Bernoulli **4**(4), 1998, 497–518

Approximate distributions for the various serial correlograms

RONALD W. BUTLER[1]* and MARC S. PAOLELLA[2]

[1]*Department of Statistics, Colorado State University, Fort Collins, CO 80525, USA. E-mail: walrus@stat.colostate.edu*
[2]*Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany*

Saddlepoint methods are used to approximate the joint density of the serial correlogram up to lag m. Jacobian transformations also lead to approximations for the related partial correlogram and inverse correlogram. The approximations consider non-circularly and circularly defined models in both the null and the non-null settings. The distribution theory encompasses the standard non-circularly defined correlogram computed from least-squares residuals removing arbitrary fixed regressors. Connections of the general theory to the approximations given by Daniels and by Durbin in the circular setting are indicated. The double-saddlepoint density and distribution approximations are given for the conditional distribution of the non-circular lag m serial correlation given the previous lags from order 1 to $m-1$. This allows for the computation of p values in conditional inference when testing that the model is AR($m-1$) versus AR(m). Numerical comparisons with the tests of Daniels and of Durbin suggest that their tests based on circularity assumptions are inadequate for short non-circular series but are in close agreement with the non-circular tests for moderately long series.

Keywords: saddlepoint approximation; serial correlation; serial correlogram

1. Introduction

We consider the joint saddlepoint density approximation for random vector $r = (r_1, \ldots, r_m)^{\mathrm{T}}$ having the form

$$r_k = \frac{\epsilon^{\mathrm{T}} A_k \epsilon}{\epsilon^{\mathrm{T}} \epsilon} \qquad (k = 1, \ldots, m), \tag{1}$$

where ϵ is an n-vector whose components have a multivariate normal density with mean 0 and covariance $\Omega^{-1} > 0$, $N_n(0, \Omega^{-1})$. The $(n \times n)$ matrices $\{A_k\}$ are symmetric, with non-zero rank, and are assumed to result in a full rank distribution for r; sufficient conditions for this are given in the Appendix.

An example of such a distribution is the density of the correlogram up to lag m

*To whom correspondence should be addressed.

1350–7265

computed from the least-squares residuals $\{z_t\colon t = 1, \ldots, n\}$ of a linear regression. In this case the variables are the serial correlations

$$r_k = \sum_{t=k+1}^{n} z_t z_{t-k} \Big/ \sum_{t=1}^{n} z_t^2 \qquad (k = 1, \ldots, m). \tag{2}$$

The details of this example are considered in Section 5. Two other examples also discussed include the density for the first m partial serial correlations or the lag m partial correlogram, and the density of the first m inverse autocorrelations or the inverse correlogram. Each of these two sequences of statistics is not of form (1) but each density results from Jacobian transformation of the joint density of serial correlations. We discuss these in Section 9.

Daniels (1956) first considered saddlepoint approximations for distributions of the form (1) when r is the circularly defined serial correlogram up to lag m, either with or without a mean correction. Using a Jacobian transformation, he also gave the distribution of the circularly defined partial correlogram and showed the approximate independence of the components in the null setting. Durbin (1980b) extended these results for the circular setting to include correlograms computed from residuals of Fourier regression.

In the non-circular setting with $m = 1$, the first saddlepoint methods were Daniels' (1956) approximation for an intraclass lag 1 serial correlation, either with or without mean correction. McGregor (1960) gave an approximation for a different lag 1 serial correlation, $r_1 = 2(1 - d_1)$ where d_1 is the lag 1 Durbin–Watson (1950, 1951) statistic, and also based it on residuals from a polynomial regression. When r_1 is the least-squares estimate of the first-order autoregressive coefficient, four different approximations have been given by Phillips (1978), Jensen (1988), Wang (1992) and Lieberman (1994).

Approximations in the non-circular setting with $m > 1$ have been given by Durbin (1980b) who considered some specially defined serial correlations that result in simultaneous diagonalizability for the matrices $\{A_k\colon k = 1, \ldots, m\}$. Section 6 gives further discussion of this. His approximations allow these serial correlations to be based on residuals from Fourier regression.

The contributions of this paper relate to this previous work in the following way. First we extend the saddlepoint distribution theory of Daniels (1956) and Durbin (1980b) for the ordinary correlogram from the circular model setting to the more commonly used non-circular setting. This development follows the same approach as Daniels (1956) by using a multivariate version of Geary's (1944) method as given in Section 3 but retains a different leading term for the saddlepoint expansion in this more general non-circular setting. The non-circular settings already considered by Durbin (1980b) do not pertain to the ordinary correlogram as in (2). His correlograms are based on modified definitions of serial correlations in order to retain the same mathematical tractability as in the circular setting, namely that the matrices $\{A_k\colon k = 1, \ldots, m\}$ are simultaneously diagonalizable. This assumption essentially assures that the likelihood is a regular exponential family admitting sufficient statistics so that the sufficiency approach of Durbin (1980a) can be used for the approximations. This approach, however, is not extendable to the more general non-circular setting with correlogram (2).

A second contribution is the construction of a double-saddlepoint approximation for the

cumulative distribution of r_m given $r_1, \ldots, r_{m-1}$. This allows for computation of approximate p values in conditional inference when testing AR($m-1$) versus AR(m) for a non-circular series. Numerical examples with non-circular data compare these p values with those computed from the tests of Daniels (1956) and Durbin (1980b) which assume that the series is circular. The examples confirm that the circular-theory tests should not be used with short non-circular series. With moderately long non-circular series, however, p values from the circular and non-circular tests are in close agreement. Confirmation of the accuracy of the Daniels and the Durbin tests would appear difficult without the use of this new double-saddlepoint procedure since it would require the approximation of conditional probabilities for r_m given $r_1, \ldots, r_{m-1}$.

Two different saddlepoint approximations are given for the density of r in the null setting in which $\Omega = I_n$. Section 2 gives a double-saddlepoint approximation and Section 3 develops a single-saddlepoint approximation for arbitrary $\Omega > 0$ in the most general setting. The two approximations are shown to agree analytically when $\Omega = I_n$ in Section 4. Such agreement usually only occurs when the likelihood admits a cut (Booth and Butler 1990); however, a cut does not occur here.

The non-circular saddlepoint methods are applicable to correlograms as in (2) constructed from least-squares residuals regressing out arbitrary variables. This is discussed in Section 5. The methods of Durbin (1980b), by contrast, allow only for the removal of Fourier independent variables since such residuals retain the requisite mathematical tractability needed with his method. In addition, the methods presented here are applicable in the non-null setting with arbitrary $\Omega > 0$ so that power calculations are straightforward.

The remainder of the paper is organized as follows. Special models that have simpler approximations in the null setting are given in Section 6. The connections of the single-saddlepoint density in the circular setting to those of Daniels (1956) and Durbin (1980b) are given in Section 7. Densities for the partial and inverse correlograms in the general setting are noted in Section 9 and comments on asymptotics are given in Section 10. Numerical work appears in Section 11.

2. Double-saddlepoint approximation

We first present a double-saddlepoint approximation for the density of r under the assumption that $\Omega^{-1} = I_n$ which we call the null setting. With a slight abuse of notation we let r stand for both the random variable as well as an argument value for its density so that $f_r(r)$ denotes the density of r at r. The double-saddlepoint approximation cannot be used in the non-null setting because its derivation relies on the use of Basu's Lemma which applies only in this null case. The double-saddlepoint method does, however, give considerable insight into the related single-saddlepoint approximation of the next section which encompasses both the null and the non-null settings.

Let $r = N/D$ where $N = (N_1, \ldots, N_m)^{\mathrm{T}}$, $N_k = \epsilon^{\mathrm{T}} A_k \epsilon$, and $D = \epsilon^{\mathrm{T}} \epsilon$. The vector r is independent of its denominator D by Basu's Lemma; so finding the joint density of r is the same as finding the joint conditional density of N given that $D = 1$. The double-saddlepoint

density approximation of Barndorff-Nielsen and Cox (1979) uses the joint cumulant generating function of (N, D) which is easily computed as

$$K(s, t) = -\tfrac{1}{2}\log\left|(1-2t)I_n - 2\sum_{i=1}^{m} s_i A_i\right| = -\tfrac{1}{2}\log|Q|, \tag{3}$$

where $s = (s_1, \ldots, s_m)$ and $K(s, t)$ is defined over the largest neighbourhood of $0 \in \Re^{m+1}$ for which the matrix Q so defined is positive definite. The double-saddlepoint density approximation for $f_r(r) = f_{N|D=1}(r)$, the conditional density of N at r given $D = 1$, is

$$\tilde{f}(r) = (2\pi)^{-m/2}\left(\frac{\partial^2 K(0, \tilde{t}_0)/\partial\tilde{t}_0^2}{|K''(\tilde{s}, \tilde{t})|}\right)^{-1/2} \exp\left[\{\tilde{t}_0 - K(0, \tilde{t}_0)\} - \{\tilde{s}^{\mathrm{T}} r + \tilde{t} - K(\tilde{s}, \tilde{t})\}\right], \tag{4}$$

where $(\tilde{s}, \tilde{t})$ solves the set of equations

$$\begin{aligned} r &= \frac{\partial K(\tilde{s}, \tilde{t})}{\partial\tilde{s}}, \\ 1 &= \frac{\partial K(\tilde{s}, \tilde{t})}{\partial\tilde{t}}, \end{aligned} \tag{5}$$

