Approximate Distributions for the Various Serial Correlograms

Abstract

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 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 of Daniels (1956) and Durbin (1980) in the circular setting are shown. 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 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.

Accordingly the following practical recommendations can be made. Routine correlogram significance should be judged using the beta approximations of Daniels (1956) and Durbin (1980) which are considerably more accurate than the asymptotic normal bands \( \pm 2/\sqrt{n} \), even for series of length \( n = 70 \). 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.

Keywords: SADDLEPOINT APPROXIMATION, SERIAL CORRELOGRAM, SERIAL CORRELATION

Ronald W. Butler
Department of Statistics
Colorado State University
Fort Collins, CO 80525

Marc S. Paolella
Institut für Statistik und Ökonometrie
der Christian-Albrechts-Universität zu Kiel
D-24098 Kiel, Germany
1. Introduction

We consider the joint saddlepoint density approximation for random vector \( r = (r_1, ..., r_m)^T \) having the form

\[
r_k = \frac{\varepsilon^T A_k \varepsilon}{\varepsilon^T \varepsilon} \quad (k = 1, ..., m)
\]

where \( \varepsilon \) is an \( n \)-vector whose components have a multivariate normal density with mean 0 and covariance \( \Sigma^{-1} > 0, N_n(0, \Sigma^{-1}) \). 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 \( m \) computed from the least squares residuals \( \{z_t : t = 1, ..., n\} \) of a linear regression. In this case the variables are the serial correlations

\[
r_k = \frac{\sum_{t=k+1}^{n} z_t z_{t-k}}{\sum_{t=1}^{n} z_t^2} \quad (k = 1, ..., m).
\]

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's (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, three different approximations have been given by Phillips(1978), Wang(1992), and Lieberman(1994).

Approximations in the non-circular setting with \( m > 1 \) have been given in Durbin(1980b) who considered some specially-defined serial correlations that result in simultaneous diagonalizability for the matrices \( \{ A_k : k = 1, \ldots, m \} \). 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 in 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 \( \{ A_k : k = 1, \ldots, m \} \) are simultaneously diagonalizable. This assumption essentially assures that the likelihood is a regular exponential family admitting sufficient statistics so that his sufficiency approach (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 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 both the circular and non-circular tests are in close agreement. Confirmation of the accuracy of the Daniels and 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 \( r_1, \ldots, r_{m-1} \).

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, only allow for the removal of Fourier independent variables since such residuals retain the requisite mathematical tractable 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.

Two different saddlepoint approximations are given for \( r \) 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 & Butler, 1990), however a cut does not occur here. We consider this further in Section 4.

The remainder of the paper is organized as follows. Special models that have simpler approximations in the null setting are given in Section 6. In Section 7, the single saddlepoint density approximation, when restricted to a circular setting and after some approximation, is shown to reduce to the saddlepoint density expressions in Daniels(1956) and Durbin(1980b). 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 \( f_r(r) \) 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 only applies 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 non-null settings.

Let \( r = N/D \) where \( N = (N_1, \ldots, N_m)^T \), \( N_k = \epsilon_i A_k \epsilon \), and \( D = \epsilon^T \epsilon \). 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 in Barndorff-Nielsen & Cox(1979) uses the joint cumulant generating function of \((N, D)\) which is easily computed as

\[
K(s, t) = -\frac{1}{2} \log \left|(1 - 2t) I_n - 2 \sum_{i=1}^{m} s_i A_i \right| = -\frac{1}{2} \log |Q| \quad (3)
\]

where \( s = (s_1, \ldots, s_m) \) and \( K(s, t) \) is defined over the largest neighborhood of \( 0 \in \mathbb{R}^{m+1} \) 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

\[
\tilde{f}(r) = (2\pi)^{-\frac{q}{2}} \left| \frac{\partial^2 K(0, \tilde{t}_0)/\partial \tilde{t}_0}{K''(\tilde{s}, \tilde{t})} \right|^{-\frac{1}{2}} \exp \left[ \{t_0 - K(0, \tilde{t}_0)\} - \{\tilde{s}^T r + \tilde{t} - K(\tilde{s}, \tilde{t})\} \right] \quad (4)
\]

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

\[
\begin{align*}
    r &= \partial K(\tilde{s}, \tilde{t})/\partial \tilde{s} \\
    1 &= \partial K(\tilde{s}, \tilde{t})/\partial \tilde{t}
\end{align*} \quad (5)
\]

\( \tilde{t}_0 \) 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

\[
\begin{align*}
    r_t &= \text{tr} \tilde{Q}^{-1} A_i \quad (i = 1, \ldots, m) \\
    1 &= \text{tr} \tilde{Q}^{-1}
\end{align*} \quad (6)
\]

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

\[
\begin{align*}
    \tilde{K}_{s_i}'' &= (\partial^2 K(\tilde{s}, \tilde{t})/\partial \tilde{s}_i \partial \tilde{s}_j) = \left( 2 \text{tr} \tilde{Q}^{-1} A_i A_{j-1} \right) \quad (i, j = 1, \ldots, m) \\
    \tilde{K}_{s_t}'' &= (\partial^2 K(\tilde{s}, \tilde{t})/\partial \tilde{s}_i \partial \tilde{t}) \quad = \left( 2 \text{tr} \tilde{Q}^{-1} A_i A_{j-1} \right) \quad (i = 1, \ldots, m) \\
    \tilde{K}_{t_t}'' &= (\partial^2 K(\tilde{s}, \tilde{t})/\partial \tilde{t}^2) \quad = 2 \text{tr} \tilde{Q}^{-1} A_{j-1}
\end{align*}
\]

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

A simplification occurs in (4) by way of (6). The saddlepoints have the following relationship:

\[
(1 - n)/2 = \tilde{t}_0 = \tilde{t} + \tilde{s}^T r \quad (7)
\]
so (4) becomes

\[
\tilde{f}(r) = (2\pi)^{-\frac{m}{2}} 2^{-\frac{3}{2}m} n^{\frac{m+1}{2}} \left| K''(\tilde{s}, \tilde{t}) \right|^{\frac{3}{2}} \exp \left\{ -K(\tilde{s}, \tilde{t}) \right\}.
\]

(8)

We show (7) in the Appendix.

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, sec.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 \( W = (W_1, ..., W_m)^T \) is a random vector with moment generating function

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

(9)

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

\[
f_r(r) = (2\pi)^{-m} \prod \int \frac{\partial^m}{\partial t^m} M_{N,D}(s, t) \bigg|_{t=-r^T s} ds
\]

(10)

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

\[
K_\Omega(s, t) = -\frac{1}{2} \log \left| \Omega - 2t I_n - 2 \sum_{i=1}^m s_i A_i \right|.
\]

The structure of the \( m^{th} \) derivative of \( M_{N,D} \) in the general non-null setting is shown in the Appendix to have the form

\[
\frac{\partial^m}{\partial t^m} M_{N,D}(s, t) \bigg|_{t=-r^T s} = |\Omega|^{\frac{1}{2}} |P_\Omega(s)|^{-\frac{1}{2}} (\text{tr} P_\Omega^{-1}(s))^m \left(1 + \sum_j c_j p_j \right)
\]

(11)

where

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

is the matrix \( K_\Omega(s, t) \) evaluated with \( t = -r^T s \). In the null situation where \( \Omega = I_n \) we drop the \( \Omega \) subscript so \( P_n(s) = P(s) \), etc. The \( \{p_j\} \) are products of terms involving the
\( \text{tr}(P_{\Omega}^{-k})/(\text{tr} P_{\Omega}^{-1})^k : k = 2, ..., m \). For the moment we shall presume \( \{p_j\} \) are small and ignore them and then return to discuss this in Section 10. The dominant term in the integrand of (10) from (11) is the log-convex function \( |P_{\Omega}(s)|^{-\frac{1}{2}} \) which we use to determine a saddlepoint through which the integral paths in (10) are deformed. The single saddlepoint solves

\[
0 = -\frac{1}{2} \frac{\partial}{\partial \tilde{s}_i} \log |P_{\Omega}(\tilde{s})| = \text{tr} \hat{P}_{\Omega}^{-1}(A_i - r_i I_n) \quad (i = 1, ..., m) \tag{12}
\]

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

\[
\hat{h}_{ij} = -\frac{1}{2} \frac{\partial^2}{\partial \tilde{s}_i \partial \tilde{s}_j} \log (|P_{\Omega}(\tilde{s})|) = 2 \text{ tr} \hat{P}_{\Omega}^{-1}(A_i - r_i I_n) \hat{P}_{\Omega}^{-1}(A_j - r_j I_n) \quad (i, j = 1, ..., m). \tag{13}
\]

A single saddlepoint density approximation is therefore

\[
\hat{f}(r) = (2\pi)^{-\frac{m}{2}} |\Omega|^{\frac{1}{2}} \left| \hat{H}_{\Omega} \right|^{-\frac{1}{2}} \left| \hat{P}_{\Omega} \right|^{-\frac{1}{2}} (\text{tr} \hat{P}_{\Omega}^{-1})^m. \tag{14}
\]

The last term in (14) satisfies the constraint \( \text{tr} \hat{P}_{\Omega}^{-1}\Omega = n \), an identity derived by multiplying equation \( i \) of (12) by \(-2\tilde{s}_i\) and summing over \( i \) to get

\[
0 + \text{tr} \hat{P}_{\Omega}^{-1}\Omega = \text{tr} \hat{P}_{\Omega}^{-1} \left\{ \left( \sum_{i=1}^{m} -2\tilde{s}_i A_i + 2r^T \tilde{s} I_m \right) + \Omega \right\} = \text{tr} \hat{P}_{\Omega}^{-1} \hat{P}_{\Omega} = n.
\]

In the null setting, \( \text{tr} \hat{P}_{\Omega}^{-1} = n \) and the last factor of (14) is \( n^m \).

