COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1077

TESTING FOR SERIAL CORRELATION AGAINST AN ARMA(1,1) PROCESS

Donald W. K. Andrews and Werner Ploberger

September 1994

Cowles Foundation Discussion Paper 1077

TESTING FOR SERIAL CORRELATION AGAINST AN ARMA(1,1) PROCESS

Donald W. K. Andrews*
Cowles Foundation for Research in Economics
Yale University

and

Werner Ploberger
Technische Universität Wien

March 1994
Revised: September 1994

ABSTRACT

This paper is concerned with tests for serial correlation in time series and in the errors of regression models. In particular, the nonstandard problem of testing for white noise against ARMA(1,1) alternatives is considered. Sup Lagrange multiplier (LM) and exponential average LM tests are introduced and are shown to be asymptotically admissible for ARMA(1,1) alternatives. In addition, they are shown to be consistent against all (weakly stationary strong mixing) non-white noise alternatives. Simulation results compare the tests to several tests in the literature. These results show that the Exp–LM_∞ test has very good all-around power properties.

1. INTRODUCTION

This paper considers tests of serial correlation that are designed for autoregressive moving average (ARMA) models of order (1,1) under the alternative hypothesis. It is natural to consider tests of this sort, because ARMA(1,1) models are known to provide parsimonious representations of a broad class of stationary time series. The most popular existing tests, such as the Durbin and Watson (1950) and Box and Pierce (1970) tests are designed for the less flexible AR(1) (and MA(1)) model and for the less parsimonious AR(p) model respectively.

Testing for serial correlation in an ARMA(1,1) model is a non-standard testing problem, because the ARMA(1,1) model reduces to a white noise model whenever the AR and MA coefficients are equal. In consequence, the testing problem is one in which a nuisance parameter is present only under the alternative hypothesis. Problems of this sort have been considered by Davies (1977, 1987), Hansen (1991), and Andrews and Ploberger (1993, 1994). The standard likelihood ratio (LR) statistic does not possess its usual chi-square asymptotic distribution or its usual asymptotic optimality properties (of the sort established by Wald (1943)) in such cases. Nevertheless, Andrews and Ploberger (1993) show that the standard LR test and an asymptotically equivalent "sup" Lagrange multiplier (LM) test do possess an asymptotic admissibility property. In addition, Andrews and Ploberger (1994) derive a class of tests, denoted average exponential tests, that possess certain asymptotic optimality properties for testing problems of the sort discussed above.

The results of Andrews and Ploberger (1993, 1994) are general results that impose "high-level" assumptions. In this paper, we first show that these results apply to the problem of testing for serial correlation in ARMA(1,1) models. We provide explicit expressions for the average exponential and sup LM test statistics for the problem at hand. We then show that the corresponding tests have the attractive feature of being consistent against all forms of serial correlation. In consequence, the average exponential tests possess asymptotic optimality properties for a parametric class of alternatives and the robustness property of consistency against all (weakly stationary strong mixing) alternatives. This robustness feature is not shared by the Durbin–Watson and

Box–Pierce tests. Third, we compare by simulation the tests discussed above. These include the average exponential, the sup LM, the Durbin–Watson, the Box–Pierce, and several point optimal invariant (POI) tests. The POI tests have been introduced recently by Rahman and King (1992). The Exp-LM_∞ test is found to have very good all-around power properties.

We note that the LM procedure for order selection of ARMA models, which has been considered by Poskitt and Tremayne (1980) and Pötscher (1983, 1985), reduces to just a two-sided Durbin–Watson test in the case of testing an ARMA(0,0) model against an ARMA(1,1) model as is considered here. Hence, the simulation results referred to above cover such LM tests. We also note that the sup LM test is asymptotically equivalent (under the null and local alternatives) to the sup LR test, which has been analyzed by Hannan (1982), also see Veres (1987). The sup LR test is not considered in the simulation experiment.

All limits below are taken as $T \to \infty$ unless specified otherwise.

2. TESTS OF SERIAL CORRELATION FOR ARMA(1,1) PROCESSES

2.1. Definition of Model and Test Statistics

The model we consider here is the ARMA(1,1) model:

(2.1) $$Y_t = (\pi + \beta)Y_{t-1} + \varepsilon_t - \pi\varepsilon_{t-1} \quad \text{for} \quad t = 2, 3, \ldots,$$

where $\{Y_t \ : \ t = 1, \ldots, T\}$ are observed random variables (rv's) and $\{\varepsilon_t \ : \ t = 1, 2, \ldots\}$ are unobserved innovations. The parameter space for π is Π and for β is B. Throughout the paper, we assume Π and B are such that the absolute value of the autoregressive coefficient $\pi + \beta$ is bounded below one, Π is closed, and B contains a neighborhood of zero. The former condition rules out unit root and explosive behavior of $\{Y_t \ : \ t = 1, 2, \ldots\}$.

We are interested in testing the null hypothesis of white noise against the alternative of serial correlation of $\{Y_t \ : \ t = 1, 2, \ldots\}$. These hypotheses are given by

(2.2) $$\mathrm{H}_0 \ : \ \beta = 0 \quad \text{and} \quad \mathrm{H}_1 \ : \ \beta \neq 0.$$

Note that when $\beta = 0$, the model (2.1) reduces to $Y_t = \varepsilon_t$ and the parameter π is no longer present. In consequence, the above testing problem is non-standard.

Under the following assumption, we derive the standard LR, sup LM, and the average exponential tests.

ASSUMPTION 1: *$\{\varepsilon_t : t = 1, 2, ...\}$ is a sequence of* iid $N(0, \sigma^2)$ *rv's for some $\sigma^2 > 0$ and* $Y_1 = \varepsilon_1$.

Assumption 1 is used to generate the test statistics of interest, but we consider the asymptotic properties of these tests below under a much more general specification of the distribution of the innovations. The assumption on Y_1 is made for simplicity. With some added complexity, we could assume $\{Y_t : t = 1, 2, ...\}$ is part of a doubly infinite sequence of stationary rv's that satisfy (2.1) for all $t = ..., 0, 1, ...$.

The standard LR statistic equals minus two times the logarithm of the likelihood ratio. Because the parameter π only appears in the denominator of the ratio, the unrestricted maximum of the likelihood function with respect to this parameter can be performed after the ratio has been computed for a given π. That is, let $LR_T(\pi)$ denote the standard LR statistic for testing H_0 versus H_1 when π is known under the alternative to equal π. Then, the standard LR test statistic is $\sup_{\pi \in \Pi} LR_T(\pi)$.

As shown in Andrews and Ploberger (1994, Proof of Theorem A–1), an asymptotically equivalent test statistic (under the null and local alternatives) is given by

$$(2.3) \quad \sup_{\pi \in \Pi} LM_T(\pi),$$

where for the present testing problem

$$(2.4) \quad \begin{aligned} LM_T(\pi) &= \left(\tfrac{1}{\sqrt{T}} \Sigma_{t=2}^{T} Y_t \Sigma_{i=0}^{t-2} \pi^i Y_{t-i-1}\right)^2 (1-\pi^2)/\widehat{\sigma}_Y^4, \\ \widehat{\sigma}_Y^2 &= \tfrac{1}{T-1} \Sigma_{t=1}^{T} (Y_t - \overline{Y}_T)^2, \text{ and } \overline{Y}_T = \tfrac{1}{T} \Sigma_{t=1}^{T} Y_t. \end{aligned}$$

(Note that the term $b_t(\pi) = \Sigma_{i=0}^{t-2} \pi^i Y_{t-i-1}$, which appears in the definition of $LM_T(\pi)$, can be computed recursively via the recursion $b_2(\pi) = Y_1$ and $b_t(\pi) = Y_{t-1} + \pi b_t(\pi)$ for $t = 3 ..., T$. In consequence, $LM_T(\pi)$ can be computed using a single do loop.)

By verifying the conditions of Theorem 1 of Andrews and Ploberger (1993), we find that the sup LM test satisfies the following asymptotic admissibility property. Let Power $(\varphi_T,\ \beta,\ \pi)$ denote the power of the test φ_T when the true parameters are β and π.

PROPOSITION 1: *Let $\{\xi_T\ :\ T \geq 1\}$ be a sequence of asymptotically level α sup LM tests. Under Assumption 1, given any sequence of asymptotically level α tests $\{\varphi_T\ :\ T \geq 1\}$ and any probability distribution $J(\cdot)$ on Π whose support is Π, there exists a constant $r_{\varphi,J} < \infty$ such that for all $r \geq r_{\varphi,J}$ we have*

$$\begin{aligned}&\overline{\lim_{T\to\infty}} \int \left[\text{Power}(\varphi_T,\ r\sqrt{1-\pi^2}/\sqrt{T},\ \pi) + \text{Power}(\varphi_T,\ -r\sqrt{1-\pi^2}/\sqrt{T},\ \pi)\right] dJ(\pi)\\ \leq\ &\underline{\lim_{T\to\infty}} \int \left[\text{Power}(\xi_T,\ r\sqrt{1-\pi^2}/\sqrt{T},\ \pi) + \text{Power}(\xi_T,\ -r\sqrt{1-\pi^2}/\sqrt{T},\ \pi)\right] dJ(\pi).\end{aligned}$$

COMMENT: The result of Proposition 1 concerns the asymptotic local power of the sup LM test since it considers parameter values β that are proportional to $1/\sqrt{T}$. Proposition 1 shows that the sup LM test beats any given test in terms of weighted average power against alternatives that are local to, but sufficiently distant from, the null. The weighting is over positive and negative β values and is with respect to an arbitrary function $J(\cdot)$ on Π.

