REGRESSION ANALYSIS WITH SECOND-ORDER AUTOREGRESSIVE DISTURBANCES

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
PETER J. SCHMIDT

1970

This is to certify that the

thesis entitled

Regression Analysis with Second-Order

Autoregressive Disturbances

presented by

Peter J. Schmidt

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Major professor

Date November 19, 1970

ABSTRACT

REGRESSION ANALYSIS WITH SECOND-ORDER
AUTOREGRESSIVE DISTURBANCES

By

Peter J. Schmidt

Autocorrelation is present in a regression equation when the unobservable random disturbances are not mutually independent over time. In the presence of autocorrelation, ordinary least squares will lead to inefficient estimators of the regression coefficients and to inconsistent estimators of their variances. Econometricians have therefore developed testing procedures to test for autocorrelation, and estimation procedures to alleviate the problems which it causes when it is present. These procedures must of necessity make some assumption about what types of autocorrelation might be present. In particular, it has usually been assumed that the disturbances follow a first-order autoregressive scheme.

This study considers autocorrelation in the more general form of a second-order autoregressive scheme. The usual testing and estimation procedures are generalized

to this case. Finally, the new procedures are compared to the original procedures in terms of their performance in the presence of various types of autocorrelation. The results obtained indicate that these generalized testing and estimation procedures may be useful, at least when one does not have strong a priori reasons for believing the autocorrelation in the sample to be of first-order form.

REGRESSION ANALYSIS WITH SECOND-ORDER AUTOREGRESSIVE DISTURBANCES

By

Peter J. Schmidt

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1970

ACKNOWLEDGEMENTS

This study would never have been begun, much less completed, without the constant encouragement, advice, and help of Jan Kmenta, my dissertation committee chairman. James Ramsey and Roy Gilbert also read the entire study and made many valuable suggestions. I also wish to thank Phoebus Dhrymes for his comments on Chapter III, and Richard Henshaw for his comments on earlier versions of Chapters IV and V.

The research on which this study is based was supported in part by the Mathematical Social Science Board Workshop on Lags in Economic Behavior, held at the University of Chicago in the summer of 1970. Marc Nerlove and G. S. Maddala, co-directors of the workshop, were kind enough to read a substantial part of an earlier draft and to suggest improvements.

Any remaining errors are of course my own responsibility.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Statement of the Problem[1]

Consider the linear regression model

$$y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + u_i, \quad i = 1,2,\ldots,N; \tag{1.1}$$

where each β_j is a parameter to be estimated; X_{ij} is the i^{th} observation on the j^{th} independent variable (regressor); y_i is the i^{th} observation on the dependent variable in the regression; and u_i is a random disturbance. This model can be rewritten in matrix form:

$$y = X\beta + u; \tag{1.2}$$

where y, β and u are vectors and X is a matrix, defined by

$$y = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{vmatrix} \quad X = \begin{vmatrix} X_{11} & X_{12} & \cdots & X_{1K} \\ X_{21} & X_{22} & \cdots & X_{2K} \\ \vdots & \vdots & & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{NK} \end{vmatrix} \quad u = \begin{vmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{vmatrix} \quad \beta = \begin{vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{vmatrix} \tag{1.3}$$

---

[1]This section closely follows [17], Section 5.4.

This model is said to satisfy the full ideal conditions (FIC)[2] if u is stochastically independent of X, if $E(u) = \underline{0}$,[3] if X has rank $K \leq N$, and if $E(uu') = \sigma^2 I$ (where I is the N-dimensional identity matrix and σ^2 is a parameter to be estimated).

On the other hand, autocorrelation is said to be present when $Cov(u_i u_j) \neq 0$ for some $i \neq j$. That is, the disturbances are said to be autocorrelated if $E(uu') = \sigma^2\Omega$, where Ω is a non-diagonal positive semi-definite matrix. Autocorrelation is typically considered in the context of time-series analysis; it is then the case in which the disturbances are correlated over time. This study will be concerned with cases in which the non-diagonality of Ω is the only violation of the FIC; that is, $E(u) = \underline{0}$ and u is distributed independently of X, but Ω has non-zero terms off the diagonal.

The ordinary least squares (OLS) estimator of β is defined by

$$\tilde{\beta} = (X'X)^{-1}X'y \tag{1.4}$$

and an associated estimator of σ^2 is

$$\tilde{\sigma}^2 = \frac{y'My}{N-K}, \text{ where } M = I - X(X'X)^{-1}X'. \tag{1.5}$$

---

[2]This terminology is due to [6].

[3]The symbol $\underline{0}$ will be used to denote an appropriately dimensioned matrix or vector of zeroes.

The covariance matrix of $\tilde{\beta}$ under the FIC is equal to $\sigma^2(X'X)^{-1}$, and can be estimated by replacing σ^2 by $\tilde{\sigma}^2$. Now, it is well known that, under the FIC, $\tilde{\beta}$ is best linear unbiased, consistent, and also asymptotically efficient if the disturbances are Normally distributed; $\tilde{\sigma}^2$ is unbiased, consistent, and also asymptotically efficient if the disturbances are Normal. In fact, these desirable properties of the OLS estimators under the FIC constitute the chief rationale for the use of the OLS estimation procedure. Unfortunately, however, these properties do not hold if the disturbances are autocorrelated. In this case the OLS estimator $\tilde{\beta}$ is still unbiased and consistent, but it is in general no longer best linear unbiased or asymptotically efficient. The estimator $\tilde{\sigma}^2$ is in general biased and inconsistent. Furthermore, the covariance matrix of $\hat{\beta}$ is no longer equal to $\sigma^2(X'X)^{-1}$.

It is equally well known that these difficulties could be avoided through the application of <u>generalized least squares</u> (GLS) if the disturbance term covariance matrix Ω were known. With Ω known, the GLS estimator of β is

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \tag{1.6}$$

while σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{y'M^*y}{N-K}, \text{ where } M^* = I - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \qquad (1.7)$$

The covariance matrix of $\hat{\beta}$ is $\sigma^2(X'\Omega^{-1}X)^{-1}$. It is also sometimes useful to note that since Ω is by assumption positive semi-definite, Ω^{-1} is also positive semi-definite, so that there must exist a (not necessarily unique) non-singular matrix V such that

$$\Omega^{-1} = V'V \qquad (1.8)$$

It should then be clear that if u has covariance matrix $\sigma^2\Omega$, Vu will have covariance matrix $\sigma^2 I$. Hence if the regression equation is rewritten

$$Vy = VX\beta + u \qquad (1.9)$$

the disturbances now satisfy the FIC, so that the OLS regression of Vy on VX is appropriate. Indeed, the OLS regression of Vy on VX is algebraically identical to the GLS procedure defined above.

