ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

Exact Maximum Likelihood Estimation of Regression Equations with a General Stationary Auto-Regressive Disturbance*

DISCUSSION PAPER NO. 343

by

Charles M. Beach
Queen's University

and

Stephen Yeo
London School of Economics and Political Science

May 1979

*The authors would like to express their thanks to David Card for programming assistance, David Hendry and James MacKinnon for useful comments and questions, and members of an econometrics seminar at University College London. Work on the paper was partly funded by the Social Sciences and Humanities Research Council of Canada while the first author was an academic visitor at the LSE.

Abstract

This paper develops an exact maximum likelihood technique for estimating regression equations with general p'th-order autoregressive disturbances. The approach appears to be computationally practical and straightforward, insures the estimated error coefficients satisfy *a priori* stationarity conditions, and insures convergence of the estimation procedure. Recent expression of the analytic inverse of the covariance matrix of a stationary AR(p) process provides the basis for the iterative algorithms, which employ a modified Gauss-Newton technique utilizing exact first and approximate second derivatives. The relationship between stationarity and the form of the objective function is examined. Empirical estimates are then presented for regression models with and without a lagged dependent variable.

I. Introduction

The estimation[1] of time series regression models with autocorrelated disturbances frequently arises in applied econometrics, and it is often appropriate or convenient to represent such autocorrelation by a stationary autoregressive error process. Autocorrelated disturbances are most often characterized in applied work by the Markov or first-order autoregressive (AR(1)) error process, but in many cases a higher-order process may be more appropriate. Such specifications allow a more flexible shape to the correlogram of the error process, so that it is not restricted to be positive with a strict geometric decline. It has also been emphasized recently in the literature (Hendry and Trivedi (1970), Pagan (1974), Godfrey (1978), and Newbold and Davies (1978)) that inference and parameter estimation can be greatly affected by the error structure assumed, and that applied work should consider a wider range of possible error structures than just an AR(1) formulation. Moreover, concern with the dynamics of model behaviour may lead to consideration of a higher-order autoregressive process since autoregressive errors can be interpreted as a special case of a general dynamic model specification (Sargan (1964), Hendry and Mizon (1978)). Such higher-order processes may be expected to arise particularly in work with quarterly (Thomas and Wallis (1971) Wallis (1972)) or monthly time series.

---

1. We wish to emphasize at the outset that this paper is essentially concerned with problems of estimation and not inference, and with estimation of regression models and not time series models.

Conventional approaches to estimation of regression models with general p'th-order autoregressive error processes have typically been based on least squares or other approximate maximum likelihood methods. On the one hand, (nonlinear) least squares procedures such as employed by Fuller and Martin (1961), Pierce (1972), Wallis (1972), Hendry (1971), or Pagan (1974) focus on a sum of squares objective function and disregard the initial p observations as asymptotically unimportant or as statistically not worth estimating as nuisance parameters. On the other hand, authors such as Kadiyala (1968), Thomas and Wallis (1971) or Box and Jenkins (1970) incorporate the initial observations into the exponential term of the (multivariate normal) likelihood function and then optimize this exponential term. While such approaches ease the optimization problems associated with such nonlinear estimation and yield estimates which are asymptotically equivalent to full maximum likelihood estimates, they are only approximations to exact or full maximum likelihood estimates. Estimates for small and even moderately sized samples can nonetheless differ fairly substantially from the exact ML estimates (Beach and MacKinnon (1978), Dent and Min (1978)) particularly if the parameter values of the error process are close to the boundaries of the stationarity region. Indeed, there is some evidence that the small-sample performance (for up to 100 and 200 observations) of ML estimators improves relative to alternative estimators as the number of autoregressive parameters increases (Dent and Min (1978)). Exact ML procedures are preferable from the point of view of testing the structure and order of the autoregressive process since tests are based on the correct maximized likelihood and the sample on which the estimates are based does not change with the order of the

process considered. In addition, optimization of the full likelihood function can also be used to impose stationarity restrictions on the error process so that the estimated process is well behaved, estimation results are more readily interpretable, and conventional inference procedures can be carried out.

In an earlier paper, Beach and MacKinnon (1978) proposed an exact ML procedure for a first-order autoregressive error model. The present paper extends this approach to a general p'th-order autoregressive error process. This extension makes use of recent advances in the statistics literature giving explicit analytic expressions for the inverse of the covariance matrices (and corresponding determinants) which arise in exact ML procedures. We also make use of results from numerical analysis which provide more robust algorithms, guaranteed to converge for exact ML problems (with associated determinantal terms). The paper thus attempts to provide economists with an exact maximum likelihood regression procedure for the general AR(p) error model that reflects recent advances in both statistical theory and computational techniques.

Exact ML estimation procedures have also been recently proposed for the general autoregressive-moving average (ARMA) error process by Newbold (1974), Prothero and Wallis (1976), Dent (1977), Ali (1977), Ansley (1979), and Harvey and Phillips (1979). But the procedures in this paper (suggested in part by Harvey and Phillips (1979)) are designed for regression models, as opposed to simple time-series models, and make more efficient use of the particular analytic structure of AR processes in the iterative algorithms, which should yield greater computational efficiency. The procedures also explicitly provide gradient algorithms

for estimating the autoregressive parameters, and take into account the computational implications of lagged dependent variables in regression equations.

The next section of the paper formally sets out the problem of exact maximum likelihood estimation in the general AR(p) regression model. Section III then examines the relationship between stationarity of the error process and the form of the objective function. The optimization algorithms are discussed in the following Section IV. The procedures are then illustrated with two regression equation estimates (one without and one with a lagged dependent variable) in Section V. Concluding remarks are presented in the final section.

II. Setting Out the Problem of Exact ML Estimation

Consider, for convenience, the linear regression model

$$y = X\beta + u \tag{2.1}$$

with t=1,...,T observations and K exogenous independent variables, where the error term follows a general p'th-order autoregressive process,

$$u_t = \alpha_1 u_{t-1} + \ldots + \alpha_p u_{t-p} + \varepsilon_t, \qquad t = p+1,\ldots,T, \tag{2.2}$$

where $\varepsilon_t \sim$ NID $(0,\sigma^2)$. Then, following Box and Jenkins (1970, p. 274), one notes that the joint probability density of the full set of autocorrelated errors $u' = (u_1, \ldots, u_T)$ can be written

$$f(u|\alpha,\sigma^2) = g(u_{p+1},\ldots,u_T|u_*,\alpha,\sigma^2)\cdot h(u_*|\alpha,\sigma^2)$$

where $u_*' = (u_1, \ldots, u_p)$. Under the assumption of normality of the ε's, the joint density of $\varepsilon_{p+1}, \ldots, \varepsilon_T$ is

$$(2\pi\sigma^2)^{-(T-p)/2}\cdot\exp\left[-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T}\varepsilon_t^2\right]$$

Conditional upon the initial values, u_*, then, the corresponding joint density of $u_{p+1}, \ldots, u_T$ becomes

$$g(u_{p+1},\ldots,u_T|u_*,\alpha,\sigma^2) = (2\pi\sigma^2)^{-(T-p)/2}\cdot\exp[\frac{-1}{2\sigma^2}\Sigma_{t=p+1}^{T}(u_t-\alpha_1 u_{t-1}-\ldots-\alpha_p u_{t-p})]$$

where the Jacobian of the transformation from the ε's to the u's is unity. If we also assume u_* is joint normal with zero mean and covariance matrix $\sigma^2 M_*^{-1}$,

$$h(u_*|\alpha,\sigma^2)=(2\pi\sigma^2)^{-p/2}|M_*|^{\frac{1}{2}}\exp[\frac{-1}{2\sigma^2}u_*' M_* u_*],$$

and the joint density of the full set of u's becomes

$$f(u|\alpha,\sigma^2) = (2\pi\sigma^2)^{-T/2}|M_*|^{\frac{1}{2}}\exp[\frac{-1}{2\sigma^2}S(\alpha)] \tag{2.3}$$

where

$$S(\alpha) = u_*' M_* u_* + \Sigma_{t=p+1}^{T}(u_t-\alpha_1 u_{t-1}-\ldots-\alpha_p u_{t-p})^2 \tag{2.4}$$

$$= u_*' M_* u_* + u'P'Pu$$

and P is the transformation matrix corresponding to (2.2):

$$\underset{(T-p)xT}{P} = \begin{bmatrix} -\alpha_p & \cdots & -\alpha_1 & 1 & & \\ & \ddots & & \ddots & \ddots & \\ & & -\alpha_p & & -\alpha_1 & 1 \end{bmatrix}.$$

Alternatively, one could assume that the initial u's and ε's are related by

$$\varepsilon_* = Ru_*$$

for R nonsingular, in which case R and M_* are related simply by $R'R = M_*$. So far R and M_* are unspecified, but they are determined below by requiring the initial error vector u_* to have the same covariance structure as the rest of the u vector, so that the entire set of u's is covariance stationary.

