Distribution-free tests for time series models specification[☆]

Miguel A. Delgado*, Carlos Velasco[1]

Departamento de Economía, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain

A R T I C L E   I N F O

Article history:
Received 19 November 2008
Received in revised form
24 August 2009
Accepted 28 September 2009
Available online 1 October 2009

JEL classification:
C12
C14
C22

Keywords:
Optimal tests
Residuals autocorrelation function
Specification tests
Time series models
Dynamic regression model

A B S T R A C T

We consider a class of time series specification tests based on quadratic forms of weighted sums of residuals autocorrelations. Asymptotically distribution-free tests in the presence of estimated parameters are obtained by suitably transforming the weights, which can be optimally chosen to maximize the power function when testing in the direction of local alternatives. We discuss in detail an asymptotically optimal distribution-free alternative to the popular Box–Pierce when testing in the direction of AR or MA alternatives. The performance of the test with small samples is studied by means of a Monte Carlo experiment.

1. Introduction

Let $\{X_t\}_{t=-\infty}^{\infty}$ be a covariance stationary time series with zero mean such that the filtered series

$$\varepsilon_t = \varphi(B) X_t, \quad t = 0, \pm 1, \pm 2, \ldots,$$

is a White Noise process, i.e. an uncorrelated process with zero mean and variance σ^2, where the linear filter φ is a prescribed function of the backshift operator B. We adopt the normalization $\varphi(0) = 1$. The series X_t might not be observable, as it happens when X_t are errors of a general regression model. The discussion of this case is postponed to Section 4.

Given a data set $\{X_t\}_{t=1}^{n}$, statistical inferences usually rely on a parametric specification of φ, which is described by means of a class of functions indexed by parameters taking values in a suitable parameter space $\Theta \subset \mathbb{R}^q$, say $\mathcal{J} = \{\varphi_\theta : \theta \in \Theta\}$, so that $\varphi_\theta(0) = 1$ for all $\theta \in \Theta$. The resulting statistical inferences are invalid when the putative specification is incorrect and, hence, testing the null hypothesis

$$H_0 : \varphi \in \mathcal{J}$$

is sorely needed before performing any statistical inference.

The null hypothesis of the correct specification can be written as

$$H_0 : \rho_{\theta_0}(j) = 0 \quad \text{for all } j \geq 1 \text{ and some } \theta_0 \in \Theta,$$

where $\rho_\theta(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda) f_\theta^{-1}(\lambda) \cos(\lambda j)\, d\lambda$ is the autocorrelation function of the residuals $\varepsilon_{\theta t} = \varphi_\theta(B) X_t$, $t = 0, \pm 1, \ldots$, $f(\lambda) = \left|\varphi\left(e^{i\lambda}\right)\right|^{-2}$ and $f_\theta(\lambda) = \left|\varphi_\theta\left(e^{i\lambda}\right)\right|^{-2}$ are the underlying normalized spectral density of $\{X_t\}_{t=-\infty}^{\infty}$ and its parametric specification counterpart, respectively.

A vast majority of test statistics for time series model specification are functions of some estimated residual autocorrelation (ERA) function, i.e. suitable estimates of ρ_{θ_0}. Portmanteau test statistics are quadratic forms of an ERA vector, e.g. Quenouille (1947), Box and Pierce (1970), Ljung and Box (1978) or Hosking (1980). Lagrange Multiplier (LM) test statistics, obtained after imposing parametric restrictions to a time series model, are quadratic forms of weighted sums of ERA vectors, e.g. Durbin (1970), Hosking (1978), or Robinson (1994) more recently.

Sometimes it is possible to compute the residuals $\{\varepsilon_{\theta t}\}_{t=1}^{n}$, and $\rho_\theta(j)$ can be estimated by the ERA, $\hat{\rho}_{n\theta}(j) = \hat{\gamma}_{n\theta}(j) / \hat{\gamma}_{n\theta}(0)$,

☆ We are very thankful to the Editor and two referees for helpful comments, which have led to improve the article. Research funded by a Spanish "Plan Nacional de I + D + i" grant number SEJ2007-62908/ECON.

* Corresponding author. Tel.: +34 916249804; fax: +34 916249875.
E-mail addresses: MiguelAngel.Delgado@uc3m.es, delgado@est-econ.uc3m.es (M.A. Delgado), Carlos.Velasco@uc3m.es (C. Velasco).
1 Tel.: +34 916249646.

where the sample autocovariance function of $\{\varepsilon_{\theta t}\}_{t=1}^{n}$ is $\hat{\gamma}_{n\theta}(j) = n^{-1}\sum_{t=j+1}^{n}(\varepsilon_{\theta t}-\bar{\varepsilon}_{\theta})(\varepsilon_{\theta t-j}-\bar{\varepsilon}_{\theta}), j = 0, 1, \ldots,$ and $\bar{\varepsilon}_{\theta} = n^{-1}\sum_{t=1}^{n}\varepsilon_{\theta t}$ is the residual sample mean. The residuals are often hard to compute, if not impossible, and it may be advisable to apply the computationally friendly autocorrelation estimates $\tilde{\rho}_{n\theta}(j) = \tilde{\gamma}_{n\theta}(j)/\tilde{\gamma}_{n\theta}(0)$, where

$$\tilde{\gamma}_{n\theta}(j) = \frac{2\pi}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{I_X(\lambda_k)}{f_\theta(\lambda_k)}\cos(j\lambda_k), \quad j = 0, 1, \ldots, \tag{1}$$

$\tilde{n} = [n/2]$, $[a]$ being the integer part of a, and for generic sequences $\{V_t\}_{t=1}^{n}$ and $\{U_t\}_{t=1}^{n}$, $I_{V,U}(\lambda_j) = (2\pi n)^{-1}\sum_{t=1}^{n}\sum_{\ell=1}^{n}V_t U_\ell' \exp\{i\lambda_j(t-\ell)\}, j = 1, \ldots, \tilde{n}$, so $I_X(\lambda_j) = I_{X,X}(\lambda_j)$ denotes the periodogram of $\{X_t\}_{t=1}^{n}$ evaluated at the Fourier frequency $\lambda_j = 2\pi j/n$ for positive integers j. We omit the zero frequency for mean correction.

Henceforth, for the sake of motivation and notational economy, we shall not distinguish between the alternative autocorrelation estimates, and we shall denote by $\rho_{n\theta}$ either $\hat{\rho}_{n\theta}$ or $\tilde{\rho}_{n\theta}$. However, the different results presented in the paper will be formally justified in the Appendix for both estimators.

Let us assume first that the hypothesis to be tested is simple, i.e. the values of the components of θ_0 are known under H_0. The most popular test for testing H_0 is the popular Box–Pierce's portmanteau test, which uses as test statistic $BP_{\theta_0}(m)$ with

$$BP_\theta(m) = n\sum_{j=1}^{m}\rho_{n\theta}(j)^2,$$

where m must be chosen by the practitioner. This test is a compromise between the classical omnibus test based on Bartlett's T_p and U_p processes and the parametric LM tests based on some restrictions on the parameters of a more or less flexible specification. Among them, the $ARFIMA(p, d, q)$ specification is the most popular, with

$$\varphi_\theta(z) = (1-z)^d\frac{\Phi_\delta(z)}{\Xi_\eta(z)}, \quad \theta = (\delta', d, \eta')',$$

such that $\Phi_\delta(z) = 1-\delta_1 z - \cdots - \delta_p z^p$ and $\Xi_\eta(z) = 1-\eta_1 z - \cdots - \eta_q z^q$ are the autoregressive and moving average polynomials, respectively. In fact, $BP_{\theta_0}(m)$ is the LM test statistic when testing that m parameters of the autoregressive part $(\delta_{01}, \ldots, \delta_{0m})$ or the moving average part $(\eta_{01}, \ldots, \eta_{0m})$ equal zero. This is also the LM statistic for testing that all the components of the vector θ_{10} are 0 in the Bloomfield's (1973) exponential spectral density specification

$$f_\theta(\lambda) = g_{\theta_2}(\lambda)\exp\left(\sum_{k=1}^{m}\theta_{1k}\cos\lambda k\right), \quad \theta = (\theta_1', \theta_2')', \tag{2}$$

for some $\theta_0 = (\theta_{10}', \theta_{20}')'$ and $\int_{-\pi}^{\pi}\log g_{\theta_2}(\lambda)\,d\lambda = 0$ for all θ_2 such that $\theta = (\theta_1', \theta_2')' \in \Theta$.

The Box–Pierce's test belongs to the class of test statistics defined by quadratic forms of weighted sums of residual autocorrelations of the form,

$$\Psi_{n\theta}(\omega) = \psi_{n\theta}(\omega)'\psi_{n\theta}(\omega)$$

with

$$\psi_{n\theta}(\omega) = n^{1/2}\left(\sum_{j=1}^{n-1}\omega(j)\omega(j)'\right)^{-1/2}\sum_{j=1}^{n-1}\omega(j)\rho_{n\theta}(j),$$

where ω is a $m \times 1$ weight function such that $\sum_{j=1}^{\ell}\omega(j)\omega(j)'$ is positive definite for each $\ell \geq m$, and for some generic $K > 0$

$$\|\omega(j)\| \leq Kj^{-1}, \quad j = 1, 2, \ldots. \tag{3}$$

Thus, $BP_{n\theta}(m) = \Psi_{n\theta}(\omega)$ with $\omega(j) = (1_{\{j=1\}}, \ldots, 1_{\{j=m\}})'$.

When ω is scalar, Theorem 1 below provides a large sample justification for the class of tests described by means of the Bernoulli random variable $\phi_{n\theta_0}^{\alpha}(\omega) = 1_{\{\psi_{n\theta_0}(\omega)>z_\alpha\}}$, when testing at the α significance level, where $1_{\{\cdot\}}$ is the indicator function and z_α is the $(1-\alpha)$th quantile of the standard normal distribution. When ω is multivariate, tests are described by $\Phi_{n\theta_0}^{\alpha}(\omega) = 1_{\{\Psi_{n\theta_0}(\omega)>\chi^2_{m\alpha}\}}$, where $\chi^2_{m\alpha}$ is the $(1-\alpha)$th quantile of the chi-squared with m degrees of freedom. The theorem refers to Class A of processes, defined in the Appendix. Class A allows for a wide range of autocorrelation patterns in $\{X_t\}_{t=-\infty}^{\infty}$, including long memory, and imposes a Martingale difference assumption on the powers of the white noise process $\{\varepsilon_t\}_{t=-\infty}^{\infty}$. This assumption is weaker than Gaussianity, or independence, which are usually assumed in the time series goodness-of-fit testing literature. See Robinson (1994) and Delgado et al. (2005) for discussion. Theorem 1 also allows to compute the efficiency of the tests in this class under the sequence of local alternatives of the form

$$H_{1n}: \rho_{\theta_0}(j) = \frac{r(j)}{\sqrt{n}} + \frac{a_n(j)}{n} \quad \text{for some } \theta_0 \in \Theta, \tag{4}$$

where $\theta_n \to_p \theta_0$ and r and a_n can depend on θ_0, and are subject to conditions specified in Class L defined in the Appendix. We assume implicitly that r and a_n are such that ρ_{θ_0} is a positive semi-definite sequence for all n. 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

$$\varepsilon_{\theta t} = \Phi_{n\theta}(B)\upsilon_{\theta t},$$

where $\{\upsilon_{\theta t}\}_{t\in\mathbb{Z}}$ are uncorrelated and

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

with $\sum_{j=1}^{\infty}\alpha_{n\theta}(j)^2 < \infty$ and $\lim_{n\to\infty}\alpha_{n\theta_0}(j) = r(j)$.

