.0	4.0	3.8	4.0	4.4	14.2	10.8	8.9	7.8
0.0	10.9	4.2	4.5	3.4	3.5	3.8	11.2	8.6	7.3	6.6
0.5	11.0	3.8	3.9	3.4	3.6	3.8	13.6	10.5	8.5	7.5
0.8	10.6	4.2	4.1	4.2	4.4	4.9	17.8	14.1	11.5	10.1
			H$_0$: MA(1)							
η_{10}										
-0.8	9.4	4.7	4.4	4.8	5.2	5.9	23.8	19.8	16.7	14.8
-0.5	13.1	4.9	4.7	5.0	5.2	5.9	13.8	10.4	8.6	7.6
0.0	10.4	3.9	4.0	3.8	3.6	3.5	10.8	8.2	6.7	6.0
0.5	10.0	4.5	4.3	4.6	4.9	5.3	13.2	10.3	8.5	7.5
0.8	9.4	4.7	4.4	4.8	5.2	5.9	23.8	19.8	16.7	14.8
			H$_0$: I(d)							
d_0										
0.0	10.6	4.0	3.9	4.1	4.3	4.6	14.0	10.8	8.8	7.7
0.2	10.7	4.1	3.9	4.2	4.4	4.7	14.0	10.8	8.7	7.6
0.4	10.5	4.2	4.1	4.3	4.6	4.8	14.1	10.8	8.6	7.5

$T=400$

	CvM	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\hat{B}_{T\hat{\theta}_T}(s)$			
$s:$			1	2	3	5	5	10	20	40
			H$_0$: AR(1)							
δ_{10}										
-0.8	38.1	5.0	4.7	5.3	5.7	6.2	38.1	33.0	25.9	19.3
-0.5	32.1	4.5	4.4	4.5	4.9	5.1	31.6	26.9	20.7	15.0
0.0	28.0	4.1	4.3	3.5	3.5	3.7	27.1	23.2	17.7	12.8
0.5	31.3	4.4	4.3	4.3	4.6	4.9	31.1	26.8	20.6	15.0
0.8	34.5	4.9	4.5	5.1	5.5	6.0	37.6	32.7	25.8	19.1
			H$_0$: MA(1)							
η_{10}										
-0.8	33.7	4.0	4.9	5.6	6.5	7.1	38.1	32.6	25.3	18.7
-0.5	30.8	5.1	4.9	5.3	5.6	6.1	31.1	26.7	20.4	14.9
0.0	28.2	4.5	4.5	4.4	4.3	4.1	28.0	23.8	18.1	13.1
0.5	30.2	4.9	4.6	5.1	5.4	5.9	30.8	26.4	20.3	14.9
0.8	33.3	5.3	4.7	5.5	6.0	6.7	37.6	32.5	25.3	19.0
			H$_0$: I(d)							
d_0										
0.0	31.0	4.8	4.5	5.0	5.4	5.9	32.8	28.5	21.9	16.1
0.2	31.1	4.8	4.5	5.1	5.5	5.9	32.8	28.6	22.0	16.1
0.4	30.8	4.9	4.6	5.2	5.6	6.1	32.9	28.6	22.0	16.1

Note: Test statistics are as in Table 1. $\hat{A}_T$ is restricted to be diagonal in the estimation and the diagonal elements have also has the MD form. ARCH Innovations generated with $\alpha_1 = 0.8$.

Table 3. Empirical size of Portmanteau tests at 5% of significance. $\hat{A}_T$ no restricted.