The log likelihood function corresponding to (2.3),

$$(2.5)\quad L = \text{const} - \frac{T}{2}\,\ell n\,\sigma^2 + \frac{1}{2}\,\ell n|M_*| - \frac{1}{2\sigma^2}\,(u_*'M_*u_* + u'P'P_u),$$

serves as a basis for comparing the principal types of estimators, conventionally suggested for the AR(p) model. Straightforward least-squares procedures essentially focus on only the conditional $g(\cdot)$ function or the sum-of-squares term, u'PPu, as the minimand, so that resulting estimates are interpreted as conditional upon the initial p disturbances. "Unconditional" least squares procedures such as suggested by Box and Jenkins (1970), on the other hand, can be seen as minimizing the full exponent term $S(\alpha) = u_*'M_*u_* + u'P'Pu$. Both approaches disregard the determinantal term $1/2\ \ell n|M_*|$ as asymptotically negligible, and both can be seen as approximations to exact ML estimators based on the full likelihood function (2.5). Consideration of the "penalty" term $1/2\ \ell n|M_*|$ in (2.5), however, greatly facilitates investigation of the structure of the autoregressive process near the boundaries of the stationarity region.

Alternatively, $S(\alpha)$ can be written as

$$(2.6)\quad \begin{aligned} S(\alpha) &= u_*'M_*u_* + u'P'Pu \\ &= u_*'R'Ru_* + u'P'Pu \\ &= u' \begin{bmatrix} R'R + P_1'P_1 & P_1'P_2 \\ P_2'P_1 & P_2'P_2 \end{bmatrix} u \quad \text{where } P = \underset{(T-p)xp}{[P_1} \,\vdots\, \underset{(T-p)x(T-p)}{P_2}] \end{aligned}$$

$$(2.7)\quad = u'Q'Qu \qquad \text{where } \underset{(TxT)}{Q} = \left[\begin{array}{c:c} R & 0 \\ \hline & P \end{array}\right]$$

and $Q'Q = \sigma^2 V(u)^{-1}$. Consequently, given α, $S(\alpha)$ may be calculated from M_* and P as in (2.6) or from Rand Pasin (2.7).

The approach[2] which is followed in this paper is to determine M_* explicitly in terms of α by imposing the condition that the full disturbance vector u be covariance stationary, and then utilizing an expression for the inverse of the covariance matrix of stationary autoregressive processes. Specifically, Galbraith and Galbraith (1974, eq. 11)[3] show that the elements of $M_* = \sigma^2 V(u_*)^{-1}$ can be expressed analytically in terms of the α's in relatively simple fashion as

$$(2.8) \quad m^*_{rs} = \sum_{j=0}^{r-1} \alpha_j \alpha_{j+s-r} - \sum_{j=p+1-s}^{p+r-s} \alpha_j \alpha_{j+s-r} \quad \text{for } 1 \leq r \leq s \leq p$$

and m^*_{rs} equal zero otherwise, with $\alpha_0 \equiv -1$.

It may be informative to examine the matrix M_* given by (2.8) for several plausible autoregressive processes. A second-order process is still relatively simple, but allows a more flexible correlogram than an AR(1) process. On the other hand, a full AR(4) or the simpler mixed first- and fourth-order process may be appropriate in quarterly regression models where the errors incorporate effects from the previous quarter as well as from the same quarter of the previous year. The corresponding matrices for these processes are presented in Table 1 for reference.

---

2. One alternative approach to maximizing (2.6) might be to replace M_* by R'R and explicitly express the elements R_{ij} of R in terms of the α's. One could follow Lempers and Kloek (1973), for example, and again require that the covariance structure of u_* be the same as that of $u_{p+1},\ldots,u_T$ expressed in terms of the α's. But this procedure rapidly becomes burdensome for AR processes beyond second order. Another approach suggested by Ali (1978) involves computing the matrix Q numerically from a procedure that requires characteristic roots of a p-by-p matrix that is substantially less straightforward to operate on than the M_* matrix used in the present paper.

3. Further results are also provided by de Gooijer (1978) and the references he cites. It should perhaps be noted that M_* is not the upper-left corner of the matrix $\sigma^2 V(u)^{-1}$.

Table 1

M_* Matrices for Illustrative AR Processes

AR(2): $u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \varepsilon_t$

$$M_* = \begin{bmatrix} 1 - \alpha_2^2 & -\alpha_1(1+\alpha_2) \\ -\alpha_1(1+\alpha_2) & 1 - \alpha_2^2 \end{bmatrix}$$

AR(1/4): $u_t = \alpha_1 u_{t-1} + \alpha_4 u_{t-4} + \varepsilon_t$

$$M_* = \begin{bmatrix} 1-\alpha_4^2 & -\alpha_1 & 0 & -\alpha_1\alpha_4 \\ -\alpha_1 & 1+\alpha_1^2-\alpha_4^2 & -\alpha_1 & 0 \\ 0 & -\alpha_1 & 1+\alpha_1^2-\alpha_4^2 & -\alpha_1 \\ -\alpha_1\alpha_4 & 0 & -\alpha_1 & 1-\alpha_4^2 \end{bmatrix}$$

AR(4): $u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \alpha_3 u_{t-3} + \alpha_4 u_{t-4} + \varepsilon_t$

$$M_* = \begin{bmatrix} 1- \alpha_4^2 & -(\alpha_1+\alpha_3\alpha_4) & -(\alpha_2+\alpha_2\alpha_4) & -(\alpha_3+\alpha_1\alpha_4) \\ -(\alpha_1+\alpha_3\alpha_4) & 1+\alpha_1^2-\alpha_3^2-\alpha_4^2 & -(\alpha_1-\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_3\alpha_4) & -(\alpha_2+\alpha_2\alpha_4) \\ -(\alpha_2+\alpha_2\alpha_4) & -(\alpha_1-\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_3\alpha_4) & 1+\alpha_1^2-\alpha_3^2-\alpha_4^2 & -(\alpha_1+\alpha_3\alpha_4) \\ -(\alpha_3+\alpha_1\alpha_4) & -(\alpha_2+\alpha_2\alpha_4) & -(\alpha_1+\alpha_3\alpha_4) & 1-\alpha_4^2 \end{bmatrix}$$

It is typically simpler to operate on the concentrated log-likelihood function obtained from solving the first-order conditions for $\tilde{\sigma}^2 = S(\alpha)/T$, substituting back into (2.6), and expressing the log-likelihood function for the regression model in terms of the full set of α and β parameters based on the sample observations for y and X:

$$(2.9) \quad L_c(\alpha,\beta;y,X) = \text{const} + \frac{1}{2}\ell n|M_*| - \frac{T}{2}\ell n\,[(y-X\beta)'Q'Q(y-X\beta)].$$

This is the objective function which will be maximized by the procedures described in Section IV. It is useful, however, to look more closely at the relationship between the stationarity of the autoregressive process and the form of the log likelihood function in (2.9).

III. Covariance Stationarity and the Likelihood Function

Consider initially a first-order process in which the Jacobian term in (2.6) is $(1/2)\ell n(1-\alpha_1^2)$. Clearly, this process is stationary and the log-likelihood function well behaved for $-1 < \alpha_1 < 1$. But as α_1 approaches a boundary of the stationarity region $(-1,1)$, $\ell n(1-\alpha_1^2) \to -\infty$, and with σ^2 fixed the likelihood function itself approaches zero. In other words, the likelihood is zero that the sample u was generated by an AR(1) process for which $\alpha_1 = \pm 1$. This should not be surprising, as the likelihood function was formulated on the assumption that the process was stationary, and therefore assigns zero likelihood to the possibility of it being nonstationary. This suggests that a similar relationship between stationarity and behaviour of the likelihood function may also hold for higher-order processes, and this is indeed the case.

Consider then the p'th-order autoregressive process defined in (2.2). The process is stationary if and only if the roots of the

associated polynomial

$$(3.1)\quad \phi(z) = z^p - \alpha_1 z^{p-1} - \ldots - \alpha_p$$

lie inside the unit circle in the complex plane (i.e., have modulus less than unity). For the first-order process, $\phi(z)=z-\alpha_1$, so that α_1 is the root of $\phi(z)$, and the Jacobian term $(1/2)\ell n|M_*| = (1/2)\ell n(1-\alpha_1^2)$ clearly involves the root of the polynomial $\phi(z)$. In fact, for higher-order processes, this simple result generalizes easily. Anderson and Mentz (1977) have demonstrated that for the p'th-order process,

$$(3.2)\quad |M_*| = \Pi_{i,j=1}^{p} (1 - r_i r_j)$$

where $r_1,\ldots,r_p$ are the proofs of $\phi(z)$. Thus the Jacobian term of the likelihood function can generally be expressed in terms of the roots of the associated polynomial of the process. Now (3.2) can be usefully rewritten as

$$(3.3)\quad |M_*| = \Pi_{i=1}^{p} (1-r_i^2) \cdot \Pi_{\substack{j,k=1 \\ j\neq k}}^{p} (1-r_j r_k)^2.$$

When all the roots of $\phi(z)$ are inside the unit circle, $r_1^2 < 1$ and $|M_*| > 0$. But as a root r_i approaches the unit circle, $1-r_i^2 \to 0$ and $|M_*| \to 0$. In addition, if the α's are chosen such that a root r_i lies outside the unit circle, $1-r_i^2 < 0$ and $|M_*|<0$. However, it is not true that $|M_*|>0$ if and only if the autoregressive process is stationary. For if an even number of roots lie outside the unit circle, $|M_*|$ will again be positive. Clearly, then $|M_*|>0$ is a necessary but not sufficient condition for the corresponding autoregressive process to be stationary.