The point of using GLS is of course that the GLS estimators (with Ω known) have the same optimal properties as do the OLS estimators under the FIC. That is, $\hat{\beta}$ is best linear unbiased, consistent, and asymptotically efficient; $\hat{\sigma}^2$ is unbiased, consistent, and asymptotically efficient. It should therefore be clear that autocorrelation is really a problem only in that Ω is generally not

known. With Ω unknown, the above GLS procedure cannot be applied, at least not directly, and the question of what to do when Ω is unknown (but suspected to be non-diagonal) is not a trivial one. It is this question to which the rest of this study will be addressed.

1.2 Types of Autocorrelation

When Ω is unknown, an obvious procedure is to find a consistent estimate of it, and then to use GLS with the estimate $\hat{\Omega}$ replacing the (unknown) true covariance matrix Ω. The statistical justification for this procedure is the well-known fact that if GLS is applied with any consistent estimator $\hat{\Omega}$ used in place of Ω, the resulting estimators $\hat{\beta}$ and $\hat{\sigma}^2$ will be consistent and asymptotically efficient.[4]

Unfortunately, however, it is in general not possible to estimate Ω consistently. After all, Ω has $\frac{1}{2}N(N-1)$ distinct elements, and one can hardly hope to estimate them all from a sample of N observations. It should therefore be clear that one can hope to proceed only by putting fairly severe restrictions on Ω. Essentially, what must be done is to make Ω depend on some fixed number of parameters that does not depend on the sample size. These parameters can then (hopefully)

---

[4] This theorem was proved for a special case in [52]. However, it holds whenever X is fully independent of u. For a full discussion see [36].

be consistently estimated, $\hat{\Omega}$ can be constructed, and GLS can be applied.

In particular, the procedure which has usually been used is to assume that the disturbances follow a first-order Markov process:

$$u_i = \rho_1 u_{i-1} + \varepsilon_i, \quad i = -\infty,\ldots,N; \tag{1.10}$$

where $-1 < \rho_1 < 1$, $E(\varepsilon_i) = 0$, $E(\varepsilon_i u_{i-s}) = 0$ for $s > 0$, $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$, and $\sigma^2 \equiv \mathrm{Var}(\varepsilon_i) = \mathrm{Var}(u_j)(1 - \rho_1^2) \equiv \sigma_u^2(1 - \rho_1^2)$ for all i and j. (This is often referred to as a first-order autocorrelation scheme.) Then it is readily verified that $\mathrm{Cov}(u_i u_j) = \sigma_u^2 \rho_1^{|i-j|}$. Hence Ω is of the form:

$$\Omega = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1^{N-1} \\ \rho_1 & 1 & \cdots & \rho_1^{N-2} \\ \vdots & \vdots & & \vdots \\ \rho_1^{N-1} & \rho_1^{N-2} & \cdots & 1 \end{vmatrix} \frac{1}{1-\rho_1^2} \tag{1.11}$$

Direct multiplication will verify that Ω^{-1} is given by the tridiagonal matrix

$$\Omega^{-1} = \begin{vmatrix} 1 & -\rho_1 & 0 & \cdots & 0 & 0 \\ -\rho_1 & 1+\rho_1^2 & -\rho_1 & \cdots & 0 & 0 \\ 0 & -\rho_1 & 1+\rho_1^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho_1^2 & -\rho_1 \\ 0 & 0 & 0 & \cdots & -\rho_1 & 1 \end{vmatrix} \qquad (1.12)$$

As noted in the last section, Ω^{-1} can be decomposed as $\Omega^{-1} = V'V$. V is in this case given by:[5]

$$V = \begin{vmatrix} a & 0 & 0 & \cdots & 0 & 0 \\ -\rho_1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho_1 & 1 \end{vmatrix}, \text{ where } a = (1-\rho_1^2)^{\frac{1}{2}}. \qquad (1.13)$$

One last fact to note is that the determinant of Ω^{-1} is equal to $a^2 = 1-\rho_1^2$.

It is clear that in this case Ω depends on only one parameter, ρ_1. Given a consistent estimator $\hat{\rho}_1$, $\hat{\Omega}$ can be formed and GLS can be applied. If the disturbances are in fact generated by a first-order Markov process, $\hat{\Omega}$ will be a consistent estimator of Ω, and the resulting GLS estimators will thus be asymptotically efficient.

---

[5]The first explicit statement of Ω^{-1} and its decomposition is given in [40].

In general, however, there is no reason to suppose that a first-order scheme is appropriate. First-order autocorrelation is generally only an approximation to the (unknown) type of autocorrelation in the sample, and it is frequently assumed because it is a particularly simple way to make estimation possible. However, while it is true that the form of Ω must be restricted to make it estimable, it seems rather drastic to make Ω depend on only one parameter. After all, a more general type of autocorrelation scheme might provide a reasonable approximation to more different types of autocorrelation, and this would seem desirable when Ω is not known *a priori* to be of any particular form.

This study will be particularly concerned with the obvious generalization of the usual first-order procedures to the case of a second-order autocorrelation scheme:

$$u_i = \rho_1 u_{i-1} + \rho_2 u_{i-2} + \varepsilon_i, \tag{1.14}$$

where $|\rho_1| + |\rho_2| < 1$ and the same assumptions about ε_i are made as in the first-order case, except that here $\mathrm{Var}(u_i)$ is equal to

$$\sigma_u^2 = \frac{\sigma^2}{1 - \rho_1^2 - \rho_2^2 - 2\rho_1^2\rho_2/(1 - \rho_2)}$$

$$= \frac{\sigma^2(1 - \rho_2)}{1 - \rho_2 - \rho_1^2 - \rho_2\rho_1^2 - \rho_2^2 + \rho_2^3} \qquad (1.15)$$

Then the following facts are readily verified:

$$E(u_i u_{i-1}) = \sigma_u^2 \rho_1/(1 - \rho_2) \qquad (1.16)$$

$$E(u_i u_{i-2}) = \sigma_u^2[\rho_2 + \rho_1^2/(1 - \rho_2)] \qquad (1.17)$$

$$E(u_i u_{i-s}) = \rho_1 E(u_{i-1} u_{i-s}) + \rho_2 E(u_{i-2} u_{i-s}),$$

$$s > 2. \qquad (1.18)$$

Ω can thus be written out, though this becomes rather tedious since the terms become horrendous as one moves away from the diagonal. The useful expression in any case in Ω^{-1}, which is given by

$$\Omega^{-1} = \begin{vmatrix} 1 & -\rho_1 & -\rho_2 & 0 & \cdots & 0 & 0 & 0 \\ -\rho_1 & 1+\rho_1^2 & \rho_1(\rho_2-1) & -\rho_2 & \cdots & 0 & 0 & 0 \\ -\rho_2 & \rho_1(\rho_2-1) & 1+\rho_1^2+\rho_2^2 & \rho_1(\rho_2-1) & \cdots & 0 & 0 & 0 \\ 0 & -\rho_2 & \rho_1(\rho_2-1) & 1+\rho_1^2+\rho_2^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\rho_1^2+\rho_2^2 & \rho_1(\rho_2-1) & -\rho_2 \\ 0 & 0 & 0 & 0 & \cdots & \rho_1(\rho_2-1) & 1+\rho_1^2 & -\rho_1 \\ 0 & 0 & 0 & 0 & \cdots & -\rho_2 & -\rho_1 & 1 \end{vmatrix} \qquad (1.19)$$