	IID innovations									GARCH innovations ($\alpha_1 = 0.4$)								
	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\tilde{Q}_s$				$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\tilde{Q}_s$			
$s:$		1	2	3	5	2	3	5	12		1	2	3	5	2	3	5	12
$T = 100$																		
δ_{10}	H$_0$: AR(1)																	
0.0	5.5	4.1	7.3	8.2	8.8	7.1	4.7	4.7	5.0	5.6	3.9	7.2	8.8	9.7	8.4	6.7	5.8	5.1
0.5	4.7	4.1	5.1	6.0	7.0	3.0	3.2	3.6	4.2	5.6	4.9	6.1	7.3	8.7	6.0	5.9	5.6	5.0
0.8	4.9	4.6	4.8	5.2	5.7	2.6	1.8	2.4	3.3	6.0	5.4	6.0	6.4	7.4	7.1	5.7	5.0	4.9
η_{10}	H$_0$: MA(1)																	
0.0	4.4	3.4	5.9	7.2	7.7	6.7	4.5	4.8	5.3	4.6	3.4	6.1	7.7	9.0	8.1	6.3	5.8	5.2
0.5	4.3	3.8	4.6	5.5	6.5	3.1	2.8	3.6	4.2	5.2	4.5	5.8	6.9	8.1	6.0	5.4	5.2	4.9
0.8	6.1	5.2	6.7	7.8	9.4	9.0	8.7	9.4	9.5	6.9	5.7	7.0	8.7	11.5	12.0	11.9	12.2	11.3
d_0	H$_0$: I(d)																	
0.0	4.5	4.2	4.8	5.6	6.3	3.0	2.8	3.2	3.8	5.3	4.6	5.7	6.6	8.0	5.8	5.4	5.1	4.9
0.4	4.5	4.2	4.8	5.6	6.3	3.0	2.8	3.2	3.8	5.3	4.6	5.7	6.6	8.0	5.8	5.4	5.1	4.9
$T = 400$																		
δ_{10}	H$_0$: AR(1)																	
0.0	5.7	4.5	7.0	7.3	8.2	9.9	5.0	4.9	5.0	5.0	4.0	6.6	7.5	8.7	11.5	9.1	8.0	6.4
0.5	5.3	4.9	5.5	6.0	7.3	3.0	3.1	3.7	3.9	4.9	4.6	5.3	5.6	6.8	8.0	8.2	7.6	6.2
0.8	5.5	5.0	5.7	6.0	6.3	2.3	2.1	2.2	2.8	5.3	5.1	5.4	5.5	5.8	9.2	8.0	6.8	5.9
η_{10}	H$_0$: MA(1)																	
0.0	5.6	4.4	6.6	7.4	8.0	9.5	5.1	4.9	5.1	4.8	3.7	6.1	7.3	7.9	11.4	9.2	8.0	6.5
0.5	5.3	4.8	5.3	5.8	7.0	3.1	3.1	3.7	4.0	5.0	4.4	5.3	5.7	7.0	7.9	8.0	7.5	6.1
0.8	5.5	5.0	5.8	6.3	6.9	3.1	2.8	2.7	3.2	5.6	5.1	5.7	5.9	6.8	9.5	8.6	7.5	6.0
d_0	H$_0$: I(d)																	
0.0	5.3	4.9	5.5	5.8	6.4	3.1	3.2	3.4	3.5	4.9	4.4	5.2	5.4	6.1	8.3	8.3	7.4	6.1
0.4	5.3	4.9	5.5	5.8	6.4	3.1	3.2	3.4	3.5	4.9	4.4	5.2	5.4	6.1	8.3	8.3	7.4	6.1

Note: $\bar{B}_{T\hat{\theta}_T}(s)$ denotes Box-Pierce test with projected residuals autocorrelations and $\tilde{Q}_s$ is the modified Box-Pierce test of Francq, Roy and Zaköian (2005). A general estimation of $\hat{A}_T$ is used for both models.

Table 4. Empirical power of CvM and Portmanteau tests at 5% of significance. IID Innovations: $\hat{A}_T = I_m$. $T = 100$.

	CvM	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\hat{B}_{T\hat{\theta}_T}(s)$			
$s:$			1	2	3	5	5	10	15	20
$H_0: AR(1)$. H_1 :ARMA(1,1), $\delta_{10}=0$.										
η_{10}										
-0.8	95.5	98.3	92.4	94.8	93.6	84.3	87.1	71.4	62.7	58.2
-0.5	38.4	65.3	52.5	43.9	36.1	28.3	29.3	23.6	21.9	21.4
0.2	4.9	7.6	7.9	6.6	5.9	5.3	5.0	5.7	6.2	6.9
0.5	37.4	65.2	51.0	45.6	37.7	29.5	30.8	25.1	22.8	22.2
0.8	83.4	95.0	91.3	95.2	93.6	84.6	88.0	71.7	63.1	58.3
$H_0: MA(1)$. H_1 : ARMA(1,1) $\eta_{10}=0$.										
δ_{10}										
-0.8	99.7	99.7	98.7	99.4	98.7	97.3	99.5	98.9	98.4	97.9
-0.5	49.3	49.5	47.1	48.9	42.3	35.6	36.7	30.1	27.3	26.5
0.2	4.6	9.5	5.1	4.5	4.2	4.0	3.8	4.6	5.3	6.0
0.5	43.9	60.9	48.1	44.9	38.7	31.8	33.7	27.9	25.4	24.9
0.8	99.0	99.8	99.2	99.6	99.4	98.9	99.1	98.2	97.2	96.5
$H_0: I(d)$. H_1 :ARFIMA$(1,d,0)$.										
δ_{10}										
d_0= 0.0										
0.2	3.7	17.8	19.7	11.5	8.3	7.3	9.3	9.3	9.5	10.0
0.5	7.5	42.5	45.0	31.1	21.9	17.8	21.3	19.4	18.1	18.0
0.8	2.9	23.6	25.5	18.2	12.1	9.9	11.3	12.9	12.7	13.1
d_0= 0.2										
0.2	3.7	17.7	19.7	11.5	8.4	7.3	9.3	9.3	9.4	10.0
0.5	7.4	42.6	45.0	31.3	21.9	17.8	21.4	19.5	18.1	18.1
0.8	2.7	27.8	28.8	23.0	17.1	12.9	13.8	15.9	15.4	15.5
$H_0: AR(1)$. H_1 :ARFIMA$(1,d,0)$.										
d_0										
δ_{10}= 0.0										
0.2	8.6	9.9	10.1	9.1	8.7	7.9	7.2	7.4	7.7	7.8
0.4	17.8	26.1	25.9	22.7	20.2	17.4	14.9	15.4	14.2	13.5
δ_{10}= 0.5										
0.2	2.1	5.4	5.7	5.1	4.8	4.5	4.1	4.9	5.5	6.2
0.4	1.8	11.5	12.1	9.2	7.9	6.9	6.7	7.0	7.2	7.7