Let N_m and I_m be the m-dimensional normal distribution and identity matrix respectively.

Theorem 1. *Assume that* $\{X_t\}_{t=-\infty}^{\infty} \in A$. *Under* $H_{1n} \in L$,

$$\psi_{n\theta_0}(\omega) \to_d N_m\left(\left(\sum_{j=1}^{\infty}\omega(j)\omega(j)'\right)^{-1/2}\sum_{j=1}^{\infty}r(j)\omega(j), I_m\right).$$

Thus, the corollary below justifies inferences based on $\Phi_{n\theta_0}^{\alpha}(\omega)$.

Corollary 1. *Under conditions in Theorem 1 and* H_{1n},

$$\Psi_{n\theta_n}(\omega) \to_d \chi_m^2(W(\omega)),$$

where

$$W(\omega) = \sum_{j=1}^{\infty}r(j)\omega(j)'\left(\sum_{j=1}^{\infty}\omega(j)\omega(j)'\right)^{-1}\sum_{j=1}^{\infty}\omega(j)r(j).$$

Thus the Pitman–Noether asymptotic relative efficiency of $\Phi_{n\theta_0}^{\alpha}(\omega)$ (Noether, 1955) is given by $W(\omega)/W(r)$, which is in $[0, 1]$ since $W(r) = \sum_{j=1}^{\infty}r(j)^2$ and $W(\omega)$ is the sum of squares of the projection of r on ω. Thus, $\Phi_{n\theta_0}^{\alpha}(r)$ is the most efficient test in its class. When ω is scalar, the asymptotic relative efficiency of $\phi_{n\theta_0}^{\alpha}(\omega)$ reduces to the squared correlation coefficient between ω and r when $\sum_{j=1}^{\infty}\omega(j)r(j) > 0$, showing that $\phi_{n\theta_0}^{\alpha}(r)$ is the most efficient test in its class. When $\sum_{j=1}^{\infty}\omega(j)r(j) < 0$, $\lim_{n\to\infty}\Pr\left(\phi_{n\theta_0}^{\alpha}(\omega) = 1\right) < \alpha$.

Parametric tests consist of assuming that $\varphi = \varphi_{\theta_0}$ and testing the hypothesis,

$$\dot{H}_0: \theta_{10} = 0,$$

where θ_{10} is a q_1-valued subvector of θ_0, $q_1 \leq q$, in the direction of the parametric local alternative,

$\dot{H}_{1n} : \theta_{10} = \gamma/\sqrt{n}.$

Testing such a hypothesis is equivalent, applying a standard mean value theorem (MVT) argument to $\rho_\theta(j)$, to test H_0 versus H_{1n} with $r(j) = \gamma' d_{1\theta_0}(j)$, where

$$d_{1\theta}(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(\lambda j)\, \frac{\partial}{\partial\theta_1} \log f_\theta(\lambda)\, \mathrm{d}\lambda,$$

assuming suitable smoothness restrictions on f_θ to be specified later. Henceforth, we always assume that it is possible to interchange the integration and differentiation operators. Then, if θ_{10} and γ are scalars, the one-sided test is $\phi^{\alpha}_{n\theta_0}(r) = 1_{\left\{\psi_{n\theta_0}\left(\mathrm{sign}(\gamma)\cdot d_{1\theta_0}\right) > z_\alpha\right\}}$. However, in parametric testing, two sided tests are required when testing that a vector of parameters is equal to zero.

Parameters are unknown in practical situations and they must be estimated. The corresponding ERA's with estimated parameters are neither asymptotically independent or distribution-free. This is why the asymptotic distribution of classical Portmanteau test statistics is not well approximated by the distribution of a chi-squared random variable, except when a suitably large number of sample autocorrelations is considered. In the next sections we develop asymptotically pivotal tests under these circumstances.

In Section 2 we propose a transformation of the weights which result in test statistics converging to a standard normal under the null. We show that a new Box–Pierce-type test based on a linear transformation of the ERA's, belongs to this class and is asymptotically distributed as a chi-squared using a fixed number of transformed ERA's. Section 3 discusses the implementation of the test with regression residuals. In Section 4, we illustrate the finite sample properties of our test by means of a Monte Carlo experiment. Conclusions and further comments on the extension of the proposed tests to different models and alternative regularity conditions are placed in a final section. Mathematical proofs are contained in an Appendix at the end of the article.

2. Asymptotically distribution free tests with estimated parameters

In order to implement the test when θ_0 is unknown under the null, we need a $\sqrt{n}$-consistent estimator, θ_n say, see, for instance, Velasco and Robinson (2000). Theorem 2 provides an asymptotic expansion of the test statistics, which depends on the "score" function

$$d_\theta(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(\lambda j)\, \frac{\partial}{\partial\theta} \log f_\theta(\lambda)\, \mathrm{d}\lambda.$$

Notice that $d_{\theta_0}(\cdot) = -\partial\rho_\theta(\cdot)/\partial\theta\rfloor_{\theta=\theta_0}$ under H_0. The statement of Theorem 2 refers to Class B, which imposes some further mild restrictions on the class of functions $\mathcal{J}$ in order to avoid some pathological behaviour of d_θ, but allowing fairly flexible specifications, including those exhibiting long-memory such as fractionally integrated ARMA and exponential models. Similar assumptions were also used by Delgado et al. (2005). Henceforth, it is assumed that the parameter estimator θ_n is $\sqrt{n}$-consistent under the sequence of local alternatives H_{1n}.

Theorem 2. *Assume that $\{X_t\}_{t=-\infty}^{\infty} \in A$ and $\mathcal{J} \in B$. Under $H_{1n} \in L$,*

$$\sum_{j=1}^{n-1} \omega(j)\, \rho_{n\theta_n}(j) = \sum_{j=1}^{n-1} \omega(j)\, \rho_{n\theta_0}(j) - \sum_{j=1}^{n-1} \omega(j)\, d_{\theta_n}(j)'(\theta_n - \theta_0) + o_p\left(n^{-1/2}\right).$$

Thus, asymptotically distribution-free tests can be obtained for any vector of weight functions ω using a sample dependent transformation $\hat{\omega}_{n,\theta_n}$ such that

$$\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_n}(j)\, d_{\theta_n}(j)' = 0. \quad (5)$$

Assuming that ω and d_{θ_n} are not perfectly collinear, the least squares residuals $\hat{\omega}_{n,\theta_n}$ satisfy (5) non trivially, where for any generic function $g : \mathbb{Z} \to \mathbb{R}^m$,

$$\hat{g}_{n,\theta}(j) = g(j) - \sum_{k=1}^{n-1} g(k)\, d_\theta(k)' \left(\sum_{k=1}^{n-1} d_\theta(k)\, d_\theta(k)'\right)^{-1} d_\theta(j), \quad j = 1, 2, \ldots. \quad (6)$$

Theorem 3. *Under the conditions in Theorem 2 and $H_{1n} \in L$,*

$$\psi_n\left(\hat{\omega}_{n,\theta_n}\right) \to_d N_m\left(\left(\sum_{j=1}^{\infty} \hat{\omega}_{\infty,\theta_0}(j)\, \hat{\omega}_{\infty,\theta_0}(j)'\right)^{-1/2} \times \sum_{j=1}^{\infty} \hat{\omega}_{\infty,\theta_0}(j)\, r(j), I_m\right).$$

Here $\hat{\omega}_{\infty,\theta_0} = \lim_{n\to\infty} \hat{\omega}_{n,\theta_0}$. Notice that $\sup_{j\in\mathbb{N}} |\hat{\omega}_{n,\theta_n}(j) - \hat{\omega}_{\infty,\theta_0}(j)| = o_p(1)$ straightforwardly using the fact that θ_n is $\sqrt{n}-$*consistent*, weights ω satisfying (3) and d_θ are smooth in θ, i.e. satisfying (v) in Class B in the Appendix. We can justify inferences based on $\Phi^{\alpha}_{n\theta_n}(\hat{\omega}_{n,\theta_n})$ with the next corollary.

Corollary 2. *Under conditions in Theorem 2 and $H_{1n} \in L$,*

$$\Psi_{n\theta_n}\left(\hat{\omega}_{n,\theta_n}\right) \to_d \chi^2_m\left(W\left(\hat{\omega}_{\infty,\theta_0}\right)\right).$$

Let $\hat{r}_{n,\theta}$ be the residual function where g in (6) is replaced by r. Now, the relative efficiency of $\Phi^{\alpha}_{n\theta_0}\left(\hat{\omega}_{n,\theta_n}\right)$ is given by $W\left(\hat{\omega}_{\infty,\theta_0}\right)/W\left(\hat{r}_{\infty,\theta_0}\right)$, where $W\left(\hat{r}_{\infty,\theta_0}\right) = \sum_{j=1}^{\infty} \hat{r}_{\infty,\theta_0}(j)^2 = \sum_{j=1}^{\infty} r(j)\, \hat{r}_{\infty,\theta_0}(j)$. Taking into account that $\sum_{j=1}^{\infty} r(j)\, \hat{\omega}_{\infty,\theta_0}(j) = \sum_{j=1}^{\infty} \hat{r}_{\infty,\theta_0}(j)\, \hat{\omega}_{\infty,\theta_0}(j)$, it is immediate that $\Psi_{n\theta_n}\left(\hat{r}_{n,\theta_n}\right)$ is also efficient relative to its class.