Next, we discuss the average exponential tests that are introduced in Andrews and Ploberger (1994). These tests are asymptotically optimal in the sense that they minimize weighted average power for specific weight functions. The weight functions for the parameter β are mean zero normal densities with variances proportional to a scalar $c > 0$. For small c, most weight is placed on alternatives that are close to the null. For large c, weight is distributed more uniformly across β values. The weight function J for the parameter π is chosen by the investigator. For the simulation results of this paper, we take it to be uniform on Π.

For each $c \in (0, \infty)$, the average exponential LM test statistic is given by

$$\text{(2.5)} \quad \text{Exp–}LM_{cT} = (1+c)^{-1/2} \int \exp\left(\tfrac{1}{2}\,\tfrac{c}{1+c} LM_T(\pi)\right) dJ(\pi),$$

where $LM_T(\pi)$ is as defined above and $J(\cdot)$ is a probability measure on Π, such as the uniform measure.

The limiting average exponential LM test statistics (after suitable normalization, see Andrews and Ploberger (1994)) as $c \to 0$ and $c \to \infty$ are given by

$$(2.6)\quad \begin{aligned} &\text{Exp--}LM_{0T} = \int LM_T(\pi) dJ(\pi) \text{ and} \\ &\text{Exp--}LM_{\infty T} = \ln \int \exp\left(\tfrac{1}{2} LM_T(\pi)\right) dJ(\pi). \end{aligned}$$

Under Assumption 1, Theorem 2 of Andrews and Ploberger (1994) can be applied to yield the following asymptotic local power optimality property for the Exp–LM_{cT} test. Let ξ_{cT} denote a test based on the test statistic Exp–LM_{cT}. Let $\phi(\beta, \omega)$ denote the density at the point β of a mean zero variance ω normal rv.

PROPOSITION 2: *Under Assumption* 1, *for any* $0 < c < \infty$ *and any sequence of asymptotically level* α *tests* $\{\varphi_T : T \geq 1\}$, *the sequence of asymptotically level* α *exponential LM tests* $\{\xi_{cT} : T \geq 1\}$ *satisfies*

$$\begin{aligned} &\overline{\lim_{T\to\infty}} \int\int \text{Power}(\varphi_T, \beta/\sqrt{T}, \pi)\phi(\beta, c/(1-\pi^2))d\beta dJ(\pi) \\ \leq\ &\underline{\lim_{T\to\infty}} \int\int \text{Power}(\xi_{cT}, \beta/\sqrt{T}, \pi)\phi(\beta, c/(1-\pi^2))d\beta dJ(\pi). \end{aligned}$$

2.2. Asymptotic Null Distribution of the Test Statistics

We establish the asymptotic null distribution of the test statistics introduced above using the following assumption.

ASSUMPTION 2: *The* rv's $\{Y_t : t = 1, 2, ...\}$ *satisfy* $E(Y_t|\mathcal{F}_{t-1}) = 0$ a.s. $\forall t \geq 1$, $E(Y_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s. $\forall t \geq 1$, *and* $\sup_{t\geq 1} E|Y_t|^{4+\delta} < \infty$ *for some* $\delta > 0$, *where* $\mathcal{F}_t$ *denotes the* σ*-field generated by* $Y_1, ..., Y_t$.

The asymptotic null distributions of the test statistics are established by showing that the sequence of stochastic processes $\{LM_T(\cdot) : T \geq 1\}$ indexed by $\pi \in \Pi$ converges weakly to a stochastic process $G(\cdot)$ and then applying the continuous mapping theorem. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes. (We define weak convergence using the uniform metric on the appropriate space of functions on Π, as in Pollard (1990).) Let $\xrightarrow{d}$ denote convergence in distribution of a sequence of rv's. Let $\{Z_i \,:\, i \geq 1\}$ be a sequence of iid $N(0,1)$ rv's. Define

$$(2.7) \quad G(\pi) = (1-\pi^2)\left(\sum_{i=0}^{\infty} \pi^i Z_i\right)^2 \quad \text{for } \pi \in \Pi.$$

THEOREM 1: *Under Assumption* 2,

(a) $LM_T(\cdot) \Rightarrow G(\cdot)$,

(b) $\sup_{\pi\in\Pi} LM_T(\pi) \xrightarrow{d} \sup_{\pi\in\Pi} G(\pi)$,

(c) Exp–$LM_{cT} \xrightarrow{d} (1+c)^{-1/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c}G(\pi)\right) dJ(\pi)$ *for all* $0 < c < \infty$,

(d) Exp–$LM_{0T} \xrightarrow{d} \int G(\pi) dJ(\pi)$, *and*

(e) Exp–$LM_{\infty T} \xrightarrow{d} \ln \int \exp\left(\frac{1}{2}G(\pi)\right) dJ(\pi)$.

COMMENT: The martingale difference condition in Assumption 2 is not essential for the results of Theorem 1 to hold. What is essential is that (i) $EY_t = 0 \ \forall t \geq 1$, (ii) $EY_t^2 = \sigma^2 > 0 \ \forall t \geq 1$, (iii) $EY_sY_t = 0 \ \forall s < t$, (iv) $EY_sY_tY_uY_v = 0 \ \forall s \leq t \leq u \leq v$ unless $s = t$ and $u = v$, and (v) $\frac{1}{\sqrt{T}}\Sigma_{t=2}^T Y_t \Sigma_{i=0}^{t-2} \pi^i Y_{t-i-1}$ satisfies a CLT. Assumption 2 implies conditions (i)–(v). An alternative to Assumption 2, which avoids the martingale difference assumption, is to assume conditions (i)–(iv) hold and $\{Y_t \,:\, t \geq 1\}$ is strong mixing (defined below) with strong mixing numbers that satisfy $\Sigma_{j=0}^{\infty} \alpha(j)^{(\kappa-2)/\kappa} < \infty$ and $\sup_{t\geq 1} E|Y_t|^{\kappa} < \infty$ for some $\kappa \geq 4$. The CLT of condition (v) holds under these conditions by Corollary 1 of Herrndorf (1984).

Asymptotic critical values for the test statistics in Theorem 1 can be simulated quite easily by truncating the series $\Sigma_{i=0}^{\infty} \pi^i Z_i$ at a large value TR. Table 1 provides such values for the sup LM, Exp–LM_0, and Exp–LM_∞ tests for the parameter space $\Pi = \{0, \pm.01, ..., \pm.79, \pm.80\}$. The critical values in Table 1 are based on $TR = 50$ and 40,000 repetitions.

2.3. Consistency Properties

In this section we show that the sup LM and average exponential LM test statistics are consistent against all deviations from the null hypothesis of white noise within a class of weakly stationary strong mixing sequences of rv's. This property illustrates the robust power properties of the tests. It is not shared by other common tests such as the Durbin–Watson and Box– Pierce tests.

We first state several definitions. The sequence of rv's $\{Y_t : t \geq 1\}$ is said to be *weakly stationary* if EY_tY_{t-i} does not depend on t for all $t \geq 1$ and $i \geq 0$. The sequence $\{Y_t : t \geq 1\}$ is said to be *strong mixing* if

$$\alpha(m) = \sup_{t \geq 1} \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+m}^\infty} |P(A \cap B) - P(A)P(B)| \to 0 \text{ as } m \to \infty,$$

where $\mathcal{F}_{-\infty}^t$ and $\mathcal{F}_{t+m}^\infty$ are the σ- fields generated by ..., Y_{t-1}, Y_t and Y_{t+m}, Y_{t+m+1}, ... respectively. A sequence of rv's $\{W_T : T \geq 1\}$ is said to *converge in probability to infinity* (denoted $W_T \xrightarrow{p} \infty$) if $P(W_T > M) \to 1 \; \forall M < \infty$.