Note that this is a band diagonal matrix, with five non-zero bands. One can again decompose Ω^{-1} as V'V, where in this case V is given by

$$V = \begin{vmatrix} a & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ b & c & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho_2 & -\rho_1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho_2 & -\rho_1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\rho_2 & -\rho_1 & 1 \end{vmatrix} \tag{1.20}$$

and where

$$\begin{aligned} a &= [1 - \rho_2^2 - \rho_1^2(1 + \rho_2)/(1 - \rho_2)]^{\frac{1}{2}} \\ b &= -\rho_1 [(1 + \rho_2)/(1 - \rho_2)]^{\frac{1}{2}} \\ c &= (1 - \rho_2^2)^{\frac{1}{2}} \end{aligned} \tag{1.21}$$

It is also clear that the determinant of Ω^{-1} is equal to $a^2c^2 = 1 - \rho_1^2 - 2\rho_2^2 - 2\rho_1^2\rho_2 - \rho_1^2\rho_2^2 + \rho_2^4$.

Finally, it should be repeated that the reason for considering this second-order scheme is not that there is necessarily any a priori reason for believing it to be common in actual data. The point is simply that the assumption of first-order autocorrelation may be unduly restrictive.

CHAPTER II

ESTIMATION IN A LINEAR MODEL

2.1 Introduction

Consider the linear model defined by (1.1). Then it is clear, under the assumption of first-order autocorrelation, that GLS can be applied to get asymptotically efficient estimates if one can obtain a consistent estimator $\hat{\rho}_1$ with which to construct $\hat{\Omega}$. Several methods of getting a consistent estimator $\hat{\rho}_1$ have been proposed in the literature. One method, suggested by Durbin,[1] is to use for $\hat{\rho}_1$ the OLS estimator of the coefficient of y_{-1} in the transformed equation

$$y_i = \rho_1 y_{i-1} + \beta_1 x_{i,1} - \beta_1 \rho_1 x_{i-1,1} + \ldots + \beta_K x_{i,K} - \beta_K \rho_1 x_{i-1,K} + (u_i - \rho_1 u_{i-1}), \quad i = 2,3,\ldots,N \tag{2.1}$$

This estimator is consistent. Hildreth and Lu[2] have suggested a modification of this procedure in which (2.1) is estimated subject to the constraints that

---

[1] [10] and [11].

[2] [24].

$$-\widehat{\beta_j \rho_1} = -\hat{\beta}_j \hat{\rho}_1, \quad j = 1,\ldots,K. \tag{2.2}$$

Clearly this is a non-linear procedure. The estimators it yields are in fact the maximum likelihood estimators, conditional on y_1. This procedure gives a consistent and asymptotically efficient estimate of ρ_1. Finally, an estimator which will be referred to as the C-O estimator (after Cochrane and Orcutt[3]) is the following:

$$\hat{\rho}_1 = \frac{\sum_{2}^{N} \tilde{u}_i \tilde{u}_{i-1}}{\sum_{1}^{N} \tilde{u}_i^2} \tag{2.3}$$

the $\tilde{u}_i$ being the OLS residuals from the regression of y on X. This estimator is consistent.

As noted, if the disturbances do in fact follow a first-order scheme, each of these estimators is consistent. Hence in this case GLS based on any of the $\hat{\rho}_1$ above will yield asyptotically efficient results.

It should also be noted that in actual econometric practice the usual procedure in applying GLS is not to actually form $\hat{\Omega}$ or $\hat{\Omega}^{-1}$, but rather to use the equivalent procedure of forming $\hat{V}$ and applying OLS to the transformed

---

[3]Actually, the estimator defined in their article differs from the one defined above by a factor of N/(N-1).

equation $\hat{V}y = \hat{V}X\beta + \hat{V}u$. Notice that disregarding the first row of V, this amounts to the following regression:

$$(y_i - \hat{\rho}_1 y_{i-1}) = \beta_1(X_{i,1} - \hat{\rho}_1 X_{i-1,1}) + \ldots$$

$$+ \beta_K(X_{i,K} - \hat{\rho}_1 X_{i-1,K}) + (u_i - \hat{\rho}_1 u_{i-1}),$$

$$i = 2,3,\ldots,N. \qquad (2.4)$$

Disregarding the first row of $\hat{V}$ is thus equivalent to discarding the first observation, and this has commonly been done. This common "approximation" to the actual GLS procedure (which would include the first observation with a "weight" of $(1 - \hat{\rho}_1^2)^{\frac{1}{2}}$) is clearly asymptotically equivalent to GLS.

2.2 Generalization to the Second-Order Case

In the case of second-order autocorrelation it is of course necessary to estimate both ρ_1 and ρ_2 in order to form Ω. Fortunately, consistent estimators ρ_1 and ρ_2 can be obtained by straightforward generalizations of the procedures of the last section. Durbin's method can be generalized to this case[4] by applying OLS to the transformed equation:

---

[4] In fact, it was presented in general q^{th} order form (q any integer) in [11].

$$y_i = \rho_1 y_{i-1} + \rho_2 y_{i-2} + \beta_1 X_{i,1} - \beta_1\rho_1 X_{i-1,1} - \beta_1\rho_2 X_{i-2,1} + \cdots$$

$$+ \beta_K X_{i,K} - \beta_K\rho_1 X_{i-1,K} - \beta_K\rho_2 X_{i-2,K}$$

$$+ (u_i - \rho_1 u_{i-1} - \rho_2 u_{i-2}), \quad i = 3,4,\ldots,N. \tag{2.5}$$

The maximum likelihood procedure (conditional on y_1 and y_2) is to estimate the above equation subject to the constraints

$$-\widehat{\beta_j\rho_q} = -\hat{\beta}_j\hat{\rho}_q, \quad j = 1,\ldots,K; \quad q = 1,2. \tag{2.6}$$

