BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-87/34

AN ITERATIVE GLS APPROACH TO MAXIMUM LIKELIHOOD ESTIMATION OF REGRESSION MODELS WITH ARIMA ERRORS

by

Mark C. Otto and William R. Bell
Statistical Research Division
U.S. Bureau of the Census
Washington, D.C. 20233

J. Peter Burman
University of Kent
at Canterbury
Canterbury, England

Recommended by: Nash J. Monsour

Report completed: November 1, 1987

Report issued: November 1, 1987

An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with ARIMA Errors

by

Mark C. Otto,
William R. Bell,

Bureau of the Census
Statistical Research Division
Washington, D.C. 20233

and

J. Peter Burman

School of Mathematical Studies
University of Kent
Canterbury, England

For Presentation at the
Annual Meetings of the American Statistical Association
San Francisco, California
August 1987

ABSTRACT

We present a method for estimating regression models with autoregressive integrated moving average (ARIMA) time series errors. The method maximizes the likelihood for different groups of parameters (AR, regression, and MA) separately within each iteration. The idea is to gain numerical efficiency by using generalized least squares (GLS) to maximize the likelihood over the regression and the autoregressive parameters, leaving only the moving average parameter estimates to be obtained by a nonlinear optimization routine. The method uses the "exact likelihood" suggested by Hillmer and Tiao (1979) that is the exact likelihood for pure moving average models. Implementing the method amounts to feeding vectors of the regression and lagged dependent variable to routines that calculate exact likelihood residuals for pure MA models, and then doing regression with these residuals to get the regression and AR parameters. In this way the same software used for exact MA likelihood estimation may be easily modified and used to estimate models with AR and regression effects.

1. INTRODUCTION

Estimating regression models with autoregressive-moving average (ARMA) errors is computationally intensive because the likelihood function is difficult to calculate, especially when the exact form of the likelihood is used. This difficulty is a major factor when numerical derivatives are used in the nonlinear estimation because the likelihood function must be calculated at least twice as many times as there are nonlinear parameters. Most computer packages calculate estimates of the model parameters by directly maximizing the likelihood or approximate likelihood jointly over both the regression (β), autoregression (ϕ), and moving average (θ) parameters (σ^2 can be solved for analytically). Since ϕ and θ are parameters of the covariance function of the data, w_t, and β are parameters of the mean function, this suggests the use of iterative generalized least squares (IGLS) to estimate β, ϕ, and θ. That is, given values for ϕ and θ at one iteration, the next value for β is obtained by GLS; then given a value for β the regression effects are removed and the likelihood is maximized over ϕ and θ only, etc. It also is possible to estimate ϕ using a three part GLS iteration.

IGLS has two advantages. First, the GLS estimation of β given ϕ and θ is a linear least squares problem, while maximizing the likelihood jointly over β, ϕ, and θ is a nonlinear one. The IGLS approach thus reduces the nonlinear optimization problem to maximizing over ϕ and θ at each iteration of β. This has the potential for significant computational savings. Second, since the regression parameter estimates are asymptotically independent of the ARMA parameter estimates (Pierce 1971), the final GLS regression yields the asymptotic covariance matrix of $\hat{\beta}$ explicitly.

Both Jones (1986) and Wincek and Reinsel (1984) have developed methods for separately estimating regression models with ARMA errors using IGLS. Their methods differ by how the covariance matrix , Σ, of the data vector, $\mathbf{w}$, is inverted. Jones uses the Kalman filter and Wincek & Reinsel use a Cholesky decomposition. The method we propose uses Hillmer & Tiao's (1979), hereafter HT, and Ljung & Box's (1979) exact likelihood residual calculation method. Our method calculates exact likelihood residuals (ELR) for both the data, $\mathbf{w}$, and the regression variables, $\mathbf{X}$, and does a GLS regression by doing an ordinary least squares regression (OLS) on the transformed variables. This method would only require slight modifications to existing ARIMA model estimation software to include models with regression terms.

The rest of this paper will, first, describe the model we are attempting to estimate; second, discuss exact maximum likelihood estimation for pure moving average models, and finally, show the IGLS estimation of regression models with ARMA errors. The IGLS section will describe how the likelihood can be rewritten in a GLS framework and give examples for specific types of models.