Necessary and sufficient conditions on the coefficients of a polynomial so that the roots of the polynomial lie within the unit circle

have been forwarded in several forms. Wise (1956) and Samuelson (1941) derive conditions based on the Routh-Hurwitz criteria, but this approach is not as simple and direct as an alternative approach that has more immediate applicability in the present context. It is perhaps natural to look at work on the stability of linear, constant coefficient, difference equations for appropriate conditions since such difference equations are stable if and only if a certain polynomial has all its roots inside the unit circle. Such results were first obtained by Schur, Cohn, and Fujiwara (see Marden (1966)), and all known as the Schur-Cohn conditions. In the present context, they may be most conveniently stated as follows:

Theorem 1: A necessary and sufficient condition for all the roots of the polynomial $\phi(z) = z^p - \alpha_1 z^{p-1} - \ldots - \alpha_p$ to have modulus less than unity is that a matrix S be positive definite, where S is the symmetric (pxp) matrix with elements

$$(3.4) \quad S_{ij} = \sum_{k=1}^{\min(i,j)} (\alpha_{i-k}\alpha_{j-k} - \alpha_{p-i+k}\alpha_{p-j+k}) \qquad i,j=1,\ldots,p$$

where $\alpha_0 = -1$ by convention (for proof and discussion, see Kalman (1965) and Folsom, Boger, and Mullikin (1976)). The matrix S has therefore a simple form and moreover its rows can be calculated in a simple recursive fashion (Power (1970)). The Schur-Cohn criterion is therefore easily implemented and less cumbersome than the results given by Wise.

The Schur-Cohn criterion given in (3.4) is also of importance because of its relationship to the covariance matrix of the auto-regressive process. In fact, on expanding the expressions in (2.6) and (3.4) and noting the symmetry in M_* and S, one can see that S is identical to the matrix M_* defined earlier. In other words, as Pagano (1973) has

noted, the autoregressive process is stationary if and only if the expression for the inverse of the covariance matrix for p consecutive observations from the process yields a positive definite matrix. One can thus see why $|M_*|>0$ is not sufficient to guarantee stationarity of the process since <u>all</u> the principal minors of M_* must be positive and not merely the principal minor of order p. The Schur-Cohn criterion thus provides a direct relationship with the covariance structure of the process.

Though quite straightforward, the Schur-Cohn criterion can be simplified further for particular applications. Jury (1962) has shown that half the determinantal conditions involving S are redundant and need not be considered.[4] A relatively simply application of Jury's simplified criterion is to a mixed first- and fourth-order AR(1/4) process refered to earlier. The process

$$u_t = \alpha_1 u_{t-1} + \alpha_4 u_{t-4} + \varepsilon_t$$

is stationary if and only if all the roots of

$$\phi(z) = z^4 - \alpha_1 z^3 - \alpha_4$$

lie inside the unit circle. Specializing Jury's results, we have the following stability criteria, after allowing for unnecessary restrictions:

$$1 - \alpha_1 - \alpha_4 > 0$$

$$1 + \alpha_1 - \alpha_4 > 0$$

$$1 + \alpha_4 - \alpha_4^2 - \alpha_4^3 + \alpha_1\alpha_4 > 0.$$

---

4. Jury's result is the analogue for the Schur-Cohn conditions of the Lienard-Chirpart theorem concerning the Routh-Hurwitz conditions (see Anderson and Jury (1973)).

The corresponding stationarity region is illustrated in Figure 1 and shown to have nonlinear boundaries on two of its sides. In contrast to an AR(2) process, for example, this region is not a simplex. Note also that movement from an interior point A across two boundary lines to a point such as B may leave M_* still positive although one is clearly in a nonstationary region.

In summary, then, it can be seen that the behaviour of the full likelihood function reflects one of the necessary conditions for stationarity, namely $|M_*|>0$, but that this alone does not generally guarantee stationarity unless M_* is also constrained to be positive definite.

IV. Modified Gauss-Newton Optimization Procedures

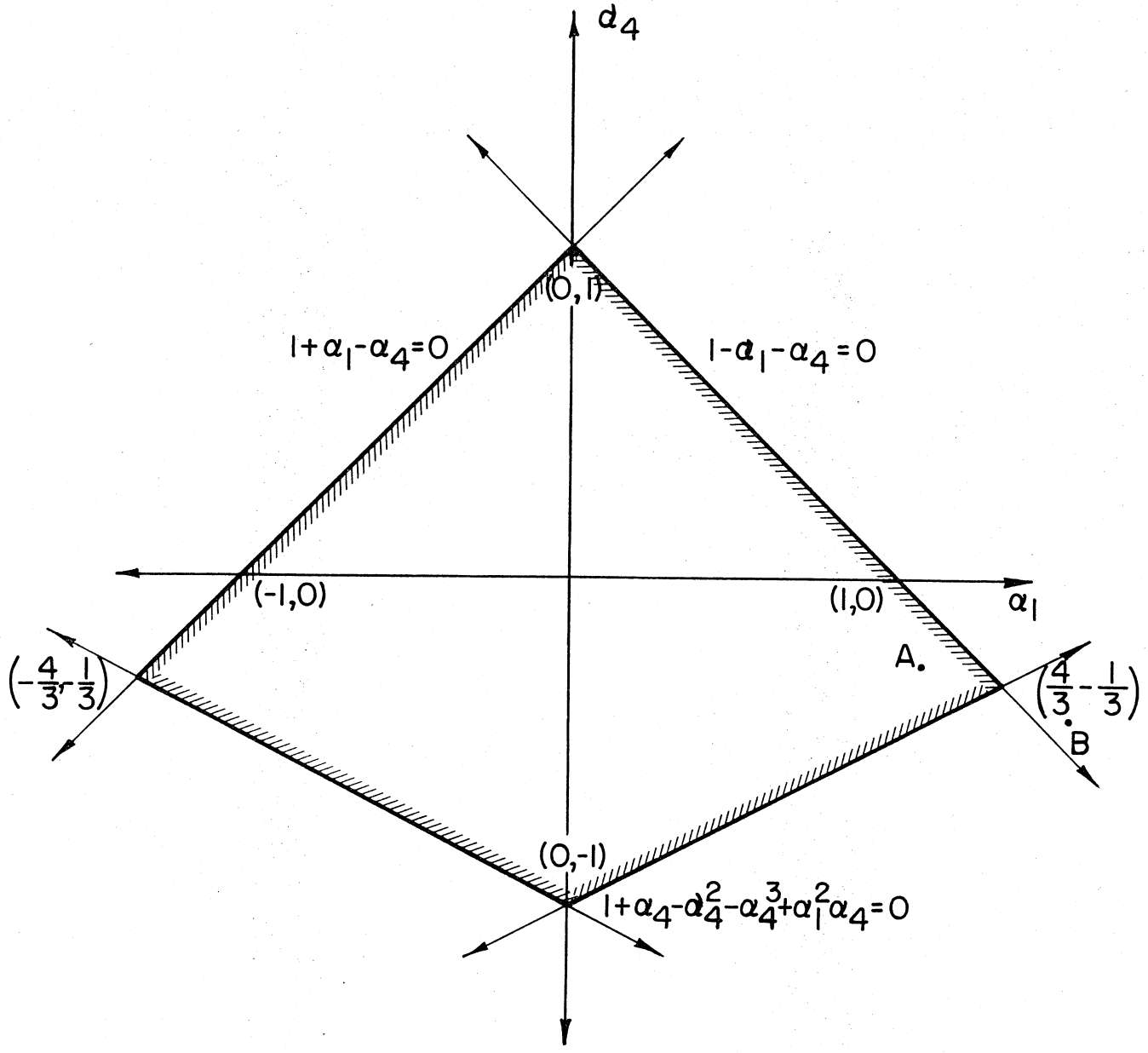
IV.1 Outline of Procedure

The objective function (2.9) is clearly nonlinear in the two sets of parameters, α and β, and the present section outlines the optimization algorithms employed to maximize the concentrated log-likelihood function. General purpose algorithms such as Davidson-Fletcher-Powell and Gill-Murray-Pitfield described, for example, in Wolfe (1978) could be applied directly to (2.9) if these are readily available to the user in packaged form. However, modified Gauss-Newton procedures are relatively straightforward to program and appear to work quite well on the present problem, and so form the basis of the approach of this paper. These procedures employ analytic first and approximate second derivatives, and can be shown to have useful convergence properties.

Perhaps a natural point of departure is to employ the same type

Figure 1

Stationarity Region for the Mixed First- and Fourth-Order AR Process



of iterative algorithm to find stationary points in the full AR(p) case as was suggested by Cochrane and Orcutt (1949) for the approximate AR(1) case, and more recently advocated by Magnus (1978) in a much more general context. Their procedure sequentially maximizes the objective function with respect to the β's holding α fixed, and then with respect to the α's holding β fixed.

