Algorithm choice for (partial) autocorrelation functions

B.D. McCullough
Federal Communications Commission, 2000 M St. NW Room 533, Washington, DC 20554, USA
Tel. +1 202 418 2019, Fax +1 202 418 2053, E-mail bmcculloch@fcc.gov

Different methods for calculating partial autocorrelation coefficients can produce different estimates, and these differences can be non-trivial. It has long been known that the Yule-Walker equations are particularly susceptible to numerical error, yet it is the most widely used method in statistical and econometric software. Two other methods, conditional maximum likelihood and Burg's algorithm, are known to be more reliable, yet are infrequently used. All three methods are applied to several datasets. A forecasting example shows that a model identified by Yule-Walker can produce inferior forecasts.

Keywords: Autocorrelation function, Burg algorithm, Yule-Walker equations

1. Introduction

The numerical performance of statistical and econometric software receives very little attention, as noted by McCullough [19]. McCullough [20] proposed a testing methodology which uncovered numerous errors in various packages when applied to statistical [21] and econometric [22] software. However, the accuracy with which specific algorithms are implemented is only one aspect of an evaluation of numerical performance. Another important aspect is the choice of algorithm [28].

As one example, consider the problem of distinguishing between empirical distribution functions. One approach is to use moment generating functions (MGFs). However, two distributions with markedly different theoretical MGFs can give rise to decidedly similar empirical MGFs [18]. The problem can be dealt with more reliably by changing the problem slightly to focus on the empirical characteristic function [33]. The difficulty arises because the real Vandermonde matrix associated with an MGF is extremely susceptible to numerical error due to ill-conditioning, whereas the complex Vandermonde matrix associated with the characteristic function is not [17].

As another example, there are many ways to solve for the coefficients of a linear least squares problem. The Cholesky decomposition is less robust to ill-conditioned data than the QR decomposition. That is, when the data are collinear, the Cholesky decomposition is more likely to give erroneous estimates than the QR decomposition. However, the QR decomposition requires more memory and is slower. If a user knows that he will only encounter well-conditioned problems, then it might not matter that his software uses the Cholesky decomposition. On the other hand, if the user can expect to encounter ill-conditioned data of the variety commonly found in...
economic applications, then the use of the Cholesky decomposition is an invitation to inaccuracy.

In both of these examples the choice of algorithm plays an important role in the quality of the final solution. In the first example there does not seem to be a reliable algorithm for computing the desired quantity (the empirical MGF), so the problem must be changed slightly (from computing the empirical MGF to computing the empirical characteristic function) to take advantage of an algorithm which will provide the necessary robustness. The second example is a situation in which there are many algebraically equivalent ways to calculate the desired quantity (the problem does not have to be changed), and these methods are not computationally equivalent. That is, two algebraically equivalent methods might give two different answers. Sometimes the difference between the two answers is so slight that some might call it rounding error, other times the difference is so pronounced that the qualitative characteristics of the two answers lead to different conclusions. Often, one method is easier to compute than another; frequently the price paid for computational convenience is less accurate results.

Thirty years ago when calculations were done by hand, twenty years ago when computer time on mainframes was expensive, or even ten years ago when memory on PCs was quite limited, such a price might have been justified. When 640K was a memory bound on PCs and the slow 8086 and 286 CPUs were common, the greater speed and lower memory requirements of the Cholesky decomposition often made it a reasonable choice of algorithm. In the present day, when computing time is free and memory limits are more a binding constraint on data sets than on program size, the decreased accuracy due to such expedient methods is harder to justify.

Such anachronistic algorithms, once written into a software program, can remain through several versions for many years. One such algorithm which has serious implications for Box-Jenkins modeling of economic series is the use of the Yule-Walker equations (YWE) to obtain partial autocorrelation coefficients. The first edition of Box and Jenkins appeared in 1970 and warned against the use of the YWE. Box and Jenkins [1, p. 65] note that regression-based methods for calculating partial correlation coefficients are more accurate than the YWE. Too, Priestley [27] lists four methods for computing partial correlation coefficients in order of decreasing accuracy: exact maximum likelihood, conditional maximum likelihood (least squares), approximate least squares, and the YWE. The method of Burg [3] can also be used. While the YWE works well for some types of data (such as the physical series examined in many engineering applications), it is unsuited to economic data. Tjostheim and Paulsen [32, p. 391] conducted simulations and observed that the Yule-Walker estimates are "inferior to least squares estimates for strongly autocorrelated processes, and that the behavior of the Yule-Walker estimates becomes progressively worse as the autocorrelation increases." Of course, many economic series are strongly autocorrelated. Stoica et al. [30] conducted a Monte Carlo study and reached a similar conclusion. Additional evidence on the inferiority of the YWE can be found in Dent and Min [7], Enser and Newton (1990) and Newton and Pagano [24].
While the inferiority of the YWE may be long-known, it certainly is not well-known. Granger and Newbold [12], Diebold [8], Enders [11], Hamilton [13], and many other texts discuss the YWE and no other method for computing the PACF Software packages such as EVIEWS, Autobox, SPSS, SHAZAM, and many others implement the YWE and no other method. Because vendors frequently do not identify the algorithm employed to compute partial autocorrelation coefficients, users often are unaware that their estimates may be grossly inaccurate and their ARMA models consequently misidentified. The purpose of this paper is to address these issues. To this end Section Two defines notation and definitions Section Three describes the YWE, least squares, and Burg methods of computing partial autocorrelation coefficients. Section Four applies these methods to eight common datasets, showing that they produce different PACFs, and the YWE identifies a different model than the other methods. Section Five offers the conclusions.

