

# An Asymptotically Pivotal Transform of the Residuals Sample Autocorrelations With Application to Model Checking

Miguel A. DELGADO and Carlos VELASCO



---

We propose an asymptotically distribution-free transform of the sample autocorrelations of residuals in general parametric time series models, possibly nonlinear in variables. The residuals autocorrelation function is the basic model checking tool in time series analysis, but it is not useful when its distribution is incorrectly approximated because the effects of parameter estimation and/or higher-order serial dependence have not been taken into account. The limiting distribution of the residuals sample autocorrelations may be difficult to derive, particularly when the underlying innovations are uncorrelated but not independent. In contrast, our proposal is easily implemented in fairly general contexts and the resulting transformed sample autocorrelations are asymptotically distributed as independent standard normals when innovations are uncorrelated, providing a useful and intuitive device for time series model checking in the presence of estimated parameters. We also discuss in detail alternatives to the classical Box–Pierce test, showing that our transform entails no efficiency loss under Gaussianity in the direction of MA and AR departures from the white noise hypothesis, as well as alternatives to Bartlett’s  $T_p$ -process test. The finite-sample performance of the procedures is examined in the context of a Monte Carlo experiment for the new 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. Additional technical details are included in a supplemental material online.

**KEY WORDS:** Higher-order serial dependence; Local alternatives; Long memory; Model checking; Nonlinear in variables models; Recursive residuals.

---

## 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; for example, 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. [Box and Pierce \(1970\)](#) and [Durbin \(1970\)](#) showed that sample autocorrelations of ARMA residuals are neither asymptotically independent or identically distributed, even when the underlying innovations are iid. It happens that the asymptotic distribution of test statistics depends on nuisance parameters when genuine innovations are substituted by estimated residuals. Exceptions can be found in the rank statistics literature where tests based on residuals can be asymptotically equivalent to tests based on genuine innovations, which have a known exact distribution for each sample size. See, for instance, [Hallin and Puri \(1994\)](#) using aligned rank test statistics or [Hallin and Jurečková \(1999\)](#) and [Hallin, Jurečková, and Koul \(1999\)](#) using regression rank score statistics. [Delgado and Velasco \(2010\)](#) consider specification tests of parametric linear processes with iid innovations where the test statistics are weighted sums of the residuals sample autocorrelations. In this case, tests statistics based on genuine innovations and estimated residuals are asymptotically equivalent, resulting in optimal tests within its class in the direction of certain

nonparametric local alternatives. Other authors have considered residuals of more general models with iid innovations; for example, [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, for example, 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, for example, [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 the underlying 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, which can be directly applied to time series model checking in the presence of estimated parameters. In particular, we consider natural alternatives to [Box and Pierce \(1970\)](#) and Bartlett’s  $T_p$ -process type tests based on these transforms. The martingale part in the asymptotic representation of the sample autocorrelations of residuals forms the basis of the proposed transformations, which are asymptotically distributed as iid standard normals, like the sample autocorrelations based on iid genuine innovations.

---

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 pseudo-residuals of a parametric model  $\{U_\theta(X_t)\}_{t \in \mathbb{Z}}$  whose correlation structure can be represented according to the linear model

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

where

$$\varphi_\theta(z) = \sum_{j=0}^{\infty} \phi_\theta(j)z^j, \quad z \in \mathbb{C},$$

and  $U_\theta: \mathbb{R}^k \rightarrow \mathbb{R}$  and  $\phi_\theta: \mathbb{Z} \rightarrow \mathbb{R}$  are known functions indexed by the vector of parameters  $\theta \in \Theta \subset \mathbb{R}^q$ . The parameter space  $\Theta$  restricts the functions  $\phi_\theta$  and  $U_\theta$  such that the process  $\{\varepsilon_{\theta t}\}_{t \in \mathbb{Z}}$  is strictly stationary,  $\phi_\theta(0) = 1$  and  $\sum_{j=0}^{\infty} |\phi_\theta(j)| < \infty$  for each  $\theta \in \Theta$ . This means that  $\{\varepsilon_{\theta t}\}_{t \in \mathbb{Z}}$  can be represented as a parametric linear filter of the residuals  $\{U_\theta(X_t)\}_{t \in \mathbb{Z}}$  with absolute summable weights. 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  $X_t$ , that is, 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, nonlinear in variables models naturally appear when variables are transformed to get more functional flexibility, for example, [Box and Cox \(1964\)](#).

The transfer function  $\varphi_\theta$  specifies the linear serial dependence behavior of the residuals. The most popular model is the ARMA( $p_1, p_2$ ) with

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

such that  $\Phi_\delta$  and  $\Xi_\eta$  are the autoregressive and moving average polynomials with coefficients  $\delta$  and  $\eta$  of orders  $p_1$  and  $p_2$ , respectively. The function  $U_\theta$  is usually not indexed by the parameters  $(\delta, \eta)$ , which are restricted in such a way that  $\Xi_\eta$  and  $\Phi_\delta$  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  $\phi_\theta$  are no longer summable; cf. Assumption A.3 in the [Appendix](#). However, when  $X_t$  is a linear process, that is,  $U_\theta(X_t) = X_t$ , the results of [Delgado, Hidalgo, and Velasco \(2005\)](#) can be straightforwardly applied for justifying the methods proposed in the 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  $\{\varepsilon_{\theta t}\}_{t \in \mathbb{Z}}$ ,

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

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

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

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

$$\hat{\rho}_{T\theta}(j) = \frac{\hat{\gamma}_{T\theta}(j)}{\hat{\gamma}_{T\theta}(0)}, \quad j \in \mathbb{Z},$$

where

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

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  $\{\sqrt{T}\hat{\rho}_{T\theta_0}(j)\}_{j=1}^m$  are asymptotically independent distributed as standard normals. This is still the case under martingale difference sequence (MDS) 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 uncorrelated raw data has 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,  $\hat{\rho}_{T\theta}^{(m)} = (\hat{\rho}_\theta(1), \dots, \hat{\rho}_\theta(m))'$ . Under  $H_0$ , but with  $\{\varepsilon_{\theta_0 t}\}_{t \in \mathbb{Z}}$  exhibiting general higher-order serial dependence conditions,

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

see, for example, [Romano and Thombs \(1996\)](#), where

$$a_{\theta_0}^{(i,j)} = \sum_{\ell=-\infty}^{\infty} \mathbb{E}[\varepsilon_{t,\theta} \varepsilon_{t+i,\theta} \varepsilon_{t+\ell,\theta} \varepsilon_{t+\ell+j,\theta}], \quad i, j = 1, \dots, m. \quad (2)$$

The asymptotic distribution of the vector  $\sqrt{T}\hat{\rho}_{T\theta_0}^{(m)}$  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, for example, kernels, unless certain restrictions on the higher serial dependence of  $\{\varepsilon_{\theta_0 t}\}_{t \in \mathbb{Z}}$  are imposed. For instance, when  $\{\varepsilon_{\theta_0 t}\}_{t \in \mathbb{Z}}$  is a MDS,  $a_{\theta_0}^{(i,j)} = \mathbb{E}[\varepsilon_{t,\theta_0}^2 \varepsilon_{t+i,\theta_0} \varepsilon_{t+j,\theta_0}]$ , which can be estimated by its sample analog without need of specifying any bandwidth or lag number. 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,

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

Thus, under  $H_0$  and any of the previously mentioned regularity conditions, we obtain that  $T^{1/2}\tilde{\rho}_{T\theta_0}^{(m)} \xrightarrow{d} N_m(0, I_m)$ . In practice, a preliminary estimator of  $\theta_0$  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,

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

and

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

Next proposition provides an asymptotic expansion for  $\sqrt{T}\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$ , which implies that under  $H_0$  and fairly general regularity conditions  $\sqrt{T}\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$  converges to a vector of independent standard normals plus a stochastic drift, which depends on the unknown parameters  $\theta_0$ , the specified model, and the particular estimation method. Define

$$\xi_\theta^{(m)} = A_\theta^{(m)-1/2} \zeta_\theta^{(m)},$$

with  $\xi_\theta^{(m)} = (\xi_\theta(1)', \dots, \xi_\theta(m)')'$  and  $\zeta_\theta^{(m)} = (\zeta_\theta(1)', \dots, \zeta_\theta(m)')'$ , where  $\zeta_\theta$  is defined by

$$\frac{\partial}{\partial \theta'} \hat{\rho}_{T\theta}(j) \xrightarrow{p} \zeta_\theta(j) \quad \text{each } j \in \mathbb{Z} \setminus \{0\}$$

under  $H_0$ .

