Western University
Scholarship@Western

Department of Economics Research Reports
Economics Working Papers Archive

1975

The Estimation of Misspecified Polynomial Distributed Lag Models

R. A. L. Carter

A. L. Nagar

P. G. Kirkham

Follow this and additional works at: https://ir.lib.uwo.ca/economicsresrpt

Part of the Economics Commons

Citation of this paper:

Carter, R. A. L., A. L. Nagar, P. G. Kirkham. "The Estimation of Misspecified Polynomial Distributed Lag Models." Department of Economics Research Reports, 7525. London, ON: Department of Economics, University of Western Ontario (1975).

20227

Research Report 7525

THE ESTIMATION OF MISSPECIFIED POLYNOMIAL DISTRIBUTED LAG MODELS

R. A. L. Carter
University of Western Ontario

A. L. Nagar
Delhi School of Economics

P. G. Kirkham
Statistics Canada

November 1975

THE ESTIMATION OF MISSPECIFIED POLYNOMIAL DISTRIBUTED LAG MODELS[1]

R. A. L. Carter
University of Western Ontario

A. L. Nagar
Delhi School of Economics

and

P. G. Kirkham
Statistics Canada

1. Introduction

Distributed lag models appear quite frequently in economics and a popular way of estimating their coefficients is to specify a polynomial lag structure. This specification is useful because any continuous lag structure can be accurately approximated by a polynomial. The Almon (1965) technique, using Lagrangian interpolation, is often employed, although this procedure gives identical results to those given by the simpler procedure discussed by Dhrymes (1971) unless endpoint restrictions are imposed (Dhrymes (1971), pp. 229-234). However, even if no endpoint restrictions are employed, the specification of a polynomial lag structure can lead to biased, inconsistent parameter estimates if the length of the lag and the degree of the polynomial are incorrectly specified (e.g., Rowley (1971), Schmidt and Waud (1973), Frost (1975)).

The aim of this paper is to give a (unbiased) method of estimating the bias in the coefficient estimates which results from misspecifying the

[1]We have benefited from conversations with W. Haessel but remaining errors are ours alone.

length of the lag and/or the degree of the polynomial. The method can be used to provide an unbiased estimate of the mean squared error of the traditional procedure.

2. Specification of the True Model

We write the true finite distributed lag model as

$$(2.1) \qquad y_t = \sum_{i=0}^{n} \beta_i x_{t-i} + u_t$$

where x_{t-i} is a value of the exogenous variable, n is the unknown length of the lag and the β_i are unknown coefficients to be estimated. Our assumptions about the random disturbance are:

$$(A.1) \qquad u_t \sim N(0,\sigma^2) \text{ for all } t \text{ and}$$

$$(A.2) \qquad E\, x_{t-i} u_t = 0 \text{ for all } t \text{ and } i,$$

that is the random disturbance is independent of all values of the exogenous variable.[2]

If T observations are available on y_t and x_t we can write the model in matrix notation as

$$(2.2) \qquad y = X\beta + u$$

where:

$$y = \begin{bmatrix} y_{n+1} \\ y_{n+2} \\ \vdots \\ y_T \end{bmatrix}, \quad u = \begin{bmatrix} u_{n+1} \\ u_{n+2} \\ \vdots \\ u_T \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_o \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \quad \text{and}$$

---

[2] If x_t is random all expectations would be conditional on the sample of x's.

$$X = \begin{bmatrix} x_{n+1} & x_n & x_q & x_{q-1} & x_1 \\ x_{n+2} & x_{n+1} & x_{q+1} & x_q & x_2 \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n+p} & x_{n+p-1} & x_{q+p} & x_{q+p-1} & x_p \\ \vdots & \vdots & \vdots & & \vdots \\ x_T & x_{T-1} & x_{T-q} & x_{T-q-1} & x_{T-n} \end{bmatrix}$$

Note that both the number of rows and the number of columns of X are affected by the lag length n. We could partition X and β into

(2.3) $X = [X_1 X_2]$, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ where X_2 contains columns 1 to q-1 of X (numbered from right to left), X_1 contains columns q to n+1, β_2 has q-1 elements, β_1 has n-q+2 elements.

Alternatively we could partition X as

(2.4) $X = \begin{bmatrix} X_3 \\ X_4 \end{bmatrix}$ where X_3 has p rows and X_4 has T-n-p rows.

3. Specification of the Estimated Model

Before β can be estimated we must specify what we think n, the lag length, is. There are, of course, two possible, mutually exclusive, errors.

a. Lag Length too Small

If s is our guess for the value of n and $s < n$ then our estimating equation is

(3.1) $y_t = \beta_o x_t + \beta_1 x_{t-1} + \cdots + \beta_s x_{t-s} + u_t$

where $e_t = \beta_{s+1} x_{t-s-1} + \ldots + \beta_n x_{t-n} + u_t$.