2. Notation and definitions

Let \( \{x_t\} \) be a covariance stationary time series. Then the autocovariance function at lag \( k \) is defined

\[
\text{cov}(x_t, x_{t-k}) = R_k
\]

and the variance of the process is given by \( R_0 \). The autocorrelation function at lag \( k \) is defined to be

\[
\rho_k = \frac{R_k}{R_0}
\]

(2)

Let \( \{x_t\} \) be a \( p \)-th order autoregressive process

\[
x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_p x_{t-p} + \epsilon_t
\]

(3)

which will be stationary if the roots of

\[
(1 - \beta_1 z - \beta_2 z^2 - \ldots - \beta_p z^p) = 0
\]

(4)

are outside the unit circle. Slight deviations in the estimates of the AR parameters can produce large changes in estimated pole locations if the poles are near the unit circle. Consequently, slight deviations in the parameter estimates can produce an estimated model which is unstable using the LS method. The YWE and Burg methods, by contrast, are guaranteed to produce stable solutions of (3); that is, none of the roots of (4) will fall inside the unit circle. This is more a concern when fitting AR models, not for identifying ARMA models. So for present purposes, the possible instability of LS estimates is not a drawback.

The \( k \)-th partial autocorrelation coefficient, \( \pi_k \) is defined to be the partial correlation between \( x_t \) and \( x_{t+k} \): that is, the correlation between \( x_t \) and \( x_{t+k} \) after
the effect of $x_{t+1}, \ldots, x_{t+k-1}$ has been removed from $x_t$. The first-order partial autocorrelation coefficient is defined to be the first-order autocorrelation coefficient, i.e., $\pi_1 \equiv \rho_1$. The second-order partial autocorrelation coefficient is the correlation between $x_t$ and $x_{t+2}$ after the effect of $x_{t+1}$ on $x_t$ has been removed. Letting $\rho_{01}$ denote the simple correlation between $x_t$ and $x_{t-1}$,

$$\pi_2 \equiv \rho_{02} = \frac{\rho_{02} - \rho_{01}\rho_{12}}{(1 - \rho_{01}^2)(1 - \rho_{12}^2)}$$

(Priestley [27, p. 121]). Generalizations of (5) to higher-order partial autocorrelations may be found in Stuart and Ord [31, ch. 27].

The autocovariances and partial correlation coefficients are linked via the Yule-Walker Equations, $R\pi = \pi$:

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_{p-1} \\ R_1 & R_0 & \cdots & R_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p-1} & R_{p-2} & \cdots & R_0 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_p \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{bmatrix}$$

(6)

so $\pi_1, \pi_2, \ldots, \pi_p$ are the coefficients which fit an AR model of order $p$, and $\pi_1, \pi_2, \ldots, \pi_p$ are the first $p$ partial autocorrelation coefficients. Casual inspection of (6) suggests that the Yule-Walker equations must be solved $p - 1$ times to obtain coefficients $2, \ldots, p$, inverting $R$ each time. Note that $R$ is a Toeplitz matrix. There are many methods for solving the YWE; Zhang and Duhamel [34] classify the various methods according to their relation to the Padé approximant. Levinson [16] obtained a general solution to solving a Toeplitz system without inverting $R$. Durbin [10] later adapted this method to the special case given by (6), the resulting method being referred to as the Durbin-Levinson recursion (DLR), a computationally expedient method for obtaining the partial autocorrelation coefficients.

Bunch [2] investigated the stability of various methods for solving Toeplitz systems, including the DLR. Cybenko [5] showed that the DLR is stable in an algorithmic sense. The sources and propagation of numerical error in the DLR have been studied by Papadyszews et al. [25]. In sum, these studies show that while the DLR solution of the YWE is numerically stable in an algorithmic sense, the use of the YWE is not robust to ill-conditioned $R$, which is often encountered in economics. To be specific, the problem is not the DLR, but the YWE. For example, the Burg method can be solved via the DLR without ill effect.