For the consistency results, we assume:

ASSUMPTION 3: *$\{Y_t : t \geq 1\}$ is a mean zero weakly stationary strong mixing sequence of random variables with $EY_t^2 = \sigma^2 > 0 \; \forall t \geq 1$ whose strong mixing numbers $\{\alpha(j) : j \geq 1\}$ satisfy $\Sigma_{j=1}^\infty \alpha(j)^{\delta/(4+\delta)} < \infty$ and for which $\sup_{t \geq 1} E|Y_t|^{4+\delta} < \infty$ for some $\delta > 0$.*

THEOREM 2: *Suppose $\{Y_t : t \geq 1\}$ satisfies Assumption* 3. *Also, in parts* (b) *and* (c) *below, suppose $\gamma_i \neq 0$ for some $i \geq 1$, where $\gamma_i = EY_tY_{t-i}$. Then,*

(a) $\sup_{\pi \in \Pi} |LM_T(\pi)/T - (\Sigma_{i=0}^\infty \pi^i \gamma_{i+1})^2(1-\pi^2)/\sigma^4| \xrightarrow{p} 0$,

(b) $\sup_{\pi \in \Pi} LM_T(\pi) \xrightarrow{p} \infty$ *provided* Π *is an infinite set, and*

(c) Exp–$LM_{cT} \xrightarrow{p} \infty \; \forall 0 \leq c \leq \infty$ *provided the support of $J(\cdot)$ is an infinite set.*

COMMENTS: 1. Theorem 2(b) and (c) show that the sup LM and average exponential LM tests are consistent against processes that have some autocovariance not equal to zero.

2. For a result analogous to Theorem 2(b) for the sup LR test, see Pötscher and Srinivasan (1994, pp. 37–38).

3. TESTS OF SERIAL CORRELATION FOR REGRESSION ERRORS

In this section, we show that the tests introduced above can be used to test whether regression errors are serially correlated. The tests are constructed using residuals rather than the errors themselves. Provided that the regressors are exogenous (defined below), the resultant sup LM and average exponential LM test statistics have the same asymptotic distribution as when the actual errors are used to construct the statistics. In consequence, the asymptotic critical values given in Section 2 above are applicable.

The model we consider is given by

$$(3.1)\quad W_t = g(X_t,\ \lambda) + Y_t \ \text{ for } \ t = 1, ..., T,$$

where $\{Y_t \ : \ t \leq T\}$ are unobserved errors, $\{X_t \ : \ t \leq T\}$ are observed regressor p–vectors, $\{W_t \ : \ t \leq T\}$ are observed dependent variables, λ is an unknown parameter, and $g(\cdot\,,\cdot)$ is a known function. We consider two cases concerning the properties of the regression function. In the first case, the regression function may be non-linear, but must be non-trending. In the second case, the regression function is linear, but may be deterministically trending. In either case, we assume we have a consistent estimator $\widehat{\lambda}$ of λ_0 that is used to define the residuals

$$(3.2)\quad \widehat{Y}_t = W_t - g(X_t,\ \widehat{\lambda}) \ \text{ for } \ t = 1, ..., T.$$

Under the null hypothesis of no serial correlation we impose one or other of the following two assumptions depending upon the nature of the regression function.

ASSUMPTION 3: (i) *Assumption* 2 *holds with* $\mathcal{F}_t$ *equal to the* σ*-field generated by* $(X_1,\ X_2,\ ...\ )$ *and* $(Y_1,\ ...,\ Y_t)$.

(ii) $g(X_t,\ \lambda)$ *is twice differentiable in* λ a.s., $\sup_{t\geq 1} E\|\frac{\partial}{\partial\lambda}g(X_t,\ \lambda_0)\|^2 < \infty$, *and* $\sup_{t\geq 1} E\|\frac{\partial^2}{\partial\lambda\partial\lambda'}g(X_t,\ \lambda)\|^{4/3} < \infty$ *for some* $\varepsilon > 0$, *and*

(iii) $T^{1/4}(\widehat{\lambda} - \lambda_0) \xrightarrow{p} 0$.

ASSUMPTION 4: (i) *Assumption* 2 *holds with* $\mathcal{F}_t$ *equal to the* σ*-field generated by* $(X_1, X_2, \dots)$ *and* $(Y_1, \dots, Y_t)$.

(ii) $g(X_t, \lambda) = X_t'\lambda \ \forall t \geq 1$, *and*

(iii) *For some sequence* $\{\Delta_T : T \geq 1\}$ *of non-stochastic* $p \times p$ *diagonal matrices,* $\Delta_T(\widehat{\lambda} - \lambda) = O_p(1)$, $\sup_{t \leq T} E\|\Delta_T^{-1} X_t\|^2 \to 0$, *and* $[\Delta_T]_{jj} \to \infty \ \forall j \leq p$.

Part (i) of Assumptions 3 and 4 requires exogeneity of $\{X_t : t \geq 1\}$ in the strong sense that the conditional mean of Y_t is zero given past values of Y_t and past and *future* values of X_t. This assumption rules out dynamic regression models that include lagged values of the dependent variable.

The least squares estimator of λ typically satisfies the consistency and rate of convergence results required in part (iii) of Assumptions 3 and 4.

If $X_t = (1, t, t^2)'$, then $\Delta_T = \text{Diag}(T^{1/2}, T^{3/2}, T^{5/2})$ in Assumption 4(iii) and $\sup_{t \leq T} E\|\Delta_T^{-1} X_t\|^2 \to 0$ as required.

The following result justifies the use of the sup LM and average exponential LM tests when constructed using residuals.

THEOREM 3: *Under Assumption* 3 *or* 4, *the results of Theorem* 1 *still hold when* $LM_T(\pi)$ *is constructed using the residuals* $\{\widehat{Y}_t : t \leq T\}$ *defined in* (3.2) *rather than the* rv's $\{Y_t : t \leq T\}$.

4. MONTE CARLO POWER COMPARISONS

In this section, we compare the finite sample power of the tests introduced above with several tests in the literature. The latter include one- and two-sided versions of the Durbin–Watson (DW) test, two versions of the Box–Pierce (BP) test, and three point optimal invariance (POI) tests introduced recently by Rahman and King (1992). The DW tests have some asymptotic optimality properties for AR(1) and MA(1) alternatives. The BP tests have some asymptotic optimality properties for higher order AR processes and are often used in practice to detect serial correlation beyond the first order. The POI tests have optimal power against particular

ARMA(1,1) alternatives and have been suggested as being suitable for testing against a variety of ARMA(1,1) alternatives.

Since our interest here is in tests with good all-around power properties, we consider a variety of different alternatives. We consider AR(1), MA(1), AR(6), AR(12), and ARMA(1,1) models. We are interested in (i) the sensitivity of the power of the Exp–LM_c tests to $c \in [0, \infty]$, (ii) power comparisons between the Exp–LM_c tests and the Sup–LM test, and (iii) power comparisons between the former tests and the DW, BP, and POI tests.

The model we consider is the location model with serially correlated errors:

(4.1) $W_t = \lambda + Y_t$ for $t = 1, ..., T$,

where the sample size T is equal to 100. The models used for the errors Y_t are:

$$
\begin{aligned}
&\text{AR}(1) : Y_t = \rho Y_{t-1} + \varepsilon_t\,,\ \rho = .2,\ .3,\ -.2,\ -.3;\\
&\text{MA}(1) : Y_t = \varepsilon_t + \phi\varepsilon_{t-1},\ \phi = .2,\ .3,\ -.2,\ -.3;\\
&\text{AR}(6) : Y_t = \rho\Sigma_{j=1}^{6}\tfrac{7-j}{6}Y_{t-j} + \varepsilon_t,\ \rho = .15,\ .2;\\
&\text{AR}(12) : Y_t = \rho\Sigma_{j=1}^{12}\tfrac{13-j}{12}Y_{t-j} + \varepsilon_t\,,\ \rho = .1,\ .2;\\
&\text{AR}(6)\pm : Y_t = \rho\Sigma_{j=1}^{6}(-1)^{j+1}\tfrac{7-j}{6}Y_{t-j} + \varepsilon_t\,,\ \rho = .3,\ .4;\\
&\text{AR}(12)\pm : Y_t = \rho\Sigma_{j=1}^{12}(-1)^{j+1}\tfrac{13-j}{12}Y_{t-j} + \varepsilon_t\,,\ \rho = .3,\ .4;\\
&\text{ARMA}(1,1) : Y_t = \rho Y_{t-1} + \varepsilon_t + \phi\varepsilon_{t-1}\,, (\rho, \phi) = (.2,\ .1),\ (.4,\ -.1),\ (.6,\ -.3),\\
&\qquad (.8,\ -.6),\ (-.2,\ .6),\ (-.4,\ .8),\ (-.6,\ 1.0),\ (-.8,\ 1.6),\ (.2,\ -.5),\\
&\qquad (.4,\ -.7),\ (.6,\ -1.2),\ (-.2,\ -.1),\ (-.4,\ .1),\ (-.6,\ .4),\ (-.8,\ .6).
\end{aligned}
\tag{4.2}
$$

The innovations ε_t are distributed iid $N(0,1)$. The AR(1), MA(1), and ARMA(1,1) models are simulated with a stationary start-up using the method described in Ansley (1979) and Rahman and King (1992). The AR(6), AR(12), AR(6)±, and AR(12)± models are simulated with an approximately stationary start-up by taking the last 100 rv's from a simulated sequence of 300 rv's that is started-up by taking the 6 or 12 pre-sample Y_t rv's to be zeroes.