2. MODEL DESCRIPTION

Regression models with autoregressive-moving average (ARMA) errors for equally spaced data can be described by,

$$\phi(B)(w_t - X_t'\beta) = \theta(B)\Theta(B^s)a_t \quad t = 1 \text{ to } n. \qquad (2.1)$$

Here w_t is a covariance stationary time series, possible differenced; X_t' is a row vector from X, an n x k matrix of possibly differenced regression variables; a_t is the innovation error--assumed to be iid $N(0,\sigma^2)$. $\phi(B)$ is a p-order autoregressive operator,

$$\phi(B)=1-\phi_1 B-\phi_2 B^2-\phi_3 B^3- \cdots -\phi_p B^p,$$

where B is the backshift or lag operator. Powers of B represent the length of the lag, $B^i y_t=y_{t-i}$. $\phi(B)$ need not include all the lags $1,\cdots,p$, so it could thus allow for seasonal lags. We are excluding multiplicative seasonal AR operators, $\phi(B)\Phi(B^s)$, however.

$\theta(B)$ is a q-order regular moving average operator,

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-\theta_3 B^3- \cdots -\theta_q B^q.$$

$\theta(B^s)$ is a Q-order seasonal moving average operator with seasonal period s,

$$\theta(B^s)=1-\theta_1 B^s-\theta_2 B^{2s}-\theta_3 B^{3s}- \cdots -\theta_Q B^{Qs}.$$

The roots of both the regular and seasonal moving average operators must lie on or outside the unit circle and have no common roots with the autoregressive operator. The roots of the autoregressive operator are unconstrained however.

Ljung and Box (1979) and HT document the importance of using exact likelihood methods for the estimation of moving average parameters. We follow HT's approach of conditioning on $w_1,\ldots,w_p$ for estimating AR parameters. We do this for three reasons. First, exact AR likelihood estimation imposes an assumption of stationarity that is unnecessary. In fact, we may wish to use the estimation procedure to check for the presence of unit or explosive AR roots. Exact AR likelihood estimation falls apart as the boundary of the stationarity region is approached. Second, the potential benefits of exact AR estimation derive from the use of the stationary distribution for the first p observations. Even when there is no concern about unit or explosive AR roots, it is not clear to us that this assumption is warranted. It certainly cannot be checked with the data, and if it is wrong it seems the results from exact AR estimation could be worse than those from conditional AR estimation.

Third, even if the stationarity assumption holds, the benefits of exact AR estimation are likely to be meager for series of moderate length, as noted by HT.

3. EXACT MA LIKELIHOOD EVALUATION

We review the method for exact likelihood estimation for MA models. In this section we will derive the exact form of the likelihood for pure MA models, as proposed by HT and Ljung and Box (1976), then, in a later section we will generalize the model to include regression terms. For simplicity, in the remaining sections we will discuss models with only regular MA terms, $\theta(B)$, not the full MA operator, $\theta(B)\Theta(B)$. See HT for details on how to handle multiplicative seasonal models efficiently. First, we will give an overview of the derivation.

The exact likelihood (density of the data) is obtained by relating the data to a set of iid innovations through a linear transformation. Let $\mathbf{a} = (a_1, \cdots, a_n)'$ be the vector of innovations shown in equation (2.1), $\mathbf{a}_* = (a_{1-q}, a_{2-q}, \cdots, a_{-1}, a_0)'$ be the initial innovations, assumed to be from the same stocastic process as the a's, so $\mathbf{a}_* \sim N(0,\sigma^2 I)$. Let $\mathbf{w} = (w_1,\ldots,w_n)'$ be the data vector that is also defined in (2.1), and let $\mathbf{w}_* = (w_{1-q}, w_{2-q}, \cdots, w_{-1}, w_0)'$, be artificial initial values prior to the period of observations. As described in Ljung and Box (1976), we can define $\mathbf{w}_*$ by linearly relating it to $\mathbf{a}_*$ by an arbitrary lower triangular system such that, $|J|$, the jacobian of the transformation between $[\mathbf{a}_*', \mathbf{a}']'$ and

$[w_*', w']'$ is 1. The transformation from a to w and our choice of triangular system will be described below. This transformation allows us to rewrite the exact likelihood in terms of w and w_*,

$$p(w_*,w) = p(a_*,a)|J| = p(a_*,a). \tag{3.1}$$

Now, the joint density, $p(w_*,w)$, can be factored as

$$p(w)p(w_*|w) = p(w_*,w). \tag{3.2}$$

We obtain the desired unconditional density, p(w), (the exact likelihood), by obtaining an expression for $p(w_*,w)$ and identifying p(w) and $p(w_*|w)$ from this expression.