A commonly used measure of the condition of a matrix $A$ is

$$\kappa(A) = \|A\| \|A^{-1}\|$$

(7)

where $\|\cdot\|$ is a suitable norm. Following Cybenko [5], the 1-norm is used:

$$\|A\| = \max_i \sum_{j=1}^{p} |a_{ij}|, j = 1, 2, \ldots, p$$

(8)
The larger is \( \kappa \), the more sensitive is the solution of \((6)\) to numerical error. Details of conditioning and matrix computations are given in Stewart [24]. Cybenko [5] showed that \( \kappa \) tends to be large when \( \{x_t\} \) is highly correlated. Additionally, de Hoon, et al. [6] showed that the roots of \((4)\) approach the unit circle as \( \kappa \) increases.

It is sometimes alleged that the poor performance of the YWE/DLR is due to roots near the unit circle. This is incorrect – the poor performance can be established as a function of ill-conditioned \( R \). In fact, some autoregressive processes have poles near the unit circle while \( R \) is well-conditioned. In such cases, YWE gives good answers.

Thus, for the types of data typically encountered in economics, highly correlated and/or seasonal series, the use of the Yule-Walker equations should be eschewed, and other methods of computing partial autocorrelation coefficients should be employed.

3. Estimation methods

There are two approaches to the calculation of autocorrelation function (ACF) and the partial autocorrelation function (PACF). In the first approach, one is calculated directly (often the ACF) from which the other is derived. In the second approach, each is calculated directly. Newton [23, p. 208] describes a “given data” algorithm for YWE estimates of the PACF which does not involve calculation of the ACF. Thus, the calculation of PACF is not inextricably intertwined with the calculation of autocorrelation coefficients.

Letting \( \bar{x} \) be the sample mean for all \( T \) observations, two common estimators of \( R_k \) are the biased estimate

\[
\hat{R}_k = \frac{1}{T} \sum_{t=k}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x})
\]

which is positive definite and the unbiased estimate

\[
\hat{R}_k = \frac{1}{T-k} \sum_{t=k}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x})
\]

which is not positive definite. Of course, positive definiteness has implications for the stability of the estimated autoregression.

There are many ways to estimate the autocorrelation function, one of the more popular being

\[
\hat{\rho}_k = \frac{\hat{R}_k}{\hat{R}_0}
\]

This is the most satisfactory of the several methods considered by Jenkins and Watts [15], a fact noted by Box and Jenkins [1, p. 32]. Under the white noise null,
a 95% confidence interval for the autocorrelation function is given by \( \pm 1.96/\sqrt{T} \). Additionally, \( \hat{\rho}_k \) can be calculated directly by regressing \( x_t \) on a constant and \( x_{t-k} \). This is not equivalent to (11). Besides the degrees of freedom adjustment, its numerator is different, based on \( (x_{t-k} - \bar{x}_{t-k}) \) instead of \( (x_{t-k} - \bar{x}_t) \). For large samples, these differences tend to be minor. Such is not the case in PACF estimation, for which three methods are presented.

3.1. Yule-Walker/DLR

This method, which minimizes the forward error, is presented in many texts. Suppose the autocorrelations \( \hat{\rho}_k \) have been estimated. Then Durbin’s formula for recursive calculation of \( \pi \) is

\[
\hat{\pi}_{m+1,j} = \hat{\pi}_{m,j} - \hat{\pi}_{m+1,m+1} \hat{\pi}_{m,m-j+1} \quad j = 1, \ldots m
\]

(12)

\[
\hat{\pi}_{m+1}^{YW} \equiv \hat{\pi}_{m+1,m+1} = \frac{\hat{\rho}_{m+1} - \sum_{j=1}^{m} \hat{\pi}_{m,j} \hat{\rho}_{m+1-j}}{1 - \sum_{j=1}^{m} \hat{\pi}_{m,j} \hat{\rho}_j}
\]

(13)

3.2 Least-Squares

The LS estimates, which minimize the residual variance, are easily obtained by linear regression.

\[
x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + b_m x_{t-m} + \epsilon_t
\]

(14)

Then \( \hat{\rho}_m^{LS} \equiv b_m \) is the partial correlation coefficient between \( X_t \) and \( X_{t+m} \) holding \( X_{t+1}, \ldots, X_{t+m-1} \) fixed. These estimates are more robust to ill-conditioned \( \mathbf{R} \) than the YWE, but are not guaranteed to be stable. In particular, the LS method can produce autocorrelation and partial autocorrelation coefficients greater than unity. Too, the regression can be run in the backward fashion

\[
x_t = b_0 + b_1 x_{t+1} + b_2 x_{t+2} + \ldots + b_m x_{t+m} + \epsilon_t
\]

(15)

which will produce different estimates.