4. Equivalence in the Null Setting

The single saddlepoint approximation in (14) when \( \Omega = I_n \) is now shown to be analytically the same as the double saddlepoint approximation in (8). We first show that the double saddlepoint \((\tilde{s}, \tilde{r})\) can be expressed in terms of the single saddlepoint \( \tilde{s} \) through the relation \( \tilde{s} = n\tilde{s} \). Using (7) we express \( \tilde{r} \) in terms of \( \tilde{s} \) and substitute this in the value of \( \tilde{Q} \) in (72) so that

\[
0 = n^{-1} \text{tr} \left\{ (1 + 2r^T \tilde{s}/n) I_n - 2 \sum_{i=1}^{m} (\tilde{s}_i/n) A_i \right\}^{-1} (A_i - r_i I_n) \quad (i = 1, ..., m).
\]

This is the same equation as (12) and, as a saddlepoint equation, yields a unique root so \( \tilde{s} = n\tilde{s} \). This equivalence means that \( \tilde{Q} = n\hat{P} \) which further gives

\[
n^{\tilde{s}} e^{-K(\tilde{s}, \tilde{r})} = \left| \hat{P} \right|^{-\frac{1}{2}}, \tag{15}
\]
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

\[ |K''(\tilde{s}, \tilde{r})| = 2n^{-2m-1} |\tilde{f}|. \]  \hspace{1cm} (16)

The argument involves computing \( \delta \tilde{s}/\delta r^T \) and \( n\delta \tilde{s}/\delta r^T \) and equating their determinants and is given in the Appendix. With (15) and (16) it is simple algebra to now 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.50-1) as described in Booth & Butler(1990). The simplest case with \( m = 1 \) helps to clarify this point. Suppose the eigenvectors of \( A_1 \) are \( \lambda_1, \ldots, \lambda_n \) so

\[ r_1 = \frac{\sum_{i=1}^{n} \lambda_i \varepsilon_i^2}{\sum_{i=1}^{n} \varepsilon_i^2} \]

puts Dirichlet(\( \frac{1}{n}, \ldots, \frac{1}{n} \)) weights on the eigenvalues and has form comparable to that of the Dirichlet bootstrap in Booth & 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 single saddlepoint Lugannani & Rice(1980) approximations, are the same. The latter approximation is computed by writing

\[ P(r_1 \leq r) = P\left( \sum_{i=1}^{n} (\lambda_i - r) \varepsilon_i^2 \leq 0 \right) \]

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 \( \{r_k : k = 1, \ldots, m \} \) are serial correlations from least squares residuals as in (2) and \( r_k = z^T B_k z / z^T z \) where \( B_k = (b_{ijk}) \) is the band matrix of the \( k^{th} \) off-diagonals for which \( b_{ijk} = 2^{-1} \{i - j = k\} \), where \( 1\{\cdot\} \) denotes the indicator function. We canonically reduce the problem and show how the methods of the previous sections can be applied.
Suppose regression $E(y) = X\beta$ leads to residuals $z = My = (I_n - X(X^T X)^{-1}X^T)y$, so we write

$$r_k = \frac{y^T MBkMy}{y^T My} = \frac{y^T L (L^T ML) L^T Bk L (L^T ML) L^T y}{y^T L (L^T ML) L^T y}$$

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

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

(17)

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

The matrix $L_1$ is not uniquely determined and can be any of a compact collection of matrices whose columns form an orthonormal basis for the residual space. Certainly the true distribution of $r$ does not depend on this choice of $L_1$, however the saddlepoint density approximation does when applied in this context through the values of $\{A_k\}$ and also through $\Omega$ in the non-null setting.

A very natural choice of $L_1$ is that portion of the Helmert transformation which would make the $(n - p)$ components of $\epsilon = L_1^T y$ the recursive residuals as in Brown et.al. (1975). Then the serial correlations based on least squares residuals have distribution theory related to the recursive residuals through the relation $r_k = \epsilon^T L_1^T B_k L_1 \epsilon / \epsilon^T \epsilon$ where $\epsilon$ is the $(n-p)$ recursive residual vector and $B_k$ is an $(n \times n)$ band matrix with lag $k$. Note that this usage of such least squares residuals is a quite different approach than basing serial correlations directly upon the recursive residuals $\epsilon$. This latter approach would take $r_k = \epsilon^T B_k \epsilon / \epsilon^T \epsilon$ where $B_k$ is instead an $(n-p) \times (n-p)$ band matrix of lag $k$.

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

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

(19)
then the support is \( \mathcal{S} = \{ r : |R_i| > 0 \quad i = 1, \ldots, m \} \). In the correlogram context, the mapping \( \hat{s} \leftrightarrow r \) through the saddlepoint equation (12) is a bijection from \( \hat{s} \in \mathbb{R}^m \) onto \( r \in \mathcal{S} \) as we now indicate. It suffices to consider the null setting since \( \mathcal{S} \) is not dependent on the value of \( \Omega \).

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

\[
\eta = \sum_{i=1}^n (\hat{z} - \nu_i)^{-1} \quad (\hat{z} > \nu_n).
\]  

(20)

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

\[
r_i = \text{tr} \left( \hat{P}^{-1} A_i \right) / \text{tr} \hat{P}^{-1} \quad (i = 1, \ldots, m)
\]

(21)

through rearrangement of (12). Substituting the decomposition of positive definite \( \hat{P}^{-1} = \sum_{i=1}^n (\hat{z} - \nu_i)^{-1} \sigma_i^T \sigma_i \) into (21) gives

\[
\begin{pmatrix}
    r_1 \\
    \vdots \\
    r_m
\end{pmatrix} = \sum_{i=1}^n \left( \frac{(\hat{z} - \nu_i)^{-1}}{\sum_{j} (\hat{z} - \nu_j)^{-1}} \right) \begin{pmatrix}
    \sigma_i^T A_1 \sigma_i \\
    \vdots \\
    \sigma_i^T A_m \sigma_i
\end{pmatrix}
\]

(22)

where \( (\sigma_1^T A_1 \sigma_i, \ldots, \sigma_m^T A_m \sigma_i)^T \in \mathcal{S} \) for each \( i \). Thus vector \( r \) is a convex combination of vectors in \( \mathcal{S} \) so \( r \in \mathcal{S} \) by its convexity. If the support of the distribution of \( r \) were not convex, then the bijection \( \hat{s} \leftrightarrow r \) would map between \( \mathbb{R}^m \) and the interior of the convex hull of the support of \( r \). Such would be the situation, for example, if \( r \) instead consisted 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 (14) of \( r_1 \) in (1) has the simple null form

\[
f(r) = (2\pi)^{-\frac{1}{2}} n \hat{h}^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \hat{c}^2 \right)
\]

(23)
with
\[ \hat{h} = 2 \sum_{i=1}^{n} w_i^2 \left(1 - 2\hat{s}w_i\right)^{-2}, \quad \hat{\zeta} = \text{sgn}(\hat{s}) \sqrt{\sum_{i=1}^{n} \log(1 - 2\hat{s}w_i)}, \]

where \text{sgn} is the sign function, \( w_i = \lambda_i - r \), \( \{\lambda_i\} \) are the eigenvalues of \( A_1 \), and \( \hat{s} \) solves
\[ 0 = \sum_{i=1}^{n} \left(1 - 2\hat{s}w_i\right)^{-1} w_i. \quad (24) \]

The Lugannani & Rice (1980) approximation for the cumulative distribution of \( r_1 \) is
\[ \text{pr}(r_1 \leq r) \simeq \Phi\left(\hat{\zeta}\right) + \phi\left(\hat{\zeta}\right) \left\{ \left(\hat{\zeta}^{-1} - \left(\hat{s}\hat{h}^{1/2}\right)^{-1}\right)^{-1}\right\} \quad (r \neq E(r_1)) \quad (25) \]

where \( \Phi \) and \( \phi \) are the standard normal distribution and density functions respectively. Expressions (23), (24), and (25) agree with those in Lieberman (1994, eqns. (1) & (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 are discussed in Anderson (1971, sec. 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 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 one treated by Durbin (1980b) and defined so \( r_1 \) is the first component of the non-circular correlogram but such that the vector itself is not the correlogram.