*Proposition 1.* Under  $H_0$ , (3), (4) and Assumptions A.1–A.3 in the Appendix,

$$\tilde{\rho}_{T\hat{\theta}_T}^{(m)} = \tilde{\rho}_{T\theta_0}^{(m)} + \xi_{\theta_0}^{(m)}(\hat{\theta}_T - \theta_0) + o_p(T^{-1/2}). \quad (5)$$

The asymptotic distribution of  $\sqrt{T}\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$  under  $H_0$  can be derived from the asymptotic joint distribution of  $\{\sqrt{T}\tilde{\rho}_{T\theta_0}^{(m)}, \sqrt{T}(\hat{\theta}_T - \theta_0)\}$ , as Li (1992) and Hwang, Basawa, and Reeves (1994) in the context of nonlinear models with iid innovations and as Francq, Roy, and Zakoian (2005) for weak ARMA residuals. Alternative models and estimators demand different derivations, which may be cumbersome to obtain in heavy nonlinear models, possibly exhibiting long memory or  $U_\theta$  nonlinear in variables and parameters. Rather than performing these derivations, we suggest an asymptotically distribution-free transform of the residuals sample autocorrelations by means of least squares projections, which are asymptotically distributed as independent standard normals. The transformed sample autocorrelations are in fact the recursive residuals of a 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. We show that the test is able to detect local alternatives  $\sqrt{T}$  distant from the null. We study in detail two types of tests. On one hand, we consider alternatives to the classical Box–Pierce test based on the sum of  $s$  squared autocorrelations, which are asymptotically equivalent to the Gaussian LM test under the considered local alternatives. On the other hand, we study an alternative to the classical Barlett's  $T_p$  test. The finite-sample performance of these tests is illustrated in Section 4 in the context of a Monte Carlo experi-

ment. 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 MARTINGALE 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 (1975) for CUSUM tests of parameter stability in the linear regression model with fixed regressors. This is in fact the martingale transform of a Gaussian process, that is, its martingale part. See Khmaladze (1981) for a formal derivation in the context of obtaining the martingale part of the standard empirical process with estimated parameters. The recursive residuals, like the martingale transform, can be either forward or backward, but there are also other possibilities. Brown, Durbin, and Evans (1975) consider backward, Khmaladze (1981) considers forward and Delgado, Hidalgo, and Velasco (2005) discuss both possibilities. In order to further motivate the transform, consider the asymptotic decomposition in Proposition 1,

$$\tilde{\rho}_{T\hat{\theta}_T}^{(m)}(j) = \tilde{V}_{T\hat{\theta}_T}^{(m)}(j) + o_p(T^{-1/2}), \quad j = 1, \dots, m,$$

with

$$\tilde{V}_{T\theta}^{(m)}(j) = \tilde{\rho}_{T\theta}^{(m)}(j) - \xi_{\theta_0}(j)(\theta - \theta_0).$$

The source of asymptotic autocorrelation in  $\{\tilde{V}_{T\hat{\theta}_T}^{(m)}(j)\}_{j=1}^m$  is due to  $(\hat{\theta}_T - \theta_0)$ . The transformation consists in a linear operator  $\mathcal{L}^{(m)}$  such that  $\{\mathcal{L}^{(m)}\tilde{V}_{T\hat{\theta}_T}^{(m)}(j)\}_{j \geq 1}$  are asymptotically uncorrelated. A possible operator, resembling the forward martingale transform of a Gaussian process, transforms any generic sequence  $\{\eta(j)\}_{j=1}^m$  in the forward recursive residuals of its least squares projection on  $\{\xi_{\theta_0}(j)\}_{j=1}^m$ , that is,

$$\begin{aligned} \mathcal{L}^{(m)}\eta(j) = \eta(j) - \xi_{\theta_0}(j) \left( \sum_{\ell=j+1}^m \xi_{\theta_0}(\ell)' \xi_{\theta_0}(\ell) \right)^{-1} \\ \times \sum_{\ell=j+1}^m \xi_{\theta_0}(\ell)' \eta(\ell). \end{aligned}$$

Notice that  $\mathcal{L}^{(m)}\tilde{V}_{T\hat{\theta}_T}^{(m)}(j) \equiv \mathcal{L}^{(m)}\tilde{\rho}_{T\theta_0}^{(m)}(j)$ ,  $j = 1, \dots, m - q$ , which does not depend on  $(\hat{\theta}_T - \theta_0)$ . In fact, since  $\{\sqrt{T}\tilde{\rho}_{T\theta_0}^{(m)}(j)\}_{j \geq 1}$  are asymptotically iid standard normals,  $\{\sqrt{T}\mathcal{L}^{(m)}\tilde{V}_{T\hat{\theta}_T}^{(m)}(j)\}_{j \geq 1}$  are asymptotically distributed as independent normal random variables centered at zero with variance

$$\sigma^2(j) = 1 + \xi_{\theta_0}(j) \left( \sum_{\ell=j+1}^m \xi_{\theta_0}(\ell)' \xi_{\theta_0}(\ell) \right)^{-1} \xi_{\theta_0}(j)'$$

In practice, a feasible transformation must be applied to  $\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$  using some consistent estimator of  $\xi_{\theta_0}$ . Under  $H_0$  and fairly general conditions

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

since  $\hat{\gamma}_{T\theta_0}(j) \rightarrow_p 0$  for all  $j \neq 0$ . Thus, standardization by  $\hat{\gamma}_{T\theta_0}(0)$  in  $\hat{\rho}_{T\theta_0}$  has no asymptotic effect on  $\zeta_{\theta_0}$  in the expansion (5). Then,  $\hat{\xi}_{\theta_0}^{(m)}$  is estimated by

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

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

$$\begin{aligned} \hat{\zeta}_{T\theta}(j) &= \frac{1}{T\hat{\gamma}_{T\theta}(0)} \sum_{t=j+1}^T \dot{\varepsilon}_{\theta t}(\varepsilon_{\theta t-j} - \bar{\varepsilon}_{\theta T}) \\ &\quad + \frac{1}{T\hat{\gamma}_{T\theta}(0)} \sum_{t=j+1}^T \dot{\varepsilon}_{\theta t-j}(\varepsilon_{\theta t} - \bar{\varepsilon}_{\theta T}) \end{aligned}$$

and  $\dot{\varepsilon}_{\theta t} = (\partial/\partial\theta')\varepsilon_{\theta t}$ . When  $\varepsilon_{\theta t} = \varphi_{\theta}(L)X_t$ , it is straightforward to obtain closed and easy to compute expressions for  $\zeta_{\theta}$  without further restrictions under  $H_0$ . In these cases, it is simpler to use  $\hat{\zeta}_{\hat{\theta}_T}$  rather than  $\hat{\zeta}_{T\hat{\theta}_T}$ . The feasible transformation consists of the operator  $\hat{L}_T^{(m)}$ , which transforms any generic sequence  $\{\eta(j)\}_{j=1}^m$  in the forward recursive residuals of its least squares projection on  $\{\hat{\xi}_{T\hat{\theta}_T}(j)\}_{j=1}^m$ , that is,

$$\begin{aligned} \hat{L}_T^{(m)}\eta(j) &= \eta(j) - \hat{\xi}_{T\hat{\theta}_T}(j) \left( \sum_{\ell=j+1}^m \hat{\xi}_{T\hat{\theta}_T}(\ell)' \hat{\xi}_{T\hat{\theta}_T}(\ell) \right)^{-1} \\ &\quad \times \sum_{\ell=j+1}^m \hat{\xi}_{T\hat{\theta}_T}(\ell)' \eta(\ell). \end{aligned}$$

The transformed sample autocorrelation with estimated parameters is the scaled transformation applied to  $\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$ , that is,

$$\bar{\rho}_{T\theta}^{(m)}(j) = \frac{\hat{L}_T^{(m)} \tilde{\rho}_{T\theta}^{(m)}(j)}{\hat{\sigma}_T(j)}, \quad j = 1, \dots, m-q,$$

where  $\hat{\sigma}_T^2(j) = 1 + \hat{\xi}_{T\hat{\theta}_T}(j) (\sum_{\ell=j+1}^m \hat{\xi}_{T\hat{\theta}_T}(\ell)' \hat{\xi}_{T\hat{\theta}_T}(\ell))^{-1} \hat{\xi}_{T\hat{\theta}_T}(j)'$  is the estimator of  $\sigma^2(j)$ . Notice that we can only transform the  $m-q$  first sample autocorrelations, given a scaling matrix  $\hat{A}_{T\hat{\theta}_T}^{(m)}$ , because there are no more available degrees of freedom when  $q$  parameters are estimated. Backward recursive residuals will also result in an asymptotically iid standard normal transformation. However, using backward residuals we cannot transform the first  $q$  sample autocorrelations, which typically are the most informative and easiest to interpret.

We prove that, under  $H_0$ ,  $\bar{\rho}_{T\hat{\theta}_T}^{(m)} = (\bar{\rho}_{T\hat{\theta}_T}^{(m)}(1), \dots, \bar{\rho}_{T\hat{\theta}_T}^{(m)}(m-q))'$  and  $\bar{\rho}_{T\theta_0}^{(m)}$  are asymptotically equivalent, and  $\sqrt{T}\bar{\rho}_{T\theta_0}^{(m)}$  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 A.1–A.4 in the Appendix and with  $\hat{\theta}_T$  satisfying (3) and (4),

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

and

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

The theorem is proved reasoning as in the seminal paper by Brown, Durbin, and Evans (1975). First, we justify (7) using standard recursive projections algebra and Proposition 1 expansion. Then, we show that the projection  $\bar{\rho}_{T\theta_0}^{(m)}$  can be approximated by a projection on  $\xi_{\theta_0}$ . Finally, we justify the central limit theorem (CLT) for  $\bar{\rho}_{T\theta_0}^{(m)}$  from that for  $\tilde{\rho}_{T\theta_0}^{(m)}$  and the projection built with the true  $\xi_{\theta_0}$ . This result forms a basis for implementing asymptotic specification tests of different nature based on the asymptotically iid sequence  $\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$ , 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

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

are indexed by a square summable weight function  $w: \mathbb{N} \rightarrow \mathbb{R}^+$ . It follows from Theorem 1 that, under  $H_0$ ,

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

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

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

$$H_{1T}: \rho_{\theta_0}(j) = \frac{r(j)}{\sqrt{T}} + \frac{\tau_T(j)}{T} \quad \text{for all } j = 1, 2, \dots,$$

where we assume that (6) holds under  $H_{1T}$ , and  $r$  and  $\tau_T$  are square summable such that  $\rho_{\theta_0}$  is a positive semidefinite 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

$$\varepsilon_{\theta t} = \Psi_{T\theta}(L)u_{\theta t}, \quad (9)$$

where  $\{u_{\theta t}\}_{t \in \mathbb{Z}}$  are uncorrelated with higher-order dependence characterized by  $a_{\theta_0}^{(i,j)}$  defined in (2) and