In matrix notation we have

(3.2) $y_* = X_* \beta_* + e$

where:

$$y_* = \begin{bmatrix} y_{s+1} \\ \vdots \\ y_n \\ y_{n+1} \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} y_{*1} \\ y \end{bmatrix}$$, y is from the left side of (2.2),

$$\beta_* = \begin{bmatrix} \beta_o \\ \beta_1 \\ \vdots \\ \beta_s \end{bmatrix} = \beta_1$$, a subvector of β from (2.3)

$$X_* = \begin{bmatrix} x_{s+1} & x_s & x_1 \\ \vdots & \vdots & \\ x_n & x_{n-1} & x_{n-s} \\ x_{n+1} & x_n & x_{n-s+1} \\ x_{n+2} & x_{n+1} & x_{n-s+2} \\ \vdots & \vdots & \vdots \\ x_T & x_{T-1} & x_{T-q} \end{bmatrix}$$, letting q = n-s+1 in (2.3),

$$= \begin{bmatrix} X_{*1} \\ X_1 \end{bmatrix}$$, X_1 is a submatrix of X defined in (2.3),

and

$$e_* = \begin{bmatrix} e_{*1} \\ e_{*2} \end{bmatrix} = \begin{bmatrix} e_{*1} \\ X_2\beta_2 + u \end{bmatrix}$$ with X_2 and β_2 defined in (2.3).

Then we can rewrite (3.2) as two equations

(3.3) $y_{*1} = X_{*1}\beta_1 + e_{*1}$ and

(3.4) $y = X_1\beta_1 + e_{*2} = X_1\beta_1 + (X_2\beta_2 + u)$

Equations (3.3) and (3.4) contain different observations on the same variables: a term analogous to $X_2\beta_2$ is contained in e_{*1}. As a simplification[3] we will concentrate on equation (3.4) which is clearly a case of misspecification by the omission of variables. One way to estimate β_1 is to use ordinary least squares (OLS) on (3.4) which gives

$$(3.5) \qquad \hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y = \beta_1 + (X_1'X_1)^{-1}X_1'(X_2\beta_2 + u).$$

Using (A.1), we find the sampling distribution of $\hat{\beta}_1$ to be

$$(3.6) \qquad \hat{\beta}_1 \sim N\left[\beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2, \sigma^2(X_1'X_1)^{-1}\right]$$

The bias and mean squared error of $\hat{\beta}_1$ are given by

$$(3.7) \qquad \text{bias}(\hat{\beta}_1) = (X_1'X_1)^{-1}X_1'X_2\beta_2 \quad \text{and}$$

$$(3.8) \qquad \text{MSE}(\hat{\beta}_1) = \sigma^2(X_1'X_1)^{-1} + (X_1'X_1)^{-1}X_1'X_2\beta_2\beta_2'X_2'X_1(X_1'X_1)^{-1}.$$

Alternatively we may try to gain precision by restricting each element β_i of β_1 to be a polynomial of degree k in i, that is,

$$(3.9) \qquad \beta_i = \sum_{j=o}^{k} i^j\alpha_j + \rho_i \quad \text{for } i=0,\cdots,s$$

In (3.9) the α_j are unknown coefficients and ρ_i is an unknown remainder to account for the possibility that this specification of the lag structure may be only approximately correct. In matrix notation this specification is

$$(3.10) \qquad \beta_1 = \Gamma\alpha + \rho$$

where:

$$\Gamma = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ 1 & 1 & 1 & & 1 \\ 1 & 2 & 4 & & 2^k \\ \vdots & \vdots & \vdots & & \\ 1 & s & s^2 & \ldots & s^k \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_o \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_k \end{bmatrix}, \quad \rho = \begin{bmatrix} \rho_o \\ \rho_1 \\ \cdot \\ \cdot \\ \cdot \\ \rho_s \end{bmatrix}$$

---

[3] This amounts to ignoring n-s observations and, even though n-s may be small, it is not something an applied econometrician would do. This omission does not fundamentally alter the results but it allows us to present them somewhat more clearly.

Then (3.4) becomes

$$(3.11) \qquad y = X_1\Gamma\alpha + X_1\rho + (X_2\beta_2 + u) = Z_1\alpha + X_1\rho + (X_2\beta_2 + u).$$

Traditionally the last two terms in (3.11) are ignored and α is estimated by

$$(3.12) \qquad \hat{\alpha} = (Z_1'Z_1)^{-1}Z_1'y = \alpha + (Z_1'Z_1)^{-1}Z_1'(X_1\rho + X_2\beta_2 + u)$$

and the restricted least squares (RLS) estimate of β_1 is

$$(3.13) \qquad \tilde{\beta}_1 = \Gamma\hat{\alpha} = \Gamma\alpha + \Gamma(Z_1'Z_1)^{-1}Z_1'(X_1\rho + X_2\beta_2 + u).$$