Suppose \( O^T A_k O = \text{diag}(\lambda_k) \) where \( \lambda_k^T = (\lambda_{1k}, ..., \lambda_{nk}) \) consists of the eigenvalues of \( A_k \) for each \( k \). Let \( \chi = (\chi_1, ..., \chi_n)^T \) consist of independent and identically distributed \( \chi_i^2 \) variables. Place the \( \lambda_k^T \) in the rows of \( (m \times n) \) matrix \( \Lambda \) so
\[ \Lambda = (\lambda_1, ..., \lambda_m)^T = (\ell_1, ..., \ell_n). \]

The vector ratio is now \( r = \Lambda \chi / 1^T \chi \). In this setting
\[ \log \left| \hat{P} \right| = \sum_{i=1}^{n} \log \left\{ 1 - 2(\ell_i - r)^T \hat{s} \right\} \quad (26) \]

and the saddlepoint \( \hat{s} \) solves
\[ 0 = \sum_{i=1}^{n} \left\{ 1 - 2(\ell_i - r)^T \hat{s} \right\}^{-1} (\ell_i - r). \quad (27) \]
The Hessian is
\[ \hat{H} = 2 \sum_{i=1}^{n} \left(1 - 2 (\ell_i - r)^T \hat{\delta}\right)^{-2} (\ell_i - r) (\ell_i - r)^T. \]  
(28)

Suppose further that the numerators of \( r \) above add up disjoint subsets of the \( \chi^2 \) variables so
\[ \Lambda = \begin{pmatrix}
1_{\alpha_1}^T & 0^T & \ldots & \ldots & 0^T \\
0^T & 1_{\alpha_2}^T & 0^T & \vdots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0^T & 0^T & \ldots & 1_{\alpha_m}^T & 0^T
\end{pmatrix}. \]  
(29)

In this case the true density of \( r \) is Dirichlet(\( \alpha_1/2, \ldots, \alpha_m/2; \alpha_{m+1}/2 \)) where \( \alpha_{m+1} = n - \sum_{i=1}^{m} \alpha_i \). As shown in the Appendix, \( \hat{f}(r) \) differs from \( f(r) \) by Stirling's approximation in the Gamma functions, i.e.
\[ \hat{f}(r) = \left\{ \hat{\Gamma}(n/2)/\Gamma(n/2) \right\} \left\{ \prod_{i=1}^{m+1} \Gamma(\alpha_i/2)/\hat{\Gamma}(\alpha_i/2) \right\} f(r). \]  
(30)

Furthermore, solution to the saddlepoint equation in (27) is explicit with
\[ \hat{s}_i = (2n)^{-1} \left( \alpha_{m+1}/r_{m+1} - \alpha_i/r_i \right) \quad (i = 1, \ldots, m) \]  
(31)

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

7. Circular Parametric Likelihood Connections

Saddlepoint density (14) encompasses the density of the first \( m \) circularly defined serial correlations considered in Daniels(1956). In such context we show how (14) can be approximated in the null setting to give the same analytical results of Daniels for the density of the first \( m \) circularly defined serial correlations and, when transformed, for the partial serial correlations. A lag \( m \) autoregressive model is used to analytically approximate the two determinant factors in the saddlepoint density of (14) which leads to the simplification. These methods are equivalent to the smoothing methods of Dixon(1944) that are discussed in Daniels.

Similar arguments show how (14) relates to approximations in Durbin(1980b) for the density of circularly-defined serial and partial serial correlations constructed from residuals that remove effects of Fourier regressors. Removal of the sample mean \( \bar{y} \) from \( \{y_t\} \) is a special case of this since the vector of ones is always a eigenvector for the collection of circulant matrices involved.
7.1 AR(m) with known mean

Consider first a circularly defined AR(m) model with known mean 0 given as

\[ y_t = \alpha_1 y_{t-1} + \cdots + \alpha_m y_{t-m} + \eta_t \quad (t = 1, \ldots, n) \quad \text{(32)} \]

where \( y_t = y_{t+n} \) for all \( t \). From Daniels, the density of \( y = (y_1, \ldots, y_n)^T \) when \( \{\eta_t\} \) consists of independent and identically distributed N(0,1) variables is

\[ f(y) = (2\pi)^{-\frac{n}{2}} |\Upsilon(\alpha)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} y^T \Upsilon(\alpha) y \right\} \quad \text{(33)} \]

where \( |\Upsilon(\alpha)|^{\frac{1}{2}} = \prod_{i=1}^{m} (1 - \theta_i^2) \), \( \{\delta_i\} \) are the roots of \( \theta^m + \alpha_1 \theta^{m-1} + \cdots + \alpha_m = 0 \), and \( \Upsilon(\alpha) \) is the \( n \times n \) circulant matrix given by

\[ \Upsilon(\alpha) = \begin{pmatrix} \delta_0 & \delta_1 & \delta_2 & \cdots & \delta_m & 0 & \cdots & 0 & \delta_m & \delta_{m-1} & \cdots & \delta_1 \\ \delta_1 & \delta_0 & \delta_1 & \cdots & \delta_{m-1} & \delta_m & 0 & \cdots & 0 & \delta_m & \delta_2 & \cdots & \delta_2 \\ \vdots & \cdots & \delta_1 & \cdots & \delta_m & \delta_m & \delta_m & \cdots & \delta_m & \cdots & \delta_m & \cdots & \delta_2 \\ \end{pmatrix} \quad \text{(34)} \]

where \( \delta_0 \) and \( \delta = (\delta_1, \ldots, \delta_m)^T \) are the inverse autocovariances defined as

\[
\begin{align*}
\delta_0 & = 1 + \alpha_1^2 + \cdots + \alpha_m^2 \\
\delta_1 & = -\alpha_1 + \alpha_1 \alpha_2 + \cdots + \alpha_{m-1} \alpha_m \\
\delta_2 & = -\alpha_2 + \alpha_1 \alpha_2 + \cdots + \alpha_{m-2} \alpha_m \\
& \vdots \\
\delta_{m-1} & = -\alpha_{m-1} + \alpha_1 \alpha_m \\
\delta_m & = -\alpha_m
\end{align*}
\quad \text{(35)}
\]

We can write \( \Upsilon(\alpha) = \delta_0 I_n + 2 \sum_{i=1}^{m} \delta_i C_i \) where \( C_i \) is the matrix for lag \( i \) circularly defined serial correlation in that \( r_i = y^T C_i y / y^T y \) as in Anderson(1971, sec. 6.5.2).

Analytic approximation to (14) is based on the common circular structure shared by the matrices \( \hat{P} = (1 + 2r^T s) I_n - 2 \sum_{i=1}^{m} s_i C_i \) and \( \Upsilon(\alpha) \). With \( r \) held fixed, consider the parameter transformations \( s \leftrightarrow \delta \leftrightarrow \alpha \) by equating

\[ \frac{s}{1 + 2r s} = -\frac{\delta}{\delta_0} \quad \text{(36)} \]
Then the two matrices are related by

\[
\Upsilon (\alpha) = \delta_0 (1 + 2r^T s)^{-1} \left\{ (1 + 2r^T s) I_n - 2 \sum_{i=1}^m s_i C_i \right\} = \delta_0 (1 + 2r^T s)^{-1} P(s). \tag{37}
\]

The scale factor on the right side of (37) can be computed by left multiplying the expression in (36) by \( r^T \) and solving to get

\[
\delta_0 (1 + 2r^T s)^{-1} = \delta_0 + 2r^T \delta = Q(\alpha, r) \tag{38}
\]

where the notation \( Q(\alpha, r) \) is that used in Daniels and is another way of writing the exponent \( y^T \Upsilon (\alpha) y \). Taking logarithms of the determinants in (37) then

\[
0 \simeq \log |\Upsilon (\alpha)| = n \log Q(\alpha, r) + \log |P(s)| \tag{39}
\]

where the approximation to 0 was also used in Daniels. Now a critical value of (39) with \( s \) as a function of \( \alpha \) occurs when \( Q(\alpha, r) \) attains its unique minimum and \( \log |P(s)| \) attains its unique maximum so that

\[
0 = \frac{\partial Q(\hat{\alpha}, r)}{\partial \hat{\alpha}} \Leftrightarrow 0 = \frac{\partial \log |P(\hat{s})|}{\partial \hat{s}^T} \frac{\partial \hat{s}}{\partial \hat{\alpha}^T}. \tag{40}
\]