Testing the hypothesis $\dot{H}_0$ in the direction $\dot{H}_{1n}$ is equivalent to test H_0 versus H_{1n} with $r(j) = \gamma' d_{1\theta_0}(j)$, where $d_\theta(j) = \left(d_{1\theta}(j)', d_{2\theta}(j)'\right)'$ is conformable with respect to $\theta = \left(\theta_1', \theta_2'\right)'$. Then, using a restricted $\sqrt{n}$-consistent estimate $\hat{\theta}_n$ of θ_0, so that $\left(\hat{\theta}_n - \theta_0\right)' d_\theta(\cdot) = \left(\hat{\theta}_{2,n} - \theta_{2,0}\right)' d_{2\theta}(\cdot) - n^{-1/2}\gamma' d_{1\theta}(\cdot)$ under $\dot{H}_{1n}$, the optimal weights are estimated by $\hat{r}_{n,\hat{\theta}_n}(j) = \gamma' \hat{d}_{n,1\hat{\theta}_n}(j)$, where

$$\hat{d}_{n,1\theta}(j) = d_{1\theta}(j) - \sum_{k=1}^{n-1} d_{1\theta}(k)\, d_{2\theta}(k)' \left(\sum_{k=1}^{n-1} d_{2\theta}(k)\, d_{2\theta}(k)'\right)^{-1} \times d_{2\theta}(j), \quad (7)$$

i.e. $\hat{d}_{n,1\theta}$ are the least squares residuals when projecting $\{d_{1\theta}(j)\}_{j=1}^{n-1}$ on $\{d_{2\theta}(j)\}_{j=1}^{n-1}$.

Interestingly, $\Phi^{\alpha}_{n\hat{\theta}_n}\left(\hat{d}_{n,1\hat{\theta}_n}\right)$ is asymptotically equivalent to generalized LM tests based on different objective functions considered in the literature, cf. Robinson (1994), such as $LM_n = n \cdot S_{1,n}\left(\tilde{\theta}_n\right)' H^{11}_n\left(\tilde{\theta}_n\right) S_{1,n}\left(\tilde{\theta}_n\right)$, where $\tilde{\theta}_n = \left(0', \tilde{\theta}_{2,n}'\right)'$ is the associated restricted (pseudo) maximum likelihood estimate (MLE) under

$\dot{H}_0$, $S_{1,n}\left(\tilde{\theta}_n\right) = -\sum_{j=1}^{n-1} \rho_{n\tilde{\theta}_n}(j)\, d_{1\tilde{\theta}_n}(j)$ and $H_n^{11}(\theta) = \left(\sum_{j=1}^{n-1} \hat{d}_{n,1\theta}(j)\, \hat{d}_{n,1\theta}(j)'\right)^{-1}$. For example, when $\rho_{n\theta}(j) = \tilde{\rho}_{n\theta}(j)$, LM_n corresponds approximately to the LM test based on the Whittle's log-likelihood objective function, which is $\tilde{\gamma}_{n\theta}(0)$ in (1), whereas with $\rho_{n\theta}(j) = \hat{\rho}_{n\theta}(j)$, it corresponds to its time domain Gaussian likelihood counterpart. Applying arguments in Robinson (1994), we conclude that $LM_n \to_d \chi^2_{q_1}\left(\gamma' H_\infty^{11}(\theta_0)^{-1}\gamma\right)$. The statistics $\Psi_{n\hat{\theta}_n}$ are asymptotically equivalent to LM_n under H_{1n} when using optimal weights, as stated in the following corollary, which is a straightforward consequence of Theorem 2.

Corollary 3. *Under conditions in Theorem 2 and $\dot{H}_{1n}$,*

$$\Psi_{n\hat{\theta}_n}\left(\hat{\omega}_{n,\hat{\theta}_n}\right) \to_d \chi^2_{q_1}\left(\gamma' \Omega_{\theta_0}\left(\hat{\omega}_{\infty,\theta_0}\right)\gamma\right),$$

where $\Omega_\theta(\omega) = \sum_{j=1}^\infty d_{1\theta}(j)\,\omega(j)'\left(\sum_{j=1}^\infty \omega(j)\,\omega(j)'\right)^{-1} \sum_{j=1}^\infty \omega(j)\, d_{1\theta}(j)'$, *and* $\Psi_{n\hat{\theta}_n}\left(\hat{d}_{n,1\hat{\theta}_n}\right) = LM_n + o_p(1)$.

The tests $\Phi^\alpha_{n\hat{\theta}_n}\left(\hat{\omega}_{n,\hat{\theta}_n}\right)$ are computed using any preliminary restricted $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ under the sequence of alternatives $\{H_{1n}\}_{n\geq 1}$. Thus, $\Psi_{n\hat{\theta}_n}\left(\hat{d}_{n,1\hat{\theta}_n}\right)$ is asymptotically locally efficient in its class for testing $\dot{H}_0$ in the direction of $\dot{H}_{1n}$, as well as asymptotically equivalent to the LM test, noticing that $\Omega_{\theta_0}\left(\hat{d}_{\infty,1\theta_0}\right) = H_\infty^{11}(\theta_0)^{-1}$ because $\sum_{j=1}^\infty d_{1\theta_0}(j)\,\hat{d}_{\infty,1\theta_0}(j)' = \sum_{j=1}^\infty \hat{d}_{\infty,1\theta_0}(j)\,\hat{d}_{\infty,1\theta_0}(j)'$.

When testing in the direction of innovations autocorrelated according to a $MA(m)$, $AR(m)$ or the autocorrelation structure described in (2),

$$d_{1\theta}(j) = \left(1_{\{j=1\}}, \ldots, 1_{\{j=m\}}\right)' \tag{8}$$

in (7), so that $S_{1,n}(\theta) = -\left(\rho_{n,\theta}(1), \ldots, \rho_{n,\theta}(m)\right)'$, and $H_n^{11}(\theta)^{-1}$ equals

$$I_m - \left(d_{2\theta}(1), \ldots, d_{2\theta}(m)\right)' \left(\sum_{j=1}^{n-1} d_{2\theta}(j)\, d_{2\theta}(j)'\right)^{-1} \times \left(d_{2\theta}(1), \ldots, d_{2\theta}(m)\right).$$

The corresponding LM statistic has the form

$$LM_n = n\left(\rho_{n,\tilde{\theta}_n}(1), \ldots, \rho_{n,\tilde{\theta}_n}(m)\right) H_n^{11}\left(\tilde{\theta}_n\right) \times \left(\rho_{n,\tilde{\theta}_n}(1), \ldots, \rho_{n,\tilde{\theta}_n}(m)\right)'$$

and, by Corollary 3, is asymptotically equivalent to $\Psi_{n,\hat{\theta}_n}\left(\hat{d}_{n,1\hat{\theta}_n}\right)$ for any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ restricted under the null.

However, in the presence of estimated parameters, tests based on the sum of the squares of the first m ERAs are not equivalent to LM tests, even asymptotically.

3. Tests based on regression residuals

When $\{X_t\}_{t=-\infty}^\infty$ are the unobserved errors of a multiple regression model, new difficulties arise because nonparametric nuisance functions appear when computing the optimal weights. Suppose that

$$Y_t = Z_t'\beta_0 + X_t, \quad t = \pm 1, \pm 2, \ldots,$$

where we assume first that $\{Y_t, Z_t\}_{t=-\infty}^\infty$ is a $1+p$-valued vector covariance stationary time series, and $\beta_0 \in \mathbb{R}^p$ is a vector of unknown parameters. We shall discuss the case when Z_t admits non-stochastic regressors later.

Let β_n be a $\sqrt{n}$-consistent estimator of β_0, e.g. the Gaussian MLE. In order to test the specification of X_t in these circumstances, consider residuals $X_t(\beta) = Y_t - \beta' Z_t$, $t = 0, \pm 1, \ldots$, i.e., $X_t = X_t(\beta_0)$ and

$$\varepsilon_t(\theta, \beta) = \varphi_\theta(B)\, X_t(\beta) = \frac{\varphi_\theta(B)}{\varphi(B)}\left\{\varepsilon_t + \varphi(B)\, Z_t'(\beta_0 - \beta)\right\}, \quad t = 0, \pm 1, \ldots,$$

i.e., $\varepsilon_t = \varepsilon_t(\theta_0, \beta_0)$. As before, the autocorrelation function of $\{\varepsilon_t(\theta,\beta)\}_{t=-\infty}^\infty$ can be estimated either by the sample autocorrelation function $\hat{\rho}_{n\theta\beta}(j) = \hat{\gamma}_{n\theta\beta}(j)/\hat{\gamma}_{n\theta\beta}(0)$, with $\hat{\gamma}_{n\theta\beta}(j) = n^{-1}\sum_{t=j+1}^n \varepsilon_t(\theta_n, \beta_n)\,\varepsilon_{t-j}(\theta_n, \beta_n)$, $j = 0, 1, \ldots$, or by, $\tilde{\rho}_{n\theta\beta}(j) = \tilde{\gamma}_{n\theta\beta}(j)/\tilde{\gamma}_{n\theta\beta}(0)$, where $\tilde{\gamma}_{n\theta\beta}(j)$ is defined as $\tilde{\gamma}_{n\theta}(j)$ with I_X being replaced by $I_{X(\beta)}$. Also in this Section, $\rho_{n\theta\beta}$ refers to either $\tilde{\rho}_{n\theta\beta}$ or $\hat{\rho}_{n\theta\beta}$.

In order to identify the parameters, assume that $\varphi_\theta(B)\, Z_t$, are predetermined, i.e. $\mathbb{E}\left(\varepsilon_0(\theta,\beta)\, Z_j\right) = 0$, $j \leq 0$, but not necessarily strictly exogenous. Then, defining the cross-spectral density function between $X_t(\beta)$ and Z_t, $f_{X(\beta),Z}$ say, by $\mathbb{E}\left(X_0(\beta)\, Z_j\right) = (2\pi)^{-1}\int_{-\pi}^{\pi} \exp(i\lambda j)\, f_{X(\beta),Z}(\lambda)\,d\lambda$, we note that

$$\eta_{\theta\beta}(j) = \frac{\mathbb{E}\left(\varepsilon_0(\theta,\beta)\cdot \varphi_\theta(B)\, Z_j\right)}{\sigma^2} = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi} \exp(i\lambda j)\,\frac{f_{X(\beta),Z}(\lambda)}{f_\theta(\lambda)}\,d\lambda,$$

is then zero for $j \leq 0$, but allowed to be nonzero for $j > 0$. We also extend Class B to Class C to incorporate equivalent conditions on $\eta_{\theta\beta}$ as on d_θ. Assuming that $\mathcal{J} \in C$, the next theorem is a straightforward extension of Theorem 3. Hence, its proof is omitted.