The joint density can also be factored into a density of the data given the initial conditions and a density of the intitial conditions,

$$p(w|w_*)p(w_*) = p(w_*,w).$$

$p(w|w_*)$ is the density used in conditional least squares (CLS) estimation. CLS estimation assumes that $p(w_*) = 1$ for some given initial values usually $w_* = 0$. The exact density is not conditional on the initial values. Now we will review the derivation in detail.

The pure MA model is defined as follows:

$$\mathbf{w}_t = \boldsymbol{\theta}(B)\mathbf{a}_t, \; t = 1, \cdots, n,$$

or (3.3)

$$\mathbf{w}_t = -\theta_q a_{t-q} - \theta_{q-1} a_{t-q-1} - \cdots - \theta_1 a_{t-1} + a_t$$

Notice that the equations for w_1 to w_q require $\mathbf{a}_*$ so we include q more equations to account for the initial conditions. We use Tunnicliffe-Wilson's (1983) choice of triangular system relating $\mathbf{w}_*$ to $\mathbf{a}_*$ which is

$$\begin{aligned}
w_{1-q} &= a_{1-q} \\
w_{2-q} &= -\theta_1 a_{1-q} + a_{2-q} \\
w_{3-q} &= -\theta_2 a_{1-q} - \theta_1 a_{2-q} + a_{3-q} \\
&\vdots \\
w_0 &= -\theta_{q-1} q_{1-q} - \theta_{q-2} a_{2-q} - \cdots - \theta_1 a_{-1} + a_0
\end{aligned} \tag{3.4}$$

Equations 3.3 and 3.4 in matrix form are,

$$\begin{bmatrix} w_{1-q} \\ w_{2-q} \\ \vdots \\ w_0 \\ \hline w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} 1 & & & & & & & & \\ -\theta_1 & 1 & & & & & & & \\ \vdots & & \ddots & & & & & & \\ -\theta_{q-1} & -\theta_{q-2} & \cdots & -\theta_2 & -\theta_1 & 1 & & & \\ -\theta_q & -\theta_{q-1} & \cdots & -\theta_3 & -\theta_2 & -\theta_1 & 1 & & \\ \vdots & & & & & & & \ddots & \\ 0 & & \cdots & 0 & \cdots & -\theta_q & \cdots & -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} a_{1-q} \\ a_{2-q} \\ \vdots \\ a_0 \\ \hline a_1 \\ \vdots \\ a_n \end{bmatrix}$$

and so in an obvious notation

$$\begin{bmatrix} \mathbf{w}_* \\ \hline \mathbf{w} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{a}_* \\ \hline \mathbf{a} \end{bmatrix}$$

The innovations and the initial innovations are related to the data and initial values by the inverse transformation,

$$\mathbf{A}^{-1} \begin{bmatrix} \mathbf{w}_* \\ \hline \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_* \\ \hline \mathbf{a} \end{bmatrix} \qquad (3.5)$$

where $\mathbf{A}^{-1}$ is a matrix of a finite series of π weights from the expansion $\boldsymbol{\pi}(B) = \sum_{j=0}^{\infty} \pi_j B^j = \boldsymbol{\theta}(B)^{-1}$, where $\pi_0 = 1$ (Tunnicliffe-Wilson 1983)

$$A^{-1} = \begin{bmatrix} 1 & & & \\ \pi_1 & 1 & & \\ \vdots & & \ddots & \\ \pi_{n+q-1} & \cdots & \pi_1 & 1 \end{bmatrix}. \qquad (3.6)$$