3.3. Burg

Burg [4] argued that the YWE assumes \( x_t = 0 \) for \( t < 1 \) and \( t > N \). The Burg estimator is constructed to operate only on actual data values, whereas the YWE is affected by the implicit use of zeroes to pad the series length. The Burg estimator further differs from the YWE in that not the forward error, but the average of the forward and backward is minimized. What follows is the “given data” implementation presented by Percival and Walden ([26], §9.5), using their notation...
For a vector \( \mathbf{V} \equiv [v_1, v_2, \ldots, v_{M-1}, v_M]' \) define a circular circular shift operator
\[
\mathcal{L}\mathbf{V}_1 \equiv [v_M, v_1, v_2, \ldots, v_{M-1}]'
\] (16)
and a subvector extraction operator
\[
\mathcal{M}_{j,k}\mathbf{V} \equiv [v_j, v_{j+1}, \ldots, v_{k-1}, v_k]'\]
(17)
where \( \mathbf{V}' \) denotes the transpose of \( \mathbf{V} \). Let \( <\mathbf{V}_1, \mathbf{V}_2> \) denote the inner product of two vectors, so the squared norm of a vector \( \mathbf{V} \) is \( ||\mathbf{V}||^2 \equiv <\mathbf{V}, \mathbf{V}> \). Then define two vectors of length \( M = N + p \)
\[
\tilde{\mathbf{e}} (0) \equiv [x_1, x_2, \ldots, x_n, 0, \ldots, 0]'
\] (18)
where the rightmost \( p \) elements are zeroes, and
\[
\hat{\mathbf{e}} (0) \equiv \mathcal{L} \tilde{\mathbf{e}} (0) \equiv [0, x_1, x_2, \ldots, x_n, 0, \ldots, 0]'
\] (19)
where the rightmost \( p - 1 \) elements are zeroes. The PAC coefficients can be computed recursively for \( k = 1, \ldots, p \)
\[
\hat{\pi}_k^B = \frac{2 < \mathcal{M}_{k+1,N} \tilde{\mathbf{e}} (k - 1), \mathcal{M}_{k+1,N} \hat{\mathbf{e}} (k - 1)>}{||\mathcal{M}_{k+1,N} \tilde{\mathbf{e}} (k - 1)||^2 + ||\mathcal{M}_{k+1,N} \hat{\mathbf{e}} (k - 1)||^2}
\] (20)
where
\[
\tilde{\mathbf{e}} (k) = \hat{\mathbf{e}} (k - 1) - \pi_k \tilde{\mathbf{e}} (k - 1)
\]
\[
\hat{\mathbf{e}} (k) = \mathcal{L}[\tilde{\mathbf{e}} (k - 1) - \pi_k \tilde{\mathbf{e}} (k - 1)]
\]

As noted already, the Burg method has the advantage that its coefficients are stable. While the YWE coefficients also are stable, it is well known that the Burg algorithm has a better ability to split spectral peaks than the YWE (Newton [22, §3 4 4]). This is an important consideration for parametric spectral analysis. Thus, any package which offers spectral methods and the YWE should offer the Burg, as well.

4. Application

The various methods are applied to Box-Jenkins Series A-G and Harvey's [14] UK Manufacturing Employment (MANEMP), 83 quarterly observations on employment in UK manufacturing (in thousands), seasonally adjusted, 1963:1–1983:3. The data are described in Table 1.

Figs 1 and 2 display the PACFs computed by the various methods for each dataset. The Yule-Walker estimates are solid lines, the least-squares estimates are dots, and the Burg estimates are dashes. Tables 3 and 4 display not only the numerical values of
Table 1