The distribution of $\tilde{\beta}_1$ is

$$(3.14) \qquad \tilde{\beta}_1 \sim N\left[\Gamma\alpha + \Gamma(Z_1'Z_1)^{-1}Z_1'(X_1\rho + X_2\beta_2), \sigma^2\Gamma(Z_1'Z_1)^{-1}\Gamma'\right]$$

so that its bias and MSE are

$$(3.15) \qquad \text{bias}(\tilde{\beta}_1) = \Gamma(Z_1'Z_1)^{-1}Z_1'(X_1\rho + X_2\beta_2) - \rho \quad \text{and}$$

$$(3.16) \qquad \text{MSE}(\tilde{\beta}_1) = \sigma^2\Gamma(Z_1'Z_1)^{-1}\Gamma' + \left[\Gamma(Z_1'Z_1)^{-1}Z_1(X_1\rho + X_2\beta_2) - \rho\right] \cdot \left[\Gamma(Z_1'Z_1)^{-1}Z_1(X_1\rho + X_2\beta_2) - \rho\right]' .$$

Whether $\tilde{\beta}_1$ is, in fact, any more precise than $\hat{\beta}_1$ can be decided comparing the MSE matrices in equations (3.8) and (3.16). The difference between these two matrices depends upon the size of ρ which in turn depends upon how closely (3.9) approximates the true lag structure. It is clear though that the precision of both $\hat{\beta}_1$ and $\tilde{\beta}_1$ is overstated by their covariance matrices. However, since X_2 is unknown[4] it is not possible to estimate the bias and MSE of $\hat{\beta}_1$ and $\tilde{\beta}_1$.

---

[4] Of course, the only reason X_2 is unknown is that n is unknown. If n were known X_2 could be constructed in the same fashion as X is constructed.

b. Lag Length too Large

The other possible error we can make in specifying the lag length is to set $s > n$. In this case we have

$$(3.17) \qquad y_t = \beta_o x_t + \beta_1 x_{t-1} + \cdots + \beta_n x_{t-n} + \beta_{n+1} x_{t-n-1} + \cdots + \beta_s x_{t-s} + u_t$$

where $\beta_i = 0$ for $i = n+1, \cdots, s$. In matrix notation

$$(3.18) \qquad y_* = X_* \beta_* + u_*$$

where

$$y_* = \begin{bmatrix} y_{s+1} \\ \vdots \\ y_T \end{bmatrix}$$, the last T-s elements from y (similarly, u_* is the last T-s elements from u),

$$\beta_* = \begin{bmatrix} \beta_o \\ \beta_1 \\ \vdots \\ \beta_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$$, the vector β with s-n zeros added,

$$X_* = \begin{bmatrix} x_{s+1} & \cdots & x_1 \\ \vdots & & \\ x_T & \cdots & x_{T-s} \end{bmatrix} = \begin{bmatrix} x_{n+p} & \cdots & x_p & x_{p-1} & \cdots & x_1 \\ & & & & & \\ x_T & \cdots & x_{T-n} & x_{T-n-1} & \cdots & x_{T-s} \end{bmatrix}$$

$$= \begin{bmatrix} X_4, X_5 \end{bmatrix} \text{ from (2.4) with } p = s+1-n \text{ and } X_5 = \begin{bmatrix} x_{p-1} & \cdots & x_1 \\ & & \\ x_{T-n-1} & \cdots & x_{T-s} \end{bmatrix}$$

Then we can rewrite (3.18) as

$$(3.19) \qquad y_* = X_4 \beta + X_5 0 + u_* = X_4 \beta + u_*$$

where X_4 is just X with the first s-n rows omitted; see (2.4).

Using OLS on (3.18) gives

$$(3.20) \qquad \hat{\beta}_* = (X_*'X_*)^{-1}X_*'y_* = \beta_* + (X_*'X_*)^{-1}X_*'u_*$$

The distribution of $\hat{\hat{\beta}}_*$ is

$$(3.21) \qquad \hat{\beta}_* \sim N\left[\beta_*, \sigma^2(X_*'X_*)^{-1}\right]$$

so that this estimator is unbiased with an MSE equal to its covariance matrix.

Specification of a polynomial lag structure of degree k implies imposition of the restrictions (3.9) which now take the form $\beta_* = \Gamma\alpha + \rho$. Note that even if β_o to β_n follow the polynomial specification exactly, so that ρ_o to ρ_n are zero, ρ_{n+1} to ρ_s will be non-zero because of the error in setting $s > n$.