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

with  $\sum_{j=1}^{\infty} \psi_{T\theta}(j)^2 < \infty$  for all  $\theta$  and  $T$ , and  $\lim_{T \rightarrow \infty} \psi_{T\theta_0}(j) = r(j)$ . The function  $\Psi_{T\theta}$  can be either parametric or nonparametric. For instance, it may correspond to 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  $\bar{\rho}_{T\hat{\theta}_T}^{(m)}$  under  $H_{1T}$  define first the vector  $\bar{\mathbf{h}}_{\theta}^{(m)} = (\bar{h}_{\theta}^{(m)}(1), \dots, \bar{h}_{\theta}^{(m)}(m-q))'$  of projected and standardized autocorrelation drifts, where

$\bar{h}_\theta^{(m)}(j) = \mathcal{L}^{(m)} h_\theta^{(m)}(j)$ , that is,

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

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

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

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

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

and

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

### 3.1 Box–Pierce Type Tests

Consider the uniform weights  $w(j) = 1_{\{j \leq s\}}$ ,  $1 \leq s \leq m - q$ , for each  $j \in \mathbb{N}$ , which correspond 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 (1970) statistic  $\hat{B}_{T\hat{\theta}_T}(s)$ , with

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

Box and Pierce (1970) showed that, when  $\{\varepsilon_{\theta_0 t}\}_{t \in \mathbb{Z}}$  are iid, and  $s$  is increasing with  $T$ , in particular  $s = o(T^{1/2})$ ,  $\hat{B}_{T\hat{\theta}_T}(s) \underset{asy}{\sim} \chi_{(s-q)}^2$ . However, Box–Pierce test is unable to detect nonparametric local alternatives like  $H_{1T}$ , since

$$\frac{\hat{B}_{T\hat{\theta}_T}(s) - s}{\sqrt{2s}} \rightarrow_d N(0, 1) \quad \text{when } \frac{1}{s} + \frac{s}{T^{1/2}} \rightarrow 0,$$

under  $H_{1T}$ , thought it can detect local alternatives of this form converging to the null at the slower rate  $\sqrt[4]{s}/\sqrt{T}$ . See Hong (1996). When  $s$  remains fixed,  $\hat{B}_{T\hat{\theta}_T}(s)$  has a limiting null distribution depending on the parameter vector  $\theta_0$  and other unknown features of the underlying data generating process. This fact prevents from using Box–Pierce test when  $s$  is small. Therefore, fairly large  $s$  are required for good level accuracy, but the smaller  $s$ , the more powerful is the test to detect  $\sqrt[4]{s}/\sqrt{T}$  local alternatives. That is, there is a trade-off between good size and power properties.

On the other hand, the test statistic  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$  is asymptotically  $\chi_{(s)}^2$  distributed, for any  $m \geq s + q$ , and equivalent to the Gaussian LM test statistic, for  $m$  large, in the directions of parametric local alternatives when  $\Psi_{T\theta}$  in (9) is an autoregressive or moving average polynomial of order  $s$  and the innovations are iid, so that  $A_{\theta_0}^{(m)} = I_m$ . We state this result in the next proposition. Let  $\chi_{(n)}^2(\sum_{i=1}^n \lambda_i^2)$  denote a noncentered chi-squared random

variable with  $n$  degrees of freedom and noncentrality parameter  $\sum_{i=1}^n \lambda_i^2$ ; that is,  $\chi_{(n)}^2(\sum_{i=1}^n \lambda_i^2) \stackrel{d}{=} \sum_{j=1}^n (Z_j + \lambda_j)^2$ .

**Proposition 2.** Under the assumptions in Theorem 2, with  $A_{\theta_0}^{(m)} = I_m$  for all  $m$ , the test based on  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$  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  $j > s$ ,

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

when  $m = m_T \rightarrow \infty$  as  $T \rightarrow \infty$ ,  $m < T$ , where  $\bar{r}_{\theta_0}^{(\infty)}(j) = \mathcal{L}^{(\infty)} r(j)$ .

While, under Gaussianity, the classical Box–Pierce test  $\hat{B}_{T\theta_0}(s)$  is asymptotically equivalent to the LM test of simple hypotheses of lack of autocorrelation of the true innovations in the direction of AR( $s$ ) or MA( $s$ ) alternatives, this proposition states a similar result for composite hypothesis and a Portman-teau test based on projected residuals autocorrelations. That is, the test  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$  is asymptotically equivalent to the LM test and, hence, optimal, for testing lack of serial correlation of residuals up to a finite order  $s$ , without need to resort to fully efficient maximum likelihood estimates (MLE) of  $\theta_0$ . This also points out that the procedure applied to eliminate the estimation effect in the sample autocorrelations  $\tilde{\rho}_{T\hat{\theta}_T}^{(m)}$  does not neglect any important information asymptotically when the dimension of the recursive projections  $m$  is large enough. However, the parameter  $m$  does not affect the rate of convergence and the asymptotic limit only depends on the fixed number of transformed autocorrelations  $s$  used in the test.

### 3.2 $T\rho$ -Process Type Tests

The sequence of weights  $w(j) = 1/j^2$  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 resemble the spectral representation of the classical  $T\rho$ -process test based on the Cramér–von Mises criterion, that is,

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

See, for example, 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 not faster, both  $\bar{T}_{T\hat{\theta}_T}^{(m)}$  and the unfeasible  $\hat{T}_{T\theta_0}$  are asymptotically distributed as  $\sum_{j=1}^{\infty} Z_j^2/j^2$  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  $H_{1T}$ ,

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

when  $m = m_T \rightarrow \infty$  as  $T \rightarrow \infty$ ,  $m < T$ , where  $\tilde{r}_{\theta_0}^{(\infty)}(j) = \mathcal{L}^{(\infty)} r(j)$  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  $\tilde{T}_{T\hat{\theta}_T}^{(m)}$  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  $H_0$ .

#### 4. MONTE CARLO SIMULATIONS

In this section, we evaluate the performance of our proposals comparing simulated percentage of rejections under  $H_0$  and  $H_1$  of alternative tests based on residuals sample autocorrelations. The comparison is made in the context of ARFIMA designs with possibly heteroscedastic innovations

$$\varepsilon_{\theta_0 t} = u_t(1 + \alpha_1 \varepsilon_{\theta_0 t-1}^2)^{1/2}, \quad (11)$$

where  $\alpha_1 \in \{0, 0.4\}$  and  $u_t \sim \text{iid } N(0, 1)$ . When  $\alpha_1 = 0$ ,  $\varepsilon_{\theta_0 t} = u_t$  is iid, while if  $\alpha_1 > 0$  then  $\varepsilon_{\theta_0 t}$  follows an ARCH(1) model, so that is a martingale difference sequence with serial dependence in second moments. We consider sample sizes  $T = 100$  and 400 and 10,000 replications in each experiment. Parameters are estimated using Whittle's likelihood method; see, for example, Velasco and Robinson (2000). We consider two null models: AR(1) with  $\zeta_{\theta_0}(j) = -\delta_{10}^{j-1}$  and ARFIMA(0,  $d$ , 0) with  $\zeta_{\theta_0}(j) = -j^{-1}$ , with parameter values  $\delta_{10} = 0.8$  and  $d_0 = 0.4$ , respectively. Results for other parameter values as well as for the MA(1) model are similar.

We concentrate on our new  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  test, which uses critical values from a chi squared distribution with  $s$  degrees of freedom. Nominal level of all tests is 5%. For the sake of comparison, we use values for  $s$  from 1 up to 15 and 24 for  $T = 100$  and 400, respectively. We set  $m = s + q$  in order to avoid studentization of unnecessary residual autocorrelations and speeding up computations. It is not reasonable to use a larger  $m$  than needed when  $A_{\theta_0}^{(m)}$  must be estimated. We use three different estimates of  $A_{\theta_0}^{(m)}$ , the first two using information on its true structure. These are  $\hat{A}_{T\hat{\theta}_T}^{(m)} = I_m$ ,  $\hat{A}_{T\hat{\theta}_T}^{(m)} = \text{diag}\{\hat{a}_{T\hat{\theta}_T}^{(1,1)}, \dots, \hat{a}_{T\hat{\theta}_T}^{(m,m)}\} / \hat{\gamma}_{T\hat{\theta}_T}(0)^2$  with  $\hat{a}_{T\hat{\theta}_T}^{(j,j)} = T^{-1} \sum_{t=1+j}^T \varepsilon_{\theta_0 t}^2 \varepsilon_{\theta_0 t-j}^2$ , and the unrestricted estimate of  $A_{\theta_0}^{(m)}$  used by Lobato, Nankervis, and Savin (2002) with preliminary bandwidth  $n = 2(T/100)^{1/3}$ , no prewhitening and Barlett's kernel. The first estimate exploits a possible asymptotic iid property of sample autocorrelations, while the second estimate is consistent under (11) with  $\alpha_1 \neq 0$ . The Supplemental Materials Appendix B justifies the consistency of these estimates based on residuals.

We compare the new tests with the popular correction by Ljung and Box (1978) of the classical Box–Pierce (B–P) test,  $\hat{B}_{T\hat{\theta}_T}(s)$ , under iid innovations (with a  $\chi_{(s-q)}^2$  approximation) and with Francq, Roy, and Zakoian (2005) test  $\tilde{Q}_s$  under general forms of nonlinear dependence. Francq, Roy, and Zakoian (2005) propose a Portmanteau test where residuals sample autocorrelations are scaled by a consistent estimate of their variance

and covariance matrix, which is derived from the joint distribution of the least squares estimator of the parameters and the sample autocorrelations of the true innovations. For this standardization we use the same  $\hat{A}_{\hat{\theta}_T}^{(m)}$  estimates as for  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$ , either diagonal or unrestricted ones, and employ Imhof's (1961) method to obtain  $p$ -values.