$\tilde{t}_0$ solves the equation $\partial K(0, \tilde{t}_0)/\partial\tilde{t}_0 = 1$ and K'' denotes the $(m+1)\times(m+1)$ Hessian matrix of second derivatives. The marginal saddlepoint value is explicit as $\tilde{t}_0 = (1-n)/2$ and differentiation shows that (5) is

$$\begin{aligned} r_i &= \operatorname{tr}\tilde{Q}^{-1}A_i \qquad (i = 1, \ldots, m), \\ 1 &= \operatorname{tr}\tilde{Q}^{-1}, \end{aligned} \tag{6}$$

where $\tilde{Q}$ is Q evaluated at the saddlepoint $(\tilde{s}, \tilde{t})$. The Hessian matrix K'' in (4) consists of

$$\tilde{K}''_{ss} = \frac{\partial^2 K(\tilde{s}, \tilde{t})}{\partial\tilde{s}_i\partial\tilde{s}_j} = 2\operatorname{tr}\tilde{Q}^{-1}A_i\tilde{Q}^{-1}A_j \qquad (i, j = 1, \ldots, m),$$

$$\tilde{K}''_{st} = \frac{\partial^2 K(\tilde{s}, \tilde{t})}{\partial\tilde{s}_i\partial\tilde{t}} = 2\operatorname{tr}\tilde{Q}^{-1}A_i\tilde{Q}^{-1} \qquad (i = 1, \ldots, m),$$

$$\tilde{K}''_{tt} = \frac{\partial^2 K(\tilde{s}, \tilde{t})}{\partial\tilde{t}^2} = 2\operatorname{tr}\tilde{Q}^{-1}\tilde{Q}^{-1},$$

and $\partial^2 K(0, \tilde{t}_0)/\partial\tilde{t}_0^2 = 2/n$.

A simplification occurs in (4) by noting that, from (6),

$$\operatorname{tr}\tilde{Q}^{-1}(A_i - r_i I_n) = 0, \tag{7}$$

so that

$$\tilde{t} = 0^{\mathrm{T}}\tilde{s} + \tilde{t} = \sum_{i=1}^{m} \tilde{s}_i \operatorname{tr} \tilde{Q}^{-1}(A_i - r_i I_n) + \tilde{t} \operatorname{tr} \tilde{Q}^{-1}$$

$$= \operatorname{tr}[\tilde{Q}^{-1}\{-\tfrac{1}{2}\tilde{Q} + (\tfrac{1}{2} - r^{\mathrm{T}}\tilde{s})I_n\}].$$

This reduces to

$$\frac{1-n}{2} = \tilde{t}_0 = \tilde{t} + \tilde{s}^{\mathrm{T}} r; \tag{8}$$

so (4) can be written as

$$\tilde{f}(r) = (2\pi)^{-m/2} 2^{-1/2} n^{(n+1)/2} |K''(\tilde{s}, \tilde{t})|^{1/2} \exp\{-K(\tilde{s}, \tilde{t})\}. \tag{9}$$

3. Single-saddlepoint approximation

A single-saddlepoint approximation for the density of r is based on first obtaining a Geary-type representation for the true density in the manner of Daniels (1956, Section 2). Daniels extends the Geary (1944) representation for the density of a scalar ratio to that for a vector ratio as occurs in $r = N/D$. Essentially the density $f_r(r) = E\{D^m\} f_W(0)$ where $W = (W_1, \ldots, W_m)^{\mathrm{T}}$ is a random vector with moment generating function

$$M_W(s) = E\{D^m\}^{-1} \frac{\partial^m}{\partial t^m} M_{N,D}(s, t)\bigg|_{t=-r^{\mathrm{T}}s}, \tag{10}$$

where $M_{N,D}$ denotes the joint moment generating function of N, D. The multivariate inversion of this expression leads to $f_W(0)$ and hence

$$f_r(r) = (2\pi \mathrm{i})^{-m} \int \cdots \int \frac{\partial^m}{\partial t^m} M_{N,D}(s, t)\bigg|_{t=-r^{\mathrm{T}}s} \mathrm{d}s, \tag{11}$$

where the integration is along deformable paths of the imaginary axes of $s_1, \ldots, s_m$. The moment generating function of N, D is given as $M_{N,D}(s, t) = |\Omega|^{1/2} \exp\{K_\Omega(s, t)\}$ where

$$K_\Omega(s, t) = -\tfrac{1}{2} \log \left| \Omega - 2tI_n - 2\sum_{i=1}^{m} s_i A_i \right|.$$

The inversion in (11) requires the mth derivative of $M_{N,D}$ whose structure we now indicate. The first derivative is

$$\frac{\partial M_{N,D}(s, t)}{\partial t}\bigg|_{t=-r^{\mathrm{T}}s} = M_{N,D}(s) \operatorname{tr} P_\Omega^{-1}(s) = |\Omega|^{1/2} |P_\Omega(s)|^{1/2} \operatorname{tr} P_\Omega^{-1}(s), \tag{12}$$

where

$$P_\Omega(s) = \Omega + 2r^{\mathrm{T}} s I_n - 2\sum_{i=1}^{m} s_i A_i$$

is the matrix in expression $K_\Omega(s, t)$ evaluated with $t = -r^{\mathrm{T}}s$. The product rule of

differentiation necessarily results in $M_{N,D}$ as the lead factor in derivatives of all orders. Furthermore, the product rule and the form of (12) result in each term of the mth derivative having exponents of P_Ω^{-1} adding to m. For example, the third derivative is

$$\left.\frac{\partial^3 M_{N,D}(s,\,t)}{\partial t^3}\right|_{t=-r^{\mathrm{T}}s} = |\Omega|^{1/2}|P_\Omega(s)|^{-1/2}[\{\operatorname{tr} P_\Omega^{-1}(s)\}^3 + 6\operatorname{tr} P_\Omega^{-1}(s)\operatorname{tr} P_\Omega^{-2}(s) + 8\operatorname{tr} P_\Omega^{-3}(s)]$$

and the powers of P_Ω^{-1} in each term add to 3; thus after factoring out $\{\operatorname{tr} P_\Omega^{-1}(s)\}^3$ the latter factor is expressed in terms of q_2 and q_3 where

$$q_k = \frac{\operatorname{tr} P_\Omega^{-k}}{(\operatorname{tr} P_\Omega^{-1})^k} \qquad (k \geqslant 2).$$

Hence, it follows that the structure of the mth derivative of $M_{N,D}$ in the general non-null setting is

$$\left.\frac{\partial^m}{\partial t^m} M_{N,D}(s,\,t)\right|_{t=-r^{\mathrm{T}}s} = |\Omega|^{1/2}|P_\Omega(s)|^{-1/2}\{\operatorname{tr} P_\Omega^{-1}(s)\}^m\left(1 + \sum_j c_j p_j\right), \tag{13}$$

where the $\{p_j\}$ are products of terms involving $\{q_k\colon k = 2, \ldots, m\}$. In the null situation where $\Omega = I_n$ we drop the Ω subscript so that $P_{I_n}(s) = P(s)$, etc. For the moment we shall presume that $\{p_j\}$ in (13) are small and can be ignored. We return to discuss this in Section 10.

The dominant term in the integrand of (11) from (13) is the log-convex function $|P_\Omega(s)|^{-1/2}$ which we use to determine a saddlepoint through which the integral paths in (11) are deformed. The single saddlepoint solves

$$0 = -\tfrac{1}{2}\frac{\partial}{\partial \hat{s}_i}\log|P_\Omega(\hat{s})| = \operatorname{tr}\hat{P}_\Omega^{-1}(A_i - r_i I_n) \qquad (i = 1, \ldots, m), \tag{14}$$

where $\hat{P}_\Omega = P_\Omega(\hat{s})$. The Hessian matrix of second derivatives we denote as $\hat{H}_\Omega = (\hat{h}_{ij})$ where

$$\hat{h}_{ij} = -\tfrac{1}{2}\frac{\partial^2}{\partial \hat{s}_i \partial \hat{s}_j}\log(|P_\Omega(\hat{s})|) = 2\operatorname{tr}\hat{P}_\Omega^{-1}(A_i - r_i I_n)\hat{P}_\Omega^{-1}(A_j - r_j I_n) \qquad (i, j = 1, \ldots, m). \tag{15}$$

A single-saddlepoint density approximation is therefore

$$\hat{f}(r) = (2\pi)^{-m/2}|\Omega|^{1/2}|\hat{H}_\Omega|^{-1/2}|\hat{P}_\Omega|^{-1/2}(\operatorname{tr}\hat{P}_\Omega^{-1})^m. \tag{16}$$

The last term in (16) satisfies the constraint $\operatorname{tr}\hat{P}_\Omega^{-1}\Omega = n$, an identity derived by multiplying equation i of (14) by $-2\hat{s}_i$ and summing over i to get

$$0 + \operatorname{tr}\hat{P}_\Omega^{-1}\Omega = \operatorname{tr}\hat{P}_\Omega^{-1}\left\{\left(\sum_{i=1}^m -2\hat{s}_i A_i + 2r^{\mathrm{T}}\hat{s} I_n\right) + \Omega\right\} = \operatorname{tr}\hat{P}_\Omega^{-1}\hat{P}_\Omega = n.$$

In the null setting, $\operatorname{tr}\hat{P}_\Omega^{-1} = n$ and the last factor of (16) is n^m.