The C-O method can be generalized in at least two ways. The first would be to estimate ρ_1 and ρ_2 by the OLS regression of $\tilde{u}_i$ on $\tilde{u}_{i-1}$ and $\tilde{u}_{i-2}$. A somewhat more informative way is to note that if one defines

$$\rho_1^* = \frac{\sum_2^N \tilde{u}_i\tilde{u}_{i-1}}{\sum_1^N \tilde{u}_i^2} \tag{2.7}$$

$$\rho_1^* = \frac{\sum_3^N \tilde{u}_i\tilde{u}_{i-2}}{\sum_1^N \tilde{u}_i^2}, \tag{2.8}$$

then ρ_1^* and ρ_2^* are consistent estimators of $\frac{\mathrm{Cov}(u_i u_{i-1})}{\sigma_u^2}$ and $\frac{\mathrm{Cov}(u_i u_{i-2})}{\sigma_u^2}$ respectively; from (1.16) and (1.17) it

is clear that these expressions are not in general equal to ρ_1 and ρ_2. However, consistent estimators can be derived by setting ρ_1^* and ρ_2^* equal to their probability limits given by (1.16) and (1.17) and then solving for ρ_1 and ρ_2. That is, solve the following equations for $\hat{\rho}_1$ and $\hat{\rho}_2$:

$$\rho_1^* = \hat{\rho}_1/(1 - \hat{\rho}_2) \tag{2.9}$$

$$\rho_2^* = \hat{\rho}_2 + \hat{\rho}_1^2/(1 - \hat{\rho}_2). \tag{2.10}$$

The solution is

$$\hat{\rho}_2 = (\rho_2^* - \rho_1^{*2})/(1 - \rho_1^{*2}) \tag{2.11}$$

$$\hat{\rho}_1 = \rho_1^*(1 - \hat{\rho}_2). \tag{2.12}$$

Once again the estimators $\hat{\rho}_1$ and $\hat{\rho}_2$ are consistent. Hence if the disturbances do follow a second-order autocorrelation scheme, GLS (using $\hat{\Omega}$ constructed from $\hat{\rho}_1$ and $\hat{\rho}_2$) will yield asymptotically efficient estimates of β and σ^2.

Finally, it is once again apt to be computationally simpler to form $\hat{V}$ (rather than $\hat{\Omega}$ or $\hat{\Omega}^{-1}$) and to apply OLS to the transformed equation $\hat{V}y = \hat{V}X\beta + \hat{V}u$. Disregarding the first two rows of $\hat{V}$, this amounts to the following:

$$(y_i - \hat{\rho}_1 y_{i-1} - \hat{\rho}_2 y_{i-2}) = \beta_1 (X_{i,1} - \hat{\rho}_1 X_{i-1,1} - \hat{\rho}_2 X_{i-2,1}) + \cdots$$

$$+ \beta_K (X_{i,K} - \hat{\rho}_1 X_{i-1,K} - \hat{\rho}_2 X_{i-2,K})$$

$$+ (u_i - \hat{\rho}_1 u_{i-1} - \hat{\rho}_2 u_{i-2}), \quad i = 3,4,\ldots,N. \tag{2.13}$$

Clearly this is asymptotically equivalent to the actual GLS procedure, which would include the first and second observations with appropriate weights given by the elements in the first two rows of $\hat{V}$.

2.3 Properties of the GLS Estimators

Let us denote by GLS1 the GLS procedure which assumes first-order autocorrelation. That is, GLS1 means GLS with $\hat{\Omega}$ formed from $\hat{\rho}_1$, with $\hat{\rho}_1$ calculated by any of the methods of section 2.1. Similarly, let GLS2 be the GLS procedure using $\hat{\Omega}$ formed from $\hat{\rho}_1$ and $\hat{\rho}_2$, with $\hat{\rho}_1$ and $\hat{\rho}_2$ calculated by any of the methods of section 2.2. In this section we will compare the efficiency of OLS, GLS1, and GLS2 under various specifications of the form of Ω.

Consider first the asymptotic properties of the various estimation procedures. These asymptotic comparisons can be easily made since, as noted in section 1.2, estimation based on a consistent estimator of Ω will yield asymptotically efficient results, while estimation based

on an inconsistent estimator of Ω will in general give asymptotically inefficient estimates of β and inconsistent estimates of σ^2.

Suppose first that the disturbances satisfy the full ideal conditions (FIC). Clearly OLS will be asymptotically efficient. But it is also clear that the FIC are just the special case of a first-order autocorrelation scheme corresponding to $\rho_1 = 0$, so that $\hat{\rho}_1$ estimated by any method of section 2.1 will be a consistent estimator of $\rho_1 = 0$. Hence Ω formed from $\hat{\rho}_1$ will be a consistent estimator of Ω, and GLS1 will also be asymptotically efficient. Similarly, $\hat{\rho}_1$ and $\hat{\rho}_2$ estimated by any of the methods of section 2.2 will both be consistent estimators of zero, so that GLS2 will also give asymptotically efficient results. In other words, under the FIC, OLS, GLS1 and GLS2 all give equally asymptotically efficient results.

Suppose next that the disturbances follow a first-order scheme with $\rho_1 \neq 0$. Then by the same type of reasoning, it is clear that OLS yields asymptotically inefficient results, while GLS1 and GLS2 are both asymptotically efficient. Finally, if the disturbances follow a second-order scheme with $\rho_2 \neq 0$, only GLS2 will be asymptotically efficient; OLS and GLS1 will both yield asymptotically inefficient results.

This can be summarized by the general statement that estimation based on the assumption of the "correct" (true) order or on the assumption of "too high" an order of autocorrelation (i.e., an order of autocorrelation greater than the true order) will lead to asymptotically efficient estimators $\hat{\beta}$ and $\hat{\sigma}^2$. On the other hand, estimation based on the assumption of "too low" an order of autocorrelation will in general yield asymptotically inefficient estimates of β and inconsistent estimates of $\hat{\sigma}^2$. In other words, in infinite samples one should always prefer a higher order of autocorrelation, since this will minimize the chance of getting inefficient estimates.[5]

Of course, in small samples none of this need be true. In fact, there are almost no analytic results on the small sample properties of GLS estimators with an estimated covariance matrix; questions of this sort are of necessity usually investigated by Monte Carlo methods.[6] The next section will describe a Monte Carlo experiment which attempts to examine the small sample properties of OLS, GLS1 and GLS2 under various specifications of Ω. In particular, the question which we attempt to answer

---

[5]This discussion has ignored differences in computational costs.

[6]For an example of an experiment comparing the small-sample properties of OLS and GLS1, see [20].

concerns the size of the loss that results in small samples (if in fact any loss does result) from assuming too low or too high an order of autocorrelation. The criteria for choosing the superior estimation procedure will be the variance of the estimator of β and the bias of the estimator of σ_u^2.[7]

2.4 The Experiment

The experiment was conducted in the context of the simple regression model:

$$y_i = \alpha + \beta X_i + u_i, \tag{2.14}$$

with α, β and σ_u^2 all being taken equal to one. The generation of the values of X and u is described below; given X and u, observations on y were created and (2.14) was estimated in each of three ways:

A. OLS

B. GLS1, $\hat{\rho}_1$ estimated by the C-O method given by (2.3) of section 2.1.

C. GLS2, $\hat{\rho}_1$ and $\hat{\rho}_2$ estimated by the C-O method given by (2.7) - (2.12) of section 2.2.[8]

---

[7]The bias of $\hat{\sigma}_u^2$ is considered rather than its variance or mean square error because one is typically not interested in $\hat{\sigma}_u^2$ per se, but only to make confidence statements, tests, etc. It is therefore most important that $\hat{\sigma}_u^2$ not be strongly biased in one direction or the other.