Combining (3.10) and (3.18) gives

$$(3.22) \qquad y_* = X_*\Gamma\alpha + X_*\rho + u_* = Z_*\alpha + X_*\rho + u_*.$$

The traditional estimation procedure produces

$$(3.23) \qquad \hat{\alpha} = (Z_*'Z_*)^{-1}Z_*'y_* = \alpha + (Z_*'Z_*)^{-1}Z_*(X_*\rho + u_*)$$

and the RLS estimator of β_* is

$$(3.24) \qquad \tilde{\beta}_* = \Gamma\hat{\alpha} = \Gamma\alpha + \Gamma(Z_*'Z_*)^{-1}Z_*'(X_*\rho + u).$$

The distribution of $\tilde{\beta}_*$ is

$$(3.25) \qquad \tilde{\beta} \sim N\left[\Gamma\alpha + \Gamma(Z_*'Z_*)^{-1}Z_*'X_*\rho, \sigma^2\Gamma(Z_*'Z_*)^{-1}\Gamma'\right]$$

so its bias and MSE are

$$(3.26) \qquad \text{bias}(\tilde{\beta}_*) = \left[\Gamma(Z_*'Z_*)^{-1}Z_*'X_* - I\right]\rho \quad \text{and}$$

$$(3.27) \qquad \text{MSE}(\tilde{\beta}_*) = \sigma^2\Gamma(Z_*'Z_*)^{-1}\Gamma' + \left[I - \Gamma(Z_*'Z_*)^{-1}Z_*'X_*\right]\rho\rho'\left[I - X_*'Z_*(Z_*'Z_*)^{-1}\Gamma'\right].$$

The biased estimator, $\tilde{\beta}_*$, may be preferred to the unbiased estimator, $\hat{\beta}_*$, if the difference between their MSE matrices is positive semi-definite. This difference is

$$(3.28)\quad MSE(\hat{\beta}_*) - MSE(\tilde{\beta}_*) = \sigma^2\left[(X_*'X_*)^{-1} - \Gamma(Z_*'Z_*)^{-1}\Gamma'\right] - \left[I - \Gamma(Z_*'Z_*)^{-1}Z_*'X_*\right]\rho\rho'\left[I - X_*'Z_*(Z_*'Z_*)^{-1}\Gamma'\right]$$

The first term on the right side of (3.28) is positive semi-definite (Dhrymes (1971), p. 226) and so is the second term. (The reason this matrix is not positive definite is discussed in the next section.) For a given X_* and Γ, the difference depends upon the size of σ^2. For large enough values of σ^2 the difference is positive semi-definite and $\tilde{\beta}_*$ is preferred on grounds of smaller MSE. If σ^2 is small enough the difference is negative semi-definite and $\hat{\beta}_*$ is preferred. For a range of σ^2 values between the two extremes the difference will be indefinite. For a given X_* and σ^2 the difference depends on ρ. Minor specification errors give small ρ values which can leave the difference positive semi-definite. Large errors can lead to this difference being negative definite while for some errors the difference is indeterminate.

4. Estimating the Bias and MSE of $\tilde{\beta}_*$

We continue to consider the case in which $s > n$. In order to estimate the bias and MSE of $\tilde{\beta}_*$ we must derive an estimate of ρ. We begin by writing

$$(4.1)\quad y_*-Z_*\hat{\alpha} = M_Z y_* = X_*\rho + u_* + Z_*(\alpha - \hat{\alpha})$$

$$= X_*\rho + u_* - Z_*(Z_*'Z_*)^{-1}Z_*'(X_*\rho + u_*),\ \text{from (3.23)},$$

$$= M_Z X_*\rho + M_Z u,$$

where $\hat{\alpha}$ is from (3.23) and $M_Z = I - Z_*(Z_*'Z_*)^{-1}Z_*'$. Since $EM_Z u = 0$ and

$EM_Z u'uM_Z = \sigma^2 M_Z$, we follow the generalized least squares procedure to write the normal equations for $\hat{\rho}$ as

$$(4.2) \qquad X_*'M_Z M_Z^+ M_Z X_* \hat{\rho} = X_*'M_Z M_Z^+ M_Z y_*$$
$$= X_*'M_Z X_* \hat{\rho} = X_*'M_Z y_*$$

where M_Z^+ is the generalized inverse of the symmetric indempotent matrix M_Z. Before attempting to solve (4.2) we must ascertain the rank of the (s+1) order square, symmetric matrix $X_*'M_Z X_*$.

$$(4.3) \qquad X_*'M_Z X_* = X_*'X_*' - X_*'X_*\Gamma(Z_*'Z_*)^{-1}\Gamma'X_*'X_*$$
$$= X_*'X_*\left[I_{s+1} - \Gamma(\Gamma'X_*'X_*\Gamma)^{-1}\Gamma'X_*'X_*\right] = X_*'X_*M_1 \ .$$

Assume $r(X_*'X_*) = s+1$ = the order of $X_*'X_*$. Since M_1 is indempotent
$r(M_1) = \text{tr } I_{s+1} - \text{tr}\left[\Gamma(\Gamma'X_*'X_*\Gamma)^{-1}\Gamma'X_*'X_*\right] = s+1 - \text{tr}\left[(\Gamma'X_*'X_*\Gamma)^{-1}\Gamma'X_*'X_*\Gamma\right]$