4. Equivalence in the null setting

The single-saddlepoint approximation in (16) when $\Omega = I_n$ is now shown to be analytically the same as the double-saddlepoint approximation in (9). We first show that the double saddlepoint $(\tilde{s}, \tilde{t})$ can be expressed in terms of the single-saddlepoint $\hat{s}$ through the relation $\tilde{s} = n\hat{s}$. Using (8) we express $\tilde{t}$ in terms of $\tilde{s}$ and substitute this in the value of $\tilde{Q}$ in (7) so that

$$0 = n^{-1} \operatorname{tr} \left\{ \left(1 + \frac{2r^{\mathrm{T}}\tilde{s}}{n} \right) I_n - 2 \sum_{i=1}^{m} \frac{\tilde{s}_i}{n} A_i \right\}^{-1} (A_i - r_i I_n) \qquad (i = 1, \ldots, m).$$

This is the same equation as (14) and, as a saddlepoint equation, yields a unique root so that $\tilde{s} = n\hat{s}$. This equivalence means that $\tilde{Q} = n\hat{P}$ which further gives

$$n^{n/2} e^{-K(\tilde{s},\tilde{t})} = |\hat{P}|^{-1/2}; \tag{17}$$

so the exponential portions determining the saddlepoints of the two approximations are equivalent.

The most difficult and unobvious portion of the argument is in showing the equivalence of the Hessian-type corrections which are related by

$$|K''(\tilde{s}, \tilde{t})| = 2n^{-2m-1} |\hat{H}|. \tag{18}$$

The argument involves separate computations of $\partial\tilde{s}/\partial r^{\mathrm{T}}$ and $n\,\partial\hat{s}/\partial r^{\mathrm{T}}$ by differentiating their respective saddlepoint equations in (6) and (14). The former yields

$$\frac{\partial \tilde{s}}{\partial r^{\mathrm{T}}} = (\tilde{K}''_{ss} - \tilde{K}''_{st} \tilde{K}''^{-1}_{tt} \tilde{K}''_{ts})^{-1},$$

while the latter is a quite long computation leading to the matrix relation

$$(I_m \operatorname{tr} \hat{P}_{\Omega}^{-1} + \hat{w}\hat{s}^{\mathrm{T}}) \frac{\partial r}{\partial \hat{s}^{\mathrm{T}}} = \hat{H}_{\Omega}, \tag{19}$$

where

$$\hat{w} = 2(\operatorname{tr} \hat{P}_{\Omega}^{-2}(A_1 - r_1 I_n), \ldots, \operatorname{tr} \hat{P}_{\Omega}^{-2}(A_m - r_m I_n))^{\mathrm{T}}.$$

The next step equates $|\partial\tilde{s}/\partial r^{\mathrm{T}}| = |n\,\partial\hat{s}/\partial r^{\mathrm{T}}|$. This implies (18) after some more computations which have been given by Butler and Paolella (1996). With (17) and (18) it is now simple algebra to show that $\tilde{f}(r) = \hat{f}(r)$.

The equivalence of these two approximations was not expected since such equivalence is usually connected with the presence of a cut (Barndorff-Nielsen, 1978, pp. 50–51) as described by Booth and Butler (1990). The simplest case with $m = 1$ helps to clarify this point. Suppose the eigenvalues of A_1 are $\lambda_1, \ldots, \lambda_n$ so that

$$r_1 = \sum_{i=1}^{n} \lambda_i \epsilon_i^2 \Big/ \sum_{i=1}^{n} \epsilon_i^2$$

puts Dirichlet $(\frac{1}{2}, \ldots, \frac{1}{2})$ weights on the eigenvalues and has a form comparable with that of the Dirichlet bootstrap of Booth and Butler (1990). The likelihood associated with these chi

squares does not admit a cut but the two density approximations are the same. Likewise, the two cumulative distribution approximations, the double-saddlepoint (Skovgaard 1987) and the single-saddlepoint (Lugannani and Rice 1980) approximations, are the same. The latter approximation is computed by writing

$$\Pr(r_1 \leqslant r) = \Pr\left(\sum_{i=1}^{n}(\lambda_i - r)\epsilon_i^2 \leqslant 0\right)$$

and evaluating the Lugannani–Rice approximation for the distribution of the latter quadratic form at 0. Further details are given in Section 6.

5. The correlogram from regression residuals

Suppose that $\{r_k\colon k = 1, \ldots, m\}$ are serial correlations from least-squares residuals as in (2) and $r_k = z^{\mathrm{T}}B_k z/z^{\mathrm{T}}z$ where $B_k = (b_{ijk})$ is the band matrix of the kth off-diagonals for which $b_{ijk} = 2^{-1}1\{|i-j| = k\}$, where $1\{.\}$ denotes the indicator function. We canonically reduce the problem and show how the methods of the previous sections can be applied.

Suppose that regression $E(y) = X\beta$ leads to residuals $z = My = \{I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\}y$, so we write

$$r_k = \frac{y^{\mathrm{T}}MB_kMy}{y^{\mathrm{T}}My} = \frac{y^{\mathrm{T}}L(L^{\mathrm{T}}ML)L^{\mathrm{T}}B_kL(L^{\mathrm{T}}ML)L^{\mathrm{T}}y}{y^{\mathrm{T}}L(L^{\mathrm{T}}ML)L^{\mathrm{T}}y}, \tag{20}$$

where $L = (L_1 L_2)$ is an orthogonal matrix that canonically reduces M as

$$L^{\mathrm{T}}ML = \begin{pmatrix} L_1^{\mathrm{T}}ML_1 & L_1^{\mathrm{T}}ML_2 \\ L_2^{\mathrm{T}}ML_1 & L_2^{\mathrm{T}}ML_2 \end{pmatrix} = \begin{pmatrix} I_{n-p} & 0 \\ 0 & 0 \end{pmatrix}. \tag{21}$$

Letting $\epsilon = L_1^{\mathrm{T}}y$ and denoting $A_k = L_1^{\mathrm{T}}B_kL_1$ as the upper left $(n-p)\times(n-p)$ principal submatrix of $L^{\mathrm{T}}B_kL$, then r_k is as specified in (1) with dimension $n-p$ for ϵ instead of n. In the null case for which y is $N_n(X\beta, I_n)$, then $\Omega^{-1} = L_1^{\mathrm{T}}L_1 = I_{n-p}$; in the non-null case with y as $N_n(X\beta, \Phi^{-1})$, then $\Omega^{-1} = L_1^{\mathrm{T}}\Phi^{-1}L_1$.

The matrix L_1 is uniquely determined when $p = 1$, but for $p > 1$ can be any member of a compact collection of matrices whose columns form an orthonormal basis for the residual space. Although the true distribution of r does not depend on the choice of L_1, the saddlepoint density approximation does for $p > 1$ through the values of $\{A_k\}$ and also through Ω in the non-null setting.

The support of the m-dimensional correlogram is an open convex set of values r identified in the following way. If

$$R_i = \begin{pmatrix} 1 & r_1 & \cdots & r_i \\ r_1 & 1 & \cdots & r_{i-1} \\ \vdots & & \ddots & \vdots \\ r_i & \cdots & r_1 & 1 \end{pmatrix} \qquad (i = 1, \ldots, m), \tag{22}$$

then the support is $\mathfrak{I} = \{r: |R_i| > 0 \quad i = 1, \ldots, m\}$. In the correlogram context, the mapping $\hat{s} \leftrightarrow r$ through the saddlepoint equation (14) is a bijection from $\hat{s} \in \mathfrak{R}^m$ onto $r \in \mathfrak{I}$ as we now indicate. It suffices to consider the null setting since $\mathfrak{I}$ is not dependent on the value of Ω.