The models above are chosen because they include a wide variety of patterns of serial correlation with both positive and negative serial correlations present. For each model considered, one or more of the tests considered has some asymptotic optimality properties. The parameter values were chosen so that the tests had power in the range (.5, .9) in most cases. The parameter values for the ARMA(1,1) model correspond (roughly) to points on diagonal lines above and below the main diagonal in the (ρ, ϕ) parameter space. The main diagonal constitutes the null hypothesis for this model.

The average exponential tests considered have $c = 0$, 1, and ∞ and $\Pi = \{-.80, -.79, ..., .79, .80\}$. Results for tests with $c = 1/3$ and $c = 3$ were also computed, but they are not reported here, because their power lies between that of the $c = 0$, 1, and ∞ tests. The parameter space Π above is chosen as a compromise between a smaller finite set, which has computational advantages, and a larger finite set, which has broader consistency properties.

The DW statistic is defined by

$$DW = \Sigma_{t=2}^{T}(\widehat{W}_t - \widehat{W}_{t-1})^2 / \Sigma_{t=1}^{T}\widehat{W}_t^2\,, \text{ where} \tag{4.3}$$
$$\widehat{W}_t = W_t - \overline{W}_T \text{ and } \overline{W}_T = \tfrac{1}{T}\Sigma_{t=1}^{T}W_t\,.$$

Three tests based on DW are considered. DW2 denotes the two-sided DW test that rejects H_0 when |DW−2| is sufficiently large. DW+ denotes the one-sided DW test that is designed for positive serial correlation. It rejects H_0 when DW is sufficiently small. DW+ is the test that is usually referred to in the literature as 'the Durbin–Watson test'. DW− denotes the one-sided DW test that is designed for negative serial correlation. It rejects H_0 if DW is sufficiently large.

Two BP tests are considered — one based on six sample autocorrelations and the other on twelve. The two test statistics are defined by

$$\text{BP6} = T(T{+}2)\Sigma_{j=1}^{6} r_j/(T{-}j) \text{ and}$$
$$\text{BP12} = T(T{+}2)\Sigma_{j=1}^{12} r_j/(T{-}j)\,, \text{ where} \tag{4.4}$$
$$r_j = \Sigma_{t=j+1}^{T}\widehat{W}_t\widehat{W}_{t-j} / \Sigma_{t=1}^{T}\widehat{W}_t^2\,.$$

The tests based on BP6 and BP12 reject when the corresponding statistics are sufficiently large.

Three POI tests are considered. Each is designed to direct power at a particular ARMA(1,1) parameter vector (ρ_0, ϕ_0). Following Rahman and King's (1992) suggestion, we consider the three vectors $(\rho_0, \phi_0) = (.5,\ 0), (0,\ .5)$, and $(.5,\ .5)$. Let $\Omega(\rho_0, \phi_0)$ denote the $T \times T$ covariance matrix of $W = (W_1, ..., W_T)'$ when $W \sim$ ARMA$(1,1)$ with parameters (ρ_0, ϕ_0). Let G be a $T \times T$

matrix such that $\Omega(\rho_0, \phi_0) = GG'$. Let W_t^* be the t–th residual from the transformed model $G^{-1}W = G^{-1}\underset{\sim}{1}\lambda + G^{-1}Y$, where $Y = (Y_1, ..., Y_T)'$ and $\underset{\sim}{1} = (1, ..., 1)' \in R^T$.

Given (ρ_0, ϕ_0), the POI test statistic is defined by

$$\text{POI}(\rho_0, \phi_0) = W^{*\prime}W^*/\widehat{W}'\widehat{W}\,, \text{ where} \tag{4.5}$$
$$\widehat{W} = (\widehat{W}_1, ..., \widehat{W}_T)' \quad \text{and} \quad W^* = (W_1^*, ..., W_T^*)'.$$

See Rahman and King (1992) for an algorithm for computing POI(ρ_0, ϕ_0). The POI(ρ_0, ϕ_0) test rejects H_0 if the POI(ρ_0, ϕ_0) statistic is sufficiently small.

Table 2 presents the power of each of the tests described above for the AR(1), MA(1), AR(6), AR(12), AR(6)±, and AR(12)± models. Table 3 does likewise for the ARMA(1,1) models. In each case, size-corrected power is presented. The finite-sample critical values required to compute size-corrected power were calculated by simulation using 25,000 repetitions. The power results were calculated by simulation using 5,000 repetitions. The $N(0,1)$ rv's were simulated using the RAN1 and GASDEV algorithms in Press, Flannery, Teukolsky, and Vetterling (1986).

Tables 2 and 3 show that the power of the Exp–LM_c test is always monotone in c. Depending upon the model, it may be monotone increasing or monotone decreasing. The difference in power of the tests for $c = 0$ and $c = \infty$ is quite small in most cases. In only six cases out of a total of thirty-one is the difference greater than .03. Furthermore, in each of the six cases where the difference is greater than .03, power is increasing in c, so that $c = \infty$ is the best test. Hence, we conclude that (i) the choice of c is not critical for most models and (ii) the choice of $c = \infty$ is preferable because it is the best choice for those models in which power is most sensitive to c.

Tables 2 and 3 show that Sup–LM has higher power than Exp–LM_∞ in only six out of thirty-one cases, whereas the reverse is true in twenty-two cases. For those cases where Sup–LM beats Exp–LM_∞, the average difference in power is .02, whereas for the cases where Exp–LM_∞ beats Sup–LM, the average difference in power is .04. Thus, we conclude that Exp–LM_∞ has higher all-around power than Sup–LM by a small, but significant, margin.

Next, we compare the power of the Exp–LM_∞ test to the DW, BP, and POI tests. The DW+, POI(.5, 0), POI(0, .5), and POI(.5, .5) tests all have no power against alternatives that exhibit negative first-order serial correlation. Hence, none of these tests is a good all-around test. On the other hand, each of these tests has higher power than Exp–LM_∞ for all of the other cases except the AR(6) and AR(12) cases. The DW– test has no power against alternatives that exhibit positive first-order serial correlation. Hence, it too is not a good all-around test.

The DW2, BP6, and BP12 tests have nondegenerate power against each of the alternatives considered. The Exp–LM_∞ test dominates the BP6 test which, in turn, domiantes the BP12 test (for the given alternatives and the sample size 100). The superiority of Exp–LM_∞ over BP6 and BP12 is substantial. Averaged over the thirty-one cases, its power is greater than that of the latter two tests by .15 and .24 respectively.

The Exp–LM_∞ and DW2 tests are the best all-around tests. DW2 is better in the AR(1) and MA(1) models with positive coefficients. The two tests are equal for these models with negative coefficients. The Exp–LM_∞ test is better than the DW2 test for the AR(6), AR(12), AR(6)±, and AR(12)± models. Averaged over the sixteen cases in Table 2, Exp–LM_∞ has higher power than DW2 by .03. For the ARMA(1,1) models, the ordering between Exp–LM_∞ and DW2 depends on the parameters. Averaged over all fifteen parameter vectors considered in Table 3, Exp–LM_∞ has higher power than DW by .04. Based on the above results, we conclude that Exp–LM_∞ and DW2 are the best all-around tests with Exp–LM_∞ being preferable by a slight margin.

5. CONCLUSION

This paper establishes the asymptotic null distribution and asymptotic admissibility for ARMA(1,1) alternatives of a class of exponential average LM tests and the sup LM test. The paper also shows that these tests are consistent against all non-white noise alternatives. Simulation results suggest that the Exp–LM_∞ test is the best of these tests and that it performs very well relative to existing tests in an all-around sense.

APPENDIX OF PROOFS

PROOF OF PROPOSITION 1: It suffices to verify Assumptions 1–7 of Andrews and Ploberger (1993), denoted Assumptions AP1–AP7. Theorem 1 of Andrews and Ploberger (1993) then gives the results of Proposition 1.

Under Assumption 1, the likelihood function is given by

$$\text{(A.1)} \quad f_T(\theta,\pi) = (2\pi\sigma^2)^{-T/2}\exp\left(-\tfrac{1}{2\sigma^2}\Sigma_{t=1}^T(Y_t - \beta\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1})^2\right),$$

where $\theta = (\beta, \sigma^2)'$. Assumption AP1(a) holds, since $f_T(\theta_0, \pi)$ does not depend on π, where $\theta_0 = (0, \sigma^2)'$. Assumption AP1(b) holds because B contains a neighborhood of zero. Assumption AP1(c) holds because $f_T(\theta, \pi)$ is infinitely differentiable in θ for all $\beta \in B$, $\pi \in \Pi$, and $\sigma^2 > 0$. Assumption AP3 (i.e., $\widetilde{\theta} \xrightarrow{p} \theta_0$ under θ_0) holds by the law of large numbers, because $\widetilde{\theta} = (0, \frac{1}{T}\Sigma_{t=1}^T Y_t^2)'$ and $\{Y_t : t \geq 1\}$ are iid $N(0,\sigma^2)$ under θ_0. Assumption AP4 holds by the definition of Π. Assumption AP6 holds by the definition of $J(\cdot)$. Assumption AP7 holds with $A_\pi(A_\pi'\mathcal{I}_\pi A_\pi)^{-1/2} = \mathcal{I}_{1\pi}^{-1/2} = \sqrt{1-\pi^2}$, since $\mathcal{I}_{2\pi} = [\mathcal{I}(\theta_0,\pi)]_{12} = 0$ as shown below.