$= s+1 - (k+1) < s+1$. Therefore, $r(X_*'M\ X) < s+1$ and (4.2) cannot be solved using the regular inverse of $X_*'M_Z X_*$. We, therefore, use the generalized inverse to obtain (Greville (1959))

$$(4.4) \qquad \hat{\rho} = (X_*'M_Z X_*)^+ X_*'M_Z y + v$$
$$= M_1^+(X_*X_*)^{-1}X_*'M_Z y + v \qquad \text{(Deutsch (1965), pp. 84, 85)},$$

where M_1^+ is the generalized inverse of the nonsymmetric indempotent matrix M_1 and v is any nonzero (s+1) order vector such that $X'M_Z\ X\ v = 0$. We will impose two additional criteria on v: v must contain only observable quantities and Ev= 0. Since $M_1\Gamma = 0$, any vector of the form $v = \Gamma w$, where w is (k+1) by 1, will satisfy the first criterion. The second criterion is met by setting $w = (I_{k+1}, 0)M_X y$ where $M_X = I_{T-s} - X_*(X_*'X_*)^{-1}X_*'$ and the matrix 0 has k+1 rows and T-s-k-1 columns. Then our estimator of ρ becomes

$$(4.5)\qquad \hat{\rho} = \left[M_1^+(X_*'X_*)^{-1}X_*'M_Z + \Gamma(I,0)M_X\right]y_*$$

$$= \rho + \left[M_1^+(X_*'X_*)^{-1}X_*'M_Z + \Gamma(I,0)M_X\right]u_*$$

whose distribution is

$$(4.6)\qquad \hat{\rho} \sim N\left\{\rho, \sigma^2\left[(X_*'M_ZX_*)^+ + \Gamma(I,0)M_X\begin{pmatrix} I \\ 0' \end{pmatrix}\Gamma'\right]\right\}$$

using $(X_*'M_ZX_*)^+(X_*'M_ZX_*)(X_*'M_ZX_*)^+ = (X_*'M_ZX_*)^+$ and $X_*'M_ZM_X = 0$. Note that $\hat{\rho}$ is unbiased.

Now from (3.26) we have

$$(4.7)\qquad \text{bias}(\tilde{\beta}_*) = - M_1\rho$$

so we estimate this bias by

$$(4.8)\qquad \widehat{\text{bias}}(\tilde{\beta}_*) = - M_1\hat{\rho}$$

$$= - M_1\left[M_1^+(X_*'X_*)^{-1}X_*'M_Z - \Gamma(I,0)M_*\right]y_*$$

$$= - (X_*'X_*)^{-1}X_*'M_Zy_*$$

$$= - \left[I - \Gamma(Z_*'Z_*)^{-1}Z_*'X_*\right]\rho + (X_*'X_*)^{-1}X_*'M_Zu_*$$

$$= \text{bias}(\tilde{\beta}_*) + (X_*'X_*)^{-1}X_*'M_Zu_* \;.$$

Our estimator of $\text{bias}(\tilde{\beta}_*)$ is unbiased and has a normal distribution with a covariance matrix of the form:

$$(4.9)\qquad V(\widehat{\text{bias}}(\tilde{\beta}_*)) = \sigma^2(X_*'X_*)^{-1}X_*'M_ZX_*(X_*'X_*)^{-1}$$

$$= \sigma^2\left[(X_*'X_*)^{-1} - \Gamma(Z_*'Z_*)^{-1}\Gamma'\right]$$

$$= \sigma^2P'\begin{bmatrix} 0 & 0 \\ 0 & I_{s-k} \end{bmatrix}P \qquad \text{(Dhrymes (1971), p. 226)}$$

where P is a nonsingular matrix such that $P'P = (X_*'X_*)^{-1}$. Since the last line of (4.9) is a positive semi-definite matrix, $V(\widehat{\text{bias}}(\tilde{\beta}_*))$ is singular.[5]

[5] This may also be the case for $V(\hat{\rho})$. Marsaglia (1964) discusses multivariate normal distributions with singular covariance matrices.

An estimate of $MSE(\tilde{\beta}_*)$ is

$$(4.10) \qquad MSE(\tilde{\beta}_*) = \hat{\sigma}^2\Gamma(Z_*'Z_*)^{-1}\Gamma' + M_1\hat{\rho}\hat{\rho}'M_1'$$

$$= \frac{y_*'M_X y_*}{T-s}\Gamma(Z_*'Z_*)^{-1}\Gamma' + (X_*'X_*)^{-1}X_*'M_Z y_* y_*'M_Z X_*(X_*'X_*)^{-1}$$

where $\hat{\sigma}^2$ is the unbiased estimate of σ^2 provided by OLS.