The π-weights are obtained by equating coefficients in $\theta(B)\pi(B) = 1$. To separate the equations relating to the initial conditions from those relating to the data, we partition A^{-1} so the first q columns are labelled **G** and the remaining n columns are labelled **H**. (3.5) can be rewritten,

$$[G|H]\begin{bmatrix} w_* \\ \hline w \end{bmatrix} = \begin{bmatrix} a_* \\ \hline a \end{bmatrix}$$

or

$$Gw_* + Hw = \begin{bmatrix} a_* \\ \hline a \end{bmatrix}$$

The transformation of the densities in (3.1) now can be described more fully.

$$p(a_*, a) = (2\pi\sigma^2)^{-(n+q)/2}\, e^{-[a_*' \mid a'] \begin{bmatrix} a_* \\ \hline a \end{bmatrix} / 2\sigma^2}$$

implies

$$p(w_*,w) = (2\pi\sigma^2)^{-(n+q)/2}\ e^{-(Gw_* + Hw)'(Gw_* + Hw)/2\sigma^2}$$

for w and w_*. We can see from (3.6) that $|A^{-1}|$ has unit Jacobian. Next, from regression theory or by completing the square the joint sum of squares can be partitioned,

$$(Gw_* + Hw)'(Gw_* + Hw) = (G\hat{w}_* + Hw)'(G\hat{w}_* + Hw) + (w_* - \hat{w}_*)'G'G(w_* - \hat{w}_*)$$

where $\hat{w}_* = -(G'G)^{-1}G'Hw$ is the conditional mean and $(G'G)^{-1}\sigma^2$ is the conditional covariance of w_* given w. Now the the joint density,

$$p(w_*,w) = (2\pi\sigma^2)^{-n/2} e^{-[(G\hat{w}_* + Hw)'(G\hat{w}_* + Hw) + (w_* - \hat{w}_*)'G'G(w_* - \hat{w}_*)]/2\sigma^2},$$

can be factored according to (3.2) and the factors are,

$$p(w) = (2\pi\sigma^2)^{-n/2}|G'G|^{-1/2} e^{-(G\hat{w}_* + Hw)'(G\hat{w}_* + Hw)/2\sigma^2} \tag{3.7}$$

and

$$p(w_*|w) = (2\pi\sigma^2)^{-q/2}|G'G|^{1/2} e^{-(w_* - \hat{w}_*)'G'G(w_* - \hat{w}_*)/2\sigma^2}$$

$p(w)$ is the unconditional density of the data which is the likelihood. Note in (3.7) that G and H are functions of θ.

The joint sum of squares, $(G\hat{w}_* + Hw)'(G\hat{w}_* + Hw)$, is a quadratic form that can be rewritten in terms of **w**, using

$$(G\hat{w}_* + Hw) = Hw - G(G'G)^{-1}G'Hw = (I - G(G'G)^{-1}G')Hw.$$

Let $C = (I - G(G'G)^{-1}G')H$ where $I - G(G'G)^{-1}G'$ is idempotent, thus

$$C'C = H'(I - G(G'G)^{-1}G')H$$

and $(C'C)^{-1}\sigma^2$ is the covariance matrix of **w**. Note that C is the linear transformation of the data to the exact likelihood residuals, ELR's,

$$\begin{bmatrix} \hat{a}_* \\ \hline \hat{a} \end{bmatrix} = Cw \;. \qquad (3.8)$$

Finally, the unconditional density, (3.7), can be written

$$p(w) = (2\pi\sigma^2)^{-n/2} |G'G|^{-1/2} e^{-w'C'Cw/2\sigma^2} . \qquad (3.9)$$

This is the likelihood that is maximized for exact likelihood estimation of pure MA models. We can maximize the exact likelihood by minimizing the deviance, a monotonically decreasing function of $p(\omega)$,

$$\text{deviance} = |G'G|^{1/n}(w'C'Cw/n).$$

Sections 1 and 2 of the appendix includes a derivation of the deviance and a step-by-step description of the pure MA model likelihood calculation.