It remains to show Assumptions AP1(d), (e), and (f), AP2, and AP5. Take $B_T = \sqrt{T}$. Then (using notation from Andrews and Ploberger (1993)),

$$-B_T^{-1}D\ell_T(\theta,\pi)B_T^{-1}$$

$$\text{(A.2)} \quad = \begin{pmatrix} \frac{1}{\sigma^2}\frac{1}{T}\Sigma_{t=1}^T(\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1})^2 & \frac{1}{\sigma^4}\frac{1}{T}\Sigma_{t=1}^T(Y_t - \beta\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}) \times \Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1} \\ \frac{1}{\sigma^4}\frac{1}{T}\Sigma_{t=1}^T(Y_t - \beta\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}) \times \Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1} & \frac{1}{\sigma^6}\frac{1}{T}\Sigma_{t=1}^T Y_t^2 - \frac{1}{2\sigma^4} \end{pmatrix}.$$

Let

$$\text{(A.3)} \quad \mathcal{I}(\theta,\pi) = \lim_{T\to\infty} E[-B_T^{-1}D^2\ell_T(\theta,\pi)B_T^{-1}] = \begin{pmatrix} \frac{1}{1-\pi^2} & \frac{-\beta}{\sigma^2(1-\pi^2)} \\ \frac{-\beta}{\sigma^2(1-\pi^2)} & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Assumption AP1(d) requires $-B_T^{-1}D^2\ell_T(\theta,\pi)B_T^{-1} \xrightarrow{p} \mathcal{I}(\theta,\pi)$ uniformly over $\pi \in \Pi$ and over θ in some neighborhood Θ_0 of θ_0. By Theorem 1 of Andrews (1992), uniform convergence is implied by pointwise convergence, stochastic equicontinuity (i.e., $\forall\varepsilon > 0\ \exists\delta > 0$ such that

$\overline{\lim}_{T\to\infty} P(\sup_{\|(\theta_1,\pi_1)-(\theta_2,\pi_2)\|<\delta} |G_T(\theta_1,\pi_1) - G_T(\theta_2,\pi_2)| > \varepsilon) < \varepsilon$, where $G_T(\theta,\pi) = -B_T^{-1} D^2 \ell_T(\theta,\pi) B_T^{-1} - \mathcal{I}(\theta,\pi))$, and total boundedness of $\Theta_0 \times \Pi$. Pointwise convergence in probability for each $(\theta,\pi) \in \Theta_0 \times \Pi$ is implied by pointwise convergence in mean square, which is straightforward, but tedious, to establish. For brevity, the proof is omitted. Since the norming is by T^{-1} (rather than $T^{-1/2}$), it is also straightforward, but tedious, to establish stochastic equicontinuity by applying Markov's inequality and standard manipulations. Again, the proof is omitted for brevity. Total boundedness of $\Theta_0 \times \Pi$ holds by definition of Θ_0 and Π. This completes the proof of Assumption AP1(d).

It is apparent that $\mathcal{I}(\theta,\pi)$ is uniformly continuous in (θ,π) over $\Theta_0 \times \Pi$. Hence, Assumption AP3 holds. In addition,

$$\text{(A.4)} \quad \inf_{\pi\in\Pi} \lambda_{\min}(\mathcal{I}(\theta_0,\pi)) = \inf_{\pi\in\Pi} \min\left\{\tfrac{1}{1-\pi^2}, \tfrac{1}{2\sigma^4}\right\} > 0,$$

since $|\pi|$ is bounded below one. In consequence, Assumption AP1(f) holds.

Assumption AP2 requires $\sup_{\pi\in\Pi} \|\widehat{\theta}(\pi) - \theta_0\| \xrightarrow{p} 0$ under θ_0, where

$$\text{(A.5)} \quad \widehat{\theta}(\pi) = \begin{pmatrix} \widehat{\beta}(\pi) \\ \widehat{\sigma}^2(\pi) \end{pmatrix} = \begin{pmatrix} \left[\frac{1}{T}\Sigma_{t=1}^T \left(\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}\right)^2\right]^{-1} \frac{1}{T}\Sigma_{t=1}^T Y_t \Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1} \\ \frac{1}{T}\Sigma_{t=1}^T \left(Y_t - \widehat{\beta}(\pi)\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}\right)^2 \end{pmatrix}.$$

Again, by Theorem 1 of Andrews (1992), it suffices to show pointwise convergence for each $\pi \in \Pi$ and stochastic equicontinuity of $\widehat{\theta}(\pi) - \theta_0$. These results are straightforward, but tedious, to establish given that $\{Y_t : t \geq 1\}$ are iid $N(0,\sigma^2)$. For brevity, the proofs are omitted.

Assumption AP5 requires that $B_T^{-1} D\ell_T(\theta_0,\cdot)$ converges weakly to a Gaussian process $G(\theta_0,\cdot)$ (as processes indexed by $\pi \in \Pi$), where $G(\theta_0,\pi)$ satisfies $EG(\theta_0,\pi)G(\theta_0,\pi)' = \mathcal{I}(\theta_0,\pi)$ and $G(\theta_0,\pi)$ has continuous sample paths (as functions of π) almost surely. Since $B_T^{-1} D\ell_T(\theta_0,\pi) = \frac{1}{\sigma^2\sqrt{T}}\Sigma_{t=1}^T Y_t \Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}$, the desired result is a special case of Theorem 1(a) with $G(\theta_0,\pi) = G(\pi)/\sigma^2$. The proof of Theorem 1(a) is given below. □

PROOF OF PROPOSITION 2: Proposition 2 follows from Theorem 2 of Andrews and Ploberger (1994) provided Assumptions 1–3 and 5 of the latter paper can be verified. These assumptions are the same as Assumptions 1–3 and 5 in Andrews and Ploberger (1993) (except for minor and insignificant differences), which have just been verified. □

PROOF OF THEOREM 1: First we establish part (a). Define

(A.6) $\nu_T(\pi) = \frac{1}{\sqrt{T}}\Sigma_{t=2}^{T} Y_t \Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}$ and $\nu(\pi) = \sigma^2 \Sigma_{i=0}^{\infty}\pi^i Z_i$.

It suffices to show that $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ and $\widehat{\sigma}^2 = \frac{1}{T-1}\Sigma_{t=1}^{T}(Y_t - \overline{Y}_T)^2 \xrightarrow{p} \sigma^2$, because the continuous mapping theorem (e.g., see Pollard (1984, p. 70)) then yields part(a). By Thm. 10.2 of Pollard (1990), $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ if (i) π is totally bounded, (ii) the finite dimensional distributions of $\nu_T(\cdot)$ converge to those of $\nu(\cdot)$, and (iii) $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous (i.e., $\forall \varepsilon > 0$ $\exists \delta > 0$ such that $\overline{\lim}_{T\to\infty} P(\sup_{|\pi_1-\pi_2|\leq\delta} |\nu_T(\pi_1) - \nu_T(\pi_2)| > \varepsilon) < \varepsilon$)). Condition (i) holds since π is a subset of $[-1, 1]$.

To establish condition (ii), we use the Cramer–Wold device and a martingale difference sequence (MDS) central limit theorem (CLT). In particular, we use the MDS CLT of Thm. 3.1 of Hall and Hyde (1980). For simplicity, we just show that $\nu_T(\pi) \xrightarrow{d} N(0, \sigma^4/(1-\pi^2))$ and $\lim_{T\to\infty} \text{Cov}(\nu_T(\pi_1), \nu_T(\pi_2)) = \text{Cov}(\nu(\pi_1), \nu(\pi_2))$. The CLT result for arbitrary linear combinations of $(\nu_T(\pi_1), ..., \nu_T(\pi_p))$ is established analogously. Let $W_{Tt} = Y_t\Sigma_{i=0}^{t-2}Y_{t-i-1}/\sqrt{T}$. $E(W_{Tt}|\mathcal{F}_{t-1}) = 0$ by Assumption 2. Hall and Hyde's condition (3.21) holds because $\mathcal{F}_t$ does not depend on T. The other two conditions of their Corollary 3.1 are:

(A.7) $\forall \varepsilon > 0,\ \Sigma_{t=1}^{T} E(W_{Tt}^2 1(|W_{Tt}| > \varepsilon)|\mathcal{F}_{t-1}) \xrightarrow{p} 0$ and

(A.8) $\Sigma_{t=1}^{T} E(W_{Tt}^2|\mathcal{F}_{t-1}) \xrightarrow{p} \sigma^4/(1-\pi^2)$.