[8]This perhaps deserves a comment. Asymptotically it makes no difference how the ρ's are estimated, as long as the estimators used are consistent. However, in small

In all cases the GLS procedures used were not actually the true GLS procedures, but rather the common approximation to GLS of applying OLS to the transformed equations (2.4) and (2.13), as described above.[9] This procedure was repeated for 100 independent trials under each of a variety of specifications, and the results of the 100 trials were used to calculate the variance of $\hat{\beta}$ and the mean of σ_u^2.

To create observations on X, X_1 was taken to be a N(0,1) deviate from a listing of random deviates prepared by the Rand Corporation.[10] The remaining X_i were generated as follows:

$$X_i = \tau X_{i-1} + (1-\tau^2)^{\frac{1}{2}}\xi_i, \quad i = 2,3,\ldots,N; \qquad (2.15)$$

where ξ_i is a N(0,1)-deviate independent of X_i and of previous ξ's. τ is thus the correlation between successive

---

samples different ways of estimating the ρ's may lead to considerably different results. This is not investigated here because the purpose of this experiment is to compare OLS and GLS under various types of autocorrelation, and to introduce different versions of GLS based on different ways of computing the ρ's would tend to confuse the issue.

[9] As just noted, this makes no difference asymptotically. The approximate procedure is used here because it is typically used in econometric practice.

[10] [42].

X_i. Two values of τ were used, 0.2 and 0.8, since it is well known that the properties of the estimators may depend on the correlation between the X_i. Hence this correlation was held constant at each of the two levels. Two values of N (sample size) were considered, 20 and 100. The results for sample size 20 are designed to show small sample properties; sample size 100 was included so as to get an idea of the results with a somewhat larger sample, and to see if the known asymptotic properties of the estimators begin to emerge.

Observations on u were obtained by reading the N(0,1)-deviates $\xi_1, \xi_2, \ldots, \xi_N$, independent of each other and of the X's, from the Rand listing, and applying the suitable transformation (to be described) for each specification. Given u, y could be constructed, and the various estimation procedures could be applied. First- and second-order autocorrelation schemes were considered, with ρ_1 and ρ_2 taking on all possible values among 0.0, 0.2, 0.4, 0.6 and 0.8, subject to the restriction that $\rho_1 + \rho_2 < 1$.

To simulate the null hypothesis of no autocorrelation, the independent N(0,1)-deviates ξ_i were simply left untransformed; that is, $u_i = \xi_i$ for all i. To simulate first order autocorrelation with parameter ρ_1, the deviates ξ_i were transformed as follows:

$$u_1 = \varepsilon_1$$

$$u_i = \rho_1 u_{i-1} + (1-\rho_1^2)^{\frac{1}{2}} \varepsilon_i, \quad i = 2,3,\ldots,N, \quad (2.16)$$

where the factor $(1-\rho_1^2)^{\frac{1}{2}}$ is included to ensure that the u_i will have a variance of 1 for all i.

Finally, second order autocorrelation was simulated by applying the following transformation to the ε_i:

$$u_i = \varepsilon_1$$

$$u_2 = \rho_1 u_1 + (1-\rho_1^2)^{\frac{1}{2}} \varepsilon_2$$

$$u_j = \rho_1 u_{j-1} + \rho_2 u_{j-2} + a_j^{\frac{1}{2}} \varepsilon_j, \quad j = 3,4,\ldots,N \quad (2.17)$$

where

$$a_j = 1 - \rho_1^2 - \rho_2^2 - 2\rho_1^2\rho_2 \sum_{r=0}^{j-3} \rho_2^r. \quad (2.18)$$

Again the a_j are taken so as to ensure that the u_i will have constant variance.[11]

2.5 Results

Table 2.1 gives the variance of the estimates of β under the various specifications of the model, and Table 2.2

[11]Note two things. First, it is σ_u^2 which is being held constant rather than σ^2. Second, this scheme is not precisely the same as that defined in section 1.2, as it does not start at $-\infty$. However, all the covariances converge to those of section 1.2 as i increases, and even with N=20 the difference should be negligible.

TABLE 2.1.--Variance of $\hat{\beta}$.

	ρ_1	ρ_2	N = 20			N = 100		
			OLS	GLS1	GLS2	OLS	GLS1	GLS2
$\tau = 0.2$	0.0	0.0	0.0636*	0.0744	0.0867	0.0105*	0.0113	0.0114
	0.2	0.0	0.0711*	0.0787	0.0916	0.0108	0.0108	0.0107*
	0.4	0.0	0.0777	0.0720*	0.0820	0.0112	0.0084	0.0083*
	0.6	0.0	0.0815	0.0545*	0.0597	0.0117	0.0053	0.0052*
	0.8	0.0	0.0712	0.0259*	0.0364	0.0118	0.0023*	0.0023
	0.0	0.2	0.0626*	0.0728	0.0817	0.0105	0.0115	0.0102*
	0.0	0.4	0.0610*	0.0710	0.0660	0.0101	0.0111	0.0078*
	0.0	0.6	0.0580	0.0644	0.0432*	0.0095	0.0103	0.0050*
	0.0	0.8	0.0490	0.0460	0.0210*	0.0090	0.0088	0.0023*
	0.2	0.2	0.0708*	0.0756	0.0843	0.0107	0.0105	0.0093*
	0.4	0.2	0.0771	0.0644*	0.0692	0.0114	0.0074	0.0067*
	0.6	0.2	0.0729	0.0373*	0.0381	0.0125	0.0039	0.0036*
	0.2	0.4	0.0692	0.0706	0.0658*	0.0108	0.0100	0.0069*
	0.4	0.4	0.0722	0.0511	0.0462*	0.0128	0.0065	0.0046*
	0.2	0.6	0.0616	0.0580	0.0393*	0.0112	0.0092	0.0042*
$\tau = 0.8$	0.0	0.0	0.0909*	0.1034	0.1179	0.0113*	0.0120	0.0125
	0.2	0.0	0.1141*	0.1278	0.1450	0.0146*	0.0160	0.0166
	0.4	0.0	0.1401*	0.1430	0.1614	0.0191	0.0190*	0.0196
	0.6	0.0	0.1667	0.1359*	0.1553	0.0257	0.0179*	0.0182
	0.8	0.0	0.1806	0.1011*	0.1187	0.0355	0.0103*	0.0105
	0.0	0.2	0.0954*	0.1064	0.1237	0.0136*	0.0143	0.0146
	0.0	0.4	0.0983*	0.1075	0.1150	0.0159	0.0166	0.0147*
	0.0	0.6	0.0993	0.1071	0.0929*	0.0181	0.0187	0.0119*
	0.0	0.8	0.0891	0.0951	0.0612*	0.0196	0.0201	0.0066*
	0.2	0.2	0.1237*	0.1306	0.1482	0.0185*	0.0191	0.0188
	0.4	0.2	0.1556	0.1398*	0.1566	0.0261	0.0207	0.0198*
	0.6	0.2	0.1818	0.1230*	0.1402	0.0393	0.0146	0.0141*
	0.2	0.4	0.1345	0.1329*	0.1372	0.0243	0.0223	0.0176*
	0.4	0.4	0.1756	0.1401*	0.1471	0.0414	0.0210	0.0158*
	0.2	0.6	0.1412	0.1315	0.1166*	0.0347	0.0248	0.0128*

TABLE 2.2.--Mean of $\hat{\sigma}_u^2$.