4. IGLS ESTIMATION OF REGRESSION MODELS WITH ARMA ERRORS

Oberhofer & Kmenta (1974) prove a theorem reguarding iterative procedures for obtaining maximum likelihood estimates when direct maximization with respect to all the parameters is difficult. The theorem applies to regression models with ARMA errors and the result shows that jointly maximizing the likelihood over β, ϕ, and θ can be done by iteratively maximizing it over β given ϕ and θ and visa-versa, i.e.

$$\max_{\beta,\phi,\theta} L(\beta,\phi,\theta) = \max_{\phi,\theta} (\max_{\beta} L(\beta,\phi,\theta))$$

4.1. REGRESSION MODELS WITH AR ERRORS

Conditional least squares estimation of regression models with autoregressive errors provides a simple example of IGLS estimation. Jointly estimating β and ϕ is a nonlinear problem but estimating each seperately is two linear problems. First, let $w_t^f = \phi(B)w_t$, and $X_t^{f}{}' = \phi(B)X_t'$, t=p+1, ··· ,n, where f denotes AR filtering. The likelihood in terms of β given ϕ is,

$$L(\beta|\phi) = (2\pi\sigma^2)^{-(n-p)/2} e^{-(w^f - X^f\beta)'(w^f - X^f\beta)/2\sigma^2} \quad (4.1)$$

and the solution for β is

$$\hat{\beta}^{(i)} = (X^{f(i)\prime}X^{f(i)})^{-1} X^{f(i)\prime}w^{f(i)} \tag{4.2}$$

where i indicates the iteration. The likelihood in terms of ϕ given β is

$$L(\phi|\beta) = (2\pi\sigma^2)^{-(n-p)/2} e^{-(z-Z\phi)'(z-Z\phi)/2\sigma^2} \tag{4.3}$$

where $z_t = w_t - X_t'\beta$ for t = 1, ⋯ ,n, and the columns of Z are the lags of $z = (z_{p+1}, \cdots, z_n)$. The solution for ϕ is

$$\hat{\phi}^{(i)} = (Z^{(i)\prime}Z^{(i)})^{-1} Z^{(i)\prime}z^{(i)} \tag{4.4}$$

where, again, i indicates the iteration. (4.1) is maximized to get a new estimate of β for a given $\hat{\phi}$, then (4.3) is maximized to get a new estimate of ϕ for the given $\hat{\beta}$. Note that in both cases the regression is calculated easily by doing an OLS regression on the transformed variables. In (4.2), the transformation is the AR filter, and in (4.4) transformation is to the regression residuals by $(I-X(X'X)^{-1}X')$. So a difficult nonlinear problem is reduced to a procedure that iterates between two simple linear regressions. Also, since the regression and AR parameter estimates are asymptotically independent (Pierce 1971) their asymptotic variances are obtained from the regressions directly after estimates have converged,

$$\mathrm{var}(\beta) = (X^{f\prime}X^{f})^{-1}\sigma^2$$

and

$$\mathrm{var}(\phi) = (Z'Z)^{-1}\sigma^2.$$

4.2. REGRESSION MODELS WITH MA ERRORS

For regression models with MA errors we modify (3.9), the exact density for pure MA models, to include regression effects,

$$p(w) = (2\pi\sigma^2)^{-n/2}|G'G|^{-1/2}e^{-(w-X\beta)'C'C(w-X\beta)/2\sigma^2} \tag{4.5}$$

For given θ this is maximized over β at

$$\hat{\beta} = (X'C'C\,X)^{-1}X'C'Cw$$

the GLS estimate. At each iteration, a new value of θ is obtained by fixing β, thus updating the inverse covariance matrix, $\Sigma^{-1} = C'C\sigma^{-2}$; then a new β is obtained with updated covariance matrix. These iterations continue until convergence.

Note, as was shown in 3.8, that $\mathbf{C}$ linearly transforms both $\mathbf{w}$ and $\mathbf{X}$ into estimates of exact likelihood residuals, ELR's, $\tilde{\mathbf{w}} = \mathbf{Cw}$ and $\tilde{\mathbf{X}} = \mathbf{CX}$. After the transformation to ELR's, the $\hat{\beta}$ can be obtained by an OLS regression of $\tilde{\mathbf{w}}$ on $\tilde{\mathbf{X}}$,

$$\hat{\beta} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{w}}.$$

In section 2 of the appendix we show how $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{X}}$ are obtained recursively. Finally, the regression estimates are asymptotically independent from the MA parameter estimates, so the q x q covariance matrix of θ is obtained from the inverse of the negative Hessian of the likelihood function (Kendall and Stuart, 1973) and the k x k covariance matrix of β is obtained from the regression,

$$\mathrm{var}(\hat{\beta}) = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\sigma^2.$$