The saddlepoint equation allows for the determination of r from $\hat{s}$ if $2r^{\mathrm{T}}\hat{s} = \hat{z}$ is first determined. Suppose that $\nu_1(\hat{s}) \leqslant \ldots \leqslant \nu_n(\hat{s})$ are the eigenvalues of $2\sum_{i=1}^m \hat{s}_i A_i - I_n$. Then the equality $\operatorname{tr} \hat{P}^{-1} = n$ gives $\hat{z}$ as the unique root of

$$n = \sum_{i=1}^{n} (\hat{z} - \nu_i)^{-1} \qquad (\hat{z} > \nu_n). \tag{23}$$

The values of $\hat{s}$ and $\hat{z}$ determine $\hat{P}$ which in turn determines r as

$$r_i = \frac{\operatorname{tr} \hat{P}^{-1} A_i}{\operatorname{tr} \hat{P}^{-1}} \qquad (i = 1, \ldots, m) \tag{24}$$

through rearrangement of (14). Substituting the decomposition of positive definite $\hat{P}^{-1} = \sum_{i=1}^n (\hat{z} - \nu_i)^{-1} o_i o_i^{\mathrm{T}}$ into (24) gives

$$r^{\mathrm{T}} = (r_1, \ldots, r_m) = \sum_{i=1}^{n} \left((\hat{z} - \nu_i)^{-1} \Big/ \sum_j (\hat{z} - \nu_j)^{-1} \right) (o_i^{\mathrm{T}} A_1 o_i, \ldots, o_i^{\mathrm{T}} A_m o_i) \tag{25}$$

where $(o_i^{\mathrm{T}} A_1 o_i, \ldots, o_i^{\mathrm{T}} A_m o_i)^{\mathrm{T}} \in \mathfrak{I}$ for each i. Thus vector r is a convex combination of vectors in $\mathfrak{I}$ so $r \in \mathfrak{I}$ by its convexity. If the support of the distribution of r were not convex, then the bijection $\hat{s} \leftrightarrow r$ would map between $\mathfrak{R}^m$ and the interior of the convex hull of the support of r. Such would be the situation, for example, if r consisted instead of the various lagged orders of the Durbin–Watson statistics.

6. Special null cases

There is substantial simplification if either $m = 1$ or the matrices $\{A_k: k = 1, \ldots, m\}$ are simultaneously diagonalizable. When $m = 1$, the saddlepoint density (16) of r_1 in (1) has the simple null form

$$\hat{f}(r) = (2\pi)^{-1/2} n \hat{h}^{-1/2} \exp(-\tfrac{1}{2}\hat{\zeta}^2), \tag{26}$$

with

$$\hat{h} = 2\sum_{i=1}^{n} w_i^2 (1 - 2\hat{s}w_i)^{-2}, \qquad \hat{\zeta} = \operatorname{sgn}(\hat{s}) \left(\sum_{i=1}^{n} \log(1 - 2\hat{s}w_i) \right)^{1/2},$$

where sgn is the sign function, $w_i = \lambda_i - r$, $\{\lambda_i\}$ are the eigenvalues of A_1, and $\hat{s}$ solves

$$0 = \sum_{i=1}^{n} (1 - 2w_i\hat{s})^{-1} w_i. \tag{27}$$

The Lugannani–Rice (1980) approximation for the cumulative distribution of r_1 is

$$\Pr(r_1 \leqslant r) \approx \Phi(\hat{\zeta}) + \phi(\hat{\zeta})\{\hat{\zeta}^{-1} - (\hat{s}\hat{h}^{1/2})^{-1}\} \qquad (r \neq E(r_1)) \tag{28}$$

where Φ and ϕ are the standard normal distribution and density functions, respectively. Equations (26), (27) and (28) agree with those of Lieberman (1994, equations (1) and (5)) with I_n in place of his G.

The setting in which $\{A_k\}$ are simultaneous diagonalizable encompasses three forms for vector r that have been discussed by Anderson (1971, Section 6.5). These are (1) circularly defined serial correlations treated by Daniels (1956) which comprise the circular serial correlogram, (2) successive differences treated by Durbin (1980b) and defined so that the Durbin–Watson (1950, 1951) statistic is $2(1 - r_1)$ but such that r_k for $k > 2$ is not related to the lag k Durbin–Watson statistic and (3) another form treated by Durbin (1980b) and defined so that r_1 is the first component of the non-circular correlogram but such that the vector itself is not the correlogram.

Suppose that $O^{\mathrm{T}} A_k O = \mathrm{diag}(\lambda_k)$ where $\lambda_k^{\mathrm{T}} = (\lambda_{1k}, \ldots, \lambda_{nk})$ consists of the eigenvalues of A_k for each k. Let $\chi = (\chi_1, \ldots, \chi_n)^{\mathrm{T}}$ consist of independent and identically distributed χ_1^2 variables. Place the λ_k^{T} in the rows of $(m \times n)$ matrix Λ so that

$$\Lambda = (\lambda_1, \ldots, \lambda_m)^{\mathrm{T}} = (\ell_1, \ldots, \ell_n).$$

The vector ratio is now $r = \Lambda\chi / 1^{\mathrm{T}}\chi$. In this setting

$$\log|\hat{P}| = \sum_{i=1}^{n} \log\{1 - 2(\ell_i - r)^{\mathrm{T}}\hat{s}\}, \tag{29}$$

and the saddle point $\hat{s}$ solves

$$0 = \sum_{i=1}^{n} \{1 - 2(\ell_i - r)^{\mathrm{T}}\hat{s}\}^{-1}(\ell_i - r). \tag{30}$$

The Hessian is

$$\hat{H} = 2\sum_{i=1}^{n} \{1 - 2(\ell_i - r)^{\mathrm{T}}\hat{s}\}^{-2}(\ell_i - r)(\ell_i - r)^{\mathrm{T}}. \tag{31}$$

Suppose further that the numerators of r above add up disjoint subsets of the χ^2 variables so that

$$\Lambda = \begin{pmatrix} 1_{\alpha_1}^{\mathrm{T}} & 0^{\mathrm{T}} & \cdots & \cdots & 0^{\mathrm{T}} \\ 0^{\mathrm{T}} & 1_{\alpha_2}^{\mathrm{T}} & 0^{\mathrm{T}} & \vdots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0^{\mathrm{T}} & 0^{\mathrm{T}} & \cdots & 1_{\alpha_m}^{\mathrm{T}} & 0^{\mathrm{T}} \end{pmatrix}. \tag{32}$$

In this case the true density of r is Dirichlet $(\alpha_1/2, \ldots, \alpha_m/2; \alpha_{m+1}/2)$ where $\alpha_{m+1} = n - \sum_{i=1}^{m} \alpha_i$. Some long computations by Butler and Paolella (1996) showed that $\hat{f}(r)$ differs from $f(r)$ by Stirling's approximation in the gamma functions, i.e.

$$\hat{f}(r) = \frac{\hat{\Gamma}(n/2)}{\Gamma(n/2)} \left(\prod_{i=1}^{m+1} \frac{\Gamma(\alpha_i/2)}{\hat{\Gamma}(\alpha_i/2)} \right) f(r). \tag{33}$$

Furthermore, the solution to the saddlepoint equation (30) is explicit with

$$\hat{s}_i = (2n)^{-1} \left(\frac{\alpha_{m+1}}{r_{m+1}} - \frac{\alpha_i}{r_i} \right) \qquad (i = 1, \ldots, m) \tag{34}$$

where $r_{m+1} = 1 - \sum_{i=1}^{m} r_i$.

7. Circular parametric likelihood connections

The saddlepoint density (10) encompasses the density of the first m circularly defined serial correlations considered by Daniels (1956). In this context, (16) can be approximated in the null setting to give the same analytical results as the density in Daniels' equation (9.14) for the first m circularly defined serial correlations. Upon transformation to partial serial correlations, these null distributions agree with the density of partial serial correlations given by Daniels' equation (10.5) when corrected for the misprint. Details of this have been given by Butler and Paolella (1996).

Similar arguments can be used to show how (16) relates to the approximations of Durbin (1980b) for the density of circularly defined serial and partial serial correlations constructed from residuals that remove effects of Fourier regressors. For details see Butler and Paolella (1996). Removal of the sample mean $\overline{y}$ from $\{y_t\}$ is a special case of this since the vector of ones is always an eigenvector for the collection of circulant matrices involved.

8. The conditional distribution of r_m given $r_1, \ldots, r_{m-1}$

Consideration of such conditional distributions is motivated by methods for optimal testing of the order of autoregressive models. For example, the circular AR(m) model in Daniels' equation (9.1) has an exponential family likelihood with canonical parameters as the inverse autocovariances $\{\delta_i\}$. Using standard theory as done by Anderson (1971, Section 6.3.2), the UMP unbiased test of $\delta_m = 0$ versus $\delta_m \neq 0$ or, equivalently, AR($m-1$) versus AR(m), rejects for values of r_{m0} falling sufficiently far out in either tail of the conditional density $f(r_m | r_{(m-1)0}; \Omega)$. One-sided p values can be computed as

$$\hat{p} = \min\{\tau, 1 - \tau\}, \tag{35}$$

where

$$\tau = \Pr(r_m > r_{m0} | r_{(m-1)} = r_{(m-1)0}; \delta_m = 0)$$

and $r_0^{\mathrm{T}} = (r_{10}, \ldots, r_{m0}) = (r_{(m-1)0}^{\mathrm{T}}, r_{m0})$ is the observed value of $r^{\mathrm{T}} = (r_{(m-1)}^{\mathrm{T}}, r_m)$. Such optimality is maintained with either mean correction or Fourier regressors added into the model (Anderson 1971, see Section 6.6). This optimality does not carry over to the non-

circular model but does provide motivation for use of the same procedure in the general testing of AR($m-1$) versus AR(m).