The unbiased estimator $\hat{\beta}_*$ ignores the restrictions (3.9). This makes it tempting to try to amend $\tilde{\beta}_*$ to produce an unbiased estimator which explicitly accounts for these restrictions. Consider then

$$(4.11) \qquad \tilde{\beta}_* + M_1\hat{\rho} = \tilde{\beta}_* + (X_*'X_*)^{-1}X_*'M_Z y_*$$

$$= \tilde{\beta}_* + (X_*'X_*)^{-1}X_*'y_* - \Gamma(Z_*'Z_*)^{-1}Z_*'y_*$$

$$= \tilde{\beta}_* + \hat{\beta}_* - \tilde{\beta}_* = \hat{\beta}_* .$$

So our attempt to amend $\tilde{\beta}_*$ to make it unbiased leaves us with the OLS estimator because our estimator of bias $(\tilde{\beta}_*)$ is simply the difference between the biased, restricted estimator $\tilde{\beta}_*$ and the unbiased unrestricted estimator $\hat{\beta}_*$.

5. Autocorrelation of x_t and Efficiency of $\hat{\beta}_*$

The motivation for the introduction of the polynomial lag specification was the increased efficiency of RLS. OLS is felt to suffer a loss of efficiency, in part, because the autoregressive nature of a typical x_t tends to make the matrix $X_*'X_*$ ill conditioned. To illustrate this point we make the simplifying assumption that x_t mimics a first order autoregressive process in that it obeys

$$(5.1) \qquad x_t = rx_{t-1} + \varepsilon_t \ ; \ |r| < 1, \text{ where}$$

(5.2) $\sum_{i=1}^{T-s} \varepsilon_{s-j+i} = 0 \text{ for } j=0,\cdots,s$

(5.3) $\sum_{t=1} \varepsilon_{s-j+1}\, x_{s-\ell+1} = 0 \text{ for } j,\ell=0,\cdots,s$

(5.4) $\sum_{i=1}^{T-s} x^2_{s-\ell+i} = V_x$, a constant scalar.

Using these relations we have

(5.5) $$X_*'X_* = \begin{bmatrix} x_{s+1} & x_{s+2} & \cdots & x_T \\ x_s & x_{s+1} & & x_{T-1} \\ \vdots & \vdots & & \vdots \\ x_1 & x_2 & \cdots & x_{T-s} \end{bmatrix} \begin{bmatrix} x_{s+1} & x_s & & x_1 \\ x_{s+2} & x_{s+1} & & x_2 \\ \vdots & \vdots & & \vdots \\ x_T & x_{T-1} & \cdots & x_{T-s} \end{bmatrix}$$

$$= V_x \begin{bmatrix} 1 & r & r^2 & \cdots & r^s \\ r & 1 & r & & r^{s-1} \\ r^2 & r & 1 & & \\ \vdots & & & & \\ r^s & & & & 1 \end{bmatrix}.$$

Then the covariance of $\hat{\beta}_*$ is

(5.6) $$V(\hat{\beta}_*) = \sigma^2 (X_*'X_*)^{-1} = \frac{\sigma^2}{V_x} \begin{bmatrix} 1 & -r & 0 & & 0 \\ -r & 1+r^2 & -r & & \\ 0 & -r & & & \\ & & & 1+r^2 & -r \\ 0 & & & -r & 1 \end{bmatrix}$$

If x_t was not autocorrelated at all r would be zero and $V(\hat{\beta}_*)$ would be a diagonal matrix with σ^2/V_x in each position. If, at the other extreme, $|r| = 1$ the variances of $\hat{\beta}_o$ and $\hat{\beta}_s$, the first and last elements of $\hat{\beta}_*$, would be unchanged but the variances of all other $\hat{\beta}_i$ in $\hat{\beta}_*$ would double.

Under more general schemes (higher orders of autocorrelation, etc.) we might find that the introduction of autocorrelation in x had an even more drastic effect on $V(\hat{\beta}_*)$. However, the difference $(X_*'X_*)^{-1} - \Gamma(Z_*'Z_*)^{-1}\Gamma$ remains positive semi-definite so long as $(X_*'X_*)^{-1}$ exists although the difference $V(\hat{\beta}_*) - MSE(\tilde{\beta}_*)$ may or may not be positive semi-definite depending on the size of ρ.

6. Examples

Two artificial populations were created to illustrate the findings of the previous sections. The first population, Model 1, has a polynomial lag structure of length three and degree two. It is described by

(6.1) $$y_t = .1\, x_t + .5\, x_{t-1} + .5\, x_{t-2} + .1\, x_{t-3} + u_t$$

and

(6.2) $$\beta = \begin{bmatrix} .1 \\ .5 \\ .5 \\ .1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} .1 \\ .6 \\ -.2 \end{bmatrix} = \Gamma\alpha\ .$$

If we set s=3=n and k=2 then ρ=0, since we have made no error in specifying the lag structure. In this case RLS is unbiased and has smaller MSE than OLS. This is illustrated in Case 1 of Table 1. The columns of this table headed OLS MSE and RLS MSE are the main diagonal elements of the matrices on the right of equations (3.21) and (3.27), respectively, where the X_* matrix has 20 rows and k columns derived from an x series whose coefficient of first order autocorrelation was .8096. In each case R^2 shows the goodness of fit for the population given X_*; that is, $R^2 = \dfrac{\beta' X_*' X_* \beta}{\beta' X_*' X_* \beta + T\sigma^2}$.