Figures 1 and 2 report the simulated size when the innovations are iid ( $\alpha_1 = 0$ ) and serially dependent according to an ARCH process with  $\alpha_1 = 0.4$ , respectively. As for the classical Box–Pierce test, we only report results for  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  with  $\hat{A}_{T\hat{\theta}_T}^{(m)} = I_m$  when innovations are iid. We can observe in Figure 1 that Box–Pierce test rejections increase steadily with  $s$  for both sample sizes, and the highest size accuracy is achieved for choices of  $s$  about  $T^{1/2}$ , which is accepted to provide fairly good level accuracy for moderate  $T$  in standard models. Tests based on recursive projections with  $\hat{A}_{T\hat{\theta}_T}^{(m)} = I_m$  underreject for  $T = 100$ , but provide size levels very close to nominal for  $T = 400$ . Unrestricted estimation of  $A_{\theta_0}^{(m)}$  introduces some size distortions when  $s$  is large, possibly due to the need of inverting a matrix of increasing dimension, but for  $s$  smaller than 5 the size accuracy is very high. The diagonality restriction, on the other hand, provides very stable and accurate simulated size for all  $s$ , both for iid and heteroscedastic innovations.  $\tilde{Q}_s$  tests using these diagonal estimates display reasonable behavior under the null for ARCH innovations, but rejections increase monotonically with  $s$  as with the classical Box–Pierce test. However, unrestricted estimation of  $A_{\theta_0}^{(m)}$  seems to introduce some noticeable underrejection in  $\tilde{Q}_s$  tests.

Figure 3 reports the percentage of rejections under the following alternative hypotheses:

$$(a) \quad H_0 : \text{AR}(1) \quad \text{vs} \quad H_1 : \text{ARMA}(1, 1),$$

$$\text{with } \delta_{10} = 0 \text{ and } \eta_{10} = 0.5$$

$$(b) \quad H_0 : \text{AR}(1) \quad \text{vs} \quad H_1 : \text{ARFIMA}(1, d, 0),$$

$$\text{with } \delta_{10} = 0 \text{ and } d_0 = 0.4.$$

Innovations are again iid and ARCH processes according to (11). We consider in Figure 3 the same tests and estimates  $\hat{A}_{T\hat{\theta}_T}^{(m)}$  for both recursive tests and  $\tilde{Q}_s$ , but only report results for the smaller sample size  $T = 100$ . It is confirmed that the classical B–P test detects better these alternatives the smaller  $s$  is, specially for the short memory ARMA alternative. There is a clear trade off between size accuracy and power for the B–P and  $\tilde{Q}_s$  tests. Our new tests  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  based on recursive autocorrelations exhibit good power performance for all  $s$  considered and both types of alternatives, together with the ability of considering small  $s$  values with the Type I error under control. Robust estimation of  $A_{\theta_0}^{(m)}$  introduces some costs in terms of power, but diagonal estimates provide a good compromise in terms of power and size for both designs considered.

#### 5. 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

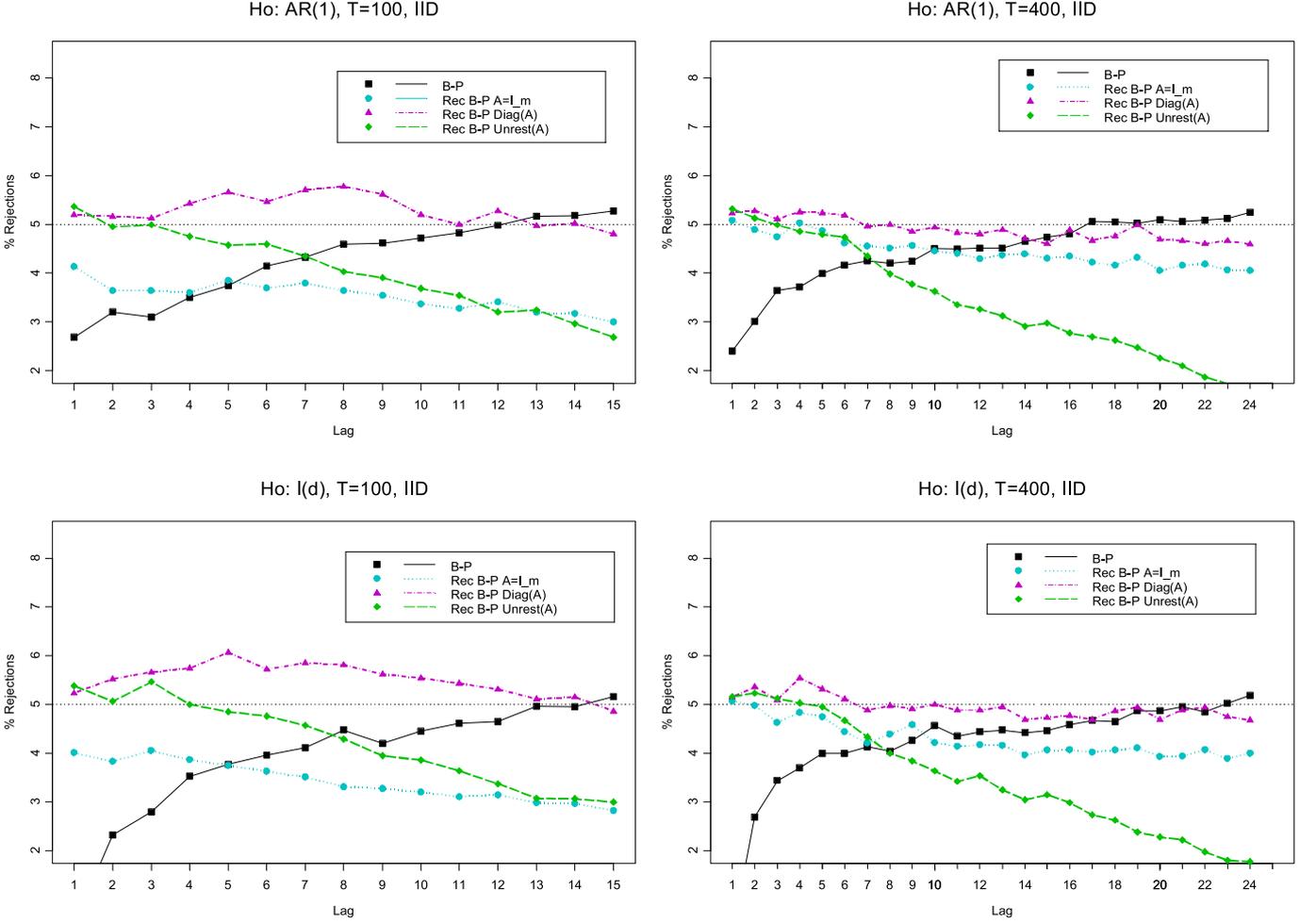


Figure 1. Percentage of rejections of Portmanteau tests in terms of the lag  $s$  under the null of correct specification. B–P is the classical Box–Pierce test  $\hat{B}_{T\hat{\theta}_T}(s)$  compared with a  $\chi^2_{(s-q)}$  critical value, Rec B–P are tests  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$  based on recursive projected residuals autocorrelations compared to a  $\chi^2_{(s)}$ . Nominal level is 5%.  $A = I_m$  means  $\hat{A}_T^{(m)} = I_m$ ,  $\text{Diag}(A)$  means  $\hat{A}_T^{(m)}$  diagonal and  $\text{Unrest}(A)$  means an unrestricted estimate of  $\hat{A}_T^{(m)}$ .  $m$  is set to  $s + q$  for each  $s$ . Models are AR(1) with  $\delta_{10} = 0.8$  in top panels and ARFIMA(0,  $d_0$ , 0) with  $d_0 = 0.4$  in bottom panels. Innovations are iid standard normals. Sample sizes are  $T = 100$  on left panels and 400 in right panels. The online version of this figure is in color.

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, which are all fitted using Whittle estimation. For checking the fit of every model we use the Box–Pierce test based on transformed residuals autocorrelations,  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$ , for  $s = 1, 2, 3, 5$  and Box and Pierce (1970) test,  $\hat{B}_{T\hat{\theta}_T}(s)$ , for  $s = 5, 10, 20, 30$ , which include all the usual lag choices in similar applications as supported by our simulations, given that  $T = 226$ . We only report the analysis with  $\hat{A}_{\hat{\theta}_T}^{(m)} = I_m$  for easier comparison with nontransformed 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 1. All models with only one short run parameter (apart from the memory parameter  $d$ ) are strongly rejected by the recursive Portmanteau test  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$  for

all lags  $s = 1, \dots, 5$ . However, 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 the new tests. However, the usual Box–Pierce test only provides strong evidence against the ARIMA(0, 1, 1) and (2, 1, 0) models.