We shall use the notation above to discuss the approximation of the conditional density of r_m given the value $r_{(m-1)0}$ in the most general context. A double-saddlepoint density as given by Barndorff-Nielsen and Cox (1979) is computed as the ratio of two single approximations:

$$\hat{f}(r_m|r_{(m-1)0};\,\Omega) = \frac{\hat{f}(r_{(m-1)0},\,r_m;\,\Omega)}{\hat{f}(r_{(m-1)0};\,\Omega)} = (2\pi)^{-1/2}\left(\frac{|\hat{H}_{m-1}|}{|H_m|}\right)^{1/2}\left[\frac{(\mathrm{tr}\,P_m^{-1})^m}{(\mathrm{tr}\,\hat{P}_{m-1}^{-1})^{m-1}}\right]\left(\frac{|P_m|}{|\hat{P}_{m-1}|}\right)^{-1/2}$$

where $\hat{H}_{m-1}$ and $\hat{P}_{m-1}$ are the $\hat{H}_\Omega$ and $\hat{P}_\Omega$ values associated with the $(m-1)$-dimensional saddlepoint $\hat{s}_{(m-1)}$ of the denominator determined by $r_{(m-1)0}$, H_m and P_m are the $\hat{H}_\Omega$ and $\hat{P}_\Omega$ values associated with the m-dimensional saddlepoint $s^{\mathrm{T}} = (s_1, \ldots, s_m)$ of the numerator determined by $(r_{(m-1)0}, r_m)$, and explicit dependence on Ω has been suppressed. For probability calculation, one-dimensional numerical integration

$$\Pr(r_m < r_{m0} r_{(m-1)0};\,\Omega) \approx \left(\int_{\mathscr{r}\cap(-1,\,r_{m0})} \hat{f}(r_m|r_{(m-1)0})\,\mathrm{d}r_m\right) \Big/ \left(\int_{\mathscr{r}} \hat{f}(r_m|r_{(m-1)0})\,\mathrm{d}r_m\right) \tag{36}$$

can be performed when $\mathscr{r}$, the conditional support of r_m given $r_{(m-1)0}$, is identifiable. In the correlogram setting $\mathscr{r} = \{r_m\colon |R_m| > 0\}$ when it is known that $|R_{(m-1)0}| > 0$, where R_m and $R_{(m-1)0}$ are R_m and R_{m-1} evaluated at $(r_{(m-1)0}^{\mathrm{T}}, r_m)$. Since

$$0 < |R_m| = |R_{(m-1)0}|\{1 - (r_{(m-1)0}^{\mathrm{T}}, r_m)R_{(m-1)0}^{-1}(r_{(m-1)0}^{\mathrm{T}}, r_m)^{\mathrm{T}}\} \tag{37}$$

then $\mathscr{r} = (a, b) \subset (-1, 1)$ where a and b are the roots to the quadratic term in r_m within the braces. Determination of $\mathscr{r}$ in settings other than the serial correlogram is not so simple. If instead r consists of the various lagged orders of the Durbin–Watson statistics, then the joint support of r is difficult to identify and so is the conditional support.

A double-saddlepoint cumulative function as used by Skovgaard (1987) is derived using the method of Temme (1982). Consider performing the integration in the numerator of (36) with the transformation

$$r_m \leftrightarrow w = \mathrm{sgn}(s_m)\left\{\log\left(\frac{|P_m|}{|\hat{P}_{m-1}|}\right)\right\}^{1/2}. \tag{38}$$

To see that w is well defined, note that saddlepoint s in

$$P_m = P_m(s) = \Omega + [2(r_{(m-1)0}^{\mathrm{T}}, r_m)s]I_n - 2\sum_{i=1}^{m} s_i A_i$$

maximizes $|P_m(\tau)|$ over $\tau \in \Re^m$ whereas $|\hat{P}_{m-1}|$ is the maximum of $|P_m(\tau_{(m-1)}, 0)|$ over $\tau_{(m-1)} \in \Re^{m-1}$; thus $|P_m| \geqslant |\hat{P}_{m-1}|$ with equality only when $s_m = 0$. The numerator of (36) is now

$$\Pr(r_m \leqslant r_{m0} | r_{(m-1)0};\, \Omega) \approx \int_{-\infty}^{w_0} h(w)\phi(w)\,\mathrm{d}w, \tag{39}$$

with

$$h(w) = \left(\frac{|\hat{H}_{m-1}|}{|H_m|}\right)^{1/2} \frac{(\operatorname{tr} P_m^{-1})^m}{(\operatorname{tr} \hat{P}_{m-1}^{-1})^{m-1}} \frac{\partial r_m}{\partial w} \tag{40}$$

and

$$w_0 = \operatorname{sgn}(\hat{s}_m) \left\{ \log\left(\frac{|\hat{P}_m|}{|\hat{P}_{m-1}|}\right) \right\}^{1/2} \tag{41}$$

where $\hat{P}_m = P_m(\hat{s})$ and $\hat{s}^{\mathrm{T}} = (\hat{s}_1, \ldots, \hat{s}_m)$ is the saddlepoint for $(r_{(m-1)0}^{\mathrm{T}}, r_{m0})$. The change in variable $\partial r_m / \partial w = w/(s_m \operatorname{tr} P_m^{-1})$ is shown in the Appendix. Using this in (40), then $h(w)$ has a removable singularity at $w = 0$ with $h(0) = 1$ as shown in the Appendix. The argument of Temme as outlined by Barndorff-Nielsen and Cox (1989, Section 3.9) now applies and leads to

$$\Pr(r_m \leqslant r_{m0} | r_{(m-1)0};\, \Omega) \approx \Phi(w_0) + \phi(w_0)\left(\frac{1}{w_0} - \frac{1}{v_0}\right) \qquad (\hat{s}_m \neq 0), \tag{42}$$

where

$$v_0 = \hat{s}_m \left(\frac{|\hat{H}_m|}{|\hat{H}_{m-1}|}\right)^{1/2} \left(\frac{\operatorname{tr} \hat{P}_{m-1}^{-1}}{\operatorname{tr} \hat{P}_m^{-1}}\right)^{m-1}. \tag{43}$$

Numerical accuracy of (42) is considered in Section 11.

9. Other correlograms

Densities for the partial correlogram vector $r_{\cdot}$ and inverse correlogram vector r_{-} are given below in the general non-null setting without the assumption of circularity. The necessary Jacobians have been given by Daniels and are specified below. Conditional distribution function approximations can also be developed for each of these correlograms, but we do not give details.

For the Jacobians, Daniels shows in his equation (10.2) that

$$\left\| \frac{\partial r}{\partial \hat{\alpha}^{\mathrm{T}}} \right\| = |R_{m-1}| \times \|\hat{D}\|^{-1}, \tag{44}$$

where $\hat{D} = (\hat{d}_{ij})$ has entries

$$\hat{d}_{ij} = (-1)1\{i = j\} + \hat{\alpha}_{i+j} 1\{i + j \leqslant m\} + \hat{\alpha}_{j-i} 1\{i < j\} \qquad (i, j = 1, \ldots, m)$$

based on $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_m)^{\mathrm{T}} = R_{m-1}^{-1} r$, the solution to the Yule–Walker equation. His equation (10.4) determines that

$$\left\| \frac{\partial \hat{\alpha}}{\partial r_{\cdot}} \right\| = \prod_{j \text{ odd}} (1 - r_{j\cdot}^2)^{(j-1)/2} \prod_{j \text{ even}} \{(1 - r_{j\cdot}^2)^{(j/2)-1}(1 - r_{j\cdot})\}.$$

Putting these together gives

$$\hat{f}(r_{\cdot}) = \hat{f}(r) \left\| \frac{\partial r}{\partial \hat{\alpha}^{\mathrm{T}}} \right\| \times \left\| \frac{\partial \hat{\alpha}}{\partial r_{\cdot}^{\mathrm{T}}} \right\| \qquad (r_{\cdot} \in (-1, 1)^m). \tag{45}$$

If we denote the lag m inverse correlation vector as $r_{-} = \hat{\delta}/\hat{\delta}_0 = (\hat{\delta}_1, \ldots, \hat{\delta}_m)^{\mathrm{T}}/\hat{\delta}_0$ where

$$\begin{aligned}
\hat{\delta}_0 &= 1 + \hat{\alpha}_1^2 + \cdots + \hat{\alpha}_m^2, \\
\hat{\delta}_1 &= -\hat{\alpha}_1 + \hat{\alpha}_1\hat{\alpha}_2 + \cdots + \hat{\alpha}_{m-1}\hat{\alpha}_m, \\
\hat{\delta}_2 &= -\hat{\alpha}_2 + \hat{\alpha}_1\hat{\alpha}_3 + \cdots + \hat{\alpha}_{m-2}\hat{\alpha}_m, \\
&\vdots \\
\hat{\delta}_{m-1} &= -\hat{\alpha}_{m-1} + \hat{\alpha}_1\hat{\alpha}_m, \\
\hat{\delta}_m &= -\hat{\alpha}_m,
\end{aligned} \tag{46}$$

then its density is

$$\hat{f}(r_{-}) = \hat{f}(r)|R_{m-1}| \times \|\hat{D}\|^{-2}(1 + r^{\mathrm{T}}R_{m-1}^{-2}r)^m|1 + 2r^{\mathrm{T}}r_{-}|^{-1}. \tag{47}$$

The Jacobian in this transformation has been computed in the Appendix. The support of r_{-} is identified through the one-to-one relationship $r_{-} \leftrightarrow r$.