To establish (A.7), let LHS_T denote the left-hand side of (A.7). Then,

$$P(LHS_T > \eta) \leq E(LHS_T)/\eta \leq \Sigma_{t=1}^{T} E|W_{Tt}|^{2+\delta}/(\varepsilon^\delta \eta)$$

$$(A.9) \quad = T^{-(1+\delta/2)}\Sigma_{t=1}^{T} E|Y_t\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}|^{2+\delta}/(\varepsilon^\delta\eta)$$

$$\leq T^{-\delta/2}\sup_{t\geq 1} E|Y_t|^{4+2\delta}(1/(1-|\pi|))/(\varepsilon^\delta\eta) \to 0.$$

The condition in (A.8) follows from

$$
\begin{aligned}
& E(W_{Tt}^2|\mathcal{F}_{t-1}) = \tfrac{\sigma^2}{T}\left(\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1}\right)^2, \\
& \frac{\sigma^4}{T}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi^{2i} \to \sigma^4/(1-\pi^2), \text{ and} \\
& E\left(\tfrac{\sigma^2}{T}\Sigma_{t=2}^{T}\left[(\Sigma_{i=0}^{t-2}\pi^i Y_{t-i-1})^2 - \sigma^2\Sigma_{i=0}^{t-2}\pi^{2i}\right]\right)^2 \\
= \; & E\left(\tfrac{\sigma^2}{T}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\Sigma_{j=0}^{t-2}\pi^i\pi^j(Y_{t-i-1}Y_{t-j-1} - \sigma^2 1(i=j))\right)^2 \\
\text{(A.10)} \quad = \; & \tfrac{\sigma^4}{T^2}\Sigma_{s=1}^{T}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\Sigma_{j=0}^{t-2}\Sigma_{k=0}^{s-2}\Sigma_{\ell=0}^{s-2}\pi^i\pi^j\pi^k\pi^\ell E(Y_{t-i-1}Y_{t-j-1} - \sigma^2 1(i=j)) \\
& \times(Y_{s-k-1}Y_{s-\ell-1} - \sigma^2 1(k=\ell)) \\
\leq \; & \tfrac{\sigma^4}{T^2}\Sigma_{t=2}^{T}\Sigma_{i=1}^{\infty}\pi^{2i}\Sigma_{k=0}^{\infty}\pi^{2k}\sup_{t\geq 1} E(Y_t^2 - \sigma^2)^2 \\
\to \; & 0,
\end{aligned}
$$

where the last equality holds because the summands on the left-hand side of the equality are zero unless $t-i-1 = t-j-1 = s-k-1 = s-\ell-1$, i.e., unless $i = j$, $k = \ell$, and $t-i = s-k$. Thus, Hall and Hyde's MDS CLT applies.

Next, we note that

$$\text{(A.11)} \quad \mathrm{Cov}(\nu(\pi_1),\, \nu(\pi_2)) = \sigma^4\Sigma_{i=1}^{\infty}\pi_1^i\pi_2^i.$$

In addition, we have

$$
\begin{aligned}
\lim_{T\to\infty}\mathrm{Cov}(\nu_T(\pi_1),\, \nu_T(\pi_2)) &= \lim_{T\to\infty}\tfrac{1}{T}\Sigma_{s=2}^{T}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\Sigma_{j=0}^{s-2}\pi_1^i\pi_2^j EY_sY_tY_{s-j-1}Y_{t-i-1} \\
\text{(A.12)} \qquad &= \lim_{T\to\infty}\tfrac{1}{T}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi_1^i\pi_2^i EY_t^2Y_{t-i-1}^2 \\
&= \mathrm{Cov}(\nu(\pi_1),\, \nu(\pi_2)),
\end{aligned}
$$

where the second equality holds because $EY_sY_tY_{s-j-1}Y_{t-i-1} = 0$ unless $s = t$ and $i = j$. This completes the proof of (ii).

We now establish the stochastic equicontinuity condition (iii). Let $\nu_T^*(\pi) = \frac{1}{\sqrt{T}}\Sigma_{t=2}^{T}Y_t \times \Sigma_{i=0}^{\infty}\pi^i Y_{t-i-1}$. We have

$$
\begin{aligned}
\text{(A.13)} \quad & E\sup_{\pi\in\Pi}|\nu_T(\pi) - \nu_T^*(\pi)| = E\sup_{\pi\in\Pi}\left|\tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=t-1}^{\infty}\pi^i Y_tY_{t-i-1}\right| \\
\leq \; & \tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=t-1}^{\infty}\pi_*^i\sup_{s\geq 1} EY_s^2 \to 0,
\end{aligned}
$$

where $\pi_* = \sup\{|\pi| : \pi \in \Pi\} < 1$. In consequence, it suffices to establish stochastic equicontinuity of $\{\nu_T^*(\cdot) : T \geq 1\}$. Let $a_i = \pi_*^{2i}$. Given $\varepsilon > 0$, we have

$$\begin{aligned}
& P(\sup_{|\pi_1-\pi_2|<\delta} |\nu_T^*(\pi_1) - \nu_T^*(\pi_2)| > \varepsilon) \\
&= P(\sup_{|\pi_1-\pi_2|<\delta} |\Sigma_{i=0}^{\infty}(\pi_1^i - \pi_2^i)\tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T} Y_t Y_{t-i-1}| > \varepsilon) \\
\text{(A.14)} \quad &\leq \varepsilon^{-2} E \sup_{|\pi_1-\pi_2|<\delta} \left(\Sigma_{i=0}^{\infty}\frac{(\pi_1^i - \pi_2^i)}{\sqrt{a_i}} \cdot \sqrt{a_i}\frac{1}{\sqrt{T}}\Sigma_{t=2}^{T} Y_t Y_{t-i-1}\right)^2 \\
&\leq \varepsilon^{-2} \sup_{|\pi_1-\pi_2|<\delta} \Sigma_{i=0}^{\infty}\frac{(\pi_1^i - \pi_2^i)^2}{a_i} \cdot \sum_{j=0}^{\infty} a_j E\left(\frac{1}{\sqrt{T}}\Sigma_{t=2}^{T} Y_t Y_{t-j-1}\right)^2 \\
&= \frac{\sigma^4}{\varepsilon^2(1-\pi_*^2)} \sup_{|\pi_1-\pi_2|<\delta} \Sigma_{i=0}^{\infty}\frac{(\pi_1^i - \pi_2^i)^2}{a_i},
\end{aligned}$$

where the first inequality uses Chebyshev's inequality and the second uses the Cauchy–Schwarz inequality. We note that an alternative method of establishing condition (iii) is to use Lemma A.1 of Bierens and Ploberger (1994).

Stochastic equicontinuity now follows if we can show that given $\xi > 0$, $\exists\delta > 0$ such that $|\pi_1 - \pi_2| < \delta$ implies $\Sigma_{i=0}^{\infty}(\pi_1^i - \pi_2^i)^2/a_i < \xi$. Given $\xi > 0$, we can take J so large that $\Sigma_{i=J+1}^{\infty}(\pi_1^i - \pi_2^i)^2/a_i < \xi/2$ (since $|\pi_1|$ and $|\pi_2|$ are bounded away from one). Next, we have $\Sigma_{i=1}^{J}(\pi_1^i - \pi_2^i)^2/a_i = \Sigma_{i=1}^{J}((\pi_1/\pi_*)^i - (\pi_2/\pi_*)^i)^2 \leq J(\pi_1/\pi_* - \pi_2/\pi_*)^2 < \xi/2$, where the last inequality holds provided $|\pi_1 - \pi_2| < \delta$ and $\delta < (\xi\pi_*^2/(2J))^{1/2}$. This completes the proof of stochastic equicontinuity.

Next, we show $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$. We have $\overline{Y}_T^2 \xrightarrow{p} 0$, since $E\overline{Y}_T^2 = \sigma^2/T \to 0$. In addition,

$$\text{(A.15)} \quad E\left(\tfrac{1}{T}\Sigma_{t=1}^{T}(Y_t^2 - \sigma^2)\right)^2 = \tfrac{1}{T^2}\Sigma_{s=1}^{T}\Sigma_{t=1}^{T} E(Y_s - \sigma^2)(Y_t - \sigma^2) = \tfrac{1}{T^2}\Sigma_{t=1}^{T} E(Y_t^2 - \sigma^2) \to 0,$$

where the second equality holds because $E(Y_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s. implies that $E((Y_s^2 - \sigma^2) \times (Y_t^2 - \sigma^2)|\mathcal{F}_{t-1}) = 0$ $\forall s < t$. These results combine to yield $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, which concludes the proof of part (a).