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 that these confidence bands are inconsistent when parameters are estimated. In Figures 4 and 5, 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,  $\bar{\rho}_{T\hat{\theta}_T}^{(m)}$ , but diagnosis based on the untransformed

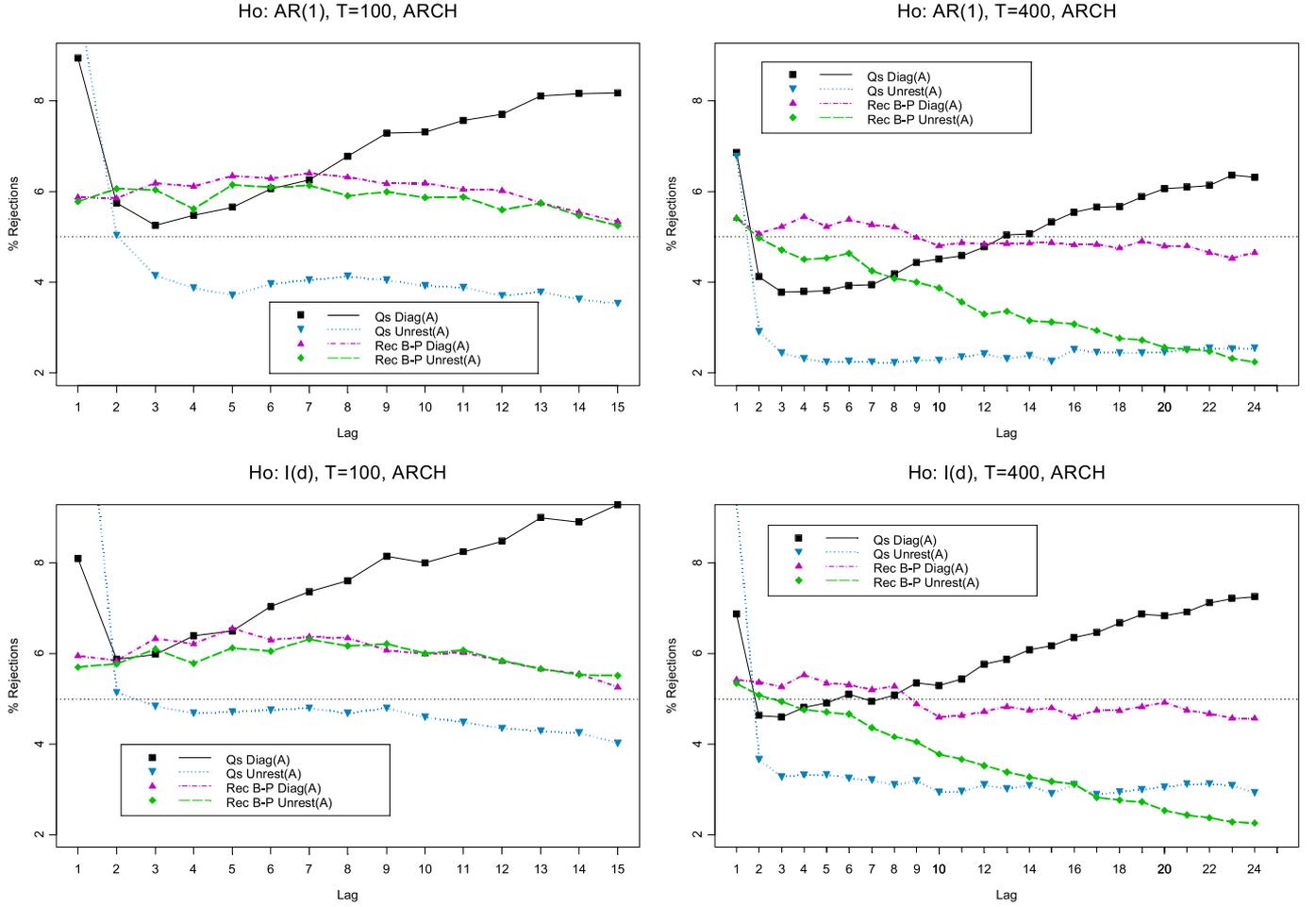


Figure 2. Percentage of rejections of Francq, Roy, and Zakoian (2005) Portmanteau test  $\tilde{Q}_s$  and of new tests  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  based on recursive projected residuals autocorrelations and compared to  $\chi_{(s)}^2$  under the null of correct specification.  $p$ -values for  $\tilde{Q}_s$  tests are computed by Imhof's method. Nominal level is 5%. Diag(A) means  $\hat{A}_T^{(m)}$  diagonal and Unrest(A) means an unrestricted estimate of  $\hat{A}_T^{(m)}$ . Models, sample sizes, and choice of  $m$  are as in Figure 1. Innovations are ARCH(1) processes with Gaussian shocks and  $\alpha_1 = 0.4$ . The online version of this figure is in color.

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.

## APPENDIX: PROOFS AND AUXILIARY RESULTS

In this appendix we present sufficient assumptions for the proof of our results and some auxiliary results that can be of independent interest. First we introduce some notation. Henceforth, for any generic function  $g_\theta$  indexed by parameters  $\theta \in \Theta$ ,

$$\dot{g}_\theta = \frac{\partial g_\theta}{\partial \theta'} \quad \text{and} \quad \ddot{g}_\theta = \frac{\partial^2 g_\theta}{\partial \theta \partial \theta'}.$$

Given the model  $\varepsilon_{\theta t} = \varphi_\theta(L)U_\theta(\mathbf{X}_t)$  denote  $\varepsilon_t = \varepsilon_{\theta t}$ .

*Assumption A.1.*  $(\mathbf{X}_t', \varepsilon_t)'$  is strictly stationary,  $\varepsilon_t$  is zero mean,  $E[\varepsilon_t^{4+2\delta}] < \infty$  for some  $\delta > 0$  and  $(\mathbf{X}_t', \varepsilon_t)'$  is strong mixing with coefficients  $\alpha_j$  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  $\sigma$  fields generated by  $\{(\mathbf{X}_t', \varepsilon_t)', t \leq 0\}$  and  $\{(\mathbf{X}_t', \varepsilon_t)', t \geq j\}$ .

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

*Assumption A.3.*  $\phi_\theta(z)$  is twice differentiable in  $\theta$ ,  $\phi_\theta(0) = 1$  with  $|\phi_\theta(j)| + \|\dot{\phi}_\theta(j)\| + \|\ddot{\phi}_\theta(j)\| \leq \phi_j$ , uniformly in  $\theta \in \Theta$ , such that  $\sum_{j=0}^{\infty} \phi_j < \infty$ .

*Assumption A.4.* For  $m > q$ ,

$$\sum_{j=m-q+1}^m \xi_{\theta_0}(j)' \xi_{\theta_0}(j)$$

is positive definite.

*Remark.* Similar conditions to Assumption A.4 are always required when using recursive residuals in different contexts, being more restrictive than the absence of multicollinearity assumption when applying ordinary least squares. See, for example, Brown, Durbin, and Evans (1975), Khmaladze (1981) or Delgado, Hidalgo, and Velasco (2005). 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

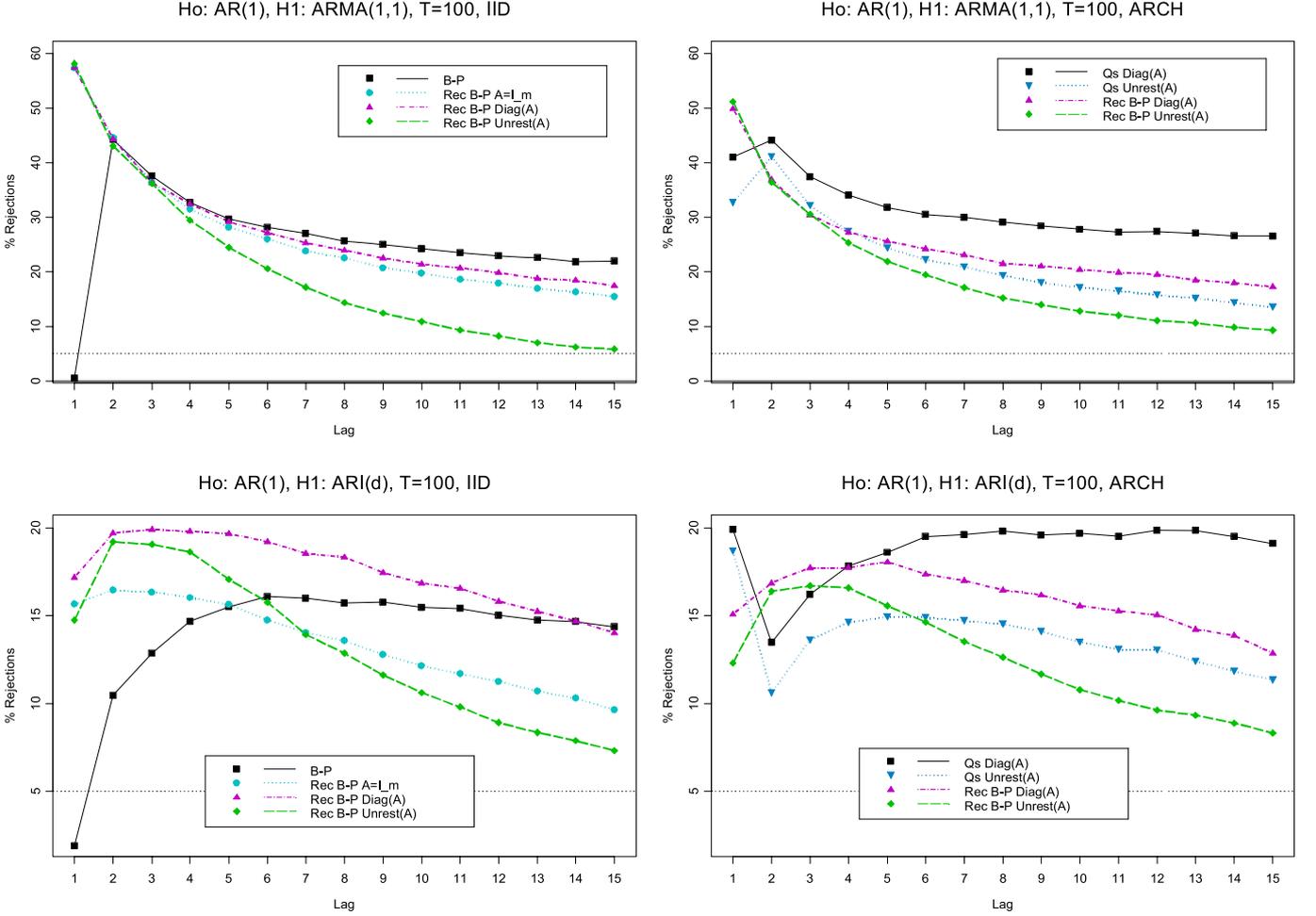


Figure 3. Percentage of rejections of classical Box–Pierce test  $\hat{B}_{T\hat{\theta}_T}(s)$ , Francq, Roy, and Zakoian (2005) Portmanteau test  $\tilde{Q}_s$  and of new tests  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  based on recursive projected residuals autocorrelations under the alternative of incorrect specification. Asymptotic critical values and  $p$ -values are obtained as in Figures 1 and 2, as well as estimation of  $\hat{A}_T^{(m)}$  and choice of  $m$ . Sample size is  $T = 100$ . Nominal level is 5%. The null model is AR(1), in top panels the data is generated under an ARMA(1, 1) model with  $\delta_{10} = 0$  and  $\eta_{10} = 0.5$ , in bottom panels the data follows an ARFIMA(1,  $d_0$ , 0) model with  $\delta_{10} = 0$  and  $d_0 = 0.4$ . Innovations are iid standard normals on left panels and ARCH(1) processes with Gaussian shocks and  $\alpha_1 = 0.4$  on the right panels. The online version of this figure is in color.