10. Null asymptotics

The asymptotic order of the error for the approximation in (16) depends on the set of matrices $\{A_k\}$ and their large-sample behaviour. We do not consider an analysis of this here but content ourselves with a heuristic discussion of the orders of magnitude.

The order of the terms eliminated from the single-saddlepoint approximation is determined by the sample moments of the eigenvalues of $\hat{P}^{-1}$. Each eliminated term p_j is a function of the $\{q_k: k \geqslant 2\}$. At the saddlepoint, these values are

$$0 < \hat{q}_k = \frac{\operatorname{tr} \hat{P}^{-k}}{(\operatorname{tr} \hat{P}^{-1})^k} = n^{-k} \operatorname{tr} \hat{P}^{-k} \qquad (k \geqslant 2) \tag{48}$$

since the eigenvalues of $\hat{P}^{-1}$ have a sample mean of $n^{-1} \operatorname{tr} \hat{P}^{-1} = 1$. If we suppose a sequence $\{A_k\}$ that produces eigenvalues for $\hat{P}^{-1}$ whose first m sample moments are $O(1)$ for large n, then

$$n^{-k} \operatorname{tr} \hat{P}^{-k} = n^{-k+1} \times n^{-1} \operatorname{tr} \hat{P}^{-k} = O(n^{-k+1}) \qquad (k \geqslant 2). \tag{49}$$

Then the errors in eliminated terms are of order $O(n^{-1})$ and smaller. There is another argument that these terms are small. Without any conditions on $\{A_k\}$, $\{\hat{q}_k\}$ decrease monotonically at an exponential rate in k since

$$\sum_{k=0}^{\infty} \hat{q}_k = \sum_{k=0}^{\infty} \operatorname{tr}\left(\frac{\hat{P}^{-1}}{n}\right)^k = \operatorname{tr}(I_n - n^{-1}\hat{P}^{-1})^{-1} < \infty. \tag{50}$$

The sum converges since the eigenvalues of $n^{-1}\hat{P}^{-1}$ are positive and add to 1.

11. Numerical examples

Table 1 compares the accuracy of various univariate approximations when the true model is a non-circular AR(1) process requiring mean correction, i.e. $y_t = \mu + \eta_t$ with $\eta_t = \alpha_1\eta_{t-1} + \vartheta_t$ for $t = 1, \ldots, n$, and $\{\vartheta_t\}$ are independent $N(0, 1)$ realizations. Displayed are approximations of $\Pr(r_1 \leqslant r_{10}|\alpha_1)$ for the various values of n and α_1 in four settings: null ($\alpha_1 = 0$), stationary ($\alpha_1 = 0.5$), unit-root ($\alpha_1 = 1$), and non-stationary ($\alpha_1 = 1.1$). The column headed Imhof refers to the approximate numerical inversion algorithm of Imhof (1961); the column headed Beta is Henshaw's (1966) beta distribution approximation matching third and fourth moments; the column headed L–R is the Lugannani–Rice approximation given in (28); the column headed NID refers to the numerically integrated single-saddlepoint density in (26) using the Romberg integration algorithm from Press *et al*.

Table 1. Approximations of $\Pr(r_1 \leqslant r_{10})$ for non-circular AR(1) models[a]

n	α_1	r_{10}	Imhof	Beta	L–R	NID	Normal
10	0	0.3471	0.9500	0.9501	0.9501	0.9496	0.8638
	0.5	0.6015	0.9500	0.9509	0.9491	0.9442	0.6445
	1.0	0.7368	0.9500	Fail[b]	0.9475	0.9448	—
	1.1	0.7460	0.9500	Fail[b]	0.9462	0.9464	—
25	0	0.2711	0.9500	0.9500	0.9501	0.9501	0.9124
	0.5	0.6467	0.9500	0.9503	0.9509	0.9501	0.8029
	1.0	0.8961	0.9500	Fail[b]	0.9484	0.9442	—
	1.1	0.8848	0.9500	1.000	0.9377	0.9239	—
70	0	0.1785	0.9500	0.9500	0.9500	0.9500	0.9323
	0.5	0.6197	0.9500	0.9501	0.9501	0.9501	0.8762
	1.0	0.9631	0.9500	Fail[b]	0.9490	0.9452	—
	1.1	0.9050	0.9500	0.9505	0.9541	Fail[c]	—

[a]The model is $y_t = \mu + \eta_t$ with $\eta_t = \alpha_1\eta_{t-1} + \vartheta_t$ where $\{\vartheta_t\}$ are i.i.d. normal innovations for $t = 1, \ldots, n$. Imhof is "exact" numerical inversion. Beta is beta-distribution approximation. L–R is the Lugannani–Rice approximation, NID is the numerically integrated single-saddlepoint density in (26), and Normal is the normal approximation.
[b]The beta approximation sometimes fails because either of its degrees-of-freedom estimates is negative.
[c]Numerical integration breaks down in this case.

(1989); the column headed Normal is the normal approximation $r_1 \sim N(\alpha_1, n^{-1})$ applicable only when $|\alpha_1| < 1$. Values of r_{10} were selected so the Imhof procedure returned a value of 0.95.

Both the beta and the Lugannani–Rice approximations are highly accurate, for all cases and sample sizes; however, the beta approximation often fails in non-stationary settings because either of its degrees-of-freedom estimates is negative. As would be expected, values obtained from the numerically integrated density are close, but not as accurate as the Lugannani–Rice values. Owing to rounding errors and the "pile-up" of the true density, numerical integration for the non-stationary case with $n = 70$ observations failed. Not surprisingly, the normal approximation is very poor for the (arguably impractical) sample size of 10, but, even with 70 observations, it breaks down as $|\alpha_1|$ increases towards unity.

Table 2 computes approximations of $\Pr(r_1 \leqslant r_{10},\ r_2 \leqslant r_{20} | \alpha_1, \alpha_2)$ when the true model is AR(0), AR(1) and AR(2), with each model requiring mean correction of the form $y_t = \mu + \eta_t$ with $\eta_t = \alpha_1 \eta_{t-1} + \alpha_2 \eta_{t-2} + \vartheta_t$ for $t = 1, \ldots, n$, and $\{\vartheta_t\}$ as independent $N(0, 1)$ realizations. Three stationary settings are considered including AR(0) with $\alpha_1 = \alpha_2 = 0$, AR(1) with $\alpha_1 = 0.5$ and $\alpha_2 = 0$, and AR(2) with $\alpha_1 = 1.2$ and $\alpha_2 = -0.8$. The column headed NID again refers to the (two-dimensional) numerically integrated density; the column headed Simulation contains results from Monte Carlo simulation using two million replications based on the "ran1" and "ran2" generators from Press *et al.* (1989); the column headed Normal is the normal approximations based on the theoretical autocovariance function and Bartlett's formula (Priestley 1981, Section 5.3.4), i.e.

$$\begin{pmatrix} r_1 - 0.5 \\ r_2 - 0.5^2 \end{pmatrix} \sim N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, n^{-1}\begin{pmatrix} 0.75 & 0.75 \\ & 1.3125 \end{pmatrix}\right]$$

for AR(1), and

$$\begin{pmatrix} r_1 - 2/3 \\ r_2 - 0 \end{pmatrix} \sim N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, n^{-1}\begin{pmatrix} 0.06173 & 0.14815 \\ & 0.46667 \end{pmatrix}\right]$$

Table 2. Approximations of $\Pr(r_1 \leqslant r_{10},\ r_2 \leqslant r_{20})$ for non-circular AR(0)–AR(2) models[a]

n	α_1	α_2	r_{10}	r_{20}	NID	Simulation	Normal
10	0	0	0	0	0.3835	0.3854	0.25
10	0.5	0	0	0	0.1476	0.1515	0.0313
10	1.2	−0.8	0.4	0.2	0.1663	0.1785	0.0^3344
25	0	0	0	0	0.3326	0.3325	0.25
25	0.5	0	0	0	0.0179	0.0176	0.00189
25	1.2	−0.8	0.4	0.2	0.0080	0.0077	0.0^7401
70	0	0	0	0	0.2983	0.2987	0.25
70	0.5	0	0	0	0.0^465	0.0^468	0.0^6680

[a]The model is $y_t = \mu + \eta_t$ with $\eta_t = \alpha_1 \eta_{t-1} + \alpha_2 \eta_{t-2} + \vartheta_t$ where $\{\vartheta_t\}$ are i.i.d. normal innovations for $t = 1, \ldots, n$. NID is the numerically integrated single-saddlepoint density in (26), Simulation is the Monte Carlo simulation, and Normal is the bivariate normal approximation.

for AR(2). The bivariate normal cumulative distribution function was approximated using the algorithm from Drezner (1978), noting the correction given by Hull (1993, p. 245).