Theorem 4. *Assume that $\{X_t\}_{t=-\infty}^\infty \in A$, $\mathcal{J} \in C$ and $H_{1n} \in L$,*

$$\sum_{j=1}^{n-1} \omega(j)\,\rho_{n\theta_n\beta_n}(j) = \sum_{j=1}^{n-1} \omega(j)\,\rho_{n\theta_0\beta_0}(j) - \begin{pmatrix} \beta_0 - \beta_n \\ \theta_n - \theta_0 \end{pmatrix}' \times \sum_{j=1}^{n-1} \omega(j) \begin{pmatrix} \eta_{\theta_0\beta_0}(j) \\ d_{\theta_0}(j) \end{pmatrix} + o_p(1).$$

Thus, asymptotically distribution free test statistics are based on weights orthogonal to both $\eta_{\theta_0\beta_0}$ and d_{θ_0}. To this end, we can consider the semiparametric estimator

$$\eta_{n\theta\beta}(j) = \frac{1}{\gamma_{n\theta\beta}(0)}\mathrm{Re}\left\{\frac{2\pi}{\tilde{n}}\sum_{k=1}^{\tilde{n}} \exp(i\lambda_k j)\,\frac{I_{X(\beta),Z}(\lambda_k)'}{f_\theta(\lambda_k)}\right\},$$

or time domain versions. This avoids to parameterize $f_{X(\beta),Z}$.

For any weight function ω and a smoothing number m, define

$$\hat{\omega}_{mn,\theta\beta}(j) = \omega(j) - \sum_{k=1}^m \omega(k)\begin{pmatrix} \eta_{n\theta\beta}(k) \\ d_\theta(k)\end{pmatrix}' \times \left[\sum_{k=1}^m \begin{pmatrix} \eta_{n\theta\beta}(k)\,\eta_{n\theta\beta}(k)' & \eta_{n\theta\beta}(k)\, d_\theta(k)' \\ d_\theta(k)\,\eta_{n\theta\beta}(k)' & d_\theta(k)\, d_\theta(k)' \end{pmatrix}\right]^{-1}\begin{pmatrix} \eta_{n\theta\beta}(j) \\ d_\theta(j)\end{pmatrix}.$$

Thus, reasoning as before, $\Psi_{mn,\theta_n\beta_n}\left(\hat{\omega}_{mn,\theta_n\beta_n}\right)$, with $\Psi_{mn,\theta\beta}(\omega) = \psi_{mn,\theta\beta}(\omega)'\,\psi_{mn,\theta\beta}(\omega)$ and

$$\psi_{mn,\theta\beta}(\omega) = n^{1/2}\left(\sum_{j=1}^m \omega(j)\,\omega(j)'\right)^{-1/2}\sum_{j=1}^m \omega(j)\,\rho_{n\theta\beta}(j),$$

is expected to be asymptotically pivotal under the null and suitable regularity conditions.

The convergence in distribution of $\psi_{mn,\theta\beta}\left(\hat{\omega}_{mn,\theta_n\beta_n}\right)$ is proved, assuming that $\left(X_t, Z_t'\right)'$ belong to Class D, a multivariate extension of Class A, but allowing $f_{X,Z}$ to be nonparametric. It is also assumed that

$$\frac{1}{m}+\frac{m}{n^{1/2}}\to 0 \quad \text{as } n\to\infty \tag{9}$$

to control the estimation effect of $\eta_{\theta_0\beta_0}(j)$ by $\eta_{n\theta_0\beta_0}(j), j = 1,\dots,m$. The trimming is needed because, unlike d_{θ_0}, $\eta_{n\theta_0\beta_0}$ depends on a sample average, but has no effect on the asymptotic properties of the tests. Notice that the trimming can be avoided by assuming a parametric function for $f_{X,Z}=f_{X(\beta_0),Z}$, which is weaker than assuming that Z_t is strictly exogenous, i.e. $\eta_{n\theta_0\beta_0}(j) = 0$ all $j\geq 1$. An study of optimal choices of m is obviously beyond the scope of this article. However, this choice seems of secondary importance in practical terms given the fast decay of the weights ω. Finally, note that our distribution free tests can be computed without resorting to smooth estimation of the cross-spectrum as considered in Delgado et al. (2009), avoiding that finite sample properties are affected by the choice of a bandwidth number.

Next theorem provides the limiting distribution of $\psi_{mn,\theta\beta}(\hat{\omega}_{mn,\theta_n\beta_n})$ under local alternatives

$$H_{1n}:\rho_{\theta_0\beta_0}(j)=\frac{r(j)}{\sqrt{n}}+\frac{a_n(j)}{n},\quad j>0 \text{ for some } \left(\theta_0',\beta_0'\right)'\in\Theta,$$

and shows that the test $\Phi^{\alpha}_{mn\theta_n\beta_n}\left(\hat{r}_{mn,\theta_n\beta_n}\right)$ is locally efficient in its class. We also omit the proof given the similarities with that of Theorem 4.

Theorem 5. *Assume that* $\left\{\left(X_t,Z_t'\right)'\right\}_{t=-\infty}^{\infty}\in D$, $\mathcal{J}\in C$, *and* (9), *under* $H_{1n}\in L$,

$$\psi_{m,n}\left(\hat{\omega}_{mn,\theta_n\beta_n}\right)\to_d N_m\left(\left(\sum_{j=1}^{\infty}\hat{\omega}_{\infty,\theta_0\beta_0}(j)\,\hat{\omega}_{\infty,\theta_0\beta_0}(j)'\right)^{-1/2}\times\sum_{j=1}^{\infty}\hat{\omega}_{\infty,\theta_0\beta_0}(j)\,r(j), I_m\right).$$

If the elements of Z_t, $t = 1,2,\ldots$., are nonstochastic, such as polynomial trends in t, and under the identifiability conditions stated in the Appendix as Class E, estimation of β does not affect the asymptotic properties of ERA's, and weights need not be orthogonalized. The reason is that the Z_t are strictly exogenous in this case, and the corresponding function $\eta_{\theta_0\beta_0}(j)$ is zero for all leads and lags. This fact, together with the assumption that β_n is (at least) $\sqrt{n}$-consistent, renders Theorems 3 and 4 valid in this set up.

We could consider general pseudo-residuals $U_{\beta_0}(Y_t,Z_t) = X_t, t = 0,\pm 1,\pm 2,\ldots$. These pseudo-residuals could be the parametrically scaled residuals $U_{\beta}(Y_t,Z_t) = Y_t/\sigma_{\beta}(Z_t)$, where σ_{β} is a known function indexed by the parameter β, e.g. a GARCH specification. The results in this Section can be straightforwardly applied towards testing the lack of autocorrelation of these pseudo-residuals.

4. A Monte Carlo Experiment

This simulation study is based on 50,000 replications of $ARFIMA(p,d,q)$ models under alternative designs. The innovations are independent standard normals. Parameters are estimated using the restricted Whittle estimator under the null hypothesis and we use time domain ERA's.

We have computed the percentage of rejections using five distribution free tests:

1. Delgado et al. (2005) omnibus test based on the transformed T_p-*process* using the Cramer-von Mises criteria, CvM.
2. The efficient LM test against different residual autocorrelation alternatives.
3. Our efficient test $\hat{\Psi}_n=\Psi_{n\theta_n}\left(\hat{d}_{n,1\theta_n}\right)$ with $\hat{d}_{n,1\theta_n}$ corresponding to different residual autocorrelation alternatives.
4. Our transformed portmanteau test (TPT) $\hat{\Psi}_n$, with $\hat{d}_{n,1\theta_n}$ corresponding to the alternative of residuals autocorrelated according to an $AR(m)$, cf. (8).
5. Box Pierce test, computed as proposed by Ljung and Box (1978), $BP_n(m)$.

Table 1 reports the percentage of rejections under the null of AR(1), MA(1) and integrated of order d process ($I(d)$), with sample sizes of 200 and 500. We have computed the BP test for $m = 10$, 20 and 30. Choices of m around $\sqrt{n}$ are expected to yield test statistics with a good size accuracy. We also provide results for $m = 5$ in order to check size accuracy and power for a small m. We report results for our TPT using small values of $m = 1,2,3,5$.

As it happens with the standard LM_n test statistic considering $AR(m)$ (or $MA(m)$, or Bloomfield (m)) departures from the innovations white noise hypothesis, the weighting matrix of the test statistic $\Psi_{n\theta_n}\left(\hat{d}_{n,1\theta_n}\right)$ becomes near idempotent as m increases. This fact prevents us from using our TPT or the LM test with large values of m in this situation. The size accuracy of the TPT is excellent for the small values reported in the three designs considered. The CvM and BP tests also perform very well for a sample size of 500, but LM_n and $\hat{\Psi}_n$ suffer very serious size distortions for some designs.

The proportion of rejections under alternative hypotheses are reported in Table 2 for $n = 200$ and different designs. All the tests detect departures from the AR(1) specification in the direction of MA(1) innovations, as well as departures from the MA(1) specification in the direction of AR(1) innovations. However, $I(d)$ departures from the white noise hypothesis are better detected by the TPT than any other test. The classical BP test rejects less than the other methods in this situation. It is worth mentioning that departures from the AR(1) specification with parameter 0.5 in the direction of $I(d)$ correlated innovations are not detected by any test for the sample sizes considered. Departures from the $I(d)$ hypothesis are better detected. However, the TPT works much better than the others in this case.

5. Further comments

This article discusses the construction of distribution free tests for general time series model specification, which include models exhibiting long memory. The resulting tests are asymptotically equivalent to Gaussian LM tests, despite using any preliminary $\sqrt{n}$-*consistent* estimator. This requires that $\left\{\sqrt{n}\rho_{n\theta_0}(j)\right\}_{j>0}$ are asymptotically independent standard normals under the null, which is provided assuming in Class A that $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ does not exhibit some form of higher order serial dependence under the null, e.g. conditional volatility.

The asymptotic distribution of $\left\{\sqrt{n}\rho_{n\theta_0}(j)\right\}_{j>0}$ has been derived under fairly general conditions on the higher order serial dependence of $\{\varepsilon_t\}_{t=-\infty}^{\infty}$, e.g. Hannan and Heyde (1972) assume a Martingale difference sequence and Romano and Thombs (1996) assumes nonparametric strong mixing dependence. In this general setting $\left\{\rho_{n\theta_0}(j)\right\}_{j>0}$ are still asymptotically normal but with asymptotic covariance function

$$AsyVar\left\{\sqrt{n}\rho_{n\theta_0}(j),\sqrt{n}\rho_{n\theta_0}(\ell)\right\}=a_{\theta_0}(j,\ell)$$

with $a_{\theta}(j,\ell) := \mathbb{E}\left(\varepsilon_{t\theta}\varepsilon_{t+j\theta}\varepsilon_{t+\ell\theta}\varepsilon_{t+j+\ell\theta}\right)$. This expression is simplified under particular circumstances. For instance, when $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a Martingale difference sequence, $a_{\theta_0}(j,\ell) =$

Table 1
Empirical size of LM and Portmanteau tests at 5% of significance.