Note: Test statistics are as in Table 1.

Table 5. Empirical power of CvM and Portmanteau tests at 5% of significance.

ARCH Innovations ($\alpha_1 = 0.5$). $\hat{A}_T$ no restricted. $T = 100$.

	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}_{T\hat{\theta}_T}(s)$				$\tilde{Q}_s$			
$s:$		1	2	3	5	2	3	5	12
		$H_0: AR(1)$. $H_1:$ ARMA(1,1), $\delta_{10} = 0$.							
η_{10}									
-0.8	88.1	84.5	83.4	76.1	64.0	94.3	91.0	84.0	62.5
-0.5	40.8	37.8	41.4	36.3	32.1	51.6	43.3	30.6	24.4
0.2	8.8	6.5	10.2	11.5	11.3	10.4	8.1	7.7	7.9
0.5	58.9	41.0	42.9	38.6	31.6	53.6	44.4	33.1	22.6
0.8	93.9	82.2	83.0	76.4	64.5	94.9	92.0	85.5	63.7
		$H_0: AR(1)$. $H_1:$ ARFIMA$(1,d,0)$.							
d_0									
		δ_{10}= 0.0							
0.1	5.8	4.5	8.3	11.2	10.1	6.9	6.8	6.5	5.8
0.2	6.9	5.2	10.0	12.3	12.7	8.0	10.0	8.9	7.8
0.3	9.8	7.4	11.9	15.4	17.0	11.3	13.2	12.9	12.0
0.4	10.6	8.8	11.5	14.3	18.4	13.5	16.3	16.6	15.4
		δ_{10}= 0.5							
0.1	6.3	5.3	6.6	8.2	8.5	5.4	5.4	5.1	4.9
0.2	6.4	5.5	6.5	7.7	8.3	5.5	5.8	4.8	5.0
0.3	6.8	6.6	6.5	8.1	9.1	5.6	6.0	5.1	5.4
0.4	8.5	7.7	9.7	9.2	4.2	8.6	7.2	6.1	6.7

Note: Test statistics are as in Table 3. The general estimation approach for $\hat{A}_T$ is imposed.

Table 6. Chemical data, $T = 226$. Goodness-of-fit analysis for Chemical Process Temperature Readings based on fractionally integrated models.