The mapping \( s \leftrightarrow \alpha \) is non-singular so the \( \hat{s} \) maximizing \( \log |P(s)| \) is the coordinate change of the \( \hat{\alpha} \) minimizing \( Q \), where the latter is shown in Daniels to be the solution to the Yule-Walker equation \( R_{m-1} \hat{\alpha} = r \) with \( R_{m-1} \) given in (19). Then

\[
Q(\hat{\alpha}, r) = \hat{\delta}_0 + 2r^T \hat{\delta} = |R_m| / |R_{m-1}| = (1 - r^T R_{m-1}^{-1} r)^{-1} \tag{41}
\]

as shown by Daniels. The saddlepoint maximizes \( \log |P(s)| \) so \( \hat{s} \) relates to the solution to the Yule-Walker equations \( \hat{\alpha} \) via

\[
\frac{\hat{s}}{(1 + 2r^T \hat{s})} = \frac{-\hat{\delta}}{\hat{\delta}_0} \quad \text{or} \quad \hat{s} = -\hat{\delta} (1 - r^T R_{m-1}^{-1} r)^{-1} \tag{42}
\]

where the inverse autocovariances \( \hat{\delta} \) are based upon the Yule-Walker estimate \( \hat{\alpha} \). Using (39) then

\[
\left| \hat{P} \right|^{-\frac{1}{2}} \simeq Q(\hat{\alpha}, r)^{\frac{n}{2}} = (1 - r^T R_{m-1}^{-1} r)^{\frac{n}{2}} \tag{43}
\]

is the main contribution to the saddlepoint density.
Further differentiation of (39) is needed to approximate the Hessian adjustment in (13). Since 
\[ \log f(y) = \text{const} - \log |Y(\alpha)| - \frac{1}{2} Q(\alpha, r) \simeq \text{const} - \frac{1}{2} Q(\alpha, r), \]
second derivatives of \( Q \) are likelihood information which we now relate to the Hessian in (14). Second derivatives of (39) are

\[ 0 \simeq -n\hat{Q}^{-2}\partial\hat{Q}/\partial\hat{s}_i \partial\hat{Q}/\partial\hat{s}_j + n\hat{Q}^{-1}\partial^2\hat{Q}/\partial\hat{s}_i \partial\hat{s}_j + \theta^2 \log |P(\hat{s})| / \partial\hat{s}_i \partial\hat{s}_j \]

\[ = 0 + 2n\hat{Q}^{-1}\gamma_{ij}(\hat{s}) - 2\hat{h}_{ij} \quad (i, j = 1, \ldots, m) \quad (44) \]

where \( \{\gamma_{ij}(\hat{s})\} \) comprise the observed information \( \iota(\hat{s}) \) about parameter \( s \). Thus \( \hat{H} = n\hat{Q}^{-1}\gamma(\hat{s}) \).

The information about \( s \) relates to the information about \( \alpha \) via

\[ |\iota(\hat{s})| = |\iota(\hat{\alpha})| \left\| \frac{\partial \hat{s}}{\partial \hat{\alpha}} \right\|^2 = |\iota(\hat{\alpha})| \left\| \frac{\partial \hat{s}}{\partial \hat{\alpha}} \right\|^2 = |\iota(\hat{\alpha})| \left\| \frac{\partial \hat{\alpha}}{\partial \alpha} \right\|^2 \hat{Q}^{2m} \quad (45) \]

from (36) and (42). In Daniels' notation \( \|\partial \hat{\alpha} / \partial \alpha\| = \|D\| = \|\hat{D}\| \) where \( \hat{D} = (\hat{d}_{ij}) \) with

\[ \hat{d}_{ij} = (-1)^i \{ i = j \} + \hat{\alpha}_{i+j}1 \{ i + j \leq m \} + \hat{\alpha}_{j-i}1 \{ i < j \} \quad (i, j = 1, \ldots, m) \]

so from (44) and (45)

\[ \left\| \hat{H} \right\|^{-\frac{1}{2}} \simeq (n\hat{Q})^{-\frac{1}{2}} \left\| \hat{D} \right\| \left| \iota(\hat{\alpha}) \right|^{-\frac{1}{2}}. \quad (46) \]

The observed information about \( \alpha \) is

\[ \iota(\alpha) \simeq \frac{1}{2} \frac{\partial^2 Q}{\partial \alpha \partial \alpha^T} = \frac{1}{2} (y^T C_{i-i} y) = \frac{1}{2} \left( \sum_{i=1}^{n} y_i y_{i+i-j} \right) \simeq \frac{1}{2} nR_{m-1}. \quad (47) \]

From (43), (46), and (47) we have

\[ \hat{f}(r) \simeq n^{-\frac{n}{2}} (1 - r^n R_{m-1}^{1-n}) \left\| \hat{D} \right\| \left| R_{m-1} \right|^{-\frac{1}{2}} \quad (48) \]

which agrees with Daniels' equation (9.14). The density for the partial serial correlations \( r = (r_1, \ldots, r_m)^T \) is therefore also the same as in Daniels (10.5) when corrected for the misprint and is

\[ \hat{f}(r.) \simeq n^{-\frac{n}{2}} \prod_{j \text{ odd}} \left\{ (1 - r_j^2) \frac{1}{2} (n-1) \right\} \prod_{j \text{ even}} \left\{ (1 - r_j^2) \frac{1}{2} n-1 \right\} (1 - r_j). \quad (49) \]

7.2 \textbf{AR}(m) with Fourier regressors

Again consider a circularly-defined \textbf{AR}(m) model with regression \( E(y) = X\beta \) where \( X \) is a known \( n \times p \) matrix and \( \beta \) is \( p \times 1 \) and unknown. Suppose furthermore that the columns of \( X \) consist
of eigenvectors of the circular matrices \( \{C_k : k = 1, \ldots, m\} \) which form the Fourier regressors of Durbin (1980b) and are given by Anderson (1971, p.280-1) in Theorem 6.5.2. The vector of ones is one such Fourier regressor. The density of \( y \) is then

\[
f(y) = (2\pi)^{-\frac{m}{2}} |\mathbf{T}_1(\alpha)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y - X\beta)^T \mathbf{T}_1(\alpha) (y - X\beta) \right\}
\]

where \( \mathbf{T}_1(\alpha) = \delta_0 I_n + 2 \sum_{i=1}^{m} \delta_i C_i \). Our arguments proceed along much the same lines as those of the previous subsection and so we shall only highlight the differences that arise in dealing with the regressors.

The first and major difference is that the treatment of serial correlations using least squares residuals is based on the distribution of the canonical reduction of these residuals which we now consider. Following the notation of Section 5, let \( L = (L_1 L_2) \) be an \( n \times n \) orthogonal matrix that canonically reduces \( M = I_n - X(X^TX)^{-1}X^T \) as in (18) with \( L_1 \) as \( (n - p) \times n \) and \( L_2 = X(X^TX)^{-\frac{1}{2}} \). The orthogonal transformation yields a density for \( L_1^T y = \varepsilon \) and \( L_2^T y = b \) as \( f(\varepsilon, b) = f(y) \). The exponent of the density simplifies since the columns of \( X \) are eigenvectors of \( \{C_k\} \) and hence also of \( \mathbf{T}_1 \) so that

\[
L^T \mathbf{T}_1(\alpha) L = \begin{pmatrix}
L_1^T \mathbf{T}_1(\alpha) L_1 & L_1^T X (X^TX)^{-\frac{1}{2}} E_p \\
E_p (X^TX)^{-\frac{1}{2}} X^T L_1 & E_p
\end{pmatrix} = \begin{pmatrix}
L_1^T \mathbf{T}_1(\alpha) L_1 & 0 \\
0 & E_p
\end{pmatrix}
\]

(50)

where \( E_p = \text{diag}(v_1, \ldots, v_p) \) consists of the eigenvalues of \( \mathbf{T}_1 \) associated with its eigenvectors in the columns of \( X \). The marginal density of \( \varepsilon \) obtained by integrating out \( db \) is

\[
f(\varepsilon) = (2\pi)^{-\frac{1}{2}(n-p)} |\mathbf{T}(\alpha)|^{\frac{1}{2}} |E_p|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \varepsilon^T \mathbf{T}(\alpha) \varepsilon \right\}
\]

where we define

\[
\mathbf{T}(\alpha) = L_1^T \mathbf{T}_1(\alpha) L_1 = \delta_0 I_{n-p} + 2 \sum_{i=1}^{m} \delta_i L_1^T C_i L_1 = \delta_0 I_{n-p} + 2 \sum_{i=1}^{m} \delta_i A_i
\]

and \( A_i = L_1^T C_i L_1 \). Recall that this form of \( \{A_i\} \) with \( n - p \) replacing \( n \) and \( \Omega^{-1} = I_{n-p} \) was the input in saddlepoint approximation based on least squares residuals of Section 5 as we have again in this situation. The matrix \( P(s) = \left( 1 + 2s^TX \right) I_{n-p} - 2 \sum_{i=1}^{m} s_i A_i \) with \( A_i = L_1^T C_i L_1 \) is matched with \( \mathbf{T}(\alpha) \) yielding the same parameter transformation as in (36) so \( \mathbf{T}(\alpha) = \)
\( \delta_0 \left( 1 + 2T \sigma^2 \right)^{-1} P(s) = Q(\alpha, r)P(s) \) where \( Q = y^T \gamma(\alpha) y \). As before we assume

\[
0 \simeq \log |Y_1(\alpha)| = \log |\gamma(\alpha)| + \log |E_p| = (n - p) \log Q(\alpha, r) + \log |P(s)| + \log |E_p| \tag{51}
\]

from (50). We ignore the less important final term in (51) when finding a critical value in \( \alpha \) which therefore minimizes \( Q \), so \( \hat{\alpha} \) is the Yule-Walker estimate, and the \( \hat{s} \) maximizing \( \log |P(s)| \) relates to \( \hat{\alpha} \) as given before in (42). Thus from (51)

\[
\left| P(\hat{s}) \right|^{-\frac{1}{2}} \simeq Q(\hat{\alpha}, r)^{\frac{1}{2}(n-p)} \left| \hat{E}_p \right|^{\frac{1}{2}}. \tag{52}
\]

According to Durbin(1980, p.340), in his expression for \( \hat{\nu}_k \), \( \left| \hat{E}_p \right| = (1 + \sum_{1}^{m} \hat{\alpha}_i)^{2p} \), but also

\[
1 + \sum_{1}^{m} \hat{\alpha}_i = \prod_{i=1}^{m} (1 - \tau_i) \tag{53}
\]

as shown on p.341-2.