			N = 20			N = 100		
	ρ_1	ρ_2	OLS	GLS1	GLS2	OLS	GLS1	GLS2
	0.0	0.0	1.0086	1.0015*	0.9578	0.9931*	0.9919	0.9891
	0.2	0.0	0.9760*	0.9648	0.9176	0.9886*	0.9860	0.9819
	0.4	0.0	0.9236*	0.8969	0.8311	0.9802*	0.9726	0.9631
	0.6	0.0	0.8354*	0.7744	0.6835	0.9607*	0.9403	0.9186
	0.8	0.0	0.6515*	0.5309	0.4456	0.8982*	0.8436	0.7916
	0.0	0.2	0.9759*	0.9629	0.9284	0.9898*	0.9882	0.9826
	0.0	0.4	0.9303*	0.9075	0.8577	0.9814*	0.9792	0.9615
$\tau = 0.2$	0.0	0.6	0.8566*	0.8174	0.7320	0.9608*	0.9576	0.9123
	0.0	0.8	0.7181*	0.6493	0.4838	0.9046*	0.8990	0.7828
	0.2	0.2	0.9268*	0.9090	0.8748	0.9856*	0.9819	0.9703
	0.4	0.2	0.8534*	0.8118	0.7641	0.9906	0.9774	0.9453
	0.6	0.2	0.7075*	0.6142	0.5538	0.0101*	0.9666	0.8810
	0.2	0.4	0.8616*	0.8333	0.7859	0.9942*	0.9879	0.9428
	0.4	0.4	0.7654*	0.6998	0.6419	1.1094	1.0798*	0.9091
	0.2	0.6	0.7457*	0.7035	0.6244	1.0398	1.0264*	0.8808
	0.0	0.0	1.0220	1.0146*	0.9688	0.9940*	0.9927	0.9897
	0.2	0.0	0.9820*	0.9693	0.9218	0.9869*	0.9845	0.9802
	0.4	0.0	0.9181*	0.8899	0.8265	0.9750*	0.9685	0.9587
	0.6	0.0	0.8115*	0.7545	0.6726	0.9502*	0.9333	0.9109
	0.8	0.0	0.5999*	0.5022	0.4230	0.8790*	0.8313	0.7783
	0.0	0.2	0.9836*	0.9706	0.9317	0.9886*	0.9871	0.9813
	0.0	0.4	0.9315*	0.9101	0.8532	0.9777*	0.9759	0.9583
$\tau = 0.8$	0.0	0.6	0.8508*	0.8142	0.7142	0.9543*	0.9518	0.9070
	0.0	0.8	0.7080*	0.6394	0.4777	0.8961*	0.8919	0.7771
	0.2	0.2	0.9231*	0.9044	0.8671	0.9804*	0.9772	0.9655
	0.4	0.2	0.8311*	0.7918	0.7478	0.9794*	0.9687	0.9370
	0.6	0.2	0.6564*	0.5813	0.5291	0.9883*	0.9506	0.8703
	0.2	0.4	0.8444*	0.8180	0.7661	0.9839*	0.9790	0.9347
	0.4	0.4	0.7181*	0.6662	0.6147	1.0862	1.0614*	0.8967
	0.2	0.6	0.7117*	0.6777	0.6023	1.0207	1.0104*	0.8688

gives the mean of the estimates of σ_u^2. For each specification of the model an asterisk (*) marks the estimated minimum variance estimator of β and the estimated least biased estimator of σ_u^2.

Consider first the null hypothesis of no autocorrelation; that is, the case $\rho_1 = \rho_2 = 0$. In terms of the estimates of β, OLS is clearly best, and GLS1 is better than GLS2. The differences are considerably larger at sample size 20 than at sample size 100, as the asymptotic equivalence of these estimators under the FIC begins to show at the larger sample size. In terms of the estimates of σ_u^2 there is little difference between the various procedures, though GLS2 seems to give somewhat inferior estimates when N = 20. Finally, the value of τ does not seem to make much difference in this case.

The next specification considered is first-order autocorrelation. Consider first the efficiency of estimation of β. GLS1 clearly dominates GLS2 at sample size 20, though the difference is not terribly great; at sample size 100 they appear to be roughly equivalent, clearly reflecting their asymptotic equivalence in this case. Both GLS1 and GLS2 gave noticeable gains in efficiency over OLS, except for "small" values of ρ_1. The minimum value of ρ_1 necessary to result in a gain in efficiency over OLS was smaller for GLS1 than for GLS2, and for either GLS1 or GLS2 it was smaller with N = 100

than with N = 20. Also, the efficiency of either GLS1 or GLS2 compared to OLS was greater when $\tau = 0.2$ than when $\tau = 0.8$.[12] The results were fairly favorable to the use of GLS in small samples in that ρ_1 of roughly 0.4 sufficed to make GLS1 more efficient than OLS, even with a sample size of only 20, while the "break-even point" for GLS2 was roughly 0.6.[13]

In terms of the bias of the estimates of σ_u^2, it is clear from Table 2 that OLS was markedly superior to either GLS procedure. It is somewhat troubling that this was true even with N = 100. It was true that the OLS estimator of σ_u^2 had the expected downward bias, but it turned out to be actually less biased than the GLS estimators. In fact, a glance at the rest of Table 2 will quickly reveal that this was also true for almost all the other specifications considered.

With respect to these last results, three points should be made. The first is that they could not hold asymptotically; apparently even sample size 100 is not large enough to reveal the asymptotic result in this case. The second point is that different results might have been obtained if $\hat{\sigma}^2$ rather than $\hat{\sigma}_u^2$ had been considered. The

---

[12] This should be expected, since it is well known that the C-O estimator $\hat{\rho}_1$ is more severely biased the larger the value of τ. See, for example, the results in [20].

[13] Asymptotically, of course, either GLS1 or GLS2 would be more efficient than OLS for any $\rho_1 \neq 0$, no matter how small.

third is that only the bias has been considered here; it is quite conceivable that the variance or even the mean square error of the GLS estimators might be smaller than that of the OLS estimator. We will return to these last two points in section 5 of the next chapter; for now the above results will simply be taken as they are.