Comparison with the simulated values shows that the numerically integrated saddlepoint approximation performs exceptionally well for the null model $\alpha_1 = \alpha_2 = 0$, even with only $n = 10$ observations. For $n \geqslant 25$, it obtains approximately two-digit accuracy for both non-null models. The asymptotic bivariate normal approximation, on the other hand, is not "too" far off for the null model with $n = 70$ observations (with value 0.25 compared with the simulated valve of 0.2987) but is otherwise completely unacceptable, being off in some cases by several orders of magnitude.

Table 3 gives p values of the form (35) under the null setting ($\Omega = I_n$) in the progressive testing of AR($m-1$) versus AR(m) for $m = 1, \ldots, 5$ using the annual pear data of Henshaw (1966) covering $n = 16$ years with $p = 5$ dependent variables inclusive of a mean location parameter. The column headed r gives the value r_{m0} and the column headed $r_{\cdot}$ the lag m partial serial correlation from the regression residuals. Interval (Low, High) is the support of the conditional distribution of r_m given $r_{(m-1)0}$. The columns headed NICD and SCDF are the approximations from the numerically integrated conditional saddlepoint density in (36), and the saddlepoint cumulative density function approximation in (42), respectively.

The column headed D–D in Table 3 is the p-value approximation based on the circular Daniels–Durbin distribution theory with $p = 5$. The conditional test that rejects for tail values of r_m given $r_{(m-1)0}$ is analytically equivalent to the partial serial correlation test that rejects for tail values of $r_{m\cdot}$ given $r_{1.0}, \ldots, r_{m-1.0}$, the observed values of the first lag $(m-1)$ partial serial correlations. In the circular null setting, $r_{m\cdot}$ is approximately independent of $r_{1\cdot}, \ldots, r_{m-1\cdot}$ and the p value when $r_{m.0} \leqslant 0$ is determined from Durbin (1980b, equation (20)) as

$$\hat{p} = \Pr\{r_{m\cdot} \leqslant r_{m.0}\} = \begin{cases} \mathrm{IB}\left[\dfrac{r_{m.0}+1}{2}; \dfrac{n-p+1}{2}, \dfrac{n+p+1}{2}\right] & (m \text{ odd}), \\ \mathrm{IB}\left[\dfrac{r_{m.0}+1}{2}; \dfrac{n-p}{2}, \dfrac{1}{2}(n+p)+1\right] & (m \text{ even}), \end{cases} \tag{51}$$

Table 3. Approximate p values in progressive AR(m) testing using the pears data set[a]

m	r	$r_{\cdot}$	Low	High	NICD	SCDF	D–D
1	0.26335	0.26335	−1	1	0.0494	0.0317	0.0096
2	−0.20321	−0.29288	−0.86130	1	0.5711	0.5665	0.6249
3	−0.42921	−0.33250	−0.99714	0.70450	0.1225	0.1239	0.4545
4	−0.24708	−0.12116	−0.91215	0.60137	0.4045	0.4339	0.8469
5	0.19903	0.17838	0.67962	0.81167	0.1178	0.1151	0.0238

[a]The data set contains $n = 16$ observations and $p = 5$ regressors. Columns r and $r_{\cdot}$ list the mth-order serial and partial serial correlations from the regression residuals respectively. The interval (Low, High) designates the range of support of the conditional distribution of r_m given $r_{(m-1)0}$. NICD is the numerically integrated conditional saddlepoint density in (36), SCDF is the saddlepoint CDF approximation (42), and D–D denotes the Daniel–Durbin approximation (51).

where IB(b, α, β) is the incomplete beta(α, β) probability up to value b. A comparable expression can be written when $r_{m.0} > 0$.

Table 4 illustrates the same sequence of conditional tests as in Table 3 using the quarterly measurements of investment data in Vinod (1973, Table 1) spanning $n = 44$ quarters and having $p = 3$ dependent variables inclusive of a mean location parameter. Whereas the Daniels–Durbin approximation based on circular models is quite different from the non-circular approximations with $n = 16$ and $p = 5$, these approximations are now quite good in this example when $n = 44$ and $p = 3$. The examples suggest that perhaps n does not need to be especially large relative to p for the approximations based on the circularity assumptions of Daniels and Durbin to be accurate.

12. Final remarks

There are other examples that might be included in the discussion above such as the sequence of Durbin–Watson statistics from lag 1 to m. However, for joint probability computation there is the problem that $\mathfrak{I}$ may be difficult to identify and also not convex. The conditional cumulative distribution function approximations avoid this difficulty and might perhaps be useful.

The points of the normalized periodogram $\{I_1, \ldots, I_{n-1})$ are also of the form (1) with $m = n - 1$ since they are values of a finite Fourier transform of the correlogram (Diggle 1990, Section 2.8) and therefore linear in $\{r_i: i = 1, \ldots, n - 1\}$. We have not addressed the issue of whether these examples can be considered. The asymptotic methods employed here and in all related work assume a fixed value of m but consideration of the periodogram should presume otherwise. It is, however, entirely possible that the saddlepoint density in (16) could retain accuracy in this setting. Evidence for this is revealed in the accuracy attained in reproducing the Dirichlet density of Section 6 as m increases with increasing n. Asymptotically this must result in degenerate categories corresponding to bounded or slowly growing values in $\{a_i\}$. However, the accuracy in (33) when renormalized is not diminished with increasing m.

If the periodogram can be approximated, then so can window-smoothed periodograms since they can be written in the form (1) based on their linearity in $\{I_1, \ldots, I_{n-1}\}$. These and other examples are worthy of further study.

Table 4. Approximate p values in progressive AR(m) testing using the Vinod data set[a]

m	r	$r_.$	Low	High	NICD	SCDF	D–D
1	0.01765	0.01765	−1	1	0.2797	0.2784	0.2846
2	0.29141	0.29119	−0.999377	1	0.00355	0.00352	0.00478
3	−0.25872	−0.29238	−0.906142	0.923709	0.05914	0.05924	0.06195
4	0.44199	0.46192	−0.781215	0.892208	0.0^4120	0.0^4112	0.0^4521
5	−0.42366	−0.60055	−0.686565	0.629792	0.0^4200	0.0^4152	0.0^4327

[a]The data set contains $m = 44$ observations and $p = 3$ regressors.

We conclude by making the following practical recommendations. Routine correlogram significance should be judged using the beta approximations of Daniels (1956) and Durbin (1980b) which are considerably more accurate than the asymptotic normal bands $\pm 2/n^{1/2}$, even for series of length $n = 44$ with $p = 3$ regressors. However, for short series, significance should be based on the new double-saddlepoint approximation in (64) that presumes non-circularity since, for such series, the approximations of Daniels and Durbin suffer from the assumption of circularity.

Appendix

A.1. Conditions for a full rank r

We show that linear independence of the matrices $\{A_i: i = 1, \ldots, m\}$ and I is sufficient to guarantee that the distribution of r has full rank. We say that $\{A_i\}$ and I are linearly independent if

$$\sum_{i=1}^{m} \ell_i A_i + \ell_{m+1} I = 0 \Rightarrow (\ell_1, \ldots, \ell_{m+1}) = 0.$$

For the distribution of r to not be of full rank, there would need to exist $(\ell_1, \ldots, \ell_m) \neq 0$ and ℓ_{m+1} such that

$$1 = \Pr\left(\sum_{i=1}^{m} \ell_i r_i = -\ell_{m+1}\right) = \Pr\left\{\epsilon^{\mathrm{T}}\left(\sum_{i=1}^{m} \ell_i A_i + \ell_{m+1} I\right)\epsilon = 0\right\}.$$

This requires that $\sum_{i=1}^{m} \ell_i A_i + \ell_{m+1} I = 0$. Under linear independence this is not possible since $(\ell_1, \ldots, \ell_m) \neq 0$.