Table 1

Effect of Misspecifying Lag Structure: Model 1

$$y_t = .1x_t + .5x_{t-1} + .5x_{t-2} + .1x_{t-3} + u_t$$

$$\begin{bmatrix} .1 \\ .5 \\ .5 \\ .1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} .1 \\ .6 \\ -.2 \end{bmatrix} = \Gamma\alpha .$$

	Case 1 s=3 k=2 σ^2=.001 R^2=.986			Case 2 s=3 k=1 σ^2=.001 R^2=.986			Case 3 s=3 k=1 σ^2=.0001 R^2=.9986		
β	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE
.1	.0607	0	.0428	.0607	.183	.0632	.00607	.183	.0364
.5	.0624	0	.0196	.0624	-.206	.0456	.00624	-.206	.0428
.5	.0631	0	.0188	.0631	-.195	.0421	.00631	-.195	.0386
.1	.0707	0	.0506	.0707	.216	.0790	.00707	.216	.0498

	Case 4 s=6 k= 2 σ^2=.001 R^2=.986			Case 5 s=6 k= 2 σ^2=.0005 R^2=.994			Case 6 s=6 k= 2 σ^2=.00001 R^2=.99988		
β	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE
.1	.0680	.118	.0624	.0340	.118	.0382	.000680	.118	.0145
.5	.0847	-.212	.0514	.0424	-.212	.0482	.000847	-.212	.0450
.5	.0925	-.193	.0574	.0462	-.193	.0474	.000925	-.193	.0376
.1	.186	.174	.0588	.0928	.174	.0445	.00186	.174	.0305
0	.225	.190	.0518	.127	.190	.0440	.00255	.190	.0363
0	.237	.0551	.0116	.119	.0551	.00730	.00237	.0551	.00313
0	.164	-.131	.0954	.821	-.131	.0563	.00164	-.131	.0179

For any polynomial lag structure a correct specification of s and a specification of k which is too large results in $\rho=0$ and RLS being unbiased. On the other hand, specifying k too small will bias RLS so that the MSE of RLS for some (Case 2) or all (Case 3) of the coefficients can be larger than those of OLS. If s is too large but k is correct the MSE of RLS may be smaller (Case 4) or larger (Case 6) than those of OLS for all coefficients or smaller for some and larger for others (Case 5). Note that in order for OLS to have smaller MSE for all coefficients it was necessary to set R^2 very close to one. This may be typical.

The second model does not have a polynomial lag structure. It is defined as

$$(6.3) \qquad y_t = .7\, x_t + .2\, x_{t-1} + .08\, x_{t-2} + .02\, x_{t-3} + u_t$$

For any choice of s and k the polynomial which approximates this lag structure most closely can be obtained by using least squares to find the value of α which minimizes $\rho'\rho$ given $\sum_{i=1}^{s+1} \rho_i = 0$. This procedure yields

$$(6.4) \qquad \alpha = (\Gamma'\Gamma)^{-1}\Gamma'\beta \quad \text{and}$$

$$(6.5) \qquad \rho = \beta - \Gamma\alpha = [(I - \Gamma(\Gamma'\Gamma)^{-1}\Gamma']\beta$$

The values of α and ρ obtained for several combinations of s and k are shown in Table 2. The ρ values shown in Table 2 were combined with the same X_* matrices as were used in Model 1 to produce the values of OLS MSE, RLS bias and RLS MSE shown in Table 3. For given values of s and k increasing R^2 has the effect of reducing the OLS MSE until they are below those of RLS, compare Cases 1 and 2. Naturally, reducing k serves to increase RLS MSE above those for OLS, compare Cases 3 and 1. If we want to increase k in order to approximate the lag structure more closely we must also increase s to ensure $s > k$.