AR(1) to a strong white noise, as Francq, Roy, and Zakoian (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 A.4. However,  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  exhibits in this case also an excellent level accuracy for the two sample sizes considered, in line with the results for other parameter values. The assumption could be relaxed by using generalized inverses when computing the recursive residuals, as proposed by Tsigroshvili (1998) in the related context of constructing chi-squared tests using innovation martingales in the classical goodness-of-fit problem. Duchesne and Francq (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

$$\hat{\rho}_{T\hat{\theta}_T}^{(m)} = \hat{\rho}_{T\theta_0}^{(m)} + \xi_{\theta_0}^{(m)}(\hat{\theta}_T - \theta_0) + o_p(T^{-1/2}), \quad (\text{A.1})$$

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

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

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

$$\frac{\partial}{\partial\theta'} \hat{\rho}_{T\theta_0}^{(m)}(j) = \frac{\partial \hat{\gamma}_{T\theta_0}(j)/\partial\theta'}{\hat{\gamma}_{T\theta_0}(0)} - \frac{\hat{\gamma}_{T\theta_0}(j)}{\hat{\gamma}_{T\theta_0}(0)} \frac{\partial \hat{\gamma}_{T\theta_0}(0)/\partial\theta'}{\hat{\gamma}_{T\theta_0}(0)}.$$

The mean correction in  $\hat{\gamma}_{T\theta_0}(j)$  has no asymptotic effect, since  $\hat{\gamma}_{T\theta_0}(j) = \gamma_{T\theta_0}(j) + O_p(T^{-1})$ , where  $\gamma_{T\theta_0}(j) = T^{-1} \sum_{t=j+1}^T \varepsilon_{\theta t} \varepsilon_{\theta t-j}$ ,  $j \in \mathbb{Z}$ , because  $\bar{\varepsilon}_{\theta_0 T} = O_p(T^{-1/2})$  under Assumption A.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  $\partial \gamma_{T\theta_0}(0)/\partial\theta' = O_p(1)$  under Assumptions A.1–A.3, as we now show, we conclude that the normalization of

Table 1. Chemical data,  $T = 226$ . Goodness-of-fit analysis for Chemical Process Temperature Readings based on fractionally integrated models

| Model                          | BIC    | $\hat{d}$<br>(s.e.) | $\bar{T}_{\hat{\theta}_T}^{(m)}$ | $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$ |          |          |          |          | $\hat{B}_{T\hat{\theta}_T}(s)$ |         |       |    |    |
|--------------------------------|--------|---------------------|----------------------------------|--------------------------------------|----------|----------|----------|----------|--------------------------------|---------|-------|----|----|
|                                |        |                     |                                  | $s =$                                | 1        | 2        | 3        | 5        | $s =$                          | 5       | 10    | 20 | 30 |
| $H_0 : \text{ARFIMA}(p, d, q)$ |        |                     |                                  |                                      |          |          |          |          |                                |         |       |    |    |
| (0, $d$ , 0)                   | 3.7949 | 0.871<br>(0.052)    | 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<br>(0.065)    | 0.73**                           | 6.88***                              | 6.92**   | 8.32**   | 9.71*    | 9.61**   | 10.87                          | 12.28   | 13.41 |    |    |
| (2, $d$ , 0)                   | 3.7101 | 1.227<br>(0.075)    | 0.17                             | 1.50                                 | 1.54     | 2.14     | 3.57     | 3.16     | 3.54                           | 4.71    | 5.81  |    |    |
| (0, $d$ , 1)                   | 3.7120 | 1.249<br>(0.159)    | 0.70**                           | 6.34**                               | 8.34**   | 8.83**   | 9.32*    | 8.00**   | 8.82                           | 9.71    | 10.76 |    |    |
| (0, $d$ , 2)                   | 3.7054 | 1.313<br>(0.126)    | 0.31                             | 1.53                                 | 1.83     | 2.00     | 2.08     | 1.55     | 1.87                           | 2.96    | 4.33  |    |    |
| (1, $d$ , 1)                   | 3.7133 | 1.326<br>(0.144)    | 0.28                             | 2.50                                 | 3.48     | 3.69     | 3.88     | 3.23     | 3.54                           | 4.51    | 5.70  |    |    |
| $H_0 : \text{ARIMA}(p, 1, q)$  |        |                     |                                  |                                      |          |          |          |          |                                |         |       |    |    |
| (1, 1, 0)                      | 3.7015 |                     | 0.21                             | 0.44                                 | 3.53     | 8.44**   | 12.43**  | 11.58**  | 14.37                          | 16.13   | 17.45 |    |    |
| (2, 1, 0)                      | 3.7236 |                     | 0.50**                           | 3.06*                                | 8.09**   | 11.15**  | 14.70**  | 12.80*** | 16.25**                        | 18.34   | 20.03 |    |    |
| (0, 1, 1)                      | 3.7162 |                     | 0.88***                          | 6.79***                              | 12.32*** | 14.93*** | 17.60*** | 15.01*** | 18.51**                        | 20.34   | 21.80 |    |    |
| (0, 1, 2)                      | 3.7104 |                     | 0.61**                           | 5.16**                               | 7.54**   | 9.10**   | 10.70*   | 9.06**   | 11.47                          | 13.45   | 14.86 |    |    |
| (1, 1, 1)                      | 3.7243 |                     | 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.  $\hat{A}_{T\hat{\theta}_T}^{(m)} = I_m$ .

$\hat{\rho}_{T\hat{\theta}_T}^{(m)}$  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

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

where  $\dot{\varepsilon}_{\theta_0 t} = \varphi_{\theta_0}(L) \dot{U}_{\theta_0}(\mathbf{X}_t) + \dot{\varphi}_{\theta_0}(L) U_{\theta_0}(\mathbf{X}_t)$ . Setting  $\zeta_{\theta_0}(j) = \zeta_{\theta_0}^{(1)}(j) + \zeta_{\theta_0}^{(2)}(j)$  where  $\zeta_{\theta_0}^{(i)}(j) := \lim_{T \rightarrow \infty} E[A_{T,i}(j)]$ , we wish to show that  $A_{T,i}(j) = \zeta_{\theta_0}^{(i)}(j) + o_p(1)$ ,  $i = 1, 2, j = 1, 2, \dots$ . We first show that

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

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

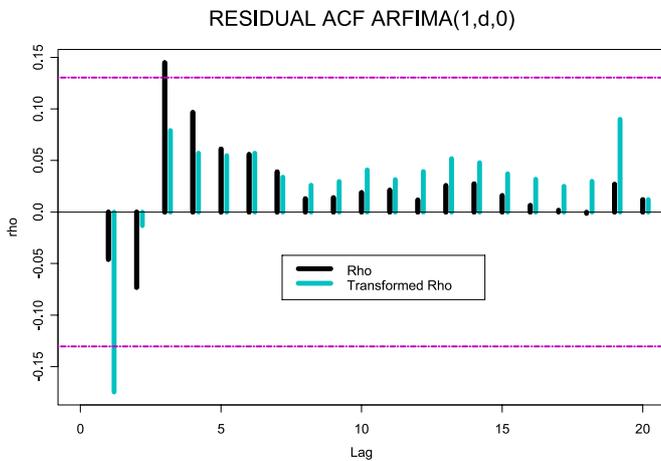


Figure 4. Autocorrelations (Rho) and projected autocorrelations (Transformed Rho) of ARFIMA(1,  $d$ , 0) residuals for Chemical Series C data,  $T = 226$ , for lags  $s = 1, \dots, 20$ . Confidence bands are plotted at  $\pm 2/\sqrt{T}$ .

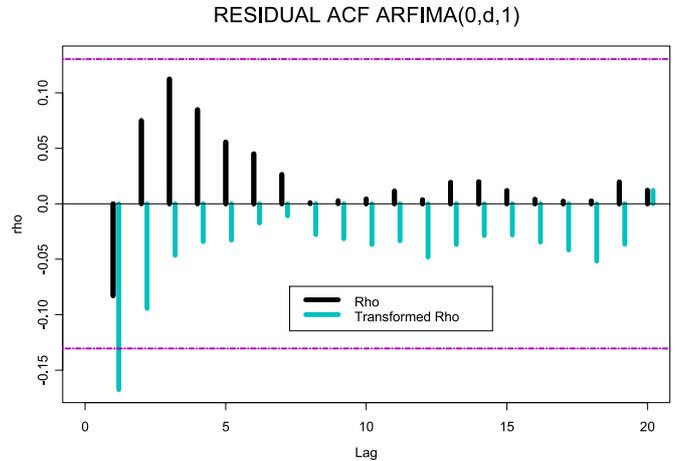


Figure 5. Autocorrelations (Rho) and projected autocorrelations (Transformed Rho) of ARFIMA(0,  $d$ , 1) residuals for Chemical Series C data,  $T = 226$ , for lags  $s = 1, \dots, 20$ . Confidence bands are plotted at  $\pm 2/\sqrt{T}$ .