4.3. REGRESSION MODEL WITH ARMA ERRORS

In this section, we discuss three approaches to estimating regression models with ARMA errors: (1), by maximizing $L(\beta,\phi,\theta)$ jointly over β, ϕ, and θ by nonlinear least squares, (2), by a two stage IGLS where ϕ and θ are estimated jointly by nonlinear least squares and β is estimated by iterative GLS regression, and (3), by a three stage process where only θ is estimated by nonlinear least squares and both β and ϕ are estimated by separate GLS regressions, i.e.

$$\max_{\beta,\phi,\theta} L(\beta,\phi,\theta) =$$

$$\max_{\phi,\theta} (\max_{\beta} L(\beta,\phi,\theta)) = \max_{\theta} (\max_{\phi} (\max_{\beta} L(\beta,\phi,\theta)))$$

The two stage procedure is similar to the regression with MA errors except now both the AR and MA parameters are jointly estimated by nonlinear least squares. By letting the n-p x n matrix, L, represent the linear AR filter,

$$L = \begin{bmatrix} -\phi_p & -\phi_{p-1} & - \cdots & -\phi_1 & 1 & & & & \\ & -\phi_p & - \cdots & -\phi_2 & -\phi_1 & 1 & & & \\ & & \vdots & & & \vdots & \ddots & & \\ 0 & & \cdots & & & -\phi_p & - \cdots & -\phi_1 & 1 \end{bmatrix},$$

the joint likelihood (4.5) can be modified to include AR terms,

$$p(w) = (2\pi\sigma^2)^{-(n-p)/2} |G'G|^{-1/2} e^{-(w-X\beta)'L'C'CL(w-X\beta)/2\sigma^2}.$$

The GLS regression estimate for β is

$$\hat{\beta} = (X'L'C'CLX)^{-1}X'L'C'CLw,$$

and the covariance matrix of $\hat{\beta}$ is

$$\text{var}(\hat{\beta}) = (X'L'C'CLX)^{-1}\sigma^2. \tag{4.6}$$

Notice that these are regression results with the data **w** and regression variables **X** filtered by both the AR filter ($\phi(B)\sim L$) and MA-ELR filter (C).

The three stage method is similar to the regression models with AR errors but at the begining of each iteration, estimates of the θ's are obtained by nonlinear least squares, and the ELR's are taken as part of the transformation for each GLS. For each iteration's β GLS the ELR's are taken after the AR filtering and for each iteration's ϕ-GLS the ELR's are taken after the regression residual transformation. Since the step-by-step procedure for this method is fairly involved it is included in section 3 of the appendix. Note that if there are no regression variables this scheme yields an interative GLS approach to maximum likelihood estimation of pure ARIMA models.

In cases where ϕ and θ might be highly correlated the three stage method may become less efficient relative to the other methods. This is likely to occur when there are ϕ and θ parameters at the same lags, such as an ARIMA (1,0,1) model. Three step estimation for models with regular AR and seasonal MA parameters or visa-versa may still be computationally efficient compared to the two stage method because we would suspect that there would be little correlation between regular and seasonal parameters. Also, the ϕ and θ parameters cannot be assumed to be independent from each other so we cannot get the covariance matrix of $\hat{\phi}$ directly, but the covariance matrix of $\hat{\beta}$ is still (4.6.)

5. CONCLUSION

The IGLS method has the potential for large computational savings because β and possibly ϕ are estimated linearly with GLS regressions. For highly correlated the AR and MA parameters, ϕ and θ, we think the IGLS computational efficiency might degrade compared to the joint nonlinear estimation of the ARMA parameters. Also, because the ARMA parameters are not asymptotically independent their variances and covariances cannot be obtained from a separate explicit formula. Finally, the IGLS method requires only slight modification of existing routines in exact likelihood estimation packages for ARIMA models since the same routines used to calculate the exact likelihood residuals of the data are used on the regression variables.