	BIC	$\underset{(s.e.)}{\hat{d}}$	CvM	$\bar{T}^{(m)}_{T\hat{\theta}_T}$	$\bar{B}^{(m)}_{T\hat{\theta}_T}(s)$				$\hat{B}_{T\hat{\theta}_T}(s)$			
s					1	2	3	5	5	10	20	30
model	H_0 : ARFIMA(p, d, q)											
$(0, d, 0)$	3.7949	.871 (.052)	4.53***	2.14***	20.87***	20.89***	21.69***	23.44***	23.58***	27.22***	29.03**	30.61
$(1, d, 0)$	3.7176	1.076 (.065)	1.37*	.73**	6.88***	6.92**	8.32**	9.71*	9.61**	10.87	12.28	13.41
$(2, d, 0)$	3.7101	1.227 (.075)	.31	.17	1.50	1.54	2.14	3.57	3.16	3.54	4.71	5.81
$(0, d, 1)$	3.7120	1.249 (.159)	.97	.70**	6.34**	8.34**	8.83**	9.32*	8.00**	8.82	9.71	10.76
$(0, d, 2)$	3.7054	1.313 (.126)	.11	.31	1.53	1.83	2.00	2.08	1.55	1.87	2.96	4.33
$(1, d, 1)$	3.7133	1.326 (.144)	.03	.28	2.50	3.48	3.69	3.88	3.23	3.54	4.51	5.70
model	H_0: ARIMA$(p, 1, q)$											
$(1, 1, 0)$	3.7015		0.31	0.21	0.44	3.53	8.44**	12.43**	11.58**	14.37	16.13	17.45
$(2, 1, 0)$	3.7236		1.21*	0.50**	3.06*	8.09**	11.15**	14.70**	12.80***	16.25**	18.34	20.03
$(0, 1, 1)$	3.7162		1.97**	.88***	6.79***	12.32***	14.93***	17.60***	15.01***	18.51**	20.34	21.80
$(0, 1, 2)$	3.7104		1.65**	0.61**	5.16**	7.54**	9.10**	10.70*	9.06**	11.47	13.45	14.86
$(1, 1, 1)$	3.7243		1.54*	0.62**	4.51**	9.27***	11.62***	14.77**	12.09***	15.20	17.06	18.48

Note: *,**,*** denote significant values at 10%, 5% and 1% respectively. Standard errors of d estimates are in parenthesis. $m = 22+$ number of estimated parameters.

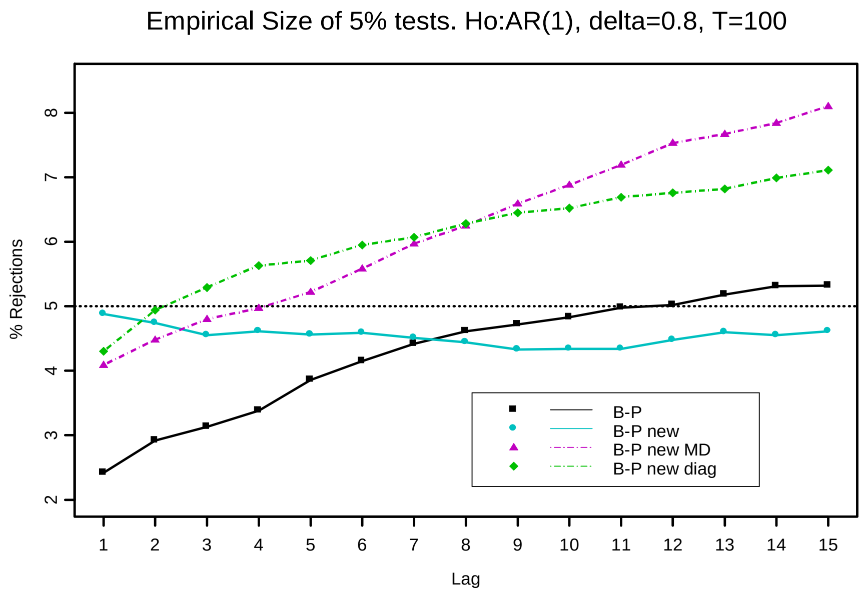

Figure 1. Empirical size of 5% tests. $\mathrm{H}_0 : \mathrm{AR}(1)$, $\delta_0 = 0.8$, $T = 100$, iid innovations.