The maximum likelihood estimator of β, given particular values of α, is seen to be the Aitken estimator

$$(4.1) \quad \tilde{\beta} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}$$

where the transformed variables $\bar{y} = Qy$ and $\bar{X} = QX$. This may be obtained by numerically computing a lower-triangular matrix R from M_* such that $R'R = M_*$ by a Cholesky decomposition algorithm (Wilkinson and Reinsch (1971)), constructing Q from (2.7), and then running ordinary least squares on the transformed variables. A particularly useful form of the decomposition (Pagano (1973)) is

$$M_* = R_*'DR_* = R_*'L'LR_* = R'R$$

where D and L are diagonal matrices and $R = LR_*$. For a real symmetric matrix M_*, such a decomposition will yield strictly positive diagonal elements on D (as characteristic roots of M_*) if and only if M_* is positive definite. Consequently, this procedure serves the purpose of insuring that the M_* matrix discussed in Theorem 1 above is indeed positive definite, so that the error process is stationary.[5]

---

5. $\ell n|M_*|$ can also then be calculated in the evaluation of the (log) likelihood function as $\Sigma_{i=1}^{p} \ell n(d_{ii})$ since $|R_*| = 1$ (Wilkinson and Reinsch (1971)).

Finding exact ML estimates of the α's given particular values for the β's, however, involves more of a problem. The first-order conditions for α given β yield a set of p nonlinear equations. In the case of a one-parameter (or "simple") AR(p) process, this equation can be solved analytically without resort to numerical methods as presented in Appendix 1 at the end of the paper. However, for general autoregressive processes of higher than first order, iterative gradiant optimization methods are used in the form of a modified Gauss-Newton procedure. If $\tilde{\alpha}^{(i)}$ denotes the i'th iteration estimate of α (for given $\tilde{\beta}$), the procedure iterates on the formula

$$(4.2) \quad \tilde{\alpha}^{(i)} = \tilde{\alpha}^{(i-1)} - \lambda^{(i)} \, G(\tilde{\alpha}^{(i-1)})^{-1} \, g(\tilde{\alpha}^{(i-1)})$$

where $g(\tilde{\alpha}^{(i-1)}) = \left(\frac{\partial L_c}{\partial \alpha}\right)\Big|_{\tilde{\alpha}^{(i-1)}}$ is the gradiant vector of (2.9) with respect to α evaluated at the previous iteration, G is a negative definite matrix representing the metric of the iteration, and $\lambda^{(i)}$ is a variable step size as suggested by Berndt, Hall, Hall, and Hausman (1974). If $\lambda^{(i)} = 1$ for all i and G is the Hessian matrix of second-order derivatives of (2.9), this reduces to the standard Newton-Raphson method (Wolfe (1978)). In the present case, we use analytic derivatives for $g(\alpha)$ based on our explicit expressions for M_* and P, but an approximation to the Hessian matrix for $G(\alpha)$ in order to gain a more robust iteration procedure, as is outlined below. This approximation to the Hessian converts the procedure to a Gauss-Newton algorithm (Bard (1974)), and the addition of a variable step size in order to assist convergence accounts for the "modified" aspect of the algorithm.

IV.2 Calculating the Components of the Algorithm

The first step in setting out the components of this algorithm then is to work out an expression for the gradient vector $g(\alpha) = (\partial L_c/\partial\alpha_1, \ldots, \partial L_c/\partial\alpha_p)'$. Using the result of Sawa (1978), Lemma A.1), one notes that

$$\frac{\partial \ell n|M_*(\alpha)|}{\partial \alpha_i} = tr\ \{ M_*^{-1} \left(\frac{\partial M_*}{\partial \alpha_i}\right)\} \qquad i = 1, \ldots, p,$$

so that now

$$(4.3)\quad g_i(\alpha) = \frac{1}{2} tr\ \{M_*^{-1} \left(\frac{\partial M_*}{\partial \alpha_i}\right)\} - \frac{1}{2\tilde{\sigma}^2} \{ \tilde{u}_*' \left(\frac{\partial M_*}{\partial \alpha_i}\right)\tilde{u}_*$$

$$- 2 \Sigma_{t=p+1}^{T} \tilde{u}_{t-i} (\tilde{u}_t - \alpha_1\tilde{u}_{t-1} - \ldots - \alpha_p \tilde{u}_{t-p})\}$$

by using (2.4) and indicating values conditional upon given $\tilde{\beta}$'s by a tilde on the u's. This can be written alternatively for the full (px1) vector $g(\alpha)$ by letting $u_* = (u_1, \ldots, u_p)'$, $u_+ = (u_{p+1},\ldots,u_T)'$, $u_{-i} = (u_{p+1-i},\ldots, u_{T-i})$ for i = 1, ..., p, and the (T-p) x p matrix $U_t = [U_{-1}, \ldots, U_{-p}]$. Then

$$g(\alpha) = (1/\tilde{\sigma}^2) (\tilde{U}_+'\tilde{u}_+ - \tilde{U}_+'\tilde{U}_+\alpha) + d$$

where the first component of this expression corresponds to the normal equations for regressing the residuals $\tilde{u}_+$ on the lagged residuals $\tilde{u}_{-1}$, ..., $\tilde{u}_{-p}$ while ignoring the initial terms of the likelihood function, and the second component, d, with

$$di = \left(\frac{1}{2\tilde{\sigma}^2}\right) tr\ \{(\tilde{\sigma}^2 M_*^{-1} - \tilde{u}_*\tilde{u}_*' \left(\frac{\partial M_*}{\partial \alpha_i}\right)\},$$

is the initial-term adjustment factor. This can be seen as adjusting the first component for any difference between the calculated covariance for p consecutive errors based on the assumption that the full sample

is covariance stationary ($\sigma^2 M_*^{-1}$) and the estimate based on the initial p observations of the sample ($\tilde{u}_* \tilde{u}_*'$).

M_*^{-1} in (4.3) can be easily obtained by numerical inversion of M_* evaluated at $\tilde{\alpha}^{(i-1)}$. The matrix of partial derivatives $(\frac{\partial M_*}{\partial \alpha_i})$ can also be efficiently computed by using the result of Pagano (1973) that M_* can be expressed as

$$M_* = KK' - LL'$$

where

$$K = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_{p-1} \\ & 1 & \ddots & \alpha_{p-2} \\ & & \ddots & \vdots \\ & & & \alpha_1 \\ & & & 1 \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} \alpha_p & \alpha_{p-1} & \cdots & \alpha_1 \\ & \alpha_p & \cdots & \alpha_2 \\ & & \ddots & \vdots \\ & & & \alpha_p \end{bmatrix}$$

Consequently,

$$(\frac{\partial M_*}{\partial \alpha_i}) = (\frac{\partial K}{\partial \alpha_i})\, K' + K(\frac{\partial K'}{\partial \alpha_i}) - (\frac{\partial L}{\partial \alpha_i})L' - L\,(\frac{\partial L'}{\partial \alpha_i})$$

and $\tilde{u}_*' (\frac{\partial M_*}{\partial \alpha_i})\, \tilde{u}_* = 2\, \tilde{u}_*' [(\frac{\partial K}{\partial \alpha_i})K' - (\frac{\partial L}{\partial \alpha_i})L']\tilde{u}_*$

where

$$(\frac{\partial K}{\partial \alpha_i})K' = \begin{bmatrix} \alpha_i & \alpha_{i-1} & \cdots & \alpha_1 & 1 & 0 & \cdots & 0 \\ \alpha_{i+1} & \alpha_i & \cdots & \alpha_1 & & 1 & \ddots & \\ \vdots & \vdots & & & & & \ddots & 0 \\ \alpha_{p-1} & \alpha_{p-2} & \cdots & \alpha_1 & & & & 1 \\ 0 & \cdots & & & & & & 0 \\ 0 & \cdots & & & & & & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \} \\ \} \end{matrix} \begin{matrix} \\ \\ \\ \\ p-i \text{ rows} \\ i \text{ rows} \end{matrix}$$

for $i = 1, \ldots, p-1$

$= 0$ for $i = p$,

and

$$\left(\frac{\partial L}{\partial \alpha_i}\right)L' = \begin{bmatrix} \alpha_i & \alpha_{i+1} & \cdots & \alpha_p & 0 & \cdots & 0 \\ \alpha_{i-1} & \alpha_i & \cdots & \alpha_{p-1} & \alpha_p & & 0 \\ \vdots & \vdots & & & & \ddots & \\ \alpha_1 & \alpha_2 & & & \alpha_{p-1} & & \alpha_p \\ 0 & \cdots & & \cdots & & \cdots & 0 \\ 0 & \cdots & & \cdots & & \cdots & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \left.\begin{matrix}\\ \\ \\ \\ \end{matrix}\right\} i \text{ rows} \\ \left.\begin{matrix}\\ \end{matrix}\right\} p-i \text{ rows} \end{matrix} \quad \text{for } i=1,\ldots,p.$$