A.2. Derivation of $\partial r_m/\partial w$ in (40) and proof that $h(0) = 1$

Differentiate $w^2 = \log(|P_m|/|\hat{P}_{m-1}|)$ so that

$$\begin{aligned} 2w\frac{\partial w}{\partial r_m} &= \operatorname{tr}\left(P_m^{-1}\frac{\partial P_m}{\partial r_m}\right) \\ &= 2\left(s_m + r^{\mathrm{T}}\frac{\partial s}{\partial r_m}\right)\operatorname{tr} P_m^{-1} - 2\sum_{i=1}^{m}\frac{\partial s_i}{\partial r_m}\operatorname{tr} P_m^{-1} A_i \\ &= 2s_m \operatorname{tr} P_m^{-1} \end{aligned} \tag{A1}$$

where the last line follows from $r_i = (\operatorname{tr} P_m^{-1} A_i)/\operatorname{tr} P_m^{-1}$.

Substitute (A1) into the expression for $h(\cdot)$ in (40) and let $w \to 0$ so that $s_m \to 0$, $s \to (\hat{s}_{(m-1)0}, 0)$, $\operatorname{tr} P_m^{-1} \to \operatorname{tr} \hat{P}_{m-1}^{-1}$, and

$$\lim_{w\to 0} h(w) = \left(\frac{|\hat{H}_{m-1}|}{|H_m(\hat{s}_{(m-1)0},\, 0)|}\right)^{1/2} \lim_{w\to 0} \frac{w}{s_m}. \tag{A2}$$

The final term $\lim_{w\to 0}(w/s_m) = \lim_{w\to 0}(\partial w/\partial s_m)$ is computed by partial differentiating the last line in (A1) with respect to s_m to get

$$w\frac{\partial^2 w}{\partial r_m^2}\frac{\partial r_m}{\partial s_m} + \left(\frac{\partial w}{\partial s_m}\right)^2 \frac{\partial s_m}{\partial r_m} = \operatorname{tr} P_m^{-1} + s_m \frac{\partial}{\partial s_m} \operatorname{tr} P_m^{-1}.$$

Taking limits then gives

$$\lim_{w\to 0} \frac{\partial w}{\partial s_m} = \left((\operatorname{tr} \hat{P}_{m-1}^{-1}) \frac{\partial r_m}{\partial s_m}\bigg|_{w=0}\right)^{1/2}.$$

The latter derivative results from rewriting (19) as

$$\frac{\partial s}{\partial r^{\mathrm{T}}} = H_m^{-1}(I_m \operatorname{tr} P_m^{-1} + ws^{\mathrm{T}})$$

so that

$$\frac{\partial s_m}{\partial r_m} = h^{mm} \operatorname{tr} P_m^{-1} + \xi_m^{\mathrm{T}} H_m^{-1} w s_m, \tag{A3}$$

where h^{mm} is the (m, m)th element of H_m^{-1} or $h^{mm} = |H_{m-1}(s_1, \ldots, s_{m-1})|/|H_m|$. Taking the limit and using this cofactor expression for h^{mm} gives

$$\lim_{w\to 0} \frac{\partial w}{\partial s_m} = \left(\frac{|H_m(\hat{s}_{m-1)0},\, 0)|}{|\hat{H}_{m-1}|}\right)^{1/2}.$$

Combine this with (A2) to show that $\lim_{w\to 0} h(w) = 1$.

A.3. Jacobian for (47)

The Jacobian $\partial r/\partial r_-^{\mathrm{T}}$ is computed in terms of $\partial r/\partial \hat{\alpha}^{\mathrm{T}} \times \partial \hat{\alpha}/\partial r_-^{\mathrm{T}}$ where the former quantity is given in (44). Since $r_- = \hat{\delta}/\hat{\delta}_0$, then the latter term can be computed by way of (46). From (46),

$$\hat{D} = \frac{\partial \hat{\delta}}{\partial \hat{\alpha}^{\mathrm{T}}} = \frac{\partial}{\partial \hat{\alpha}^{\mathrm{T}}}(\hat{\delta}_0 r_-) = \hat{\delta}_0 \frac{\partial r_-}{\partial \hat{\alpha}^{\mathrm{T}}} + r_- \frac{\partial \hat{\delta}_0}{\partial \hat{\alpha}^{\mathrm{T}}}.$$

Since $\partial \hat{\delta}_0/\partial \hat{\alpha}^{\mathrm{T}} = 2\hat{\alpha}^{\mathrm{T}}$, then we can solve for

$$\left\| \frac{\partial r_-}{\partial \hat{\alpha}^{\mathrm{T}}} \right\| = \hat{\delta}_0^{-m} \|\hat{D} - 2r_- \hat{\alpha}^{\mathrm{T}}\|$$

$$= (1 + r^{\mathrm{T}} R_{m-1}^{-2} r)^{-m} \|\hat{D}\| \times |1 - 2\hat{\alpha}^{\mathrm{T}} \hat{D}^{-1} r_-|,$$

since $\hat{\alpha} = R_{m-1}^{-1} r$. However, now $\hat{\alpha} = -\hat{D}^{\mathrm{T}} r$ as given in Daniels' equation (9.13) so that

$$\left\| \frac{\partial \hat{\alpha}}{\partial r_-^{\mathrm{T}}} \right\| = (1 + r^{\mathrm{T}} R_{m-1}^{-2} r)^m \|\hat{D}\|^{-1} \times |1 + 2r^{\mathrm{T}} r_-|^{-1} \tag{A4}$$

and the density in (47) follows from the Jacobians in (44) and (A4).

Acknowledgements

Much of the work was done while the first author was visiting the Department of Statistics, University of Oxford. The authors acknowledge the financial support of the National Science Foundation.

References

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. New York: Wiley.

Barndorff-Nielsen, O.E. (1978) *Information and Exponential Families in Statistical Theory*. New York: Wiley.

Barndorff-Nielsen, O.E. and Cox, D.R. (1979) Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. B*, **41**, 279–312.

Barndorff-Nielsen, O.E. and Cox, D.R. (1989) *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall.

Booth, J.G. and Butler, R.W. (1990) Randomization distributions and saddlepoint approximations in generalized linear models. *Biometrika*, **77**, 787–796.

Butler, R.W. and Paolella, M.S. (1996) Approximate distributions for the various serial correlograms. Technical Report 96/11. Department of Statistics, Colorado State Universisty, Fort Collins, CO.

Daniels, H.E. (1956) The approximate distribution of serial correlation coefficients. *Biometrika*, **43**, 169–185.

Diggle, P.J. (1990) *Time Series, A Biostatistical Introduction*. Oxford: Oxford University Press.

Drezner, Z. (1978) Computation of the bivariate normal integral. *Math. Comput*., **32**, 277–279.

Durbin, J. (1980a) Approximations for densities of sufficient statistics. *Biometrika*, **67**, 311–333.

Durbin, J. (1980b) The approximate distribution of serial correlation coefficients calculated from residuals on Fourier series. *Biometrika*, **67**, 335–350.

Durbin, J. and Watson, G.S. (1950) Testing for serial correlation in least squares regression I. *Biometrika*, **37**, 409–428.

Durbin, J. and Watson, G.S. (1951) Testing for serial correlation in least squares regression II. *Biometrika*, **38**, 159–178.

Geary, R.C. (1944) Extension of a theorem by Harald Cramér on the frequency distribution of the quotient of two variables. *J. Roy. Statist. Soc*., **17**, 56–57.

Henshaw, R.C. (1966) Testing single equation least squares regression models for autocorrelated disturbances. *Econometrica*, **34**, 646–660.

Hull, J.C. (1993) *Options, Futures, and Other Derivative Securities*. London: Prentice-Hall.

Imhof, J.P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 417–426.

Jensen, J.L. (1988) Uniform saddlepoint approximations. *Adv. Appl. Probab*., **20**, 622–634.

Lieberman, O. (1994) Saddlepoint approximation for the least squares estimator in first-order autoregression. *Biometrika*, **81**, 807–811.

Lugannani, R. and Rice, S.O. (1980) Saddlepoint approximations for the distribution of sum of independent random variables. *Adv. Appl. Probab.*, **12**, 475–490.

McGregor, J.R. (1960) An approximate test for serial correlation in polynomial regression. *Biometrika*, **47**, 111–119.

Phillips, P.C.B. (1978) Edgeworth and saddlepoint approximations in the first-order non-circular autoregression. *Biometrika*, **65**, 91–98.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1989) *Numerical Recipes in Pascal: The Art of Scientific Computing*. Cambridge: Cambridge University Press.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*. New York: Academic Press.

Skovgaard, I.M. (1987) Saddlepoint expansions for conditional distributions. *J. Appl. Probab.*, **24**, 275–287.

Temme, N.M. (1982) The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.*, **13**, 239–253.

Vinod, H.D. (1973) Generalization of the Durbin–Watson statistic for higher order autoregressive processes. *Commun. Statist. A*, **2**, 115–144.

Wang, S. (1992) Tail probability approximations in the first-order non-circular autoregression. *Biometrika*, **79**, 431–434.

Received September 1996 and revised April 1997