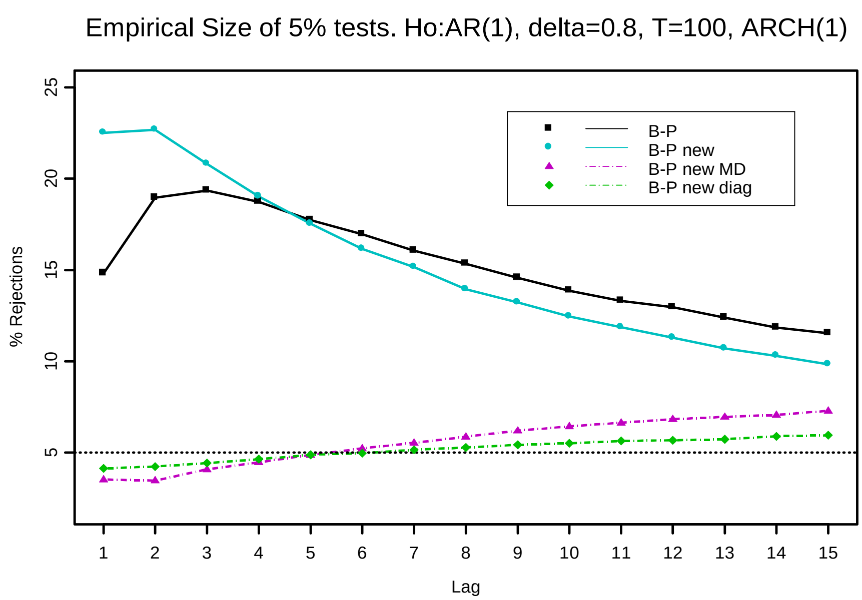

Figure 2. Empirical size of 5% tests. $\mathrm{H}_0 : \mathrm{AR}(1)$, $\delta_0 = 0.8$, $T = 100$, $ARCH\,(1)$ innovations.

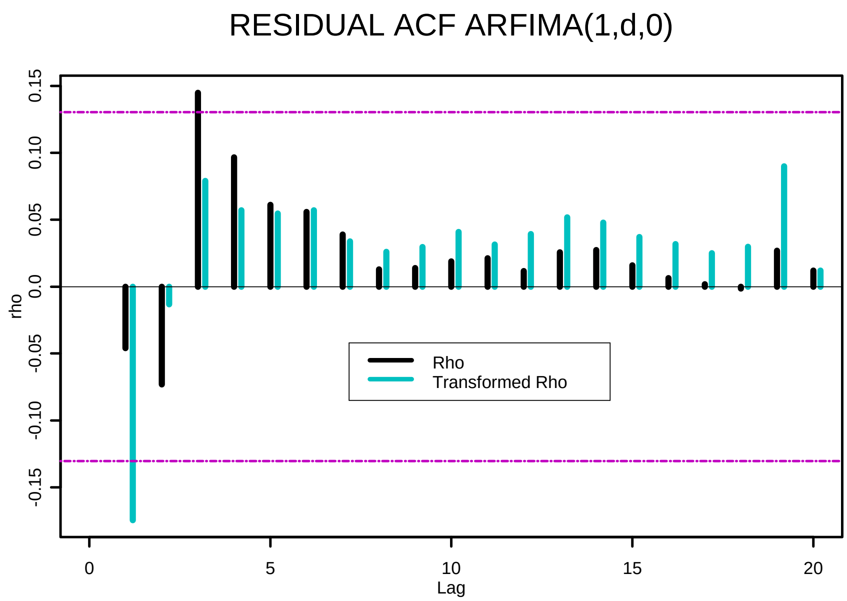

Figure 3 Residual ACF of ARFIMA$(1, d, 0)$ residuals for Chemical Series C data, $T = 226$. Confidence bands are plotted at $\pm 2/\sqrt{T}$.

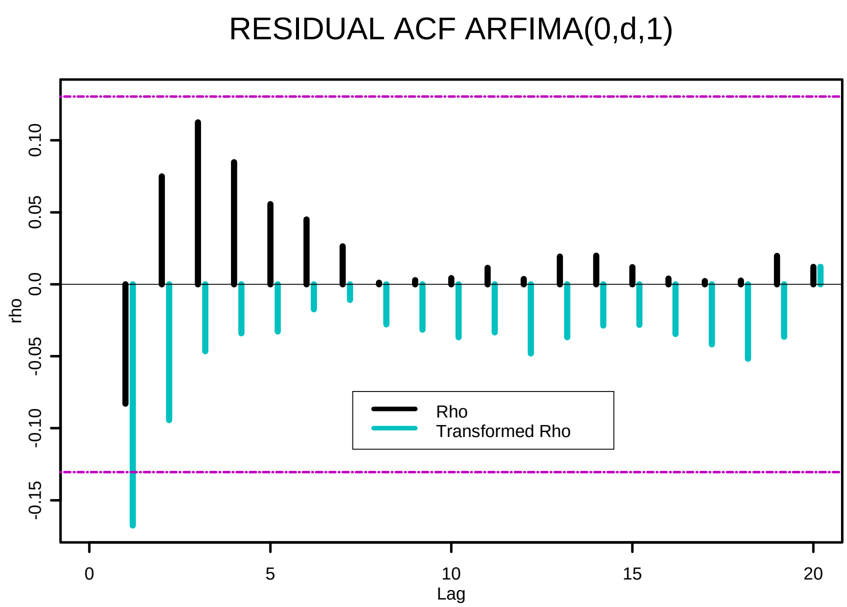

Figure 4. Residual ACF of ARFIMA$(0, d, 1)$ residuals for Chemical Series C data, $T = 226$. Confidence bands are plotted at $\pm 2/\sqrt{T}$.