The third specification considered was second-order autocorrelation with $\rho_1 = 0$, a special case of the general second-order scheme. Considering the variance of the estimators of β, GLS2 is more efficient than OLS or GLS1 except for "small" values of ρ_2, a small value of ρ_2 being 0.4 or less at sample size 20 and 0.2 or less at sample size 100. As before, the relative efficiency of the most efficient estimator is greater with the smaller value of τ. Comparisons of OLS and GLS1 shows OLS to be generally superior, with very few exceptions. The difference was usually quite small, however. The superiority of OLS over GLS1 was slightly more noticeable in the samples of size 20; this is reasonable since OLS and GLS1 are asymptotically equivalent in this case.[14]

The last specification considered is second-order autocorrelation with both ρ_1 and ρ_2 non-zero. Consider the estimates of β. GLS2 was typically most efficient, as would be expected, though GLS1 does quite well when

---

[14] This is true since $\rho_1 = 0$. Thus plim $\hat{\rho}_1 = \rho_1/(1 - \rho_2) = 0 = \rho_1$.

$\tau = 0.8$. With N = 100 GLS2 was always more efficient than GLS1, and GLS2 was more efficient than OLS in all cases except one. With sample size 20 GLS2 was more efficient than OLS in all cases except $\rho_1 = \rho_2 = 0.2$, and it was also more efficient than GLS1 except when $\rho_2 = 0.2$ and in a few cases when $\rho_2 = 0.4$ and $\tau = 0.8$. Again it appears that the relative efficiency of the most efficient estimator is somewhat less with the larger value of τ. Finally, GLS1 is typically more efficient than OLS, especially when ρ_1 is large. This is especially noticeable at sample size 100.

2.6 Summary

One implication of the last section is that GLS unfortunately does not seem to give less biased estimates of σ_u^2 than OLS, even for fairly large sample sizes. As noted earlier, this point will be considered again in section 5 of the next chapter.

In terms of the variance of the estimates of β, however, GLS performed quite well. This was true even for samples as small as 20. In particular, the loss of efficiency in assuming too high an order of autocorrelation was fairly small, while the penalty for assuming too low an order was in many cases quite large. These are of course essentially the asymptotic results, and they showed through quite well in small samples.

One implication of these results is that GLS2 might seem to be a useful procedure, at least if one is primarily interested in efficient estimation of β. Asymptotically, there is no loss in using it unnecessarily, and one will gain by using it if autocorrelation is of second-order form. Even in small samples the gains from its use may be substantial, and the loss in using it unnecessarily (for example, if autocorrelation were of first-order form) is typically small. This makes the currently almost universal use of GLS1 in cases of suspected autocorrelation seem perhaps unjustified. After all, there is frequently no particular reason to suppose that first-order autocorrelation is typically present in real data. The assumption of first-order autocorrelation is generally just a simplifying assumption made in order to make estimation possible. Second-order autocorrelation is a less restrictive assumption, and a second-order scheme ought to provide a reasonable approximation to more different types of autocorrelation than will a first-order scheme. Hence when autocorrelation is not known a priori to be of first-order form, GLS2 might be useful.

Finally, it cannot be overemphasized that the small-sample results obtained here are specific to the particular model used. Limited evidence is better than none, however, and these results may at least be useful in pointing the way for further analytical work in this area.

CHAPTER III

ESTIMATION IN A DISTRIBUTED LAG MODEL

3.1 Introduction

In the last chapter we have introduced methods of estimation in a linear model in the context of second-order autocorrelation of the disturbances, and we analyzed the properties of the resulting estimators. In this chapter we will extend these results to the case of a common type of non-linear model, the distributed lag model.

The simplest distributed lag model is a model of the form

$$y_i = \beta \sum_{j=0}^{\infty} X_{i-j} \lambda^j + u_i, \quad i = 1,2,\ldots,N \tag{3.1}$$

where u_i, $i = 1,2,\ldots,N$, is an unobserved random disturbance; X_i is either a fixed (non-stochastic) number or a random variable independent of the disturbances, with observed values $X_i,\ldots,X_N$; β is a parameter to be estimated; λ is a parameter to estimated, $0 \leq \lambda \leq 1$; and y_i is the observed dependent variable in the model. Clearly the model in this form is not amenable to estimation; it

is usually rewritten in one of two ways. Lagging (3.1) by one observation, multiplying by λ and subtracting yields

$$y_i = \beta X_i + \lambda y_{i-1} + (u_i - \lambda u_{i-1}), \quad i = 2,3,\ldots,N \tag{3.2}$$

This is the so-called "Koyck transformation."[1] Alternately, defining

$$\eta_0 = \beta \sum_{j=0}^{\infty} X_{-j} \lambda^j; \quad W_i(\lambda) = \sum_{j=1}^{i} X_j \lambda^{i-j}; \tag{3.3}$$

(3.1) can be rewritten as

$$y_i = \beta W_i(\lambda) + \eta_0 \lambda^i + u_i. \tag{3.4}$$

This transformation was suggested by Klein.[2]

The model as written in (3.2) has a certain amount of attractiveness since it can apparently be estimated directly. However, it has long been realized that ordinary least squares applied to (3.2) will in general yield inconsistent results, as the disturbance $(u_i - \lambda u_{i-1})$ is correlated with the regressor y_{i-1}. Koyck[3] suggested a method for obtaining consistent estimates of β and λ;

---

[1] [32].

[2] Appendix to [28].

[3] [32].

this method was reinterpreted by Klein[4] in an errors in the variables framework. Liviatan[5] has also suggested a method for obtaining consistent estimates. His procedure is essentially an instrumental variable one, with X_{i-1} serving as the instrument for y_{i-1}.

Both the Koyck-Klein procedure and the Liviatan procedure are fairly straightforward; their main disadvantage is that the resulting estimates are asymptotically inefficient. Assuming that the disturbances in (3.1) are normally distributed and meet the classical conditions (the FIC), asymptotically efficient estimates of β and λ can be obtained by maximum likelihood estimation. Following Klein,[6] (3.1) is rewritten as (3.4). Then the log likelihood function is

$$L = \frac{-N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{1}^{N} [y_i - \beta W_i(\lambda) - \eta_0 \lambda^i]^2 \tag{3.5}$$

Now since the estimator of σ^2 turns out to be independent of the estimators of β and λ (as will be shown in the next section), maximizing L is equivalent to minimizing

---

[4] [28].

[5] [34].

[6] Appendix to [28].

$$L^* = \sum_{1}^{N} [y_i - \beta W_i(\lambda) - \eta_0 \lambda^i]^2 \qquad (3.6)$$

with respect to β, λ and η_0. Clearly the resulting normal equations will be highly non-linear. However, it was noted by Dhrymes[7] and by Zellner and Geisel[8] that if one knew λ, one could form $W_i(\lambda)$ and λ^i and calculate the maximum likelihood estimates $\hat{\beta}$ and $\hat{\eta}_0$ by a simple regression of y_i on $W_i(\lambda)$ and λ^i. When λ is unknown, the procedure is to "search" over the admissable range of λ, picking the value of λ which minimizes the sum of squared errors. The resulting values $\hat{\lambda}$, $\hat{\beta}$, and $\hat{\eta}_0$ are then the maximum likelihood estimators; the maximum likelihood estimator of σ^2 is the sum of squared errors divided by N. It is well known that the estimators $\hat{\beta}$, $\hat{\lambda}$, and $\hat{\sigma}^2$ are consistent and asymptotically efficient; their asymptotic covariance matrix is the inverse of the so-called "information matrix," which will be written out in the next section. The estimator $\hat{\eta}_0$ is not consistent.