<table>
<thead>
<tr>
<th>name</th>
<th>description</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series A</td>
<td>chemical process concentration</td>
<td>197</td>
</tr>
<tr>
<td>Series B</td>
<td>IBM daily closing price</td>
<td>369</td>
</tr>
<tr>
<td>Series C</td>
<td>chemical process temperature</td>
<td>226</td>
</tr>
<tr>
<td>Series D</td>
<td>chemical process viscosity</td>
<td>310</td>
</tr>
<tr>
<td>Series E</td>
<td>Wolfen Sunspots</td>
<td>100</td>
</tr>
<tr>
<td>Series F</td>
<td>chemical process yields</td>
<td>70</td>
</tr>
<tr>
<td>Series G</td>
<td>monthly Airline Data</td>
<td>144</td>
</tr>
<tr>
<td>MANEMP</td>
<td>UK Manufacturing Employment</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>model</th>
<th>( b_0 )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
<th>( b_4 )</th>
<th>( b_5 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>YW AR(1)</td>
<td>-270.0</td>
<td>1031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.99606</td>
</tr>
<tr>
<td>(4.8)</td>
<td>(142.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS AR(5)</td>
<td>-78.5</td>
<td>1874</td>
<td>-0.704</td>
<td>-0.282</td>
<td>-0.134</td>
<td>0.255</td>
<td>0.99903</td>
</tr>
<tr>
<td>(-1.95)</td>
<td>(16.42)</td>
<td>(-2.86)</td>
<td>(-1.10)</td>
<td>(-0.55)</td>
<td>(2.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burg AR(4)</td>
<td>-49.4</td>
<td>1952</td>
<td>-0.814</td>
<td>-0.477</td>
<td>0.344</td>
<td></td>
<td>0.99897</td>
</tr>
<tr>
<td>(-1.28)</td>
<td>(18.15)</td>
<td>(-3.34)</td>
<td>(-1.95)</td>
<td>(3.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

the PACFs at each lag, but also the condition number of the Toeplitz matrix underlying
the YWE. Graphically, Series A and D stand out by virtue of the close agreement of
the various methods. From the Tables it can be seen that the corresponding condition
numbers are low, explaining the close agreement. The highest condition numbers
belong to Series C, and Fig. 1c shows that the YW estimates are much too smooth.
However, all the condition numbers for all the series are low. This suggests that
even extremely mild ill-conditioning can lead to a divergence between YWE and the
other methods. It is obvious that, at the preliminary stage of model identification, the
choice of algorithm used to compute the PACF makes a difference. Do the differently
identified models lead to different forecasts?

For the series MANEMP, the ACF decays in the fashion of a pure AR model, and
the PACFs are presented in Table 4. YW, LS, and Burg indicate AR models of orders
one, five, and four lags, respectively. Note the first coefficient for LS for UKMAN in
Table 4 is greater than unity. This is evidence of the fact that least squares estimates
do not always deliver stable coefficients. For the full sample, the estimated models
are presented in Table 2, with \( t \)-statistics in parentheses

It is clear the YW misidentified the lag structure of the series. To test the efficacy
of the three models, a series of 12 one-step forecasts was made, the first of which
estimated over 1963 1-1979 3 and predicted 1979 4. The resulting mean-squared
errors: YW = 3537.1; LS = 926 1; Burg = 777.7. Let “msl” denote “marginal
significance level.” The usual \( t \)-test on the correlation between the sum and the
difference of the forecast errors (Granger and Newbold [11, p. 279]) shows that LS
is preferable to YW (\( t \)-stat = 6.8, msl = 0.00), Burg is preferable to YW (\( t \)-stat = 9.4,
msl = 0.00), and that the Burg may be preferable to LS (\( t \)-stat = 21, msl = 0.06).
Fig 1 Yule-Walker (solid), least squares (dots) and Burg (dash) PACFs with 95% confidence interval
Table 3
Results of PACF calculations, including condition number ($\kappa$) asterisk denotes significance at 5% level