$E\|A_{T,1}(j) - E[A_{T,1}(j)]\|^2$  is

$$\begin{aligned} & \frac{1}{T^2} \sum_{t=1+j}^T E[e(t, t-j)'e(t, t-j)] \\ & + \frac{2}{T^2} \sum_{t=1+j}^T \sum_{t-n-j \leq r < t} E[e(t, t-j)'e(r, r-j)] \\ & + \frac{2}{T^2} \sum_{t=1+j}^T \sum_{1+j \leq r < t-n-j} E[e(t, t-j)'e(r, r-j)]. \quad (\text{A.2}) \end{aligned}$$

The first two terms of (A.2) are  $O(T^{-1}) = o(1)$  since involve at most  $T+n$  elements with bounded absolute expectation because by Assumptions A.1–A.3 and Minkowski and Hölder inequalities,

$$\begin{aligned} E\|\dot{\varepsilon}_{\theta t}\|^4 & \leq E\|\dot{\varphi}_{\theta}(L)U_{\theta}(\mathbf{X}_t)\|^4 + E\|\varphi_{\theta}(L)\dot{U}_{\theta}(\mathbf{X}_t)\|^4 \\ & \leq 2 \left( \sum_{j=1}^{\infty} |\phi_j| \right)^4 E|U_{*}(\mathbf{X}_t)|^4 < \infty. \quad (\text{A.3}) \end{aligned}$$

Now write  $\dot{\varepsilon}_{\theta t} = \dot{\varepsilon}_{\theta t}^{(0,n)} + \dot{\varepsilon}_{\theta t}^{(n+1,\infty)}$ ,  $\dot{\varepsilon}_{\theta t}^{(r,s)} = \sum_{j=r}^s (\varphi_{\theta,j}\dot{U}_{\theta}(\mathbf{X}_{t-j}) + \dot{\varphi}_{\theta,j}U_{\theta}(\mathbf{X}_{t-j}))$  and  $e_0^s(t, t-j)' = \varepsilon_{\theta 0t}\dot{\varepsilon}_{\theta 0t-j}^{(r,s)}$ . Then  $e_0^n(t, t-j)$  is mixing with mixing coefficients  $\beta_k \leq \alpha_{k-j-n}$ . The third term in (A.2) is then equal to

$$\begin{aligned} & \frac{2}{T^2} \sum_{t=1+j}^T \sum_{r < t-n-j} E[e_0^n(t, t-j)'e(r, r-j)] \\ & + \frac{2}{T^2} \sum_{t=1+j}^T \sum_{r < t-n-j} E[e_{n+1}^{\infty}(t, t-j)'e_j(r, r-j)]. \quad (\text{A.4}) \end{aligned}$$

Using Assumptions A.2 and A.3, the first term in (A.4) is bounded in absolute value by

$$\begin{aligned} & \frac{C}{T^2} (E\|e_0^n(t, t-j)\|^{2+\delta} E\|e(r, r-j)\|^{2+\delta})^{1/(2+\delta)} \\ & \times \sum_{t=1+j}^T \sum_{r < t-n-j} \alpha_{t-n-j-r}^{\delta/(2+\delta)} = O(T^{-1}) = o(1) \end{aligned}$$

by Roussas and Ioannides (1987) and Cauchy–Schwarz inequality.

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

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

and

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

because of the same reasoning as for (A.3). Then we conclude that the second term in (A.4) and the third term in (A.2) are  $o_p(1)$ .

On the other hand  $\xi_{\theta_0}^{(2)}(j)$  is  $\lim_{T \rightarrow \infty} E[A_{T,2}(j)] = E[\dot{\varepsilon}_{\theta_0 t} \varepsilon_{\theta_0 t-j}]$ , which is different from zero if  $\varphi_{\theta}(L)$  contains lags and/or if  $U_{\theta}(\mathbf{X}_t)$  contains lagged nonstrictly exogenous explanatory variables. The proof that  $E\|A_{T,2}(j) - E[A_{T,2}(j)]\|^2 = o(1)$  follows as that for  $A_{T,1}$ .

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

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

where  $\ddot{\varepsilon}_{\theta t} = \varphi_{\theta}(L)\dot{U}_{\theta}(\mathbf{X}_t) + \dot{\varphi}_{\theta}(L)\dot{U}_{\theta}(\mathbf{X}_t) + \dot{U}_{\theta}(\mathbf{X}_t)\dot{\varphi}_{\theta}(L) + \ddot{\varphi}_{\theta}(L)U_{\theta}(\mathbf{X}_t)$ , and we can show that  $\ddot{\rho}_{T\theta^*}^{(m)}(j) = O_p(1)$ ,  $j = 1, \dots, m$ , since  $E[\varepsilon_{\theta^* t}^2] + E\|\dot{\varepsilon}_{\theta^* t}\|^2 + E\|\ddot{\varepsilon}_{\theta^* t}\|^2 < \infty$  using Assumptions A.2, A.3 and similar techniques.

**Proof of Theorem 1**

We first find using standard algebra and Proposition 1 that  $\tilde{\rho}_{T\hat{\theta}_T}^{(m)} = \tilde{\rho}_{T\theta_0}^{(m)} + o_p(T^{-1/2})$ , because from Assumption A.4,  $\tilde{\beta}_{T\hat{\theta}_T}^{(j)}[\tilde{\rho}_{T\hat{\theta}_T}^{(m)}] = \tilde{\beta}_{T\theta_0}^{(j)}[\tilde{\rho}_{T\theta_0}^{(m)}] + (\hat{\theta}_T - \theta_0) + o_p(T^{-1/2})$ ,  $j = 1, \dots, m - q$ , where  $\tilde{\beta}_{T\theta_0}^{(j)}[\rho] = (\sum_{\ell=j+1}^m \hat{\xi}_{T\theta}(\ell)' \hat{\xi}_{T\theta}(\ell))^{-1} \times \sum_{\ell=j+1}^m \hat{\xi}_{T\theta}(\ell)' \rho(\ell)$  and  $\hat{\xi}_{T\hat{\theta}_T}^{(j)} \rightarrow_p \xi_{\theta_0}(j)$ . This last statement can be proved with the methods in the proof of Proposition 1 noting that replacing  $\theta_0$  by  $\hat{\theta}_T$  in the definition of  $A_{T,1}(j)$  and  $A_{T,2}(j)$  has no asymptotic effect because of Assumptions A.2, A.3, and (3).

Then, similarly, we can show that  $\tilde{\rho}_{T\theta_0}^{(m)}(j) = \tilde{\rho}_{T\theta_0}^{(m)}(j) - \xi_{\theta_0}(j)\beta_{\theta_0}^{(j)} \times [\tilde{\rho}_{T\theta_0}^{(m)}] + o_p(T^{-1/2})$ , where  $\beta_{\theta_0}^{(j)}[\rho] = (\sum_{\ell=j+1}^m \xi_{\theta_0}(\ell)' \xi_{\theta_0}(\ell))^{-1} \times \sum_{\ell=j+1}^m \xi_{\theta_0}(\ell)' \rho(\ell)$ ,  $j = 1, \dots, m - q$ .

Finally, the CLT for  $\tilde{\rho}_{T\theta_0}^{(m)}$  follows from the CLT for  $\tilde{\rho}_{T\theta_0}^{(m)}$  under Assumptions A.1, (4),  $H_0$ , and from the fact that the projections  $\tilde{\rho}_{T\theta_0}^{(m)}$  are standardized by construction if  $\tilde{\rho}_{T\theta_0}^{(m)}$  is already standardized, that is,  $\tilde{\rho}_{T\theta_0}^{(m)}$  has asymptotic variance  $I_m$ , as can be showed by immediate calculations, see the Supplemental Materials Appendix A for details.

**Proof of Theorem 2**

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

**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  $\tilde{B}_{T\hat{\theta}_T}^{(m)}(s)$  where  $s$  depends on the alternative against the LM test is directed to.

Set the sequence of  $1 \times s$  row vectors  $d_s(j) = (1_{\{j=1\}}, \dots, 1_{\{j=s\}})$  for  $j = 1, 2, \dots, s$  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

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

where  $S_{T,1}(\theta) = \sum_{j=1}^{T-1} d_s(j)' \hat{\rho}_{T\theta}(j) = (\hat{\rho}_{T,\theta}(1), \dots, \hat{\rho}_{T,\theta}(s))'$  and  $H_T^{11}(\theta) = \{A_T^{-1}(\theta)\}_{11}$ , 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  $\delta_{\theta}$ , while  $\tilde{\theta}_T$  is any restricted

estimate of  $\theta_0$  that asymptotically behaves as the MLE, that is, admits this stochastic expansion under  $H_{1T}$ ,

$$T^{1/2}(\hat{\theta}_T - \theta_0) = -T^{1/2}A_{T,22}(\theta_0)^{-1}S_{T,2}(\theta_0) + o_p(1), \quad (\text{A.5})$$

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

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  $\mathbb{X}_j^k = (\zeta_\theta(j)', \dots, \zeta_\theta(k)')$ ,  $k \geq j$ ,

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

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

$$\Psi_{T,\hat{\theta}_T}^{(T-1)}(\hat{d}_{s,\hat{\theta}_T}^{(T-1)}) = \Psi_{T,\theta_0}^{(T-1)}(\hat{d}_{s,\theta_0}^{(T-1)}) + o_p(1) = \text{LM}_T(s) + o_p(1)$$

for any  $\sqrt{T}$ -consistent estimator  $\hat{\theta}_T$  of  $\theta_0$ , while the second equality follows because of (A.5) and noting that  $H_T^{11}(\theta)^{-1} = \sum_{j=1}^{T-1} \hat{d}_{s,\theta}^{(T-1)}(j)' \hat{d}_{s,\theta}^{(T-1)}(j) = I_s - \mathbb{X}_1^s (\mathbb{X}_1^s)' \mathbb{X}_1^s)^{-1} \mathbb{X}_1^s = I_s - (\zeta_\theta(1)', \dots, \zeta_\theta(s)')' (\sum_{k=1}^{T-1} \zeta_\theta(k)' \zeta_\theta(k))^{-1} (\zeta_\theta(1)', \dots, \zeta_\theta(s)')$ .