Approximation to \( \left| \hat{H} \right| \) is exactly as before and follows from differentiation of (51) without the final term to give

\[
\left| \hat{H} \right|^{-\frac{1}{2}} \simeq \left\{ (n - p) \hat{Q} \right\}^{-\frac{1}{2}m} \left\| \hat{D} \right\| \left| \frac{1}{2} (n - p) R_{m-1} \right|^{-\frac{1}{2}}. \tag{54}
\]

The saddlepoint density of \( r \) is

\[
\hat{f}(r) \simeq \pi^{-\frac{1}{2}m} (1 - r^T R_{m-1}^{-1} r)^{\frac{1}{2}(n-p-m)} \left\| \hat{D} \right\| \left| R_{m-1} \right|^{-\frac{1}{2}} \prod_{i=1}^{m} (1 - \tau_i)^p, \tag{55}
\]

and the transformation to the density of \( r \), gives

\[
\hat{f}(r) \simeq \pi^{-\frac{1}{2}m} \prod_{j \ odd} \left\{ (1 - r_j^2)^{\frac{1}{2}(n-p-1)} (1 - r_j) \right\} \prod_{j \ even} \left\{ (1 - r_j^2)^{\frac{1}{2}(n-p)-1} (1 - r_j)^{p+1} \right\}. \tag{56}
\]

In the setting \( p = 1 \) where only the mean is removed, this agrees with Daniels' equation (11.5) when corrected for the misprint. In the general Fourier regression setting, it agrees with Durbin's equation (32).

8. The Conditional Distribution of \( r_m \) given \( r_1, ..., r_{m-1} \)

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 (32) has an exponential
family likelihood with canonical parameters as the inverse autocovariances \( \{ \delta_i \} \) in (35). Using standard theory as in Anderson (sec. 6.3.2), the UMP unbiased test of \( \delta_m = 0 \) vs. \( \delta_m \neq 0 \), or equivalently \( AR(m) \) vs. \( AR(m-1) \), rejects for small and large values of \( r_m \). P-values are therefore computed as

\[
\hat{p} = \begin{cases} 
\Pr (r_m > r_{m0} \mid r_{(m-1)} = r_{(m-1)0} ; \delta_m = 0) & (r_{m0} > 0) \\
\Pr (r_m < r_{m0} \mid r_{(m-1)} = r_{(m-1)0} ; \delta_m = 0) & (r_{m0} \leq 0)
\end{cases}
\] (57)

where \( r^T_0 = (r_{10}, ..., r_{m0}) = \left( r^T_{(m-1)0}, r_{m0} \right) \) is the observed value of \( r^T = \left( r^T_{(m-1)}, r_m \right) \). Such optimality is maintained with either mean correction or Fourier regressors added into the model (Anderson, sec. 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 \( AR(m) \) vs. \( AR(m-1) \).

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 in Barndoff-Nielsen & Cox (1979) is computed as the ratio of two single approximations

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

where \( \hat{H}_{m-1} \) and \( \hat{P}_{m-1} \) are the \( \hat{H}_1 \) and \( \hat{P}_1 \) values associated with the \( (m-1) \)-dimensional saddlepoint \( \delta_{(m-1)} \) of the denominator determined by \( r_{(m-1)0} \), \( H_m \) and \( P_m \) are the \( \hat{H}_1 \) and \( \hat{P}_1 \) values associated with the \( m \)-dimensional saddlepoint \( s^T = (s_1, ..., s_m) \) of the numerator determined by \( (r_{(m-1)0}, r_m) \), and explicit dependence on \( \Omega \) has been suppressed. For probability calculation, one dimensional numerical integration

\[
\Pr (r_m < r_{m0} \mid r_{(m-1)0}; \Omega) \simeq \left\{ \int_{pr(-1, r_{m0})} \hat{f}(r_m \mid r_{(m-1)0}) \, dr_m \right\} / \left\{ \int_{pr(-1, r_{m0})} \hat{f}(r_m \mid r_{(m-1)0}) \, dr_m \right\}
\] (58)

can be performed when \( \rho \), the conditional support of \( r_m \) given \( r_{(m-1)0} \) is identifiable. In the correlogram setting \( \rho = \{ r_m : |R_m| > 0 \} \) 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 \( (r^T_{(m-1)0}, r_m) \). Since

\[
0 < |R_m| = |R_{(m-1)0}| \left\{ 1 - \left( r^T_{(m-1)0}, r_m \right) R^{-1}_{(m-1)0} \left( r^T_{(m-1)0}, r_m \right)^T \right\}
\] (59)
then \( \rho = (a, b) \subset (-1, 1) \) where \( a \) and \( b \) are the roots to the quadratic term in \( r_m \) within the square brackets. Determination of \( \rho \) 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 distribution function as in Skovgaard (1987) is derived using the method of Temme (1982). Consider performing the integration in the numerator of (58) with the transformation

\[
  r_m \leftrightarrow w = \text{sgn}(s_m) \sqrt{\log \left( \frac{|P_m|}{|\hat{P}_{m-1}|} \right)}.
\]  

(60)

To see that \( w \) is well-defined, note that saddlepoint \( s \) in

\[
  P_m = P_m(s) = \Omega + \left\{ 2 \left( r^T_{(m-1)0}, r_m \right) s \right\} I_n - 2 \sum_{i=1}^m s_i A_i
\]

maximizes \( |P_m(\tau)| \) over \( \tau \in \mathbb{R}^m \) whereas \( \hat{P}_{m-1} \) is the maximum of \( |P_m(\tau_{(m-1)}, 0)| \) over \( \tau_{(m-1)} \in \mathbb{R}^{m-1} \); thus \( |P_m| \geq |\hat{P}_{m-1}| \) with equality only when \( s_m = 0 \). The numerator of (58) is now

\[
  \text{pr} \left( r_m \leq r_{m0} | r_{(m-1)0}; \Omega \right) \simeq \int_{-\infty}^{w_0} h(w) \phi(w) \, dw
\]  

(61)

with

\[
  h(w) = \left( \frac{|\hat{P}_{m-1}|}{|H_m|} \right)^{\frac{1}{2}} \left\{ \frac{\left( \text{tr} P_m^{-1} \right)^m}{\left( \text{tr} \hat{P}_{m-1}^{-1} \right)^{m-1}} \right\} \frac{\partial r_m}{\partial w}
\]  

(62)

and

\[
  w_0 = \text{sgn}(s_m) \sqrt{\log \left( \frac{|\hat{P}_m|}{|\hat{P}_{m-1}|} \right)}
\]  

(63)

where \( \hat{P}_m = P_m(s) \) and \( \hat{s}^T = (s_1, ..., s_m) \) is the saddlepoint for \( \left( r^T_{(m-1)0}, r_{m0} \right) \). The change of variable \( \partial r_m / \partial w = w / (s_m \text{ tr } P_m^{-1}) \) as shown in the Appendix. Using this in (62), 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 in Barndorff-Nielsen & Cox (1989, sec. 3.9) now applies and leads to

\[
  \text{pr} \left( r_m \leq r_{m0} | r_{(m-1)0}; \Omega \right) \simeq \Phi(w_0) + \phi(w_0) \left( \frac{1}{w_0} - \frac{1}{v_0} \right) \quad (s_m \neq 0)
\]  

(64)

where

\[
  v_0 = \hat{s}_m \left( \frac{|\hat{P}_m|}{|\hat{P}_{m-1}|} \right)^{\frac{1}{2}} \left\{ \frac{\left( \text{tr} \hat{P}_{m-1}^{-1} \right)^{m-1}}{\left( \text{tr} \hat{P}_m^{-1} \right)^{m-1}} \right\}
\]  

(65)

Numerical accuracy of (64) is considered in Section 11.
9. Other Correlograms

Densities for the partial correlogram and inverse correlogram are given below in the general non-null setting without the assumption of circularity. The necessary Jacobians have been given in Daniels. Conditional distribution function approximations as in the previous section can be developed for each of these correlograms but we do not give details.