Table 2

Values of α and ρ for Various Values of r and k

Model 2

s=3 k=1 α	ρ
.574	.126
-.216	-.158
	-.0620
	.0940

s=3 k=2 α	ρ
.684	.0160
-.546	-.0480
.110	.0480
	.0160

s=6 α	k=1 ρ
.419	.281
-.0921	-.127
	-.155
	-.123
	-.0507
	.0414
	.134

s=6 k=2 α	ρ
.609	.0914
-.319	-.127
.0379	-.0414
	.0286
	.0629
	.0414
	-.0557

s=6 k=3 α	ρ
.679	.0214
-.553	-.0571
.143	.0286
-.0117	.0286
	-.00714
	-.0286
	.0143

Table 3

Effect of Misspecifying Lag Structure: Model 2

$$y_t = .7x_t + .2x_{t-1} + .08x_{t-2} + .02x_{t-3} + u_t$$

$$\beta = \Gamma\alpha + \rho$$

	Case 1			Case 2			Case 3		
	s=3, k=2			s=3, k=2			s=3, k=1		
	σ^2 = .0001, R^2 = .998			σ^2 = .000025, R^2 = .995			σ^2 = .0001, R^2 = .998		
β	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE
.7	.00607	-.0280	.00506	.00152	-.0280	.00185	.00607	-.129	.0197
.2	.00624	.0433	.00384	.00156	.0433	.00237	.00624	.157	.0251
.08	.00631	-.0441	.00382	.00158	-.0441	.00241	.00631	.0639	.00448
.02	.00707	.0298	.00594	.00177	.0298	.00215	.00707	.0896	.0113

	Case 4			Case 5			Case 6		
	s=6, k=2			s=6, k=3			s=6, k=1		
	σ^2 = .0001, R^2 = .998			σ^2 = .0001, R^2 = .998			σ^2 = .0001, R^2 = .998		
β	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE	OLS MSE	RLS Bias	RLS MSE
.7	.00680	-.0689	.00958	.00680	-.0323	.00623	.00680	-.246	.0621
.2	.00847	.131	.0179	.00847	.0508	.00491	.00847	.149	.0229
.08	.00925	.0337	.00313	.00925	-.0265	.00364	.00925	.165	.0274
.02	.0186	-.0417	.00459	.0186	-.0185	.00334	.0186	.121	.0146
0	.0255	-.0749	.00717	.0255	.0203	.00433	.0255	.0364	.00155
0	.0237	-.0458	.00295	.0237	.0354	.00382	.0237	-.0679	.00541
0	.0164	.0655	.0121	.0164	-.0275	.0108	.0164	-.172	.0314

The beneficial effect of increasing k, after an increase in s, is shown in Cases 4, 5 and 6. These comparisons show that OLS has lower MSE than RLS when k is low (Cases 3 and 6) and/or R^2 is very high (Case 2).

7. Conclusions

When a distributed lag model is specified to have a polynomial lag structure of length s and degree k several errors may have been committed. The true lag structure may be a polynomial of different length or degree or it may not be a polynomial at all. However, since any continuous finite lag structure can be accurately approximated by a polynomial this specification is understandably popular. In this paper we derive the distribution of the restricted least squares estimator, which explicitly embodies the polynomial specification, and compare it to the ordinary least squares estimator, which ignores this specification. We find that although RLS is, in general, biased it may have smaller mean squared errors than OLS unless the polynomial approximation is very bad and/or the variance of the disturbances is very small. We present estimators for the bias and MSE of RLS in cases where the specified length of the lag is greater than or equal to the true length.

A good procedure for applied econometrics seems to be, first, set the length of the lag slightly larger than what prior notions suggest; six periods or more are suggested. Second, set the degree of the polynomial high enough, at least three, so that it is a fairly accurate approximation. Third compute the RLS estimate and estimate its bias and MSE using equations (4.7) and (4.10). Hypotheses about the bias of RLS can be tested using the fact that the estimator given in (4.7) is normally distributed with a covariance given in (4.9). Such tests may lead to a second round of estimation with a higher degree and, perhaps, a longer lag in an attempt to reduce estimated bias. However, if the first round produces estimates with low estimated bias the lag length should not be reduced in case it falls below the true lag length. The precision of the estimates should be judged by the estimated MSE, not the estimated variances. If the goodness of fit is very high, above .99 say, it is worth computing OLS to see whether its estimated MSE are less than those for RLS.

References

Almon, S. (1965), "The Distributed Lag Between Capital Appropriations and Expenditures," Econometrica, 30, pp. 178-196.

Deutsch, R. (1965), Estimation Theory, Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Dhrymes, P. (1971), Distributed Lags: Problems of Estimation and Formulation, San Francisco: Holden Day.

Frost, P. (1975), "Some Properties of the Almon Lag Technique When One Searches for Degree of Polynomial and Lag," Journal of the American Statistical Association, 70, pp. 606-612.

Greville, T. N. E. (1959), "The Pseudoinverse of a Rectangular or Singular Matrix and its Application to the Solution of Systems of Linear Equations," SIAM Review, 1, pp. 38-43.

Marsaglia, G. (1964), "Conditional Means and Covariances of Normal Variables with Singular Covariance Matrix," Journal of the American Statistical Association, 59, pp. 1203-1204.

Rowley, J. C. R. (1971), "The Almon Technique," mimeo.

Schmidt, P. and R. N. Waud (1973), "The Almon Lag Technique and the Monetary Versus Fiscal Policy Debate," Journal of the American Statistical Association, 68, pp. 11-19.