Second. We now show that the Box–Pierce statistic  $\bar{B}_{T\hat{\theta}_T}^{(m)}(s)$  provides an alternative way of computing  $\Psi_{T,\theta}^{(m)}(\hat{d}_{s,\theta}^{(m)})$  for any  $\theta$ ,  $m$  and  $s$ ,  $s + q \leq m \leq T - 1$ , under Assumption A.4, that is,

$$T \sum_{j=1}^s \bar{\rho}_{T,\theta}^{(m)}(j)^2 = \Psi_{T,\theta}^{(m)}(\hat{d}_{s,\theta}^{(m)}), \quad s = 1, \dots, m - q. \quad (\text{A.6})$$

For that we note that  $\sum_{j=1}^s \bar{\rho}_{T,\theta}^{(m)}(j)^2 = S_m^{(m)} - S_{m-s}^{(m)}$  using equation (5) in Brown, Durbin, and Evans (1975), where

$$S_{m-s}^{(m)} = \hat{\rho}_{T,\theta}^{(m)'} \left( \begin{pmatrix} \mathbf{0}_s & \mathbf{0} \\ \mathbf{0} & I_{m-s} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbb{X}_{s+1}^m \end{pmatrix} (\mathbb{X}_{s+1}^m)' \mathbb{X}_{s+1}^m)^{-1} \begin{pmatrix} \mathbf{0} & \mathbb{X}_{s+1}^m \end{pmatrix} \right) \hat{\rho}_{T,\theta}^{(m)}$$

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

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

$$\Psi_{T,\theta}^{(m)}(\hat{d}_s^{(m)}) = T \hat{\rho}_{T,\theta}^{(m)'} G_s^{(m)} \hat{\rho}_{T,\theta}^{(m)},$$

where  $G_s^{(m)} = P(m) V_s^{(m)'} H_s^{(m)} V_s^{(m)} P(m)$ , with  $V_s^{(m)} = (d_s(1)', \dots, d_s(m)')$ ,  $H_s^{(m)} = (I_s - \mathbb{X}_1^s (\mathbb{X}_1^s)' \mathbb{X}_1^s)^{-1} \mathbb{X}_1^s = I_m - \mathbb{X}_1^s (\mathbb{X}_1^s)' \mathbb{X}_1^s)^{-1} \mathbb{X}_1^s$ . Then we can use the facts that  $H_s^{(m)} = I_s + \mathbb{X}_1^s (\mathbb{X}_{s+1}^m)' \mathbb{X}_{s+1}^m)^{-1} \mathbb{X}_1^s$  and that  $\mathbb{X}_1^m \mathbb{X}_1^m = \mathbb{X}_1^s \mathbb{X}_1^s + \mathbb{X}_{s+1}^m \mathbb{X}_{s+1}^m$  to show that (A.6) follows after standard algebraic manipulations because  $G_s^{(m)}$  is equal to the difference in the weight matrices of  $S_m^{(m)}$  and  $S_{m-s}^{(m)}$ ; see the Supplementary Materials Appendix A for details.

Then the proof of the proposition is completed letting  $m$  to increase with  $T$  to show that  $\Psi_{T,\hat{\theta}_T}^{(m)}(\hat{d}_{s,\hat{\theta}_T}^{(m)}) = \Psi_{T,\hat{\theta}_T}^{(T-1)}(\hat{d}_{s,\hat{\theta}_T}^{(T-1)}) + o_p(1)$ , exploiting the definition of  $d_s(j)$  and that  $\zeta_\theta(j)$  is squared summable by Assumption A.3; see the proof of theorem 3 in Delgado and Velasco (2010).

### 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  $T$ ,  $\bar{T}_{\hat{\theta}_T}^{(m)} \xrightarrow{d} \sum_{j=1}^m (Z_j + \bar{r}_{\theta_0}^{(m)}(j))^2 / j^2$  as  $T \rightarrow \infty$  by Theorem 2. Finally, using theorem 4.2 in Billingsley (1999), we only need to show that

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

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

## SUPPLEMENTARY MATERIALS

**Appendixes:** *Appendix A:* contains an extended version of the technical Appendix of the paper. *Appendix B:* contains a discussion of consistent estimation of the asymptotic variance-covariance matrix of sample residuals autocorrelations. (Supplemental Material\_An asymptotically pivotal transform.pdf)

[Received April 2010. Revised November 2010.]

## REFERENCES

- Anderson, T. W. (1993), "Goodness of Fit Tests for Spectral Distributions," *The Annals of Statistics*, 21, 830–847. [950]
- Beran, J. (1995), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short- and Long-Memory ARIMA Models," *Journal of the Royal Statistical Society, Ser. B*, 57, 659–672. [951]
- Billingsley, P. (1999), *Convergence of Probability Measures* (2nd ed.), New York: Wiley. [957]
- Box, G. E. P., and Cox, D. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, Ser. B*, 26, 211–252. [947]
- 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. [946,950,952]
- Box, G. P., and Jenkins, M. J. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden-Day. [948,951]
- 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, Ser. B*, 37, 149–192. [948,949,953,957]
- Delgado, M. A., and Velasco, C. (2010), "Distribution-Free Test for Time Series Model Specification," *Journal of Econometrics*, 155, 128–137. [946,957]
- Delgado, M. A., Hidalgo, J., and Velasco, C. (2005), "Distribution Free Goodness-of-Fit Tests for Linear Processes," *The Annals of Statistics*, 33, 2568–2609. [947,948,953]
- Duchesne, P., and C. Francq (2008), "On Diagnostic Checking Time Series Models With Pormanteau Test Statistics Based on Generalized Inverses and {2}-Inverses," in *COMSTAT 2008, Proceedings in Computational Statistics*, ed. P. Brito, Heidelberg: Physica-Verlag, pp. 143–154. [954]
- Durbin, J. (1970), "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables," *Econometrica*, 38, 410–421. [946]
- Francq, C., Roy, R. and Zakoian, J.-M. (2005), "Diagnostic Checking in ARMA Models With Uncorrelated Errors," *Journal of the American Statistical Association*, 100, 532–544. [946,948,951,953,954]
- Hallin, M., and Jurečková, J. (1999), "Optimal Tests for Autoregressive Models Based on Autoregression Rank Scores," *The Annals of Statistics* 27, 1385–1414. [946]
- Hallin, M., and Puri, M. L. (1994), "Aligned Rank Tests for Linear Models With Autocorrelated Error Terms," *Journal of Multivariate Analysis*, 50, 175–237. [946]
- Hallin, M., Jurečková, J., and Koul, H. L. (1999), "Serial Autoregression and Regression Rank Score Statistics," in *Advances in Statistical Modeling and Inference; Essays in Honor of Kjell Doksum's 65th birthday*, ed. V. Nair, Singapore: World Scientific, pp. 335–362. [946]

- Hannan, E. J., and Heyde C. C. (1972), "On Limit Theorems for Quadratic Functions of Discrete Time Series," *Annals of Mathematical Statistics*, 43, 2058–2066. [946,947]
- Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica*, 64, 837–864. [950]
- Hosking, J. R. M. (1980), "Lagrange-Multiplier Tests of Time-Series Models," *Journal of the Royal Statistical Society, Ser. B*, 42, 170–181. [956]
- Hwang, S. Y., Basawa, I. V., and Reeves, J. (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. [946,948]
- Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426. [951]
- Khmaladze, E. V. (1981), "Martingale Approach to the Goodness of Fit Tests," *Probability Theory and Its Applications*, 26, 246–265. [948,953]
- Li, W. K. (1992), "On the Asymptotic Standard Errors of Residual Autocorrelations in Nonlinear Time Series Modelling," *Biometrika*, 79, 435–437. [946,948]
- Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of a Lack of Fit in Time Series Models," *Biometrika*, 65, 297–303. [951]
- Lobato, L., Nankervis, J. C., and Savin, N. E. (2002), "Testing for Zero Autocorrelation in the Presence of Statistical Dependence," *Econometric Theory*, 18, 730–743. [947,951]
- Romano, J. P., and Thombs, L. A. (1996), "Inference for Autocorrelations Under Weak Assumptions," *Journal of the American Statistical Association*, 91, 590–600. [946,947]
- Roussas, G. G., and D. A. Ioannidis (1987), "Moment Inequalities for Mixing Sequences of Random Variables," *Stochastic Analysis and Applications*, 5, 61–120. [956]
- Tsigroshvili, Z. (1998), "Some Notes on Goodness-of-Fit Tests and Innovation Martingales," *Proceedings of A. Razmadze Mathematical Institute*, 117, 89–102. [954]
- 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. [951,952]