	CvM	LM	$\hat{\Psi}_n$	$\hat{\Psi}_n\ [d_{1\theta}:AR(m)]$				$BP_{n\theta_n}(m)$			
m				1	2	3	5	5	10	20	30
$n=200$											
H_0: AR(1)											
δ_{10}	$[d_{1\theta}:I(d)]$										
−0.8	4.7	3.4	3.4	4.9	4.8	4.6	4.3	5.5	5.5	6.0	6.6
−0.5	4.4	3.2	3.3	4.8	4.7	4.5	4.2	5.1	5.2	5.7	6.3
0.0	4.1	2.5	2.5	5.0	4.6	4.4	4.2	4.9	5.0	5.7	6.3
0.5	3.6	1.1	0.7	4.9	4.7	4.5	4.2	4.8	5.1	5.6	6.3
0.8	3.1	4.9	3.0	4.8	4.6	4.6	4.4	5.0	5.2	5.8	6.3
H_0: MA(1)											
η_{10}	$[d_{1\theta}:I(d)]$										
−0.8	4.2	3.5	3.3	4.5	4.4	4.2	4.1	6.7	6.3	6.4	7.0
−0.5	4.2	3.0	3.1	4.5	4.5	4.4	4.1	5.1	5.1	5.7	6.3
0.0	4.1	2.3	2.3	4.7	4.4	4.4	4.1	4.8	5.0	5.6	6.2
0.5	3.6	3.3	0.6	4.6	4.4	4.2	4.1	4.8	5.0	5.5	6.2
0.8	3.1	24.5	3.6	4.6	4.4	4.3	4.3	6.3	5.9	6.1	6.6
H_0: I(d)											
d_0	$[d_{1\theta}:\mathrm{AR}(1)]$										
0.0	3.5	4.9	4.3	4.3	3.8	3.5	3.4	5.0	5.2	5.7	6.4
0.2	3.5	4.9	4.3	4.3	3.8	3.4	3.3	5.0	5.2	5.7	6.3
0.4	3.6	5.1	4.2	4.2	3.7	3.4	3.2	5.0	5.1	5.6	6.2
$n=500$											
H_0: AR(1)											
δ_{10}	$[d_{1\theta}:I(d)]$										
−0.8	5.1	4.3	4.3	5.1	5.0	5.0	4.8	5.4	5.3	5.5	5.8
−0.5	5.0	4.1	4.1	5.0	5.0	4.9	4.7	5.1	4.9	5.4	5.7
0.0	4.6	3.6	3.6	5.0	5.1	4.8	4.8	5.1	4.9	5.4	5.6
0.5	4.5	2.0	2.1	5.0	5.0	4.9	4.8	5.1	5.0	5.3	5.7
0.8	4.3	4.2	3.8	5.1	4.8	5.0	4.9	5.3	5.1	5.4	5.7
H_0: MA(1)											
η_{10}	$[d_{1\theta}:I(d)]$										
−0.8	4.9	4.3	4.2	5.0	4.8	4.8	4.6	6.1	5.6	5.7	6.0
−0.5	4.9	4.0	4.1	4.9	5.0	4.8	4.7	5.2	5.0	5.4	5.7
0.0	4.6	3.5	3.5	4.8	5.0	4.8	4.6	5.0	4.9	5.3	5.7
0.5	4.5	3.2	1.8	4.9	4.8	4.8	4.7	5.0	5.0	5.3	5.6
0.8	4.3	17.4	3.8	4.9	4.7	4.8	4.7	5.8	5.4	5.5	5.8
H_0: I(d)											
d_0	$[d_{1\theta}:\mathrm{AR}(1)]$										
0.0	4.5	5.0	4.7	4.7	4.4	4.3	4.1	5.3	5.1	5.4	5.7
0.2	4.5	4.9	4.6	4.6	4.4	4.3	4.1	5.2	5.1	5.4	5.7
0.4	4.6	5.3	4.5	4.5	4.3	4.2	4.0	5.3	5.1	5.4	5.7

$\mathbb{E}\left(\varepsilon_{t\theta_0}^2\varepsilon_{t+j\theta_0}\varepsilon_{t+\ell\theta_0}\right)$ (Hannan and Heyde, 1972) and when de serial dependence of $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ can be modeled according to a Gaussian GARCH model, $a_{\theta_0}(j,\ell)=0$ for $j\neq\ell$ (Lobato, 2001; Lobato et al., 2001).

Under general serial dependence, it is expected that under H_0,

$$\psi_n(\omega)\rightarrow_d N_m\left(0,\Omega_{\theta_0}\right),$$

with

$$\Omega_{\theta_0}=\left(\sum_{j=1}^{\infty}\omega(j)\,\omega(j)'\right)^{-1/2}\left(\sum_{j=1}^{\infty}\sum_{\ell=1}^{\infty}\omega(j)\,\omega(\ell)'\,a_{\theta_0}(j,\ell)\right)\times\left(\sum_{j=1}^{\infty}\omega(j)\,\omega(j)'\right)^{-1/2},$$

which can be estimated truncating the summations in the middle term and exploiting the decay of the function ω, as in a Newey and West (1987) type estimator. We could obtain asymptotically distribution-free tests, robust to unknown higher order serial dependence of the innovations using the test statistic,

$$\Psi_{n\theta_n}\left(\hat{\omega}_{n\theta_n}\right)=\psi_n(\omega)'\,\Omega_{n\theta_n}^{-1}\psi_n(\omega),$$

where $\Omega_{n\theta_n}$ is a suitable consistent estimator of Ω_{θ_0}. Though the resulting estimator is expected to be efficient within its class, it is not possible to make comparisons with the corresponding optimal *LM* test.

Assuming that the serial dependence of $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ can be modeled according to a GARCH specification, we could test that the parametric scaled innovations are not autocorrelated using the test proposed in this article by a fairly straightforward extension of the results in Section 3 to parametric models nonlinear in variables. However, justifying such procedures in the presence of long range dependence is out of the scope of this paper.

Appendix A. Tests using frequency domain autocorrelation estimates

Class A. The process $\{X_t\}_{t=-\infty}^{\infty}$ defined by $\varphi(B)X_t=\varepsilon_t$ belongs to Class A if:

(i) The process $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ satisfies that $\mathbb{E}\left(\varepsilon_t^r\middle|\mathcal{F}_{t-1}\right)=\mu_r$ with μ_r constant ($\mu_1=0$ and $\mu_2=\sigma^2$) for $r=1,\ldots,4$ and all $t=0,\pm1,\ldots$, where $\mathcal{F}_t$ is the sigma algebra generated by $\{\varepsilon_s, s\leq t\}$.

Table 2
Empirical power of LM and Portmanteau tests at 5% of significance.

	CvM	LM	$\hat{\Psi}_n$	$\hat{\Psi}_n\ [d_{1\theta} : \mathrm{AR}(m)]$				$BP_{n\theta_n}\ (m)$			
m				1	2	3	5	5	10	20	30
$H_0 : \mathrm{AR}(1).\ H_1 : \mathrm{MA}(1).\ n = 200$											
η_{10}	$[d_{1\theta} : \mathrm{MA}(1)]$										
−0.8	100.	99.8	99.8	99.8	100.	100.	100.	100.	99.6	94.9	89.1
−0.5	80.8	83.6	80.6	80.6	78.9	71.4	59.9	66.7	49.9	38.3	33.8
0.2	7.1	12.9	9.7	9.7	8.0	7.1	6.1	7.3	6.7	6.9	7.5
0.5	70.8	75.9	80.8	80.8	79.2	73.0	61.8	68.7	51.7	39.2	34.7
0.8	99.6	99.5	99.8	99.8	100.	100.	100.	100.	99.6	95.2	89.3
$H_0 : MA(1).\ H_1 : \mathrm{AR}(1).\ n = 200$											
δ_{10}	$[d_{1\theta} : \mathrm{AR}(1)]$										
−0.8	100.	100.	100.	100.	100.	100.	100.	100.	100.	100.	100.
−0.5	84.4	78.1	81.2	81.2	82.3	77.3	69.7	74.2	61.9	50.4	44.9
0.2	7.2	25.0	6.9	6.9	6.1	5.6	4.9	5.9	5.6	6.1	6.7
0.5	77.1	86.9	81.5	81.5	80.4	75.1	66.9	72.1	59.3	48.2	43.0
0.8	100.	100.	100.	100.	100.	100.	100.	100.	100.	100.	100.
$H_0 : I(d).\ H_1 : \mathrm{ARFIMA}(1, d_0, 0).n = 200$											
$d_0 = 0.0$											
δ_{10}	$[d_{1\theta} : \mathrm{AR}(1)]$										
0.2	11.3	37.2	34.3	34.3	23.2	6.1	13.0	17.5	14.3	12.5	12.4
0.5	26.8	79.8	77.7	77.7	68.3	56.8	43.7	47.4	41.2	31.7	28.6
0.8	9.8	55.4	51.4	51.4	46.4	36.7	24.4	24.4	26.4	21.4	20.2
$d_0 = 0.2$											
0.2	11.1	36.7	34.2	34.2	23.1	17.1	13.0	17.4	14.3	12.5	12.4
0.5	26.7	79.1	77.7	77.7	68.2	56.8	43.6	47.3	41.2	31.6	28.4
0.8	9.6	61.1	53.7	53.7	49.4	40.6	28.3	24.8	26.6	21.5	19.9
$H_0 : AR(1).\ \ H_1 : \mathrm{ARFIMA}(1, d_0, 0).\ n = 200$											
$\delta_{10} = 0.0$											
d_0	$[d_{1\theta} : I\ (d)]$										
0.1	8.2	10.2	8.7	8.4	8.1	7.8	7.1	8.0	7.5	7.5	7.8
0.2	19.9	29.9	26.5	22.4	21.8	21.1	19.3	20.4	18.4	15.8	15.0
0.3	36.0	47.5	42.5	42.5	42.3	40.6	37.8	37.2	35.0	30.0	26.8
0.4	48.8	46.1	38.8	60.5	60.0	57.6	53.7	49.1	48.4	41.8	37.3
$\delta_{10} = 0.5$											
0.1	3.6	2.7	1.0	5.0	4.8	4.6	4.3	5.0	5.1	5.8	6.4
0.2	3.3	4.7	1.5	5.5	5.3	5.2	5.3	5.5	5.7	6.2	6.7
0.3	3.6	8.3	2.6	7.8	6.9	6.8	6.5	7.0	6.8	7.1	7.5
0.4	5.7	16.2	7.1	14.8	11.6	10.9	9.9	11.7	9.6	8.9	9.1

(ii) $f(\lambda) = |\varphi(e^{i\lambda})|^{-2}$ is positive and continuously differentiable on $(0, \pi]$, and $|(\mathrm{d}/\mathrm{d}\lambda)\log f(\lambda)| = O(|\lambda|^{-1})$ as $|\lambda| \to 0$.