APPENDIX

1. Derivation of the deviance of the exact likelihood for pure MA models

$$L(\theta,\sigma^2) = (2\pi\sigma^2)^{-n/2}|G'G|^{-1/2}e^{-w'C'Cw/2\sigma^2}$$

Take the derivative of L with respect to σ^2, set it equal to zero, and solve. The result is

$$\sigma^2 = w'C'Cw/n.$$

Substitute this back into the likelihood,

$$L(\theta) = (2\pi(w'C'Cw/n))^{-n/2}|G'G|^{-1/2}e^{-n/2}.$$

Then get a transformation of the likelihood that is as free a possible from unnecessary constants; this is called the deviance,

$$(2\pi e/N)L(\theta)^{-2/N} = (w'C'Cw)|G'G|^{1/n}.$$

Now, the deviance is a monotonically decreasing function of $L(\theta)$, so minimizing the deviance will maximize the likelihood.

2. Step-by-step calculation of the exact likelihood (or deviance) for pure MA models. Also, steps A through F calculate the ELR's of **w** and the columns of **X**, so if **w** is **w** in A then $\hat{\mathbf{a}}$ is $\tilde{\omega}$ in step F and is a $\mathbf{X}_{\cdot t}$, a column of **X** is **w** in step a, then $\tilde{\mathbf{X}}_{\cdot i}$ is $\hat{\mathbf{a}}$ in step F.

A. Obtain the CLS residuals, $\mathbf{a}_0 = \mathbf{Hw}$, by the recursion,

$$a_{0t} = 0,\ t = 1-q,\ldots,0$$

$$\theta(B)a_{0t} = w_t,\ t = 1,\ \cdots\ ,n$$

B. Obtain elements of **G'G** by (1) calculating the π-weights, (2) calculating sum of squares and cross products of the π-weights which are the first row and column of **G'G**, and (3) calculating the remaining elements by subtracting single π-weight products from the upper back diagonal element, starting from the first row or column elements. Now in more detail, the n+q π-weights are obtained by the recursion,

$$(1-\theta_1 b-\theta_2 b^2- \cdots -\theta_q b^q)(\pi_0+\pi_1 b+\pi_2 b^2+ \cdots +\pi_{n+q}b^{n+q}) = 1,$$

for all b. The sum of squares and cross products of the π weights, $\gamma_{ij} = \sum_{k=i}^{n+q-j} \pi_k \pi_{k-i}$, are obtained from the recursion,

$$(1-\theta_1 b^{-1}-\theta_2 b^{-2}- \cdots -\theta_q b^{-q})(\gamma_{0,0}+\gamma_{0,1}b^1+\gamma_{0,2}b^2+ \cdots +\gamma_{0,n+q-1}b^{n+q-1})$$
$$= (\pi_0+\pi_1 b+\pi_2 b^2+ \cdots +\pi_{n+q-1}b^{n+q-1}),$$

for all b, where the first q γ_{ij}'s, $\gamma_{0,0}$ to $\gamma_{0,q-1}$, are the first row and column elements (since $\gamma_{ij} = \gamma_{ji}$ and in particular $\gamma_{0,j} = \gamma_{j,0}$) of $\mathbf{G'G} = [\gamma_{ij}]$. Then the remaining elements are obtained by a diagonal recursion,

$$\gamma_{i,j} = \gamma_{i-1,j-1} - \pi_{n+q-i}\pi_{n+q-j},$$
$$i = 1,\ \cdots\ ,\ q-1,$$
$$j = 1,\ \cdots\ ,\ q-1$$

C. Obtain G'Hw = G'a_0 by a recursion,

$$(1-\theta_1 b-\theta_2 b^2- \cdots -\theta_q b^q)(\eta_0+\eta_1 b^{-1}+\eta_2 b^{-2}+ \cdots +\eta_{n+q-1} b^{-n+q-1}) =$$
$$(a_{0,1-q}+a_{0,2-q}b+a_{0,3-q}b^2+ \cdots +a_{0,n}b^{n+q-1}),$$

for all b, and $[\eta_t]$ = G'Hw = G'a_0 for η_t, t= 0 $\cdots$ q-1.