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\pi}_p$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LS}$</th>
<th>$\hat{\pi}_p^{LS}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57*</td>
<td>57*</td>
<td>57*</td>
<td>-</td>
<td>99*</td>
<td>99*</td>
<td>99*</td>
<td>-</td>
<td>98*</td>
<td>99*</td>
<td>99*</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>25*</td>
<td>25*</td>
<td>27*</td>
<td>10</td>
<td>07</td>
<td>09</td>
<td>09</td>
<td>10</td>
<td>-26*</td>
<td>82*</td>
<td>82*</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>07</td>
<td>08</td>
<td>08</td>
<td>11</td>
<td>02</td>
<td>01</td>
<td>00</td>
<td>16</td>
<td>-16*</td>
<td>-03</td>
<td>-03</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>07</td>
<td>09</td>
<td>09</td>
<td>13</td>
<td>05</td>
<td>06</td>
<td>05</td>
<td>22</td>
<td>-09</td>
<td>-03</td>
<td>-03</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>07</td>
<td>07</td>
<td>07</td>
<td>14</td>
<td>02</td>
<td>03</td>
<td>02</td>
<td>26</td>
<td>-06</td>
<td>-10</td>
<td>-10</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>15*</td>
<td>14*</td>
<td>15</td>
<td>05</td>
<td>02</td>
<td>01</td>
<td>32</td>
<td>-05</td>
<td>-08</td>
<td>-07</td>
<td>55</td>
</tr>
<tr>
<td>7</td>
<td>16*</td>
<td>19*</td>
<td>19*</td>
<td>15</td>
<td>03</td>
<td>-13*</td>
<td>-13*</td>
<td>59</td>
<td>-01</td>
<td>-01</td>
<td>-01</td>
<td>66</td>
</tr>
<tr>
<td>8</td>
<td>03</td>
<td>-04</td>
<td>-04</td>
<td>14</td>
<td>03</td>
<td>-05</td>
<td>-05</td>
<td>46</td>
<td>-04</td>
<td>-03</td>
<td>-03</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>01</td>
<td>01</td>
<td>01</td>
<td>15</td>
<td>05</td>
<td>-03</td>
<td>-03</td>
<td>53</td>
<td>-02</td>
<td>05</td>
<td>04</td>
<td>87</td>
</tr>
<tr>
<td>10</td>
<td>01</td>
<td>-01</td>
<td>-01</td>
<td>16</td>
<td>08</td>
<td>06</td>
<td>06</td>
<td>62</td>
<td>-01</td>
<td>04</td>
<td>04</td>
<td>98</td>
</tr>
<tr>
<td>11</td>
<td>07</td>
<td>-09</td>
<td>-09</td>
<td>18</td>
<td>01</td>
<td>-05</td>
<td>-05</td>
<td>62</td>
<td>-04</td>
<td>-15*</td>
<td>-15*</td>
<td>108</td>
</tr>
<tr>
<td>12</td>
<td>-02</td>
<td>-04</td>
<td>-05</td>
<td>21</td>
<td>03</td>
<td>-09</td>
<td>-10</td>
<td>69</td>
<td>-04</td>
<td>10</td>
<td>10</td>
<td>120</td>
</tr>
<tr>
<td>13</td>
<td>06</td>
<td>04</td>
<td>04</td>
<td>22</td>
<td>00</td>
<td>-03</td>
<td>-03</td>
<td>76</td>
<td>-04</td>
<td>-14*</td>
<td>-14*</td>
<td>134</td>
</tr>
<tr>
<td>14</td>
<td>09</td>
<td>08</td>
<td>08</td>
<td>22</td>
<td>01</td>
<td>07</td>
<td>07</td>
<td>82</td>
<td>-02</td>
<td>01</td>
<td>01</td>
<td>148</td>
</tr>
<tr>
<td>15</td>
<td>-12</td>
<td>-15*</td>
<td>-15*</td>
<td>22</td>
<td>-02</td>
<td>-08</td>
<td>-08</td>
<td>87</td>
<td>-04</td>
<td>-10</td>
<td>-10</td>
<td>160</td>
</tr>
<tr>
<td>16</td>
<td>05</td>
<td>07</td>
<td>06</td>
<td>26</td>
<td>-02</td>
<td>06</td>
<td>05</td>
<td>93</td>
<td>-03</td>
<td>-02</td>
<td>-02</td>
<td>174</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>13</td>
<td>13</td>
<td>26</td>
<td>-07</td>
<td>-15*</td>
<td>-15*</td>
<td>101</td>
<td>-03</td>
<td>-07</td>
<td>-07</td>
<td>187</td>
</tr>
<tr>
<td>18</td>
<td>07</td>
<td>09</td>
<td>09</td>
<td>25</td>
<td>-07</td>
<td>-10*</td>
<td>-10*</td>
<td>114</td>
<td>-07</td>
<td>11</td>
<td>11</td>
<td>201</td>
</tr>
<tr>
<td>19</td>
<td>-07</td>
<td>-06</td>
<td>-06</td>
<td>24</td>
<td>01</td>
<td>-01</td>
<td>-01</td>
<td>130</td>
<td>-05</td>
<td>11</td>
<td>11</td>
<td>225</td>
</tr>
<tr>
<td>20</td>
<td>05</td>
<td>07</td>
<td>06</td>
<td>26</td>
<td>-06</td>
<td>-08</td>
<td>-08</td>
<td>135</td>
<td>-02</td>
<td>-14*</td>
<td>-14*</td>
<td>245</td>
</tr>
</tbody>
</table>

Table 4
Results of PACF calculations, including condition number ($\kappa$) asterisk denotes significance at 5% level