Parts (b)–(e) follow from part (a) and the continuous mapping theorem. □

PROOF OF THEOREM 2: First we prove part (a). Define

$$\text{(A.16)}\quad \begin{aligned} &A_T(\pi) = B_T(\pi) - C_T(\pi),\; B_T(\pi) = \tfrac{1}{T}\Sigma_{t=2}^T \Sigma_{i=0}^\infty \pi^i Y_t Y_{t-i-1},\\ &C_T(\pi) = \tfrac{1}{T}\Sigma_{t=2}^T \Sigma_{i=t-1}^\infty \pi^i Y_t Y_{t-i-1}, \quad \text{and}\quad B(\pi) = \Sigma_{i=0}^\infty \pi^i \gamma_{i+1}.\end{aligned}$$

The left-hand side (lhs) of part (a) equals

$$\text{(A.17)}\quad \sup_{\pi\in\Pi} |A_T^2(\pi)(1-\pi^2)/\widehat{\sigma}_Y^4 - B^2(\pi)(1-\pi^2)/\sigma^4|.$$

Below we show that

$$\text{(A.18)}\quad \begin{aligned} &\text{(i)}\quad E\sup_{\pi\in\Pi} C_T^2(\pi) \to 0,\\ &\text{(ii)}\quad E(B_T(\pi) - EB_T(\pi))^2 \to 0 \;\; \forall \pi \in \Pi\\ &\text{(iii)}\quad \{B_T(\cdot) - EB_T(\cdot) \,:\, T \geq 1\} \text{ is stochastically equicontinuous on } \Pi,\\ &\text{(iv)}\quad \sup_{\pi\in\Pi} |EB_T(\pi) - B(\pi)| \to 0, \text{ and}\\ &\text{(v)}\quad \widehat{\sigma}_Y^2 \xrightarrow{p} \sigma^2.\end{aligned}$$

Parts (ii) and (iii) combine to yield $\sup_{\pi\in\Pi} |B_T(\pi) - EB_T(\pi)| \xrightarrow{p} 0$ by Theorem 1 of Andrews (1992). This result and parts (i), (iv), and (v) combine to establish part (a) of the Theorem.

To show (i), we write the lhs of (i) as

$$\text{(A.19)}\quad \begin{aligned} &E\sup_{\pi\in\Pi} \tfrac{1}{T^2}\Sigma_{s=2}^T \Sigma_{t=2}^T \Sigma_{i=t-1}^\infty \Sigma_{j=s-1}^\infty \pi^i \pi^j Y_t Y_{t-i-1} Y_s Y_{s-j-1}\\ &\leq \tfrac{1}{T^2}\Sigma_{s=2}^T \Sigma_{t=2}^T \Sigma_{i=t-1}^\infty \pi_*^i \Sigma_{j=s-1}^\infty \pi_*^j \sup_{m\geq 1} EY_m^4\\ &\leq \tfrac{1}{T^2}\left(\Sigma_{t=2}^\infty \pi_*^{t-1}\right)^2 \left(\tfrac{1}{1-\pi_*}\right)^2 \sup_{m\geq 1} EY_m^4 \to 0,\end{aligned}$$

where $\pi_* = \sup\{|\pi| \,:\, \pi \in \Pi\}$.

To show (ii), we write the lhs of (ii) as

$$\text{(A.20)}\quad \begin{aligned} &\tfrac{1}{T^2}\Sigma_{s=2}^T \Sigma_{t=2}^T \Sigma_{i=0}^\infty \Sigma_{j=0}^\infty \pi^i \pi^j E(Y_t Y_{t-i-1} - EY_t Y_{t-i-1})(Y_s Y_{s-j-1} - EY_s Y_{s-j-1})\\ &\leq \tfrac{2}{T^2}\sum\sum_{2\leq s\leq t\leq T} \Sigma_{i=0}^\infty \pi_*^i \Sigma_{j=0}^\infty \pi_*^j 8 \sup_{m\geq 1} E|Y_m|^{4+\delta} \alpha(\max\{t-i-1-s,\, 0\})^{\delta/(4+\delta)}\\ &\leq \tfrac{16}{1-\pi_*} \sup_{m\geq 1} E|Y_m|^{4+\delta} \left[\Sigma_{i=0}^I \pi_*^i \tfrac{1}{T^2}\sum\sum_{2\leq s\leq t\leq T} \alpha(\max\{t-i-1-s,\, 0\})^{\delta/(4+\delta)} + \Sigma_{i=I+1}^\infty \pi_*^i\right],\end{aligned}$$

where the first inequality holds by a standard strong mixing inequality, see Hall and Hyde (1980, Cor. A.2, p. 278), and the second inequality holds for all positive integers I since $\alpha(m) \leq 1$.

The second summand on the rhs of (A.20) can be made arbitrarily small by taking I sufficiently large. Given I, the first summand on the rhs of (A.20) can be made arbitrarily small by taking T sufficiently large, since $1/T$ times the double sum over s and t is bounded over $T \geq 1$ for each fixed i. This establishes (ii).

Part (iii) is established in exactly the same way as the stochastic equicontinuity condition (iii) of the proof of Theorem 1. Part (iv) holds, because $EB_T(\pi) = B(\pi)(T-1)/T$.

Part (v) holds using the strong mixing assumption via standard arguments.

Parts (b) and (c) of Theorem 2 follow from part (a) because the function $h(\pi) = \Sigma_{i=0}^{\infty}\pi^i\gamma_i$ has only a finite number of zeros if $\gamma_i \neq 0$ for some i. This follows because the function $h(\pi)$ for π complex and $|\pi| < 1$ is analytic and analytical functions are either identically zero or have finite numbers of zeros, e.g., see Ahlfors (1966, p. 127). □

PROOF OF THEOREM 3: Let $\widehat{LM}_T(\pi)$, $\widehat{\nu}_T(\pi)$, and $\widehat{\widehat{\sigma}}^2$ denote $LM_T(\pi)$, $\nu_T(\pi)$, and $\widehat{\sigma}^2$, respectively, when the latter are constructed using the residuals $\{\widehat{Y}_t \; : \; t \leq T\}$. It suffices to show that

$$\text{(A.21)} \quad \sup_{\pi\in\Pi} |\widehat{LM}_T(\pi) - LM_T(\pi)| \xrightarrow{p} 0$$

under H_0 and Assumptions 3 or 4. The latter follows from

$$\text{(A.22)} \quad \sup_{\pi\in\Pi} |\widehat{\nu}_T(\pi) - \nu_T(\pi)| \xrightarrow{p} 0 \quad \text{and} \quad \widehat{\widehat{\sigma}}^2 - \widehat{\sigma}^2 \xrightarrow{p} 0.$$

Let $g_t = g(X_t,\, \lambda_0)$ and $\widehat{g}_t = g(X_t,\, \widehat{\lambda})$. Then,

$$\text{(A.23)} \quad \begin{aligned} & \widehat{\nu}_T(\pi) - \nu_T(\pi) \\ = \; & \tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-1}\pi^i[(g_t - \widehat{g}_t)Y_{t-i-1} + Y_t(g_{t-i-1} - \widehat{g}_{t-i-1}) + (g_t - \widehat{g}_t)(g_{t-i-1} - \widehat{g}_{t-i-1})]. \end{aligned}$$

As above, uniform convergence in probability of $\widehat{\nu}_T(\pi) - \nu_T(\pi)$ to zero is implied by pointwise convergence for all $\pi \in \Pi$ and stochastic equicontinuity, using Theorem 1 of Andrews (1992).

Pointwise convergence is obtained under Assumption 3, by taking a two-term Taylor expansion of $g(X_t,\, \lambda_0)$ about $\widehat{\lambda}$:

$$
\text{(A.24)}\quad
\begin{aligned}
&\left|\tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi^i(g_t-\widehat{g}_t)Y_{t-i-1}\right|\\
&= \left|\tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi^i\Big[(\lambda_0-\widehat{\lambda})'\tfrac{\partial}{\partial\lambda}g(X_t,\lambda_0)Y_{t-i-1}+\tfrac{1}{2}(\widehat{\lambda}-\lambda_0)'\tfrac{\partial^2}{\partial\lambda\partial\lambda'}g(X_t,\lambda_t^*)(\widehat{\lambda}-\lambda_0)Y_{t-i-1}\Big]\right|\\
&\le \|T^{1/4}(\widehat{\lambda}-\lambda_0)\|\cdot\left\|T^{-3/4}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi^i\tfrac{\partial}{\partial\lambda}g(X_t,\lambda_0)Y_{t-i-1}\right\|\\
&\quad+\tfrac{1}{2}\|T^{1/4}(\widehat{\lambda}-\lambda_0)\|^2\tfrac{1}{T}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi_*^i\sup_{\|\lambda-\lambda_0\|<\varepsilon}\left\|\tfrac{\partial^2}{\partial\lambda\partial\lambda'}g(X_t,\lambda)\right\|\cdot|Y_{t-i-1}|,
\end{aligned}
$$

where λ_t^* is a rv that lies between λ_0 and $\widehat{\lambda}$, $\pi_* = \sup\{|\pi| \;:\; \pi \in \Pi\}$, and the inequality holds with probability that goes to one. The first summand on the rhs of (A.24) is $o_p(1)$ because $T^{1/4}(\widehat{\lambda}-\lambda_0) = o_p(1)$ and the term involving $\Sigma_{t=2}^{T}$ has mean zero and variance that goes to zero as $T \to \infty$. The second summand is $o_p(1)$ because $T^{1/4}(\widehat{\lambda}-\lambda_0) = o_p(1)$ and the term involving $\Sigma_{t=2}^{T}$ has mean bounded away from infinity over $T \geq 1$ and, hence, is $O_p(1)$ by Markov's inequality. The proof that the second and third summands of (A.23) are $o_p(1)$ $\forall \pi \in \Pi$ is analogous.