D. Solve the regression equation, $(G'G)\hat{w}_*$ = G'Hw, (e.g. using a Cholesky decomposition of G'G) and get |G'G| at the same time.

E. Obtain the ELR's by the recursion,

$$\theta(B)\hat{a}_t = w_t \ , \ t = 1-q, \cdots ,n$$

where the first q elements of w_t are $\hat{w}_*$ as shown in (3.4), and $\hat{a}_t = 0$ for $t < 1-q$.

F. Obtain the deviance = $|G'G|^{1/n} \sum_{t=1-q}^{n} \hat{a}_t^2$.

3. Step-by-step description of the estimation of regression models with ARMA errors using a three stage procedure with GLS regressions to obtain estimates of both β and ϕ. The initial regression parameter estimates, $\hat{\beta}^{(0)}$, are OLS regression estimates ($\phi = 0$, $\theta = 0$); the initial AR parameter estimates, $\hat{\phi}^{(0)}$, are regression estimates of $z^{(0)} = w - X'\hat{\beta}^{(0)}$ on Z.

A. Estimate the MA parameters, $\hat{\theta}^{(i)}$, using

$z_t^{(i-1)} = w_t - X_t'\beta^{(i-1)}$ is ARMA(p,q)

$z_t^{f(i-1)} = \phi^{(i-1)}(B) z_t^{(i-1)}$ is MA(q)

Estimate $\hat{\theta}^{(i)}$ by nonlinear least squares

B. Estimate the regression parameters, $\hat{\beta}^{(i)}$

$w_t^{f(i-1)} = \phi^{(i-1)}(B) w_t$ is regression + MA(q)

$X_t^{f(i-1)\prime} = \phi^{(i-1)}(B) X_t'$

$\tilde{w}^f = C^{(i)} w^{f(i-1)}, \ \tilde{X}^f = C^{(i)} X^f$ is regression

$$\tilde{X}^{f\prime}\tilde{X}^f\hat{\beta}^{(i)} = \tilde{X}^{f\prime}\tilde{w}^f$$

C. Estimate the autoregression parameters, $\hat{\phi}^{(i)}$

$z_t^{(i)} = w_t - X_t'\hat{\beta}^{(i)}$, $t=p+1, \cdots, n$ is ARMA(p,q)

$Z^{(i)}$ = columns that are the lags of $z^{(i)}$

$\tilde{z}^{(i)} = C^{(i)} z^{(i)}, \ \tilde{Z}^{(i)} = C^{(i)} Z^{(i)}$

$$\tilde{Z}^{(i)\prime}\tilde{Z}^{(i)}\hat{\phi}^{(i)} = \tilde{Z}^{(i)\prime}\tilde{z}^{(i)}$$

D. Check for convergence and repeat steps A - C until iteration converges.

BIBLIOGRAPHY

Hillmer, S. H. and G. C. Tiao (1979) "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," Journal of the American Statistical Association, Vol. 74, No. 367, pp 652-660.

Jones, R. H. (1986) "Time Series Regression with Unequally Spaced Data," Essays in Time Series and Allied Processes, a special volume, No. 23A of the Journal of Applied Probability, pp. 89-98.

Kendall, M.G. & A. Stuart (1973) *The Advanced Theory of Statistics, 3rd Ed., Vol 2,* . New York: Hafner Publishing Co., p. 44.

Ljüng, G. M. and G. E. P. Box (1976) "Studies in the Modelling of Discrete Time Series," Technical Report No. 476, Department of Statistics, University of Wisconsin--Madison.

Pierce, D. A. (1971) "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors," Biometrika, Vol. 58, pp. 299-312.

Tunnicliffe-Wilson, G. (1983) "The Estimation for Time Series Models Part I. Yet Another Algorithm for the Exact Likelihood of ARMA Models," Technical Report No. 2528, Mathematics Research Center, University of Wisconsin--Madison.

Oberhofer, W. and J. Kmenta (1974) "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," Econometrica, Vol. 42, No. 3, pp. 579-590.

Wincek, M. A. and G. C. Reinsel (1984) "An Exact Maximum Likelihood Estimation Procedure for Regression-ARMA Time Series Models with Possibly Nonconsecutive Data," Technical Report No. 749, Department of Statistics, University of Wisconsin--Madison.