Class B. The parametric model $\mathcal{J}$ belongs to Class B if:

(i) $f_\theta(\lambda)$ is continuously differentiable in $\theta \in \Theta, \lambda \in (0, \pi]$, with derivative $\mu_\theta(\lambda) := (\partial/\partial\theta)\log f_\theta(\lambda)$, so that $\mu_{\theta_0}(\lambda)$ is continuously differentiable on $(0, \pi]$.

(ii) $\left\|\partial\mu_{\theta_0}(\lambda)/\partial\lambda\right\| = O(|\lambda|^{-1})$ as $|\lambda| \to 0$.

(iii) $\sup_{\theta\in\Theta}\|\mu_\theta(\lambda)\| = O(\log|\lambda|)$ as $|\lambda| \to 0$.

(iv) For all $\lambda \in (0, \pi]$ and $0 < \delta < 1$ there exists some $K < \infty$ such that

$$\sup_{\{\theta:\|\theta-\theta_0\|\le\delta/2\}} \frac{1}{\|\theta-\theta_0\|^2} \times \left|\frac{f_{\theta_0}(\lambda)}{f_\theta(\lambda)} - 1 + (\theta-\theta_0)'\mu_{\theta_0}(\lambda)\right| \le \frac{K}{|\lambda|^\delta}\log^2|\lambda|.$$

(v) For $d_\theta(j) = (2\pi)^{-1}\int_{-\pi}^{\pi}\mu_\theta(\lambda)\cos(j\lambda)\,\mathrm{d}\lambda$ and $\dot{d}_\theta(j) = \partial d_\theta(j)/\partial\theta, j = 1, 2, \ldots,$

$$\sum_{j=1}^{\infty} d_{\theta_0}(j)\, d_{\theta_0}(j)' \text{ is finite and positive definite;} \tag{10}$$

$$\sup_{\theta\in\Theta}\|d_\theta(j)\| + \sup_{\theta\in\Theta}\left\|\dot{d}_\theta(j)\right\| \le Cj^{-1}, \quad j = 1, 2, \ldots. \tag{11}$$

Class C. The parametric model $\mathcal{J}$ described in Section 5 belongs to Class C if:

(i) All conditions of Class B hold.

(ii) Conditions (ii)–(iii) of Class B hold replacing $\mu_\theta(\lambda)$ by $f_{X(\beta)Z}(\lambda)/f_\theta(\lambda), (\theta', \beta')' \in \Theta$.

(iii) Condition (v) of Class B holds with d_θ replaced by $\left(\eta'_{\theta\beta}, d'_\theta\right)'$, $(\theta', \beta')' \in \Theta$.

Class D. The $(1+p)$-process $\{V_t\}_{t=-\infty}^{\infty}$, $\Psi(B)V_t = U_t$, belongs to Class D if:

(i) The process $\{U_t\}_{t=-\infty}^{\infty}$ satisfies that $\mathbb{E}(U_t|\mathcal{F}_{t-1}) = 0, \mathbb{E}(U_t U_t'|\mathcal{F}_{t-1}) = \Sigma, \mathbb{E}\left(U_{t,a}U_{t,b}U_{t,c}\middle|\mathcal{F}_{t-1}\right) = \mu_{abc}, \mathbb{E}(U_{t,a}U_{t,b}U_{t,c}U_{t,d}|\mathcal{F}_{t-1}) = \mu_{abcd}$ with μ_{abc} and μ_{abcd} bounded, all $a, b, c, d = 1, \ldots, 1+p$ and all $t = 0, \pm1, \ldots$, where $\mathcal{F}_t$ is the sigma algebra generated by $\{U_s, s \le t\}$.

(ii) $f_V(\lambda) = |\Psi(e^{i\lambda})|^{-2}$ is continuously differentiable on $[-\pi, 0) \cup (0, \pi]$, and $\|(\mathrm{d}/\mathrm{d}\lambda)\log f_V(\lambda)\| = O(|\lambda|^{-1})$ as $|\lambda| \to 0$.

(iii) The elements of $f_V(\lambda)/f(\lambda)$ are bounded on $[-\pi, \pi]$, where $f = \{f_V\}_{[1,1]} \in A$.

Class E. The nonstochastic regressors $\{Z_t\}_{t=-\infty}^{\infty}$ belongs to Class E if $D_n = \sum_{t=1}^{n} W_t W_t'$ is positive definite for a large enough n, $W_t = \varphi(B)Z_t, Z_t = 0, t \le 0$.

Class L. The sequence of local alternatives $\{H_{1n}\}_{n\geq 1}$ in (4) satisfies

$$\sum_{j=1}^{\infty} r(j)^2 < \infty \quad \text{and} \quad \sum_{j=1}^{n} a_n(j)^2 = O(1) \quad \text{as } n \to \infty. \qquad (12)$$

(i) The function l defined as $l(\lambda) = (2\pi)^{-1}\sum_{j=1}^{\infty} r(j)\cos(\lambda j)$, satisfies $|l(\lambda)| \leq K|\log\lambda|$ and is differentiable in $(0,\pi]$ so that $|(\partial/\partial\lambda)\, l(\lambda)| \leq K|\lambda|^{-1}$, all $\lambda > 0$.

(ii) The absolute value of $g_n(\lambda) = (2\pi)^{-1}\sum_{j=1}^{\infty} a_n(j)\cos(\lambda j)$ is dominated by an integrable function not depending on n for all $n > n_0$.

We consider now the frequency domain case, where $\rho_{n\theta}(j) = \tilde{\rho}_{n\theta}(j)$, and ω scalar throughout the appendix, to simplify exposition, since asymptotic expansions have to be worked out element by element and multivariate convergence in distribution results would be followed by a routine application of the Cramer-Wold device.

Proof of Theorem 1. Define

$$\psi_{n,k}(\omega) = n^{1/2}\left(\sum_{j=1}^{k}\omega(j)^2\right)^{-1/2}\sum_{j=1}^{k}\rho_{n\theta_0}(j)\,\omega(j).$$

By Lemma 1 in Appendix C,

$$\psi_{n,k}(\omega) \to_d N\left(\left(\sum_{j=1}^{k}\omega(j)^2\right)^{-1/2}\sum_{j=1}^{k} r(j)\,\omega(j),\, 1\right) \quad \text{as } n\to\infty$$

for k fixed. Then, using Theorem 3.2 in Billingsley (1999) we only need to show that

$$\lim_{k\to\infty}\limsup_{n\to\infty}\Pr\left(\left|\psi_n(\omega) - \psi_{n,k}(\omega)\right| > \epsilon\right) = 0 \qquad (13)$$

for any $\epsilon > 0$. We first note that the innovation variance estimate is the same in both $\psi_{n,k}(\omega)$ and $\psi_n(\omega)$ so we concentrate on the autocovariance estimates $\tilde{\gamma}_{n\theta_0}(j), j = 0, 1, \ldots$. Then we show that, under H_{1n}, $\mathbb{E}n^{1/2}|\delta_n(j)| = O\left(n^{-\delta}\right)$ for some $\delta > 0$ and for each $j = 1, \ldots, k$, where $\delta_n(j) = \tilde{\gamma}_{n\theta_0}(j) - n^{-1/2}\sigma^2 r(j) - \tilde{\gamma}_{n\varepsilon}(j)$ and $\tilde{\gamma}_{n\varepsilon}(j)$ is defined as $\tilde{\gamma}_{n\theta_0}(j)$ but replacing $I_X(\cdot) f_{\theta_0}^{-1}(\cdot)$ by $I_\varepsilon(\cdot)$. Proceeding as in the proof of Lemma 1,

$$\tilde{\gamma}_{n\theta_0}(j) = \frac{2\pi}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{I_X(\lambda_k)}{f(\lambda_k)}\cos(j\lambda_k) \times\left\{1 + n^{-1/2} l(\lambda_k)\right\} + n^{-1}V_n(j),$$

where $\mathbb{E}|V_n(j)| = O(1)$ because g_n is uniformly integrable. Then, using Lemma 4 in in Delgado et al. (2005), DHV henceforth, for both $s = 1$ and $s = l$,

$$\mathbb{E}\left| n^{1/2}\frac{2\pi}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\left(\frac{I_X(\lambda_k)}{f(\lambda_k)} - I_\varepsilon(\lambda_k)\right) s(\lambda_k)\cos(j\lambda_k)\right| = O\left(n^{-\delta}\right)$$

for some $\delta > 0$, uniformly in j, while $\mathbb{E}|(2\pi/\tilde{n})\sum_{k=1}^{\tilde{n}} I_\varepsilon(\lambda_k) l(\lambda_k) \cos(j\lambda_k) - \sigma^2 r(j)| = O\left(n^{-1}\log n\right)$ using Lemma 2 and Lemma 1 in DHV with r and l satisfying conditions of $H_{1n} \in L$. Next, this shows that

$$\sup_k\left| n^{1/2}\sum_{j=k+1}^{n-1}\delta_n(j)\,\omega(j)\right| \leq n^{1/2}\sum_{j=1}^{n-1}|\delta_n(j)|\,|\omega(j)|$$

is $o_p(1)$ as $n\to\infty$, uniformly in k, using (3). Finally, using again (3) and Lemma 2,

$$\mathbb{E}\left| n^{1/2}\sum_{j=k+1}^{n-1}\tilde{\gamma}_{n\varepsilon}(j)\,\omega(j)\right|^2$$

$$= O\left(\sum_{j=k+1}^{n-1}\omega^2(j) + n^{-1}\sum_{j=k+1}^{n-1}\sum_{j'=k+1}^{n-1}|\omega(j)|\,|\omega(j')|\right)$$

and $\left|\sum_{j=k+1}^{n-1} r(j)\,\omega(j)\right|$ are both $o(1)$ as $k\to\infty$, so (13) holds by Markov's inequality. □