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\pi}_p$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
<th>$\hat{\pi}_p^{LY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81*</td>
<td>81*</td>
<td>82*</td>
<td>-</td>
<td>-39*</td>
<td>-42*</td>
<td>-41*</td>
<td>-</td>
<td>95*</td>
<td>96*</td>
<td>96*</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>05</td>
<td>63*</td>
<td>-71*</td>
<td>10</td>
<td>18</td>
<td>19</td>
<td>18</td>
<td>23</td>
<td>-23*</td>
<td>-33*</td>
<td>-33*</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>08</td>
<td>21*</td>
<td>21*</td>
<td>40</td>
<td>00</td>
<td>01</td>
<td>01</td>
<td>37</td>
<td>04</td>
<td>20*</td>
<td>19*</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>06</td>
<td>-15</td>
<td>-15</td>
<td>43</td>
<td>00</td>
<td>-07</td>
<td>-07</td>
<td>38</td>
<td>09</td>
<td>15</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>00</td>
<td>10</td>
<td>10</td>
<td>52</td>
<td>-07</td>
<td>-07</td>
<td>-07</td>
<td>44</td>
<td>07</td>
<td>26*</td>
<td>23*</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>10</td>
<td>11</td>
<td>53</td>
<td>-12</td>
<td>-15</td>
<td>-14</td>
<td>50</td>
<td>01</td>
<td>-03</td>
<td>-04</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>18</td>
<td>18</td>
<td>47</td>
<td>02</td>
<td>05</td>
<td>05</td>
<td>58</td>
<td>13</td>
<td>20*</td>
<td>17*</td>
<td>34</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>24*</td>
<td>25*</td>
<td>46</td>
<td>01</td>
<td>-00</td>
<td>-01</td>
<td>65</td>
<td>09</td>
<td>16</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>9</td>
<td>08</td>
<td>01</td>
<td>00</td>
<td>43</td>
<td>00</td>
<td>-00</td>
<td>-10</td>
<td>67</td>
<td>23*</td>
<td>57*</td>
<td>48*</td>
<td>34</td>
</tr>
<tr>
<td>10</td>
<td>08</td>
<td>01</td>
<td>00</td>
<td>41</td>
<td>00</td>
<td>05</td>
<td>05</td>
<td>69</td>
<td>17*</td>
<td>29*</td>
<td>17*</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>07</td>
<td>13</td>
<td>15</td>
<td>40</td>
<td>14</td>
<td>18</td>
<td>16</td>
<td>74</td>
<td>17*</td>
<td>84*</td>
<td>56*</td>
<td>28</td>
</tr>
<tr>
<td>12</td>
<td>03</td>
<td>-16</td>
<td>-17</td>
<td>39</td>
<td>-01</td>
<td>-06</td>
<td>-03</td>
<td>73</td>
<td>-14</td>
<td>61*</td>
<td>-12</td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>08</td>
<td>12</td>
<td>15</td>
<td>44</td>
<td>09</td>
<td>09</td>
<td>07</td>
<td>77</td>
<td>-54*</td>
<td>-67*</td>
<td>-82*</td>
<td>32</td>
</tr>
<tr>
<td>14</td>
<td>04</td>
<td>03</td>
<td>02</td>
<td>44</td>
<td>17</td>
<td>18</td>
<td>16</td>
<td>76</td>
<td>-03</td>
<td>-38*</td>
<td>-31*</td>
<td>74</td>
</tr>
<tr>
<td>15</td>
<td>03</td>
<td>-08</td>
<td>-08</td>
<td>45</td>
<td>00</td>
<td>01</td>
<td>02</td>
<td>76</td>
<td>09</td>
<td>08</td>
<td>08</td>
<td>81</td>
</tr>
<tr>
<td>16</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>48</td>
<td>22</td>
<td>43*</td>
<td>32*</td>
<td>79</td>
<td>02</td>
<td>-03</td>
<td>-10</td>
<td>78</td>
</tr>
<tr>
<td>17</td>
<td>16</td>
<td>-07</td>
<td>-07</td>
<td>56</td>
<td>05</td>
<td>01</td>
<td>00</td>
<td>80</td>
<td>03</td>
<td>-06</td>
<td>-13</td>
<td>81</td>
</tr>
<tr>
<td>18</td>
<td>11</td>
<td>-11</td>
<td>-13</td>
<td>69</td>
<td>-10</td>
<td>-14</td>
<td>-18</td>
<td>81</td>
<td>07</td>
<td>-04</td>
<td>-10</td>
<td>83</td>
</tr>
<tr>
<td>19</td>
<td>02</td>
<td>00</td>
<td>02</td>
<td>81</td>
<td>04</td>
<td>11</td>
<td>11</td>
<td>87</td>
<td>05</td>
<td>28*</td>
<td>09</td>
<td>81</td>
</tr>
<tr>
<td>20</td>
<td>01</td>
<td>04</td>
<td>03</td>
<td>84</td>
<td>05</td>
<td>18</td>
<td>08</td>
<td>88</td>
<td>-05</td>
<td>-04</td>
<td>-16*</td>
<td>81</td>
</tr>
</tbody>
</table>

Since $\rho_1 = \pi_1$ by definition, for sake of completeness mention must be made of the Burg method for computing autocorrelations, an algorithm for which is presented.
in the Appendix. The Burg ACF substantially agrees with the usual method, though some discrepancy was observed for nonstationary and seasonal series. In no case did the difference affect qualitative conclusions for the series in Table 1.