Daniels shows in his equations (10.2) and (10.4) that

\[ \left\| \frac{\partial r}{\partial \delta^T} \right\| = |R_{m-1}| \times \left\| \delta \right\|^{-1}, \]  

(66)

and

\[ \left\| \frac{\partial \delta}{\partial r} \right\| = \prod_{j \text{ odd}} (1 - r_j^2)^{\frac{1}{2}(j-1)} \prod_{j \text{ even}} \left\{ (1 - r_j^2)^{\frac{1}{2}j-1} (1 - r_j) \right\}, \]

so that

\[ \hat{f}(r) = \hat{f}(r) \left\| \frac{\partial r}{\partial \delta^T} \right\| \times \left\| \frac{\partial \delta}{\partial r} \right\| \quad (r \in (-1,1)^m). \]  

(67)

If we denote the lag m inverse correlation vector as \( r = \delta / \delta_0 = (\delta_1, \ldots, \delta_m)^T / \delta_0 \) then its density is

\[ \hat{f}(r) = \hat{f}(r) |R_{m-1}| \times \left\| \delta \right\|^{-2} (1 + r^T R_{m-1}^{-2} r)^m |1 + 2r^T r|^{-1}, \]  

(68)

where the Jacobian 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 (14) depends on the set of matrices \( \{A_k\} \) and their large sample behavior. 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 \( \{q_k : k \geq 2\} \). At the saddlepoint, these values are

\[ 0 < q_k = \text{tr}(\hat{P}^{-k}) / (\text{tr} \hat{P}^{-1})^k = n^{-k} \text{tr}(\hat{P}^{-k}). \quad (k \geq 2) \]  

(69)
since the eigenvalues of \( \hat{P}^{-1} \) have a sample mean of \( n^{-1} \text{tr} \hat{P}^{-1} = 1 \). 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

\[
n^{-k} \text{tr} \hat{P}^{-k} = n^{-k+1} \times n^{-1} \text{tr} \hat{P}^{-k} = O(n^{-k+1}) \quad (k \geq 2). \tag{70}
\]

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

\[
\sum_{k=0}^{\infty} \hat{q}_k = \sum_{k=0}^{\infty} \text{tr} \left( \hat{P}^{-1}/n \right)^k = \text{tr} \left( I_n - n^{-1} \hat{P}^{-1} \right)^{-1} < \infty. \tag{71}
\]

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

11. Numerical Examples

Table 1 below compares the accuracy of various univariate approximations when the true model is \( y_t = \mu + \eta_t \) with \( \eta_t = \alpha_1 \eta_{t-1} + \vartheta_t \) for \( t = 1, \ldots, n \) and \( \{\vartheta_t\} \) as i.i.d. \( N(0,1) \); a non-circular AR(1) process requiring mean correction as described in Section 5. The table computes approximations of \( \Pr(r_1 \leq r_{10} | \alpha_1) \) for the various values of \( n \) and \( \alpha_1 \) in four settings including null (\( \alpha_1 = 0 \)), stationary (\( \alpha_1 = 0.5 \)), unit-root (\( \alpha_1 = 1 \)), and non-stationary (\( \alpha_1 = 1.5 \)). Column "L-R" is the Lugannani-Rice approximation in (25), "NID" is the numerically integrated single saddlepoint density in (23), Imhof(1961) is approximate numerical inversion, "Sim" is Monte Carlo simulation based on two million repetitions, "Beta" is Henshaw’s(1966) beta approximation matching third and fourth moments, and "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. In one instance it failed so the L-R procedure was used in its place. The Beta approximation sometimes fails because either of its degrees of freedom estimates is negative. The NID procedure used a Romberg integration from Press et.al.(1989) and "Sim" used the "ran1" and "ran2" generators also from Press.

Table 2 computes approximations of \( \Pr(r_1 \leq r_{10}, r_2 \leq r_{20} | \alpha_1, \alpha_2) \) 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 i.i.d. \( N(0, 1) \). Three stationary settings are considered including \( AR(0) \) with \( \alpha_1 = 0 = \alpha_2 \), \( AR(1) \) with \( \alpha_1 = 0.5 \) and \( \alpha_2 = 0 \), and \( AR(2) \) with \( \alpha_1 = 1.2 \) and \( \alpha_2 = -0.8 \). The simulations were again based on two million repetitions. Computations of NID used nested Romberg routines from Press. The Normal approximations are given in Priestley (1981 sec. 5.3.5) as

\[
\begin{pmatrix}
  r_1 - 0.5 \\
  r_2 - 0.5^2
\end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.9375 & 0.46875 \\ 0.46875 & 1.1133 \end{pmatrix}\right)
\]

for the \( AR(1) \), and

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

for the \( AR(2) \).

Table 3 gives p-values of the form (57) under the null setting (\( \Omega = I_n \)) in the progressive testing of \( AR(m) \) vs. \( AR(m - 1) \) 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. Column \( r \) gives the value \( r_{m0} \) and column \( r \), the lag-\( m \) partial serial correlation from the regression residuals. Interval (lo,hi) is the support of the conditional distribution of \( r_m \) given \( r_{(m-1)0} \). Columns NICD and SCDF are the approximations from the numerically integrated conditional saddlepoint density in (58) and the saddlepoint CDF approximation in (64).

Column D/D in Table 3 is the p-value approximation based on the Daniels/Durbin distribution theory in (56) with \( p = 5 \). The conditional test that rejects for large and small \( r_m \) given \( r_{(m-1)0} \) is analytically equivalent to the test rejecting for large and small \( r_m \) 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, \( r_m \) is approximately independent of \( r_{1, \ldots, r_{m-1}} \), so the p-value when \( r_{m,0} \leq 0 \) is determined from (56) as

\[
\hat{\beta} = \Pr \{ r_m \leq r_{m,0} \} = \begin{cases} 
  IB \left\{ (r_{m,0} + 1) / 2; (n - p + 1) / 2, (n + p + 1) / 2 \right\} & (m \text{ odd}) \\
  IB \left\{ (r_{m,0} + 1) / 2; (n - p) / 2, (n + p) / 2 + 1 \right\} & (m \text{ even})
\end{cases}
\]

where \( IB (b, \alpha, \beta) \) is the incomplete Beta(\( \alpha, \beta \)) 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 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 $\mathcal{C}$ may be difficult to identify and also not convex. The conditional CDF 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, sec. 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 (14) 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 $n$. Asymptotically this must result in degenerate categories corresponding to bounded or slowly growing values in $\{\alpha_i\}$. However, the accuracy in (30) 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.
13. Acknowledgments

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

14. Appendix

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 \( \{ A_i \} \) 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 \( \ell_{m+1} \) such that

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

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

Proof of (7): Take the first equation in (6) and subtract \( r_i \) times the second equation to show

\[
\text{tr} \, \tilde{Q}^{-1}(A_i - r_i I_n) = 0. \tag{72}
\]

Using this and (6),

\[
\tilde{t} = 0^T \tilde{s} + \tilde{t} = \sum_{i=1}^{m} \tilde{s}_i \text{tr} \, \tilde{Q}^{-1}(A_i - r_i I_n) + \tilde{t} \text{tr} \, \tilde{Q}^{-1} = \text{tr} \, \tilde{Q}^{-1} \left\{ -\frac{1}{2} \tilde{Q}^{-1} + \left( \frac{1}{2} - r^T \tilde{s} \right) I_n \right\}
\]

which reduces to (7).

Verification of (11): The first derivative of \( M_{N,D} \) is

\[
\frac{\partial M_{N,D}(s,t)}{\partial t} \bigg|_{t=-r^T s} = M_{N,D}(s) \text{tr} \, P_{\tilde{\Omega}}^{-1}(s) = |\Omega|^{\frac{1}{2}} |P_{\tilde{\Omega}}(s)|^{-\frac{1}{2}} \text{tr} \, P_{\tilde{\Omega}}^{-1}(s). \tag{73}
\]

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 (73) result in each term of the \( m^{th} \)
derivative having exponents of \( P_{\Omega}^{-1} \) adding to \( m \). For example, the third derivative is
\[
\partial^3 M_{N,D}(s,t) / \partial t^3 \mid _{t=\cdots-r_{s}} = \left| \Omega \right|^2 \left| P_{\Omega}(s) \right|^{-\frac{1}{2}} \left[ \left\{ \text{tr} P_{\Omega}^{-1}(s) \right\}^3 + 6 \text{tr} P_{\Omega}^{-1}(s) \text{tr} P_{\Omega}^{-2}(s) + 8 \text{tr} P_{\Omega}^{-3}(s) \right]
\]
and the powers of \( P_{\Omega}^{-1} \) in each term add to 3; thus after factoring out \( \left\{ \text{tr} P_{\Omega}^{-1}(s) \right\}^3 \) the latter factor is in terms of \( q_2 \) and \( q_3 \) only.