Proof of Theorem 2. Write

$$\sum_{j=1}^{n-1}\omega(j)\,\rho_{n,\theta_n}(j) = \sum_{j=1}^{n-1}\omega(j)\,\rho_{n\theta_0}(j) - (\theta_n - \theta_0)' \times\sum_{j=1}^{n-1}\omega(j)\,d_{\theta_n}(j) + \sum_{j=1}^{5} R_{nj},$$

where $R_{n1} = (\theta_n - \theta_0)'\sum_{j=1}^{n-1}\omega(j)\left\{d_{\theta_n}(j) - d_{\theta_0}(j)\right\}$, $R_{n2} = (\theta_n - \theta_0)'\sum_{j=1}^{n-1}\omega(j)\times\left\{d_{\theta_0}(j) - d_{n\theta_0}(j)\right\}$, $R_{n3} = \sum_{j=1}^{n-1}\omega(j)\,\dot{d}_{n\theta_n}(j)$, and

$$R_{n4} = \left[\frac{1}{\sigma^2} - \frac{1}{\tilde{\gamma}_{n\theta_0}(0)}\right]\sum_{j=1}^{n-1}\omega(j)\,\tilde{\gamma}_{n\theta_0}(j),$$

$$R_{n5} = \left[\frac{1}{\tilde{\gamma}_{n\theta_n}(0)} - \frac{1}{\sigma^2}\right]\sum_{j=1}^{n-1}\omega(j)\,\tilde{\gamma}_{n\theta_n}(j),$$

with $d_{n\theta}(j) = \left(2\pi/\tilde{n}\right)\sigma^{-2}\sum_{i=1}^{\tilde{n}} I_X(\lambda_i) f_\theta^{-1}(\lambda_i)\,\mu_\theta(\lambda_i)\cos(\lambda_i j)$, and

$$\dot{d}_{n\theta}(j) = \frac{2\pi}{\tilde{n}\sigma^2}\sum_{i=1}^{\tilde{n}}\frac{I_X(\lambda_i)}{f_{\theta_0}(\lambda_i)} \left\{\frac{f_{\theta_0}(\lambda_i)}{f_\theta(\lambda_i)} - 1 + (\theta_n - \theta_0)'\mu_{\theta_0}(\lambda_i)\right\}\cos(\lambda_i j).$$

Thus, it suffices to prove that $R_{nj} = o_p\left(n^{-1/2}\right), j = 1, \ldots, 5$. Applying (12), (3), and taking into account that θ_n is $\sqrt{n}$-consistent, $R_{n1} = o_p\left(n^{-1/2}\right)$. Write

$$R_{n2} = (\theta_n - \theta_0)'\sum_{j=1}^{n-1}\omega(j)\left\{d_{\theta_0}(j) - \frac{2\pi}{\tilde{n}}\sum_{i=1}^{\tilde{n}}\mu_{\theta_0}(\lambda_i)\cos(j\lambda_i)\right\} + (\theta_n - \theta_0)'\sum_{j=1}^{n-1}\omega(j)\left\{\frac{2\pi}{\tilde{n}\sigma^2}\sum_{i=1}^{\tilde{n}} \left[\frac{\sigma^2}{2\pi} - \frac{I_X(\lambda_i)}{f_{\theta_0}(\lambda_i)}\right]\mu_{\theta_0}(\lambda_i)\cos(j\lambda_i)\right\}.$$

The first term on the left hand side is $O\left(n^{-1}\log n^2\right)$ applying Lemma 1 in DHV and (2), and the second term can be written as

$$(\theta_n - \theta_0)'\frac{2\pi}{\tilde{n}\sigma^2}\sum_{i=1}^{\tilde{n}}\left(\frac{\sigma^2}{2\pi} - I_\varepsilon(\lambda_i)\right)\mu_{\theta_0}(\lambda_i)\sum_{j=1}^{n-1}\omega(j)\cos(j\lambda_i) \qquad (14)$$

$$+ (\theta_n - \theta_0)'\frac{2\pi}{\tilde{n}\sigma^2}\sum_{i=1}^{\tilde{n}}\left(I_\varepsilon(\lambda_i) - \frac{I_X(\lambda_i)}{f_{\theta_0}(\lambda_i)}\right)\mu_{\theta_0}(\lambda_i)\cos(j\lambda_i). \qquad (15)$$

Applying (3), $\left|\sum_{j=1}^{n-1}\omega(j)\cos(j\lambda_i)\right| = O(\log n)$ uniformly in i. Thus, after applying Markov's inequality, $\theta_n - \theta_0 = O_p\left(n^{-1/2}\right)$ and (iii) of Class B, (14) is an $o_p\left(n^{-1/2}\right)$, whereas (15) $= o_p\left(n^{-1}\right)$ by DHV's Lemma 4. Hence, $R_{n2} = o_p\left(n^{-1/2}\right)$. Applying condition (iv) in Class B,

$$\left\|\dot{d}_{n\theta_n}(j)\right\| \leq \|\theta - \theta_0\|^2\frac{C}{\tilde{n}}\sum_{i=1}^{\tilde{n}}|\log\lambda_i|^2\frac{I_X(\lambda_i)}{f_{\theta_0}(\lambda_i)}$$

because θ_n is $\sqrt{n}$-consistent, and we can take $\delta = Kn^{-1/2}$ in, so that $|\lambda_i| \le K$ when $i \ge 1$, reasoning as in the proof of Lemma 8 of DHV. Therefore,

$$\|R_{n3}\| \le \|\theta_n - \theta_0\|^2 \sum_{j=1}^{n-1} |\omega(j)| \frac{C}{\tilde{n}} \times \sum_{i=1}^{\tilde{n}} |\log \lambda_i|^2 \frac{I_X(\lambda_i)}{f_{\theta_0}(\lambda_i)} = o_p\left(n^{-1/2}\right)$$

on taking expectations and using $\|\theta_n - \theta_0\| = O_p\left(n^{-1/2}\right)$. Finally note that replacing $\tilde{\gamma}_{n\theta_n}(0)$ by $\tilde{\gamma}_{n\theta_0}(0)$, and this by σ^2, makes no difference by (50) in DHV, which proves that $R_{n4} = o_p\left(n^{-1/2}\right)$ and $R_{n5} = o_p\left(n^{-1/2}\right)$. □

Proof of Theorem 3. We note that by Theorem 2 and because of the exact orthogonality of $\hat{\omega}_{n,\theta_n}$ and d_{θ_n}, $\psi_n\left(\hat{\omega}_{n,\theta_n}\right) = \bar{\psi}_n\left(\hat{\omega}_{n,\theta_n}\right) + o_p(1)$, with $\bar{\psi}_n\left(\hat{\omega}_{n,\theta_n}\right) = n^{1/2}\left(\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_n}(j)^2\right)^{-1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)$ $\hat{\omega}_{n,\theta_n}(j)$. So, we can apply Theorem 2, with ω substituted by $\hat{\omega}_{n,\theta_n}$, after noticing that $\sum_{j=1}^{\infty} \hat{\omega}_{n,\theta_n}(j)^2 < \infty$, because of (3), (v) in the definition of Class B, and using $\hat{\omega}_{n,\theta_n}(j) = \omega(j) - d_{\theta_n}(j)' \beta_{n\theta_n}$, with $\beta_{n\theta} = \left(\sum_{j=1}^{n-1} d_\theta(j) d_\theta(j)'\right)^{-1} \sum_{j=1}^{n-1} d_\theta(j) \omega_\theta(j)$, and where $\beta_{n,\theta_n} = O_p(1)$, cf. Lemma 3.

By Lemma 1,

$$\bar{\psi}_n\left(\omega_{\infty,\theta_0}\right) \to_d N\left(\left(\sum_{j=1}^{\infty} \omega_{\infty,\theta_0}(j)^2\right)^{-1/2} \sum_{j=1}^{\infty} \omega_{\infty,\theta_0}(j)\, r(j), 1\right),$$

because $0 < \sum_{j=1}^{\infty} \omega_{\infty,\theta_0}(j)^2 < \infty$ since ω and d_{θ_0} are not perfectly collinear, (3) and (v) of Class B. Then the theorem follows if we show that $\bar{\psi}_n\left(\hat{\omega}_{n,\theta_n}\right) - \bar{\psi}_n\left(\omega_{\infty,\theta_0}\right) = \bar{\psi}_n\left(\hat{\omega}_{n,\theta_n}\right) - \bar{\psi}_n\left(\hat{\omega}_{n,\theta_0}\right) + \bar{\psi}_n\left(\hat{\omega}_{n,\theta_0}\right) - \bar{\psi}_n\left(\omega_{\infty,\theta_0}\right)$ is $o_p(1)$. First,

$$\begin{aligned}
&\bar{\psi}_n\left(\hat{\omega}_{n,\theta_n}\right) - \bar{\psi}_n\left(\hat{\omega}_{n,\theta_0}\right) \\
&= n^{1/2} \frac{\sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\left\{\hat{\omega}_{n,\theta_n} - \hat{\omega}_{n,\theta_0}(j)\right\}}{\left(\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_n}(j)^2\right)^{1/2}} \\
&\quad + n^{1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\, \hat{\omega}_{n,\theta_0}(j) \left\{\left(\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_n}(j)^2\right)^{-1/2} - \left(\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_0}(j)^2\right)^{-1/2}\right\},
\end{aligned}$$

where $\hat{\omega}_{n,\theta_n}(j) - \hat{\omega}_{n,\theta_0}(j) = d_{\theta_0}(j)'\left\{\beta_{n\theta_0} - \beta_{n\theta_n}\right\} + \{d_{\theta_0}(j) - d_{\theta_n}(j)\}'\beta_{n\theta_n}$. Using a MVT argument and (11), $\left\|d_{\theta_0}(j) - d_{\theta_n}(j)\right\| \le C\|\theta_n - \theta_0\| j^{-1}$, and $\left\|\beta_{n\theta_0} - \beta_{n\theta_n}\right\| = O_p(\|\theta_n - \theta_0\|)$ using the rates of decay of ω, d and $\dot{d}$. Then

$$\begin{aligned}
&n^{1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\left\{\hat{\omega}_{n,\theta_n} - \hat{\omega}_{n,\theta_0}(j)\right\} \\
&= n^{1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\, d_{\theta_0}(j)'\left\{\beta_{n\theta_0} - \beta_{n\theta_n}\right\} \\
&\quad + n^{1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\left\{d_{\theta_0}(j) - d_{\theta_n}(j)\right\}' \beta_{n\theta_n}
\end{aligned}$$

is $o_p(1)$, using the MVT, that $n^{1/2}\sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\, d_{\theta_0}(j) = O_p(1)$, $\left\|\beta_{n\theta_0} - \beta_{n\theta_n}\right\| = O_p(\|\theta_n - \theta_0\|)$, and

$$\begin{aligned}
&\left\| n^{1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\left\{d_{\theta_0}(j) - d_{\theta_n}(j)\right\}\right\| \\
&\le C\|\theta_n - \theta_0\|\, n^{1/2} \sum_{j=1}^{n-1} |\rho_{n\theta_0}(j)|\, j^{-1},
\end{aligned}$$

which is $O_p\left(n^{-1/2}\log n\right) = o_p(1)$, proceeding as in the proof of Theorem 1.