5. Conclusions

Theoretical and simulation studies are adduced to show that the Yule-Walker equation (YWE) estimates of the partial autocorrelation function (PACF) are inferior to estimates produced by least squares (LS) and Burg methods. The three methods are applied to several series. While LS and Burg substantially agree, the PACF computed by these methods can differ markedly from YWE estimates, leading to different models being identified. Thus, ARMA forecasts can depend on the algorithm used to compute the PACF. Since the YWE is the most popular method offered by statistical and econometric software packages, users of such packages should consider the distinct possibility that their ARMA models are misspecified.

6. Appendix on computational details

All programming was done in RATS v4.3 [9] on a PC running Windows 95. RATS implements the Yule-Walker estimates of partial autocorrelation coefficients. The accuracy of these estimates was checked twice using RATS' programming language. First, the estimates were calculated by direct inversion of the sequence of Toeplitz matrices. Second, they were calculated by the DLR. All three sets of estimates agreed to several decimal places. The least squares estimates were calculated using the linear regression command in RATS, whose accuracy was verified by McCullough [19]. The Burg algorithm was written in RATS, and its output agreed to 12 decimals with a benchmark subsequently provided by D. Percival of the Applied Physics Laboratory at the University of Washington. Percival's data are 20 observations: 71, 63, 70, 88, 99, 90, 110, 135, 128, 154, 156, 141, 131, 132, 141, 104, 136, 146, 124, 129. After subtracting the mean of 117.4 from each observation, to eight decimals, the first four Burg estimates of the partial autocorrelation coefficients are π₁ = 0.80005569, π₂ = 1.05602226, π₃ = 0.15582353, π₄ = −0.44248131.

It is possible to derive the sequence of autocovariances s₀, s₁, ..., sₚ from the partial autocorrelation coefficients, whence the autocorrelation coefficients can be obtained in the usual fashion, ρₖ = sₖ/s₀. References to equation numbers from Percival and Walden [26] are denoted with a 'PW'.

To obtain the autocovariances, first define a symmetric matrix Φ whose main diagonal is the partial autocorrelation coefficients (which are known) and whose off-diagonal element φₜₖ is the j-th coefficient in an autoregression of order k (these are to be calculated). The upper off-diagonal elements can be computed according to PW-Eq 402b for k = 2, ..., p

\[ φ_{j,k} = φ_{j,k-1} - φ_{k,k}φ_{j,k-1} \quad 1 \leq j \leq k - 1 \]
e.g. \( \phi_{3,4} = \phi_{3,3} - \phi_{4,4}\phi_{1,3} \).

Next, using PW-Eq 403 calculate a sequence \( P_0, P_1, \ldots, P_p \) where \( P_0 = s_0 \) is the sample variance and

\[
P_k = P_{k-1}(1 - \phi_{k,k}^2)
\]

(22)

Define \( s_1 = \phi_{1,1}P_0 \) and for \( k = 2, \ldots, p \) use the inverted version of PW-Eq. 402a to obtain

\[
s_k = \phi_{k,k}P_{k-1} + \sum_{j=1}^{k-1} \phi_{j,k-1}s_{k-j}
\]

(23)

To eight decimals, the first four Burg autocorrelation coefficients for the Percival data are: \( \rho_1 = 0.80005569, \rho_2 = 0.67811811, \rho_3 = 0.62519921, \rho_4 = 0.40410897 \).

Finally, since many packages do not name the method used to compute partial autocorrelation coefficients, it is useful to have a simple method to determine whether or not a package uses the YWH. Consider the sequence \{1, 2, \ldots, 10\}. Application of (5) yields \( \hat{\pi}_2 = 0 \) and, by inspection, \( \pi_1 = \rho_1 = \rho_2 = 1 \). These are the solutions produced by the LS method. By contrast, if the program returns \( \hat{\pi}_1 = \hat{\rho}_1 = 0.7, \hat{\pi}_2 = -0.153 \) and \( \hat{\rho}_2 = 0.412 \) then the program uses a standard implementation of the YWH.

Acknowledgement

Thanks to C. Chatfield, C. Cummins, T. Doan, A. Koehler and the referees for comments and useful suggestions. The views expressed herein are those of the author and do not necessarily reflect those of the Commission.

References