**Proof of (16):** We first compute \( |\partial \tilde{s} / \partial r^T| \). It is easiest to directly differentiate the general saddlepoint expressions in (5) to show that
\[
\begin{pmatrix}
\tilde{K}_{ss}' & \tilde{K}_{st}' \\
\tilde{K}_{ts}' & \tilde{K}_{tt}'
\end{pmatrix}
\begin{pmatrix}
\partial \tilde{s} / \partial r^T \\
\partial \hat{t} / \partial r^T
\end{pmatrix}
= \begin{pmatrix}
I_s & 0 \\
0 & 0
\end{pmatrix}
\]
so that very generally
\[
\frac{\partial \tilde{s}}{\partial r^T} = \left( \tilde{K}_{ss}' - \tilde{K}_{st}' \tilde{K}_{tt}'^{-1} \tilde{K}_{ts}' \right)^{-1}
\]
and
\[
\left| \frac{\partial \tilde{s}}{\partial r^T} \right| = \left| \tilde{K}_{ss}' - \tilde{K}_{st}' \tilde{K}_{tt}'^{-1} \tilde{K}_{ts}' \right|^{-1} = \frac{\tilde{K}_{tt}''}{\tilde{K}''(\tilde{s}, \tilde{t})}.
\]  
(74)

We now compute \( |\partial \tilde{s} / \partial r^T| \) by differentiating \( \partial / \partial \tilde{s}_j \) the saddlepoint equation in (12). For general \( \Omega \),
\[
0 = -\frac{\partial r_i}{\partial \tilde{s}_j} \text{tr} \hat{P}_{\Omega}^{-1} - 2 \text{tr} \hat{P}_{\Omega}^{-1} \left\{ \left( \frac{\partial r^T}{\partial \tilde{s}_j} \tilde{s} + r_j \right) I_n - A_j \right\} \hat{P}_{\Omega}^{-1} (A_i - r_i I_n) \quad (i, j = 1, \ldots, m).
\]
With index \( j \) held fixed, the \( m \) equations for \( i = 1, \ldots, m \) are a system of linear equations in \( \partial r_i / \partial \tilde{s}_j, \ldots, \partial r_m / \partial \tilde{s}_j \) which we can write in matrix form as
\[
\begin{pmatrix}
I_m \text{tr} \hat{P}_{\Omega}^{-1} + \tilde{w} \tilde{s}^T
\end{pmatrix}
\begin{pmatrix}
\partial r^T / \partial \tilde{s}_j
\end{pmatrix}
= \begin{pmatrix}
\hat{h}_{ij}, \ldots, \hat{h}_{mj}
\end{pmatrix}^T \quad (j = 1, \ldots, m). 
\]  
(75)
Combining all \( m \) equations in (75) into one gives the matrix relation
\[
\begin{pmatrix}
I_m \text{tr} \hat{P}_{\Omega}^{-1} + \tilde{w} \tilde{s}^T
\end{pmatrix}
\begin{pmatrix}
\partial r^T / \partial \tilde{s}^T
\end{pmatrix}
= \hat{H}_{\Omega}.
\]  
(76)
Taking determinants yields
\[
\left| \frac{\partial r}{\partial \tilde{s}^T} \right| = \left| \hat{H}_{\Omega} \right| \left( \text{tr} \hat{P}_{\Omega}^{-1} \right)^{-m} \tilde{u}^{-1}
\]  
(77)
where
\[
\tilde{u} = \left| 1 + \tilde{s}^T \tilde{u} / \text{tr} \hat{P}_{\Omega}^{-1} \right|, \quad \tilde{w} = 2 \left( \text{tr} \hat{P}_{\Omega}^{-2}(A_1 - r_1 I_n), \ldots, \text{tr} \hat{P}_{\Omega}^{-2}(A_m - r_m I_n) \right)^T.
\]
In the null setting this becomes

$$\left| \frac{\partial \hat{\sigma}}{\partial r^T} \right| = n^m \left| \hat{H} \right|^{-1} \left( 1 + \hat{u}^T \hat{\sigma} / n \right)$$

(78)

and the last factor reduces as follows:

$$1 + n^{-1} \hat{u}^T \hat{\sigma} = 1 + 2n^{-1} \sum_{i=1}^m \text{tr} \hat{P}^{-2}(A_i - r_i I_n) \hat{\sigma}_i$$

$$= 1 + 2n \text{tr} \hat{Q}^{-1} \left( \sum_{i=1}^m A_i \hat{\sigma}_i - r^T \hat{\sigma} I_n \right) \hat{Q}^{-1}$$

$$= 1 + n \text{tr} \hat{Q}^{-1} (I_n - n^{-1} \hat{Q}) \hat{Q}^{-1}$$

$$= n \hat{K}_{ii} / 2.$$ 

Substituting this into (78), equating it with $n^{-m} \left| \partial \hat{\sigma} / \partial r^T \right|$ where $\left| \partial \hat{\sigma} / \partial r^T \right|$ is from (74), proves (16).

**Proof of (30):** In the saddlepoint equation (27), $\alpha_i$ of the $\ell_i$ vectors are $\xi_i$, the indicator vector of component $i$, and $\alpha_{n+1}$ are 0 vectors. This leads to saddlepoint equation

$$\sum_{i=1}^m \alpha_i \xi_i / \hat{d}_i = r \left\{ \sum_{i=1}^m \alpha_i / \hat{d}_i + \alpha_{m+1} / (1 + 2r^T \hat{\sigma}) \right\}$$

(79)

where $\hat{d}_i = 1 - 2\hat{\sigma}_i + 2r^T \hat{\sigma}$. Summing all components gives

$$\sum_{i=1}^m \alpha_i / \hat{d}_i = \frac{(1 - r_{n+1}) \alpha_{m+1}}{r_{m+1} (1 + 2r^T \hat{\sigma})}.$$ 

(80)

Combining (79) and (80) shows $1 + 2r^T \hat{\sigma} = \alpha_{m+1} / (n r_{m+1})$ which, along with (79), gives $\hat{\sigma}_i$ in (31).

The only remaining part of the calculation that is not obvious is the computation of $\left| \hat{H} \right|$ with

$$\hat{H} = 2n^2 \left\{ \sum_{i=1}^{m+1} r_i^2 / \alpha_i (\xi_i - r) (\xi_i - r)^T \right\}$$

(81)

where $\xi_{m+1} = 0$. Note that (81) resembles a weighted least squares matrix sum of squares in MANOVA. We add and subtract $\bar{\xi} = (\sum_i r_i^2 \xi_i / \alpha_i) / (\sum_i r_i^2 / \alpha_i)$ inside each term of form $\xi_i - r$ and reduce the expression to

$$\hat{H} = 2n^2 \left\{ D + q (\bar{\xi} - r) (\bar{\xi} - r)^T - q \bar{\xi} \bar{\xi}^T \right\}$$

$$= 2n^2 \left\{ D + q (\bar{\xi} - r) (\bar{\xi} - r)^T \right\}^{-1} \times$$(82)

$$\left[ I - \left\{ D + q (\bar{\xi} - r) (\bar{\xi} - r)^T \right\}^{-1} q \bar{\xi} \bar{\xi}^T \right].$$
where \( D = \text{diag}(r^2_1/\alpha_1, \ldots, r^2_m/\alpha_m) \) and \( q = \sum_i r^2_i/\alpha_i \). Inversion and determinant formulas can be applied to (82) and after much computation

\[
|\hat{H}| = 2^m n^{2m+1} \prod_{i=1}^{m+1} r^2_i/\alpha_i.
\]

Now computation of \( \hat{f}(r) \) in (14) shows (30).

**Derivation of \( \partial r_m/\partial w \) in (62) and proof that \( h(0) = 1 \):**

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

\[
2w \partial w / \partial r_m = \text{tr} \left( P^{-1}_m \partial P_m / \partial r_m \right) = 2 \left( s_m + r^T \partial s/\partial r_m \right) \text{tr} P^{-1}_m - 2 \sum_{i=1}^{m+1} \partial s_i/\partial r_m \text{tr} P^{-1}_m A_i.
\]

(83)

where the last line follows from \( r_i = \text{tr} P^{-1}_m A_i/ \text{tr} P^{-1}_m \).