Stochastic equicontinuity of $\{\widehat{\nu}_T(\cdot) - \nu_T(\cdot) \;:\; T \geq 1\}$ under Assumption 3 is established by combining the stochastic equicontinuity argument given in the proof of Theorem 1 with the argument given immediately above. For brevity, the details are omitted.

Pointwise convergence of $\widehat{\nu}_T(\pi) - \nu_T(\pi)$ to zero is obtained under Assumption 4 as follows: The first summand of (A.23) is

$$
\text{(A.25)}\quad \left|\tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi^i(g_t-\widehat{g}_t)Y_{t-i-1}\right| = \left|\left(\tfrac{1}{\sqrt{T}}\Sigma_{t=2}^{T}\Sigma_{i=0}^{t-2}\pi^iY_{t-i-1}X_t\Delta_T^{-1}\right)\Delta_T(\lambda_0-\widehat{\lambda})\right|.
$$

The rhs of (A.25) is $o_p(1)$ because $\Delta_T(\widehat{\lambda}-\lambda_0) = O_p(1)$ and the expression in parentheses has mean zero and variance that converges to zero as $T \to \infty$ using Assumption 4(iii). The second and third summands of (A.23) are $o_p(1)$ by analogous arguments. Stochastic equicontinuity is established under Assumption 4 in the same manner as under Assumption 3.

The proof that $\widehat{\widehat{\sigma}}^2 - \widehat{\sigma}^2 \xrightarrow{p} 0$ under Assumption 3 or 4 is similar to the proofs given above. For brevity, it is omitted. □

FOOTNOTE

*The authors thank B. M. Pötscher for providing several references. Andrews gratefully acknowledges the research support of the National Science Foundation through grants SES–9121914 and INT–9215258. Ploberger gratefully acknowledges research support from Fonds zur Foerderung der wissenschaftlichen Forsehung under Project Number P2279–SOZ and Ausseninstitut der TU–Wien.

TABLE 1

	Asymptotic Critical Values		
Test Statistic	10%	5%	1%
Sup-LM, $\pi \in [-.8, .8]$	4.62	6.01	9.09
Exp-LM_0, $\pi \in [-.8, .8]$	2.41	3.34	5.62
Exp-LM_∞, $\pi \in [-.8, .8]$	1.42	1.97	3.36

TABLE 2

AR(1)		MA(1)		AR(6)±		AR(12)±		Model	AR(1)		MA(1)		AR(6)		AR(12)	
ρ=-.2	ρ=-.3	ϕ=-.2	ϕ =-.3	ρ=.3	ρ=.4	ρ=.3	ρ=.4	Test	ρ=.2	ρ=.3	ϕ=.2	ϕ=.3	ρ=.15	ρ=.2	ρ=.1	ρ=.2
.51	.84	.47	.79	.57	.82	.47	.71	Exp-LM_0	.43	.78	.39	.72	.60	.86	.45	.83
.51	.83	.46	.78	.59	.85	.49	.75	Exp-LM_1	.42	.77	.38	.72	.64	.88	.49	.85
.49	.82	.45	.77	.60	.87	.50	.77	Exp-LM_∞	.41	.76	.37	.71	.65	.89	.52	.86
.44	.79	.39	.72	.57	.87	.47	.77	Sup-LM	.36	.72	.32	.65	.68	.90	.57	.88
.48	.82	.45	.77	.56	.77	.45	.66	DW2	.50	.83	.47	.78	.51	.79	.35	.76
.00	.00	.00	.00	.67	.86	.58	.77	DW+	.61	.89	.59	.86	.60	.84	.42	.80
.62	.90	.59	.86	.00	.00	.00	.00	DW-	.00	.00	.00	.00	.00	.00	.00	.00
.29	.61	.26	.51	.40	.63	.34	.52	BP6	.25	.55	.22	.47	.57	.86	.46	.84
.23	.49	.20	.39	.31	.47	.26	.38	BP12	.20	.45	.18	.37	.50	.81	.43	.82
.00	.00	.00	.00	.64	.84	.54	.74	POI(.5,0)	.58	.88	.55	.84	.57	.83	.40	.79
.00	.00	.00	.00	.94	.99	.88	.98	POI(0,.5)	.49	.79	.53	.86	.34	.61	.24	.62
.00	.00	.00	.00	.94	1.00	.89	.98	POI(.5,.5)	.51	.81	.56	.87	.36	.62	.26	.63

TABLE 3

	ρ	.2	.4	.6	.8	−.2	−.4	−.6	−.8	.2	.4	.6	−.2	−.4	−.6	−.8
	ϕ	.1	−.1	−.3	−.6	.6	.8	1.0	1.6	−.5	−.7	−1.2	−.1	.1	.4	.6
Exp–LM_0		.76	.80	.86	.69	.89	.82	.52	.69	.75	.68	.40	.82	.86	.66	.79
Exp–LM_1		.75	.80	.86	.71	.90	.84	.55	.71	.75	.69	.41	.81	.86	.66	.81
Exp–LM_∞		.74	.79	.86	.71	.91	.86	.57	.72	.75	.69	.40	.81	.85	.66	.81
Sup–LM		.69	.75	.84	.71	.88	.84	.54	.73	.68	.63	.34	.76	.82	.64	.82
DW2		.82	.84	.87	.65	.90	.80	.50	.57	.71	.60	.33	.81	.84	.57	.68
DW+		.88	.90	.92	.73	.95	.87	.62	.00	.00	.00	.00	.00	.00	.00	.00
DW−		.00	.00	.00	.00	.00	.00	.00	.67	.82	.73	.46	.89	.91	.69	.77
BP6		.52	.60	.71	.58	.69	.61	.39	.59	.47	.43	.25	.56	.65	.46	.70
BP12		.42	.49	.62	.51	.54	.46	.29	.52	.35	.32	.20	.46	.55	.38	.63
POI(.5,0)		.87	.88	.91	.71	.94	.86	.59	.00	.00	.00	.00	.00	.00	.00	.00
POI(0,.5)		.81	.77	.76	.47	.99	.99	.92	.00	.00	.00	.00	.00	.00	.00	.00
POI(.5,.5)		.83	.79	.77	.50	.99	.99	.93	.00	.00	.00	.00	.00	.00	.00	.00

REFERENCES

Ahlfors, L. V. (1966): *Complex Analysis,* 2nd edition. New York: McGraw–Hill.

Andrews D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, D. W. K. and W. Ploberger (1993): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter Is Present Only Under the Alternative," Cowles Foundation Discussion Paper No. 1058, Yale University.

Andrews, D. W. K. and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, forthcoming.

Ansley, C. F. (1979): "An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process," *Biometrika*, 66, 59–65.

Bierens, H. J. and W. Ploberger (1994): "Asymptotic Theory of Integrated Conditional Moment Tests," unpublished working paper, Department of Economics, Southern Methodist University, Dallas, TX.

Box, G. E. P. and D. A. Pierce (1970): "Distribution of Residual Autocorrelation in Autoregressive-integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 64, 247–254.

Davies, R. B. (1987): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 74, 33–43.

Durbin, J. and G. S. Watson (1950): "Testing for Serial Correlation in Least Squares Regression, I," *Biometrika*, 37, 409–428.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Applications.* New York: Academic Press.

Hannan, E. J. (1982): "Testing for Autocorrelation and Akaike's Criterion," in *Essays in Statistical Science*, ed. by J. Gani and E. J. Hannan, *Journal of Applied Probability*, Special Volume 19A.

Hansen, B. E. (1991): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," Working Paper No. 296, Rochester Center for Economic Research, University of Rochester.

Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 12, 141–153.

Pollard, D. (1984): *Convergence of Stochastic Processes.* New York: Springer–Verlag.

Pollard, D. (1990): *Empirical Processes: Theory and Applications. CBMS Conference Series in Probability and Statistics*, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Poskitt, D. S. and A. R. Tremayne (1980): "Testing the Specification of a Fitted Autoregressive Moving Average Model," *Biometrika*, 67, 359–363.

Pötscher, B. M. (1983): "Order Estimation in ARMA-Models by Lagrangian Multiplier Tests," *Annals of Statistics*, 11, 872–885.

Pötscher, B. M. (1985): "The Behaviour of the Lagrangian Multiplier Test in Testing the Orders of an ARMA-Model," *Metrika*, 32, 129–150.

Pötscher, B. M. and S. Srinivasan (1994): "A Comparison of Order Estimation Procedures for ARMA Models," *Statistica Sinica*, 4, 29–50.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1986): *Numerical Recipes: The Art of Scientific Computing.* Cambridge: Cambridge University Press.

Rahman, S. and M. King (1992): "Testing for ARMA(1,1) Disturbances in the Linear Regression Model," Working Paper No. 1/92, Department of Econometrics, Monash University, Australia.

Veres, S. (1987): "Asymptotic Distributions of Likelihood Ratios for Overparameterized ARMA Processes," *Journal of Time Series Analysis*, 8, 345–357.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426–482.