Turning to the second-order matrix G, one notes that minus the Hessian matrix (worked out as H_{11} in Appendix 2) will not always be positive definite (i.e., the (log) likelihood function will not always be concave) for parameter values that are not within a close neighbourhood of the full ML estimates. Consequently, a standard Newton-Raphson algorithm based on the full Hessian matrix for G runs the risk of breaking down if a characteristic root of the Hessian becomes zero or negative as the computer iterates toward a maximum of the likelihood function. A more robust procedure involves choosing a consistent approximation to the Hessian that will be positive definite. It can be seen from the appendix that the dominant term asymptotically corresponds to the set of inner products $u_{-i}' u_{-j}$ for $i,j = 1, \ldots, p$, so that a natural choice for G is

$$(4.4) \quad G = -(1/\tilde{\sigma}^2) \begin{bmatrix} \tilde{u}_{-1}' \tilde{u}_{-1} & \cdots & \tilde{u}_{-1}' \tilde{u}_{-p} \\ \vdots & & \vdots \\ \tilde{u}_{-p}' \tilde{u}_{-1} & \cdots & \tilde{u}_{-p}' \tilde{u}_{-p} \end{bmatrix} = -(1/\sigma^2)\tilde{U}_+'\tilde{U}_+.$$

Convergence to an optimum may not be as rapid as for some procedures using the exact Hessian matrix, but the algorithm is safeguarded against G not

being negative definite.[6] Once convergence has been obtained, $-G^{-1}$ can also serve as the basis for a consistent estimate of the covariance matrix of the $\tilde{\alpha}$'s, although one may wish alternatively to use the actual Hessian matrix in Appendix 2.[7]

Finally, $\lambda^{(i)}$ is computed by a simple quadratic line-search method. Start with $\lambda = 1$ and check whether the (log) likelihood function has been increased. If not, continue trying values of $\lambda = 2^{-j}$ for $j = 1,2,\ldots,J$ until one finds a value of j- for which the (log) likelihood function has increased and then decreased. Then choose the value of λ corresponding to the maximum of a quadratic approximation over this interval (Wolfe (1978)).[8]

The resulting algorithm can be seen to satisfy Berndt, Hall, Hall, and Hausman's (1974) Convergence Theorem conditional on the $\tilde{\beta}$'s, so that in the limit the process is guaranteed to converge to a stationary point which in general will be a local maximum. The convergence criterion for the $\tilde{\alpha}$'s conditional on the $\tilde{\beta}$'s is that

$$\Sigma_{j=1}^{p} \left| \frac{\tilde{\alpha}_j^{(i)} - \tilde{\alpha}_j^{(i-1)}}{\tilde{\alpha}_j^{(i-1)}} \right| < C_1$$

for some chosen value C_1.

---

6. It may be noted that (4.4) is not the only possible choice for G that safeguards against G not being negative definite. One alternative would be to use the outer product of the gradient vectors gg' following Berndt, Hall, Hall, and Hausman (1974), or a truncated version disregarding the initial-time factor (g-d)(g-d)'. However, (4.4) appears to work very well in practice.

7. Its negative definiteness at the optimum insures the second-order conditions for a maximum obtain.

8. If the resulting values of $\tilde{\alpha}^{(i)}$ from (4.2) yields an M_* matrix which is not positive definite, then a smaller value of $\lambda^{(i)}$ is sought by repeating this procedure with a reduced starting value.

Combining the conditional estimates of $\tilde{\beta}$ given $\tilde{\alpha}$ from (4.1) and of $\tilde{\alpha}$ given $\tilde{\beta}$ from the procedure of (4.2), one obtains an iterative algorithm that Oberhofer and Kmenta (1974) show will converge in the limit to a solution of the joint first-order maximizing conditions under fairly general conditions. The convergence criterion for the joint maximization is that

$$(4.5) \quad \text{i)} \quad \sum_{j=1}^{p} \left| \frac{\tilde{\alpha}_j(i) - \tilde{\alpha}_j^{(i-1)}}{\tilde{\alpha}^{(i-1)}} \right| < C_2$$

$$\text{and} \quad \text{ii)} \quad \left| \frac{L_c(\tilde{\beta}^{(i)},\tilde{\alpha}^{(i)}) - L_c(\tilde{\beta}^{(i-1)},\tilde{\alpha}^{(i-1)})}{L_c(\tilde{\beta}^{(i-1)},\tilde{\alpha}^{(i-1)})} \right| < C_3$$

for selected values C_2 and C_3, where $\tilde{\beta}^{(i)}$, $\tilde{\alpha}^{(i)}$ and $\tilde{\beta}^{(i-1)}$, $\tilde{\alpha}^{(i-1)}$ correspond to successive conditional estimates of β given $\tilde{\alpha}$ and of α given $\tilde{\beta}$. Of course, to check that the point of convergence is indeed a global maximum, several different initial values for the parameters can be tried.

IV.3 Extension for Lagged Dependent Variables

In some circumstances, however, the above hemstitching approach of Cochrane and Orcutt may be rather slow to converge if there is marked covariance between the $\tilde{\alpha}$'s and some of the $\tilde{\beta}$'s so that the iso-likelihood contours in α,β-space are very elongated. This is particularly likely to occur in models with lagged dependent variables (Hendry(1976)). The appropriate procedure to follow in this case is to incorporate information on the $\tilde{\alpha}$-$\tilde{\beta}$ covariances and to replace the sequential Cochrane-Orcutt iterative algorithm by one that estimates α and β jointly within each iteration. Indeed, the modified Gauss-Newton algorithm introduced above for the conditional estimates of the α's can be extended very simply to handle the estimation of α and β jointly according to the formula

$$\tilde{\delta}(i) = \tilde{\delta}(i-1) - \lambda(i)\, G(\tilde{\delta}(i-1)^{-1}\; g(\tilde{\delta}(i-1))$$

where now $\delta = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ represents the combined parameter vector. The first-order (gradiant) vector becomes

$$g(\alpha,\beta) = (1/\tilde{\sigma}^2)\begin{bmatrix} U_+'u_+ - U_+'U_+\alpha + d_1 \\ \bar{X}'\bar{y} - \bar{X}'\bar{X}\beta \end{bmatrix} = (1/\tilde{\sigma}^2)\begin{bmatrix} U_+'u_+ - U_+'U_+\alpha + d_1 \\ X_+'y_+ - X_+'X_+\beta + d_2 \end{bmatrix}$$

where d_1 is the vector d defined in IV.2, $d_2 = X_*'(y_* - X_*\beta)$ is the adjustment factor for the degree to which initial residuals and transformed X's are not orthogonal, and

$$\bar{X} = \begin{pmatrix}X\\ X_*\end{pmatrix} \qquad \bar{y} = \begin{pmatrix}y\\ y_*\end{pmatrix}$$

$$X_+ = PX \qquad y_+ = Py$$

$$X_* = [R\,|\,0]X \qquad y_* = [R\,|\,0]y.$$

Corresponding to (4.4), the extended second-order matrix G now becomes

$$G = -(1/\tilde{\sigma}^2)\begin{bmatrix} U_+'U_+ & U_+'X_+ \\ X_+'U_+ & X_+'X_+ \end{bmatrix} = -(1/\tilde{\sigma}^2)Z'Z$$

where the (T-p)x(K+p) matrix $Z = (U_+\,|\,X_+)$. Again the matrix is robust against singularities and also incorporates covariances between the $\tilde{\alpha}$'s and $\tilde{\beta}$'s in its off-diagonal blocks. $\lambda^{(i)}$ is chosen in the same way as before, and the convergence criterion is analogous to (4.5) except that summation is now over all K+p elements in δ. Similarly, Berndt, Hall, Hall, and Hausman's Convergence Theorem again applies for the full δ vector.[9]

---

9. It may be worth remarking that in applying a stationary auto-regressive transform to a dynamic equation, one is treating the regression parameters, α, and the error parameters, β, asymmetrically. One is imposing stationarity on the error process, but not necessarily on the coefficient structure of the lagged dependent variables. However, this may be with good reason. The regression parameters are behavioural coefficients with known

Initial parameter values used for the two maximization algorithms also differ for static and dynamic regression models. In the Cochrane-Orcutt sequential optimization procedure, initial $\tilde{\alpha}$'s were obtained simply by in effect regressing least-squares residuals on their lagged values, $\tilde{U}_+$ and initial $\tilde{\beta}$'s by running the GLS or transformed regression of $\bar{y} = Qy$ on $\bar{X} = QX$. In the joint optimization procedure applied to dynamic models, however, a consistent set of initial values for α were obtained by regressing least-squares residuals on their own lagged values as well as the full set of X regressors for the model. A corresponding set of consistent (and asymptotically efficient) initial estimates for β can then be gotten by regressing the transformed $y_+ = Py$ on the transformed $X_+ = PX$ as well as the full set of lagged least-squares residuals, $\tilde{U}_+$ (Hendry (1976), p. 81).

The various statistical properties of maximum likelihood estimates as yielded by the algorithms of this paper have been well documented elsewhere (for example, Dhrymes (1971) and Magnus (1978)). It is sufficient to note that under fairly general conditions, the final estimates $\hat{\delta}$ will be consistent and asymptotically efficient, and $\sqrt{T}(\hat{\delta}-\delta)$ is asymptotically normal with mean zero and covariance matrix given by $\underset{T\to\infty}{\text{plim}} [-\frac{1}{T}H]$ where H is the Hessian matrix in Appendix 2.

---

economic meaning and one may very well wish to allow nonstationarity in the y's if the data so indicate. The error structure, on the other hand, is parameterized on the basis of relatively little knowledge and generally does not reflect behavioural economic content, so that an assumption of stationarity is a convenient way of making more efficient use of the data available.