Substitute (83) into the expression for \( h(\cdot) \) in (62) and let \( w \to 0 \) so that \( s_m \to 0, s \to (\hat{s}_{m-1}^T, 0), \text{tr} P^{-1}_m \to \text{tr} \hat{P}^{-1}_{m-1} \), and

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

(84)

The final term \( \lim_{w \to 0} (w/s_m) = \lim_{w \to 0} (\partial w / \partial s_m) \) is computed by differentiating (83) \( \partial / \partial s_m \) to get

\[
w \frac{\partial^2 w}{\partial r^T_m \partial s_m} + \left( \frac{\partial w}{\partial s_m} \right)^2 \frac{\partial s_m}{\partial r_m} = \text{tr} P^{-1}_m + s_m \frac{\partial}{\partial s_m} \text{tr} P^{-1}_m.
\]

Taking limits then

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

The latter derivative results from rewriting (76) as

\[
\frac{\partial s}{\partial \hat{r}_m} = H^{-1}_m (I_m \text{tr} P^{-1}_m + ws^T)
\]

so

\[
\frac{\partial s_m}{\partial r_m} = h_{mm} \text{tr} P^{-1}_m + \xi^T_{m} H^{-1}_m w s_m.
\]

(85)

where \( h_{mm} \) is the \((m, m)^{th}\) element of \( H^{-1}_m \) 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

\[
\lim_{w \to 0} \frac{\partial w}{\partial s_m} = \left( |H_m (\hat{s}_{m-1}^T, 0)| / |\hat{H}_{m-1}| \right)^{1/2}.
\]
Combine this with (84) to show \( \lim_{w \to 0} h(w) = 1. \)

**Jacobian for (68):** The Jacobian \( \partial r / \partial \dot{\alpha}^T \) is computed in terms of \( \partial r / \partial \dot{\alpha}^T \times \partial \dot{\alpha} / \partial \dot{r}^T \) where the former quantity is given in (66). Since \( r_\dot{\alpha} = \dot{\delta} / \dot{\delta}_0 \), then the latter term can be computed by way of (35). From (35)

\[
\dot{D} = \frac{\delta \dot{\delta}}{\partial \dot{\alpha}^T} = \frac{\partial}{\partial \dot{\alpha}^T} \left( \delta_0 \dot{r}_{-} \right) = \delta_0 \frac{\partial r_{-}}{\partial \dot{\alpha}^T} + \dot{r}_{-} \frac{\delta_0}{\partial \dot{\alpha}^T}.
\]

Since \( \delta \dot{\delta}_0 / \partial \dot{\alpha}^T = 2 \dot{\alpha}^T \) then we can solve for

\[
\left\| \frac{\partial r_{-}}{\partial \dot{\alpha}^T} \right\| = \delta_0 \left\| \dot{D} - 2r_{-} \dot{\alpha}^T \right\|
= (1 + r^T \dot{R}^{-1}_{m-1} r)^{-m} \left\| \dot{D} \right\| \times \left\| 1 - 2 \dot{\alpha}^T \dot{D}^{-1} r_{-} \right\|
\]

since \( \dot{\alpha} = \dot{R}^{-1}_{m-1} \). But now \( \dot{r} = -\dot{D}^T r \) as given in Daniels equation (9.13) so

\[
\left\| \frac{\delta \dot{\delta}}{\partial r_{-}} \right\| = (1 + r^T \dot{R}^{-1}_{m-1} r)^m \left\| \dot{D} \right\|^{-1} \times \left\| 1 + 2 r^T r_{-} \right\|^{-1}
\]

and the density in (68) follows from the Jacobians in (66) and (86).

**REFERENCES**


Table 1. Approximations of $\Pr (r_1 \leq r_{10})$ for various non-circular AR(1) models requiring mean correction.

<table>
<thead>
<tr>
<th>n</th>
<th>$\alpha_1$</th>
<th>$r_{10}$</th>
<th>Imhof</th>
<th>Sim</th>
<th>Beta</th>
<th>L-R</th>
<th>NID</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0.3471</td>
<td>0.9500</td>
<td></td>
<td>0.9501</td>
<td>0.9501</td>
<td>0.9496</td>
<td>0.8638</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.6015</td>
<td>0.9500</td>
<td></td>
<td>0.9509</td>
<td>0.9491</td>
<td>0.9442</td>
<td>0.6445</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.7368</td>
<td>0.9500</td>
<td>fail</td>
<td>0.9475</td>
<td>0.9448</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.62495</td>
<td>0.9500</td>
<td>0.9500</td>
<td>0.9340</td>
<td>0.9325</td>
<td>0.9187</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>0.2711</td>
<td>0.9500</td>
<td></td>
<td>0.9500</td>
<td>0.9501</td>
<td>0.9501</td>
<td>0.9124</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.6476</td>
<td>0.9500</td>
<td></td>
<td>0.9503</td>
<td>0.9509</td>
<td>0.9501</td>
<td>0.8029</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.89614</td>
<td>0.9500</td>
<td>fail</td>
<td>0.9484</td>
<td>0.9442</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.656854</td>
<td>fail</td>
<td>0.9574</td>
<td>0.9462</td>
<td>0.9500</td>
<td>fail</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0</td>
<td>0.17847</td>
<td>0.9500</td>
<td></td>
<td>0.9500</td>
<td>0.9500</td>
<td>0.9500</td>
<td>0.9323</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.61970</td>
<td>0.9500</td>
<td></td>
<td>0.9501</td>
<td>0.9501</td>
<td>0.9501</td>
<td>0.8762</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.96305</td>
<td>0.9500</td>
<td>fail</td>
<td>0.9490</td>
<td>0.9452</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Approximations of $P(r_1 \leq r_{10}, r_2 \leq r_{20})$ for various non-circular $AR(0) - AR(2)$ models requiring mean correction.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$r_{10}$</th>
<th>$r_{20}$</th>
<th>NID</th>
<th>Sim</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3835</td>
<td>0.3854</td>
<td>0.25</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1476</td>
<td>0.1515</td>
<td>0.03168</td>
</tr>
<tr>
<td>10</td>
<td>1.2</td>
<td>-0.8</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1663</td>
<td>0.1785</td>
<td>0.00034</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3326</td>
<td>0.3325</td>
<td>0.25</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0179</td>
<td>0.0176</td>
<td>0.00190</td>
</tr>
<tr>
<td>25</td>
<td>1.2</td>
<td>-0.8</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0080</td>
<td>0.0077</td>
<td>0.00740</td>
</tr>
<tr>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2983</td>
<td>0.2987</td>
<td>0.25</td>
</tr>
<tr>
<td>70</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0465</td>
<td>0.0688</td>
<td>0.0688</td>
</tr>
</tbody>
</table>
Table 3. Approximate p-values in progressive AR(m) testing using the pears data set with n = 16 and p = 5.

<table>
<thead>
<tr>
<th>m</th>
<th>r</th>
<th>r.</th>
<th>lo</th>
<th>hi</th>
<th>NICD</th>
<th>SCDF</th>
<th>D/D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.26335</td>
<td>0.26335</td>
<td>-1</td>
<td>1</td>
<td>0.0494</td>
<td>0.0317</td>
<td>0.0096</td>
</tr>
<tr>
<td>2</td>
<td>-0.20321</td>
<td>-0.29288</td>
<td>-0.86130</td>
<td>1</td>
<td>0.5711</td>
<td>0.5665</td>
<td>0.6249</td>
</tr>
<tr>
<td>3</td>
<td>-0.42921</td>
<td>-0.33250</td>
<td>-0.99714</td>
<td>0.70450</td>
<td>0.1225</td>
<td>0.1239</td>
<td>0.4545</td>
</tr>
<tr>
<td>4</td>
<td>-0.24708</td>
<td>-0.12116</td>
<td>-0.91215</td>
<td>0.60137</td>
<td>0.4045</td>
<td>0.4339</td>
<td>0.8469</td>
</tr>
<tr>
<td>5</td>
<td>0.19903</td>
<td>0.17838</td>
<td>-0.67962</td>
<td>0.81167</td>
<td>0.1178</td>
<td>0.1151</td>
<td>0.0238</td>
</tr>
</tbody>
</table>

Table 4. Approximate p-values in progressive AR(m) testing using the Vinod data set with n = 44 and p = 3.

<table>
<thead>
<tr>
<th>m</th>
<th>r</th>
<th>r.</th>
<th>lo</th>
<th>hi</th>
<th>NICD</th>
<th>SCDF</th>
<th>D/D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01765</td>
<td>0.01765</td>
<td>-1</td>
<td>1</td>
<td>0.2797</td>
<td>0.2784</td>
<td>0.2846</td>
</tr>
<tr>
<td>2</td>
<td>0.29141</td>
<td>0.29119</td>
<td>-0.999377</td>
<td>1</td>
<td>0.00355</td>
<td>0.00352</td>
<td>0.00478</td>
</tr>
<tr>
<td>3</td>
<td>-0.25872</td>
<td>-0.29238</td>
<td>-0.906142</td>
<td>0.923709</td>
<td>0.05914</td>
<td>0.05924</td>
<td>0.06195</td>
</tr>
<tr>
<td>4</td>
<td>0.44199</td>
<td>0.46192</td>
<td>-0.781215</td>
<td>0.892208</td>
<td>0.04^120</td>
<td>0.04^112</td>
<td>0.04^521</td>
</tr>
<tr>
<td>5</td>
<td>-0.42366</td>
<td>-0.60055</td>
<td>-0.686565</td>
<td>0.629792</td>
<td>0</td>
<td>0.04^152</td>
<td>0.04^327</td>
</tr>
</tbody>
</table>

31