Next, $\bar{\psi}_n\left(\hat{\omega}_{n,\theta_0}\right) - \bar{\psi}_n\left(\omega_{\infty,\theta_0}\right)$ is

$$n^{1/2} \frac{\sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\left\{\hat{\omega}_{n,\theta_0}(j) - \omega_{\infty,\theta_0}(j)\right\}}{\left(\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_0}(j)^2\right)^{1/2}} \tag{16}$$

$$+\left\{\left(\sum_{j=1}^{n-1} \hat{\omega}_{n,\theta_0}(j)^2\right)^{-1/2} - \left(\sum_{j=1}^{n-1} \omega_{\infty,\theta_0}(j)^2\right)^{-1/2}\right\} \times n^{1/2} \sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\, \omega_{\infty,\theta_0}(j) \tag{17}$$

and we find that, cf. Lemma 3,

$$\begin{aligned}
&\mathbb{E}\left(n^{1/2} \sum_{j=1}^{n-1} \tilde{\gamma}_{n\theta_0}(j)\left\{\hat{\omega}_{n,\theta_0}(j) - \omega_{\infty,\theta_0}(j)\right\}\right)^2 \\
&\le \sum_{j=1}^{n-1}\left\{\hat{\omega}_{n,\theta_0}(j) - \omega_{\infty,\theta_0}(j)\right\}^2 \\
&\quad + \frac{C}{n} \sum_{j=1}^{n-1}\sum_{j'=1}^{n-1}\left|\hat{\omega}_{n,\theta_0}(j) - \omega_{\infty,\theta_0}(j)\right|\left|\hat{\omega}_{n,\theta_0}(j') - \omega_{\infty,\theta_0}(j')\right|
\end{aligned}$$

which is $o\left(\sum_{j=1}^{n-1}\left\|d_{\theta_0}(j)\right\|^2\right) + n^{-1} o\left(\sum_{j=1}^{n-1}\left\|d_{\theta_0}(j)\right\|\right)^2 = o(1)$ as $n \to \infty$, so that (16) is $o_p(1)$.

On the other hand, using Lemma 3, the term in braces in (17) is $o(1)$ as $n \to \infty$, so (17) is also $o_p(1)$ and the theorem follows. □

Proof of Corollary 3. The first part follows as Theorem 3 whereas the second one, follows noticing that $n^{1/2}\sum_{j=1}^{n-1} \rho_{n\hat{\theta}_n}(j)\, \hat{d}_{n,1\hat{\theta}_n}(j) = n^{1/2}\sum_{j=1}^{n-1} \rho_{n\theta_0}(j)\, \hat{d}_{n,1\hat{\theta}_n}(j) + o_p(1)$ using Theorem 2 and that $\hat{d}_{n,1\hat{\theta}_n}(j)$ and $d_{n,2\hat{\theta}_n}(j)$ are orthogonal. □

Appendix B. Tests using time domain autocorrelation estimates

For time domain analysis we only describe the main differences. We use the simplifying assumption that $X_t = \varepsilon_t = 0$ for $t \le 0$, cf. (2) in Robinson (1994), so that Lemmas 1 and 2 follow at once for $\hat{\gamma}_{n\theta}$ under H_0 using the Martingale property of ε_t. Then assuming that the sequence of alternatives $\{H_{1n}\}_{n\ge1}$ belongs to Class L^*, we can show Lemma 1 and then Theorem 1 under H_{1n}:

Class $\mathbf{L}^*$. $H_{1n} \in L$ and $\zeta(z) = \sum_{j=0}^{\infty} \zeta_j z^j := \varphi_{\theta_0}(z)\, \varphi^{-1}(z)$ satisfies $\zeta(0) = 1$ and $\zeta_j = n^{-1/2} r(j) + n^{-1} a_n(j)$, $j = 1, 2, \ldots$, where $|r(j)| \le K j^{-1}$, $j = 1, 2, \ldots$, and for all n sufficiently large $|a_n(j)| \le K j^{\epsilon - 1}$, $j = 1, 2, \ldots$, for all $\epsilon > 0$.

Regularity conditions on $\mathcal{J}$ for the analysis of tests based on time domain autocorrelations $\hat{\rho}_{n\theta_n}$ are similar to those for frequency domain, since, assuming that $\varphi_\theta\left(e^{i\lambda}\right)$ is differentiable so that $\xi_\theta(z) = (\partial/\partial\theta)\log\varphi_\theta(z)$, $\xi_\theta(0) = 0$ all θ, and expanding $\xi_\theta(z) = \sum_{j=1}^{\infty}\xi_{\theta,j}z^j$, we find that

$$d_\theta(j) = -\frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{Re}\left\{\xi_\theta\left(e^{i\lambda}\right)\right\}\cos(j\lambda)\,d\lambda = -\xi_{\theta,j}.$$

Theorems 2 and 3 for $\hat{\rho}_{n\theta_n}$ follow replacing condition (iv) in Class B by (iv*):

(iv*) For all $0 < \delta < 1$ there exists some $K < \infty$ such that $\psi_\theta(z) = \sum_{j=0}^{\infty}\psi_{\theta,j}z^j := \varphi_\theta(z)/\varphi_{\theta_0}(z) - 1 - (\theta-\theta_0)'\xi_{\theta_0}(z)$ satisfies that $\sup_{\{\theta:\|\theta-\theta_0\|\le\delta/2\}}\|\theta-\theta_0\|^{-2}\left|\varphi_{\theta,j}\right| \le Kj^{\delta-1}\log^2 j$, $j = 1, 2, \ldots$.

Appendix C. Lemmata

Lemma 1. *$n^{1/2}(\tilde{\rho}_{n,\theta_0}(1), \ldots, \tilde{\rho}_{n,\theta_0}(k))' \to_d N((r(1), \ldots, r(k))', I_k)$, under $H_{1n} \in L$, for k fixed and $\{X_t\}_{t=-\infty}^{\infty} \in A$.*

Proof. We only consider the asymptotic distribution of $n^{1/2}\left(\tilde{\gamma}_{n\theta_0}(1), \ldots, \tilde{\gamma}_{n\theta_0}(k)\right)'$, since $\tilde{\gamma}_{n\theta_0}(0) \to_p \sigma^2$ under H_{1n}, see e.g. (51) in the proof of Theorem 2 of DHV. First, we write $f_{\theta_0}(\lambda)^{-1} = f(\lambda)^{-1}\left\{1 + n^{-1/2}h_n(\lambda)\right\}$, where $h_n(\lambda) = l(\lambda) + n^{-1/2}g_n(\lambda)$ satisfies that $\int_0^\pi h_n(\lambda)\cos(\lambda j)\,d\lambda = r(j) + n^{-1/2}a_n(j)$. Then, under H_{1n},

$$\tilde{\gamma}_{n\theta_0}(j) = \frac{2\pi}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{I_X(\lambda_k)}{f(\lambda_k)}\cos(\lambda_k j)\left\{1 + \frac{l(\lambda_k)}{n^{1/2}} + \frac{g_n(\lambda_k)}{n}\right\}$$

Now, reasoning as in the proof of Theorem 5 of DHV and using that g_n is integrable, $\tilde{\gamma}_{n\theta_0}(j) = \tilde{\gamma}_{n\varepsilon}(j) + n^{-1/2}\sigma^2 r(j) + o_p\left(n^{-1/2}\right)$, cf. Also the proof of Theorem 1. The convergence then follows as in Lemma 7(b) of DHV using Lemma 2. □

Lemma 2. *Assume that $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is as in Class A. Then $n\mathbb{E}\left[\tilde{\gamma}_{n\varepsilon}^2(j)\right] = \sigma^4 + O\left(n^{-1}\right)$, $j = 1, 2, \ldots$, and $n\mathbb{E}\left[\tilde{\gamma}_{n\varepsilon}(j)\,\tilde{\gamma}_{n\varepsilon}(j')\right] = O\left(n^{-1}\right)$, $j \neq j'$, as $n \to \infty$.*

Proof. It follows by direct calculation of the moments of $I_\varepsilon\left(\lambda_j\right)$, cf. Brillinger (1980, Theorem 4.3.1) and approximation of sums by integrals. □

Lemma 3. *Under (3), (10) and (11), uniformly in $j = 1, 2, \ldots$, $|\hat{\omega}_{n,\theta_0}(j) - \omega_{\infty,\theta_0}(j)| = o\left(\left\|d_{\theta_0}(j)\right\|\right)$ and $\left|\hat{\omega}_{n,\theta_0}(j)^2 - \omega_{\infty,\theta_0}(j)^2\right| = o\left(\left\|d_{\theta_0}(j)\right\|^2 + \left\|d_{\theta_0}(j)\right\||\omega(j)|\right)$, as $n \to \infty$.*

Proof. Follows using standard ordinary least squares algebra. □

References

Billingsley, P., 1999. Convergence of Probability Measures, second edition. Wiley, New York.

Bloomfield, P.J., 1973. An exponential model for the spectrum of an scalar time series. Biometrika 60, 217–226.

Box, G.E.P., 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.

Brillinger, D.R., 1980. Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New York.

Delgado, M.A., Hidalgo, J., Velasco, C., 2005. Distribution free goodness-of-fit tests for linear processes. Annals of Statistics 33, 2568–2609.

Delgado, M.A., Hidalgo, J., Velasco, C., 2009. Distribution free specification tests for dynamic linear models. Econometrics Journal 12, 105–134.

Durbin, J., 1970. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. Econometrica 38, 410–421.

Hannan, E.J., Heyde, C.C., 1972. On limit theorems for quadratic functions of discrete time series. Annals of Mathematical Statistics 43, 2058–2066.

Hosking, J.R.M., 1978. An unified derivation of the asymptotic distributions of goodness-of-fit statistics for autoregressive time-series models. Journal of the Royal Statistical Society, Series B 40, 341–349.

Hosking, J.R.M., 1980. Lagrange-multiplier tests of time-series models. Journal of the Royal Statistical Society, Series B 42, 170–181.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.

Lobato, I., 2001. Testing that a dependent process is uncorrelated. Journal of the American Statistical Association 96, 1066–1076.

Lobato, I., Nankervis, J.C., Savin, N.E., 2001. Testing autocorrelation using a modified Box–Pierce Q test. International Economic Review 42, 187–205.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Noether, G.E., 1955. On a theorem of Pitman. Annals of Mathematical Statistics 26, 64–68.

Quenouille, M.H., 1947. A large-sample test for the goodness-of-fit of autoregressive squemes. Journal of the Royal Statistical Society, Series B 110, 123–129.

Robinson, P.M., 1994. Efficient tests of non-stationary hypothesis. Journal of the American Statistical Association 89, 1420–1437.

Romano, J.P., Thombs, L.A., 1996. Inference for autocorrelations under weak assumptions. Journal of the American Statistical Association 91, 590–600.

Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimates of non-stationary time series. Journal of the American Statistical Association 95, 1229–1243.