V. Some Illustrative Estimates

V.1 Weekly Earnings Equation

The procedures of Section IV are now applied to two regression equations. The first, from Pesaran (1973) and originally from Godley and Nordhaus (1972), is an earnings equation for male and female employees in United Kingdom manufacturing industries using semi-annual data over 1953-69 (34 observations).

$$\log(AWE) = \beta_1 + \beta_2 T + \beta_3 \log(BHR) + \beta_4 \log(\bar{H}) + u$$

where AWE is average weekly earnings, T is a time trend, BHR is the basic hourly wage rate, and the standard-hour-equivalents variable $\bar{H} = HS + \lambda(H-SN)$, where H denotes actual hours worked per week, HS is normal or standard hours per week, and λ is an overtime premium assumed for present convenience to take a value of 1.5. Table 2 presents regression estimates by ordinary least squares and by exact ML procedures for the AR(4) and AR(1) cases. The figures in brackets in the table are "t-ratios" given by dividing the estimated coefficient by an estimate of the asymptotic standard error of the coefficient. As can be seen, the estimated regression coefficients are not insensitive to the assumed error structure. The hours elasticity varies between .95 and .80, and the intercept also varies substantially. Indeed, the estimated hours elasticity on the OLS regression would not appear significantly less than unity, whereas the estimates based on the full AR(4) adjustments appear as 1.87 standard errors less than unity (statistically significant at the 95% confidence level), rather more in keeping with expectations (Godley and Nordhaus (1973), p. 858).

Table 2

Earnings Equation Estimates

	OLS	AR(4)	AR(2)	AR(1)
Reg Coefs				
Const.	-1.2087(2.017)	-.7740(1.481)	-.7660(1.196)	-1.1373(1.880)
T	.0250(6.723)	.0231(8.152)	.0231(7.018)	.0220(5.883)
log(BHR)	.7115(9.044)	.7475(12.05)	.7467(10.37)	.7724(9.767)
log($\bar{H}$)	.9535(7.312)	.8002(7.480)	.7990(6.168)	.8642(7.481)
AR Coefs				
α_1	0	.7252(4.265)	.9775(5.200)	.6852(5.485)
α_2	0	.0583(0.269)	-.3903(1.810)	0
α_3	0	-.2651(1.180)	0	0
α_4	0	-.2343(1.159)	0	0
Summary Statistics				
SSR	$.2328\times10^{-2}$	$.0787\times10^{-2}$	$.1056\times10^{-2}$	$.1224\times10^{-2}$
s	$.828\times10^{-2}$	$.480\times10^{-2}$	$.557\times10^{-2}$	$.602\times^{10-2}$
L	114.77	132.16	127.70	125.38
R^2	.9991	.9990	.9990	.9990
$\chi^2(H_0:AR(0))$		34.78	25.86	21.22

The treatment of autocorrelation also increases the log likelihood function a good deal, and affects the sum of squared residuals and standard error of the regression quite dramatically. Between the OLS and AR(4) estimates, SSR and s decline by 66% and 42% respectively. Since the (log) likelihood functions for exact ML estimation do not involve dropping initial observation in the data, they are readily comparable for likelihood ratio tests on the structure of the autocorrelation. Thus against the null hypothesis of no autocorrelation, twice the difference in log likelihood values is 22.22 for an AR(1) process, 25.86 for an AR(2) process, and 34.78 for an AR(4) prcess, all highly significant on conventional standards. On a test of H_0: AR(2) vs. H_s: AR(4), the (asymptotic) χ^2 statistic is 8.92, while the corresponding statistic for H_0: AR(1) vs H_1: AR(4) is 13.56, both also significant at conventional levels. With a convergence criterion of C_i = .005, the full AR(4) estimates took five iterations to converge and 0.18 seconds of central processor time on the CDC 7600 University of London computer. In summary, then, the exact ML regression estimates appear to be fairly straightforward to calculate and readily adaptable to inference on the autoregressive structure of the disturbance term.

V.2 Demand for Money Equation

The second example that is considered involves a lagged dependent variable, a larger number of regression coefficients, and more than twice the number of observations. The example is a demand for money function from Clinton (1973) estimated with quarterly Canadian data for the period 1956I to 1977IV (88 observations):

$$\log(M/P) = \beta_1 + \Sigma_{j=1}^{3} \beta_{j+1} Q_j + \beta_5 \log (Y) + \beta_6 \log (R) + \beta_7 \log (M_{-1}/P)+u$$

where M is currency plus non-government demand deposits (Clinton's MO, a narrow definition of money), P is the implicit GNE price deflator, the Q_i's are quarterly seasonal dummies (with the fourth quarter dummy omitted), Y is a real GNE income variable, R is an opportunity cost interest rate term related to the 90 day finance paper rate, and M_{-1} is a lagged stock of money term. Estimates for the equation based on OLS and on AR(4) error process are presented in Table 3 along with the standard summary statistics (Clinton's analysis in contrast was based solely on OLS procedures).

As can be seen, autocorrelation in this example is not particularly severe. Consequently, the regression coefficient estimates are not greatly different between the two procedures except for some reduction in the lag coefficient from .888 to .859. As a result, the mean lag of adjustment is reduced from the rather lengthy value of two years to one and a half (and the lag distribution variance from 72 quarters to 43). The short-run or impact elasticities on income and interest rates have been slightly increased (in absolute value) as a result of the autocorrelation adjustment, but the long-run or cumulative elasticities have been reduced -- from 1.056 to .8882 for Y, and from -.3540 to -.3008 for R. Thus it appears that the overall pattern of distributed lag effects has been substantially shifted forward in time. In addition, the long-run income elasticity of demand for money is in accordance with the more traditional view that it is less than unity (and indeed statistically significantly less than unity), while the OLS estimates suggest an approximate unit elastic effect.

Table 3

Money Demand Equation Estimates

	OLS		AR(4)	
Reg. Coefs.				
const	-.5562(6.457		-.5778(6.292)	
Q_1	-.0452(8.690)		-.0448(10.23)	
Q_2	.0000(0.005)		-.0004(0.107)	
Q_3	-.0082(1.563)		-.0091(2.032)	
log(Y)	.1174(3.771)	LR Elas 1.056	.1255(3.893)	LR Elas .888
log R	-.0394(5.655)	-.354	-.0425(5.681)	- .301
$\log(M_{-1}/P)$	.8882(16.67)	Av.Lag 24.0mths	.8587(15.89)	Av.Lag 18.2mths
AR Coefs.				
α_1	0		.1989(1.865)	
α_2	0		-.1238(1.191)	
α_3	0		.2991(2.877)	
α_4	0		-.1618(1.521)	
Summary Statistics				
SSR	1.633×10^{-2}		1.439×10^{-2}	
s	1.367×10^{-2}		1.279×10^{-2}	
L	253.17		258.56	
R^2	.9944		.9944	

The autocorrelation, though not severe, is still statistically significant at the five percent level. In a test of H_0: AR(0) against H_1:AR(4), twice the difference in the log likelihood functions comes to 10.78 compared to the critical value of a χ^2 variate with four degrees of freedom at the 95% level of confidence of 9.49, so that the null hypothesis is just rejected. This corroborates the results of Wallis (1972) that, with quarterly data, seasonal dummies may not be sufficient to take account fully of autocorrelation and a fourth-order autocorrelation adjustment may be appropriate as well. Indeed, null hypotheses of an AR(1/4) and of a simple AR(4) process are also each rejected against the alternative of a general AR(4) error process, so that fairly simple processes may not be as effective as a general fourth-order AR process in adjusting for quarterly autocorrelation. The standard error and sum of squared residuals are again reduced by autoregressive estimation but rather more moderately this time, and both regression fits appear to be fairly close. With the same convergence factor as before, the full AR(4) estimates took nine iterations to converge and 0.85 seconds of CPU time on a CDC 7600 machine.

Finally, it may be worth pointing out that both the OLS and AR(4) regression estimates appear to be noticeably different from those obtained by Clinton over the period 1955-1970. The income and interest rate elasticities in Table 3 are lower (in magnitude) while the lagged adjustment coefficients are higher. This suggests that there may have been some significant structural change in the money demand function recently that would be worth further investigation. This contrasts with one of Clinton's principal conclusions that money, narrowly defined, has a stable demand function for Canada, at least over the 1970's period.

VI. Conclusions

Several conclusions may be made on the basis of the work in this paper. First of all, exact maximum likelihood estimation of regressions with general autoregressive error processes appears to be computationally quite practical and readily lends itself to inference on the autoregressive structure of the disturbances. Indeed, making inferences about the autoregressive error structure near the stationarity boundary is a particular advantage of the method. Secondly, with quarterly data being heavily used nowadays to estimate regression models, it may be appropriate to consider more general autoregressive processes for the disturbances than a simple one-parameter AR(4) process, even when a set of seasonal dummy variables is used as well in the equation. It has been found, for example, that even when regression estimates appear to be fairly robust to the treatment of autocorrelation, distributed lag patterns may still be quite sensitive to the appropriate treatment of autocorrelation, so that fairly careful efforts should be made to identify jointly the appropriate error structure and dynamics of a regression model.

Appendix 1

Estimation of "Simple" AR(p) Processes

In the case of a simple (i.e., one-parameter) AR process, the estimation procedure can be substantially simplified from that given in Section IV of the paper in that an analytic solution for the autoregressive parameter estimates conditional on the β's can now be obtained, so that iterative numerical techniques (such as Gauss-Newton) need not be resorted to.[1]

If the simple AR error process is of p'th order,

$$u_t = \alpha\, u_{t-p} + \varepsilon_t$$

and the log-likelihood function for a given sample of the u's simplifies to

$$L(\alpha,\sigma^2;u) = \text{const} - \frac{T}{2}\,\ell n\,\sigma^2 + \frac{p}{2}\,\ell n\,(1-\alpha^2)$$

$$\frac{-1}{2\sigma^2}\left[(1-\alpha^2)\,\Sigma_1^p\, u_t^2 + \Sigma_{p+1}^T\,(u_t - \alpha u_{t-1})^2\right].$$

Maximizing the concentrated log-likelihood function with respect to α conditional on the β regression coefficients yields the first-order condition

$$\frac{-\alpha p}{(1-\alpha^2)} + \frac{T[\alpha\Sigma_1^p\,\tilde{u}_t^2 + \Sigma_{p+1}^T\,\tilde{u}_{t-p}\,(\tilde{u}_t - \alpha\tilde{u}_{t-p})]}{(1-\alpha^2)\,\Sigma_1^p\,\tilde{u}_t^2 + \Sigma_{p+1}^T\,(\tilde{u}_t - \alpha\tilde{u}_{t-p})^2} = 0.$$

Rearrangement leads to the cubic equation

---

1. This result is in contrast to work by Thomas and Wallis (1971) where approximate methods of ML estimation are employed for the simple AR(4) model, and Pagan (1974, p. 275) where numerical optimization methods are employed at each stage again for the simple AR(4) model.

$$f(\alpha) \equiv \alpha^3 + u\alpha^2 + b\alpha + c = 0$$

where $a = -(1/d)(T-2p)\ \Sigma_{p+1}^{T}\ \tilde{u}_t\ \tilde{u}_{t-p}$

$$b = (1/d)[(T-p)\Sigma_1^p\ \tilde{u}_t^2 - T\Sigma_{p+1}^{T}\ \tilde{u}_{t-p}^2 - p\ \Sigma_{p+1}^{T}\ \tilde{u}_t^2]$$

$$c = (1/d)\ T\ \Sigma_{p+1}^{T}\ \tilde{u}_t\ \tilde{u}_{t-p}$$

and $d = (T-p)\ \Sigma_{p+1}^{T-p}\ \tilde{u}_t^2$.

This equation can be shown to have three real roots, the middle of which always lies in the unit interval and is found by setting

$$g = b - a^2/3 \qquad q = c = ab/3 + 2a^3/27$$

$$\phi = \text{Cos}^{-1}\left[\frac{\sqrt{27}\ q}{2g\sqrt{-g}}\right]$$

and $\tilde{\alpha} = -2\sqrt{\frac{-g}{3}}\ \text{Cos}\ (\frac{\phi}{3} + \frac{\pi}{3}) - \frac{a}{3}$.

This provides an exact solution for $\tilde{\alpha}$, given β, which is computationally much simpler than the iterative procedure of Section IV. The procedure of Beach and MacKinnon (1978) is a special case of this corresponding to p = 1.

An estimate of the standard error of the final estimate $\hat{\alpha}$ once convergence has been attained between $\tilde{\alpha}$ and $\tilde{\beta}$ assuming fixed X's is given by the square root of the inverse of

$$\frac{-\partial^2 L}{\partial \alpha^2}\bigg|_{\hat{\alpha},\hat{\beta}} \frac{p(1+\hat{\alpha}^2)}{(1-\hat{\alpha}^2)} + \frac{1}{\hat{\sigma}^2}\ \Sigma_{p+1}^{T-p}\ (y_t - X_t\hat{\beta})^2$$

Appendix 2

Hessian Matrix for the Full Likelihood Function

It is useful to recall the matrix derivative results that, if A and B are matrices whose elements are functions of the scalar x, then

$$\frac{\partial A^{-1}}{\partial x} = - A^{-1} \left(\frac{\partial A}{\partial x}\right) A^{-1}$$

and $$\frac{\partial tr}{\partial x} (AB) = tr \left\{\left(\frac{\partial A}{\partial x}\right) B\right\} + tr \left\{A\left(\frac{\partial B}{\partial x}\right)\right\}.$$

Also, let the first-order condition for σ^2 be denoted

$$\sigma^2 = q(\alpha,\beta) = u'Mu/T \text{ where } M = Q'Q$$

and where hats and tildes are omitted for convenience. Then the second-order derivative Hessian matrix for $\delta = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ may be calculated as

$$H = \left[\frac{\partial^2 L_c}{\partial \delta_i \delta_j}\right] = \begin{bmatrix} \partial^2 L_c/\partial\alpha_i\partial\alpha_j & \partial^2 L_c/\partial\alpha_i\beta_j \\ \partial^2 L_c/\partial\beta_i\,\partial\alpha_j & \partial^2 L_c/\partial\beta_i\partial\beta_j \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

Turning to the upper-left submatrix, one recalls from Section IV.2 that

$$\frac{\partial L_c}{\partial \alpha_i} = \frac{1}{2} tr \left\{ M_*^{-1} \left(\frac{\partial M_*}{\partial \alpha_i}\right)\right\} - \frac{1}{2\sigma^2} \frac{\partial u'Mu}{\partial \alpha_i}$$

$$= \frac{1}{2} tr \left\{M_*^{-1} \left(\frac{\partial M_*}{\partial \alpha_i}\right)\right\} - \frac{1}{2\sigma^2} \left\{u_*' \left(\frac{\partial M_*}{\partial \alpha_i}\right) u_* \right.$$

$$\left. - 2 \sum_{t=p+1}^{T} u_{t-i}(u_t - \alpha_1 u_{t-1} - \ldots - \alpha_p u_{t-p})\right\}$$

$$\therefore (H_{11})_{ij} = \frac{\partial^2 L_c}{\partial\alpha_i \partial\alpha_j} = \frac{1}{2}\frac{\partial tr}{\partial\alpha_j}\{M_*^{-1}\left(\frac{\partial M_*}{\partial\alpha_i}\right)\} - \frac{1}{2\sigma^2}\{u_*'\left(\frac{\partial^2 M_*}{\partial\alpha_i\partial\alpha_j}\right)u_*$$

$$+ 2\,\Sigma u_{t-i}u_{t-j}\,\}$$

$$- \frac{1}{2}\frac{\partial}{\partial\sigma^2}\{\frac{1}{\sigma^2}\,\frac{\partial u'Mu}{\partial\alpha_i}\}\,\frac{\partial q}{\partial\alpha_j}$$

$$= \frac{1}{2}\,tr\,\{(\sigma^2 M_*^{-1} - u_*u_*')\left(\frac{\partial^2 M_*}{\partial\alpha_i\partial\alpha_j}\right) - N_jN_i\} - (1/\sigma^2)u_{-i}'u_{-j}$$

$$+ \frac{1}{2T(\sigma^2)^2}\left(\frac{\partial u'Mu}{\partial\alpha_i}\right)\left(\frac{\partial u'Mu}{\partial\alpha_j}\right)$$

where $N_i = M_*^{-1}\left(\frac{\partial M_*}{\partial\alpha_i}\right)$ and $\left(\frac{\partial u'Mu}{\partial\alpha_i}\right) = u_*'\left(\frac{\partial M_*}{\partial\alpha_i}\right)u_* - 2u_{-i}'(u_+ - U_+\alpha)$.

The only new expression introduced here that will not have already been computed for the first-order condition is $\left(\frac{\partial^2 M_*}{\partial\alpha_i\partial\alpha_j}\right)$. Using the result of Pagano (1973) referred to in the text, one can obtain a fairly simple expression for this. Let $k = \max(i,j)$, $\ell = \min(i,j)$, $s = \min(i,p-i)$, and $q = |i-j|$. Then, if $i=j$,

$$\left(\frac{\partial^2 M_*}{\partial\alpha_i\partial\alpha_j}\right) = \text{Diag}\,(\underbrace{0,\ldots,0}_{s};\ \underbrace{d,\ldots,d}_{p=2_i};\ \underbrace{0,\ldots,0}_{s})$$

where $d = -2$ if $p > 2_i$

$d = 0$ if $p = 2_i$

and $d = 2$ if $p < 2i$.

If $i \neq j$ and $p = i + j$,

$$\left(\frac{\partial^2 M_*}{\partial\alpha_i\partial\alpha_j}\right) = 0_p$$

If $i \neq j$ and $p > i+j$,

$$\left(\frac{\partial^2 M_*}{\partial\alpha_i \partial\alpha_j}\right) = \begin{array}{c} \\ q \\ \\ \ell \\ \\ p-(1+j) \\ \\ \ell \end{array} \overbrace{\begin{array}{cc} 0 & 0 \\ 0 & \\ 0 & \end{array}}^{q} \overbrace{\begin{array}{ccc} 0 & & \\ & 1 & \\ & & 0 \\ & 1 & \\ & & 1 \\ & & 0 \end{array}}^{\ell} \overbrace{\begin{array}{c} 0 \\ \\ 1 \\ \\ \\ 0 \end{array}}^{p-(i+j)} \overbrace{\begin{array}{c} 0 \\ \\ \\ \\ \\ 0 \end{array}}^{\ell}$$

If $i \neq j$ and $p < i+j$,

$$\left(\frac{\partial^2 M_*}{\partial\alpha_i \partial\alpha_j}\right) = \begin{array}{c} \\ q \\ \\ p-k \\ \\ (i+k)-p \\ \\ p-k \end{array} \overbrace{\begin{array}{cc} 0 & 0 \\ 0 & \\ 0 & \end{array}}^{q} \overbrace{\begin{array}{ccc} 0 & & \\ & -1 & \\ & & 0 \\ -1 & & \\ & & -1 \\ & 0 & \end{array}}^{p-k} \overbrace{\begin{array}{c} 0 \\ \\ -1 \\ \\ \\ 0 \end{array}}^{(1+j)-p} \overbrace{\begin{array}{c} 0 \\ \\ \\ \\ \\ 0 \end{array}}^{p-k}$$

Much simpler to compute, on the other hand, is the submatrix

$$H_{22} = \frac{\partial^2 L_c}{\partial\beta\partial\beta} = -\left(\frac{1}{\sigma^2}\right) \bar{X}'\bar{X} = -\left(\frac{1}{\sigma^2}\right) X'MX.$$

Similarly,

$$H_{21} = \frac{\partial^2 L_c}{\partial\beta\partial\alpha} = \left(\frac{1}{\sigma^2}\right) X' \left(\frac{\partial M}{\partial\alpha}\right)u \quad \text{and} \quad H_{12} = H_{21}'.$$

Bibliography

Ali, M.M., "Analysis of Autoregressive-Moving Average Models: Estimation and Prediction", *Biometrika*, vol. 64(1977), pp. 535-545.

Ali, M.M., "Estimation of Regression Models with ARMA Errors", Discussion Paper (1978).

Anderson, B.D.O. and E.I. Jury, "A Simplified Schur-Cohn Test", *IEEE Trans. Aut. Control*, vol. 18 (1973), pp. 157-163.

Anderson, T.W. and R.P. Mentz, "The Generalized Variance of a Stationary Autoregressive Process", *Journal of Multivariate Analysis*, vol. 7 (1977), pp. 584-588.

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

Beach, Charles M. and James G. MacKinnon, "A Maximum Likelihood Procedure for Regressions with Autocorrelated Disturbances", *Econometrica*, vol. 45 (1978), pp. 51-58.

Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman, "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement*, vol. 3 (1974), pp. 653-665.

Box, George E.P. and Gwilym M. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day, 1970).

Clinton, Kevin, "The Demand for Money in Canada, 1955-70: Some Single-Equation Estimates and Stability Tests", *Canadian Journal of Economics*, vol. 6 (1973), pp. 53-61.

Cochrane, D. and G.H. Orcutt, "Applications of Least Squares Regressions to Relationships Containing Auto-correlated Error Terms", *Journal of the American Statistical Association*, vol. 44 (1949), pp. 32-61.

de Gooijer, J.G., "On the Inverse of the Autocovariance Matrix for a General Mixed Autoregressive Moving Average Process", *Statistische Hefte*, vol. 19 (1978), pp. 114-123.

Dent, Warren, T., "Computation of the Exact Likelihood Function of an ARIMA Process", *Journal of Statistical Computation and Simulation*, vol. 5 (1977), pp. 193-206.

Dent, Warren and An-Sik Min, "A Monte Carlo Study of Autoregressive Integrated Moving Average Processes", *Journal of Econometrics*, vol. 7 (1978), pp. 23-55.

Dhrymes, Phoebus J., *Distributed Lags: Problems of Formulation and Estimation* (San Francisco: Holden-Day, 1971).

Folsom, R.N., D.C. Boger and H.C. Mullikin, "Stability Conditions for Linear Constant Coefficient Difference Equations in Generalized Differenced Form", *Econometrica*, vol. 44 (1976), pp. 575-591.

Fuller, W.A. and J.E. Martin, "The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models", *Journal of Farm Economics*, vol. 43 (1961), pp. 71-82.

Galbraith, R.F. and J.F. Galbraith, "On the Inverse of Some Patterned Matrices Arising in the Theory of Stationary Time Series", *Journal of Applied Probability*, vol. 11 (1974), pp. 63-71.

Godfrey, L.G., "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables", *Econometrica*, vol. 46 (1978), pp. 1303-1310.

Hannan, E.J., "The Estimation of ARMA Models", *Annals of Statistics*, vol. 3 (1975), pp. 975-981.

Harvey, A.C. and G.D.A. Phillips, "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances", *Biometrika*, vol. 66 (1979), pp. 49-58.

Hendry, David F., "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process", *International Economic Review*, vol. 12 (1971), pp. 257-272.

_____, "The Structure of Simultaneous Equations Estimators", *Journal of Econometrics*, vol. 4 (1976), pp. 51-88.

Hendry, David F. and G.E. Mizon, "Serial Correlation as a Convenient Simplification Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England," *The Economic Journal*, vol. 88 (1978), pp. 549-563.

Hendry, David F. and P.K. Trivedi, "Maximum Likelihood Estimation of Difference Equations with Moving Average Errors: A Simulation Study", *The Review of Economic Studies*, vol. 39 (1972), pp. 117-145.

Jury, E.I. "A Simplified Stability Criterion for Linear Discrete Systems", *Proc IRE*, vol. 50 (1962), pp. 1493-1500.

Kadiyala, K.R., "A Transformation Used to Circumvent the Problem of Autocorrelation", *Econometrica*, vol. 36 (1968), pp. 93-96.

Kalman, R.E., "On the Hermite Fajuvara Theorem in Stability Analysis", *Quarterly Journal of Applied Mathematics*, vol. 23 (1965), pp. 279-282.

Lempers, F.B. and T. Kloek, "On a Simple Transformation for a Second-Order Autocorrelated Disturbance in Regression Analysis", *Statistic Neerlandica*, vol. 27 (1973), pp. 69-75.

Magnus, Jan R., "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", *Journal of Econometrics*, vol. 7 (1978), pp. 281-312.

Marden, M., *The Geometry of Polynomials*. Mathematical Surveys No. 3 (Providence, R.I.: American Mathematical Society, 1966).

Newbold, Paul, "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process", *Biometrika*, vol. 61 (1974), pp. 423-437.

Newbold, Paul and N. Davies, "Error Misspecification and Spurious Regressions", *International Economic Review*, vol. 19 (1978), pp. 513-519.

Oberhofer, W. and J. Kmenta, "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models", *Econometrica*, vol. 42 (1974), pp. 579-590.

Pagan, Adrian, "A Generalized Approach to the Treatment of Autocorrelation", *Australian Economic Papers*, vol. 13 (1974), pp. 267-280.

Pagano, M., "When is an Autoregressive Scheme Stationary?", *Communications in Statistics*, vol. (1973), pp. 533-544.

Pesaran, M.H., "Exact Maximum Likelihood Estimation of a Regressive Equation with a First-Order Moving-Average Error", *Review of Economic Studies*, vol. 40 (1973), pp. 529-535.

Pierce, D.A., "Least Squares Estimation in Dynamic Disturbance Time Series Models", *Biometrika*, vol. 59 (1972), pp. 73-78.

Power, H.M., "A Note on the Hermite-Fujuvana Theorem", *Electronics Letters*, vol. 6 (1970), pp. 39-40.

Prothero, L. and K.F. Wallis, "Modelling Macroeconomic Time Series (with Discussion)", *Journal of the Royal Statistical Society*, Series A, vol. 139 (1976), pp. 468-500.

Samuelson, P.A., "Conditions That the Roots of a Polynomial be Less Than One in Absolute Value", *Annals of Mathematical Statistics*, vol. 21 (1941), pp. 360-364.

Sawa, Takamistsu, "The Exact Moments of the Least Squares Estimator for the Autoregressive Model", *Journal of Econometrics*, vol. 8 (1978), pp. 159-172.

Sargan, J.D., "Wages and Prices in the United Kingdom: A Study in Econometric Methodology" in *Econometric Analysis for National Economic Planning* edited by P.E. Hart, G. Mills and J.K. Whitaker (London: Butterworth and Co. Ltd., 1964), pp. 25-54.

Thomas, J.J. and K.F. Wallis, "Seasonal Variation in Regression Analysis", *Journal of the Royal Statistical Society*, Series A, vol. 134 (1971), pp. 54-72.

Wallis, Kenneth F., "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations", *Econometrica*, vol. 40 (1972), pp. 617-636.

Wilkinson, J.H. and C. Reinsch, *Linear Algebra: Handbook for Automatic Computation* (Springer-Verlag, 1971).

Wise, J., "Stationarity Conditions for Stochastic Processes of the Autoregressive and Moving Average Type", *Biometrika*, vol. 43 (1956), pp. 215-219.

Wolfe, M.A., *Numerical Methods for Unconstrained Optimization* (London: Van Nostrand Reinhold, 1978).