If the errors u_i are autocorrelated, the procedure outlined above must of course be modified somewhat. The usual case considered in the literature is once again the case in which the u_i follow a first-order autocorrelation scheme. This paper will treat the case of second-order autocorrelation; the usual results for the first-order case can be obtained by letting $\rho_2 = 0$.

---

[7] [8].

[8] [53].

For convenience, define $Z_1(\lambda)' = [W_1(\lambda) \ldots W_N(\lambda)]$, $Z_2(\lambda)' = [\lambda \ \lambda^2 \ \lambda^3 \ldots \lambda^N]$, $Z(\lambda) = [Z_1(\lambda) \ Z_2(\lambda)]$ and $\gamma' = (\beta \ \eta_0)$. Then the log likelihood function is

$$L = \frac{-N}{2} \log 2\pi - \tfrac{1}{2}\log |\sigma^2\Omega| - \frac{1}{2\sigma^2} [y-Z(\lambda)\gamma]'\Omega^{-1} [y-Z(\lambda)\gamma]. \tag{3.7}$$

Now letting $Z^*(\lambda) = VZ(\lambda)$ and $y^* = Vy$ (V defined as in (1.15)), and recalling that $V'V = \Omega^{-1}$, and defining Q = determinant of $\Omega^{-1} = 1 - \rho_1^2 - 2\rho_2^2 - 2\rho_1^2\rho_2 - \rho_1^2\rho_2^2 + \rho_2^4$, the log likelihood function can be rewritten as

$$L = \frac{-N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 + \tfrac{1}{2} \log Q - \frac{1}{2\sigma^2}[y^*-Z^*(\lambda)\gamma]' [y^*-Z^*(\lambda)\gamma]. \tag{3.8}$$

Given λ, ρ_1 and ρ_2, it is clear that $\hat{\gamma}$ can be calculated by the least squares regression of y^* on $Z^*(\lambda)$; call this estimator $\hat{\gamma}(\lambda,\rho_1,\rho_2)$. By searching over λ and choosing the $\hat{\lambda}$ that minimizes the sum of squared errors, one gets the maximum likelihood estimators of γ and λ, conditional on ρ_1 and ρ_2; denote these by $\hat{\gamma}(\rho_1,\rho_2)$ and $\hat{\lambda}(\rho_1,\rho_2)$. Also $\hat{\sigma}^2(\rho_1,\rho_2)$ can be calculated as the sum of squared errors divided by N. Now substitute $\hat{\gamma}(\rho_1,\rho_2)$ and $\hat{\sigma}^2(\rho_1,\rho_2)$ for γ and σ^2 in (3.8) above to get

$$L(\rho_1,\rho_2) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \hat{\sigma}^2(\rho_1,\rho_2)$$

$$+ \tfrac{1}{2}\log Q - \frac{N\hat{\sigma}^2(\rho_1,\rho_2)}{2\hat{\sigma}^2(\rho_1,\rho_2)} \qquad (3.9)$$

which simplifies to

$$L(\rho_1,\rho_2) = -\frac{N}{2}(\log 2\pi + 1) - \frac{N}{2}\log [\hat{\sigma}^2(\rho_1,\rho_2)Q^{-1/N}]. \qquad (3.10)$$

Finally, to calculate the maximum likelihood estimates one then searches over ρ_1 and ρ_2 and selects those values $\hat{\rho}_1$ and $\hat{\rho}_2$ that minimize $\hat{\sigma}^2(\rho_1,\rho_2)\ Q^{-1/N}$.

Several comments are in order here. First, since $\hat{\sigma}^2(\rho_1,\rho_2)$ is itself determined by a search over λ, what is required is a three-dimensional search. That is, given λ, ρ_1, and ρ_2, y^* and $Z^*(\lambda)$ are formed and y^* is regressed on $Z^*(\lambda)$. The maximum likelihood estimates are obtained by choosing those values $\hat{\lambda}$, $\hat{\rho}_1$, and $\hat{\rho}_2$ that minimize the sum of squared errors divided by $Q^{1/N}$. Secondly, since $\lim_{N\to\infty} Q^{1/N} = 1$, the division by $Q^{1/N}$ will not affect the results asymptotically. Hence asymptotically the estimators obtained by merely minimizing the sum of squared errors are equivalent to the maximum likelihood estimators. Thirdly, it would clearly make no difference

asymptotically if the first two rows of the transformation matrix V were eliminated.

Finally, this entire procedure is clearly a straightforward generalization of the usual procedure for the case of first order autocorrelation, which may be found in Dhrymes[9] or Zellner and Geisel.[10]

3.2 Asymptotic Properties of the Estimators

This section will begin by showing that maximum likelihood estimators have desirable asymptotic properties in the case of normally distributed, second-order autoregressive disturbances. Ignoring the initial conditions,[11] the model can be written

$$y_1 = \beta W_i(\lambda) + u_i; \quad u_i = \rho_1 u_{i-1} + \rho_2 u_{i-2} + \varepsilon_i \tag{3.11}$$

Now for the moment assume that ρ_1 and ρ_2 are known. Then the log likelihood function (conditional on y_1 and y_2) can be written in terms of the ε_i as follows:

$$L = \frac{-N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \Sigma [y_i^* - \beta W_i^*(\lambda)]^2, \tag{3.12}$$

---

[9] [8].

[10] [53].

[11] It is well known that ignoring the initial conditions makes no difference asymptotically. This should be intuitively clear since the contribution of the initial value $\eta_o = E(y_o)$ becomes less and less important as N increases.

where $y_i^* = y_i - \rho_1 y_{i-1} - \rho_2 y_{i-2}$ and $W_i^*(\lambda) = W_i(\lambda)$ $- \rho_1 W_{i-1}(\lambda) - \rho_2 W_{i-2}(\lambda)$. But in this case, with ρ_1 and ρ_2 known, the above transformation has removed the autocorrelation from the disturbance, so that the likelihood function is written in terms of mutually independent variables. If β, λ, and σ^2 are now estimated by maximum likelihood, the consistency and asymptotic efficiency of the resulting estimates is thus clear. It is instructive to write out the information matrix $-E[\frac{\partial^2 L}{\partial\theta_i \partial\theta_j}]$ (where the θ_i are the parameters to be estimated, namely β, λ, and σ^2), which is as follows:

$$\begin{array}{c|ccc|} & & & \\ (\beta) & \frac{1}{\sigma^2}\Sigma W_i^*(\lambda)^2 & \frac{1}{\sigma^2}\Sigma W_i^*(\lambda)R_i^*(\lambda) & 0 \\ & & & \\ (\lambda) & \frac{1}{\sigma^2}\Sigma W_i^*(\lambda)R_i^*(\lambda) & \frac{1}{\sigma^2}\Sigma R_i^*(\lambda)^2 & 0 \\ & & & \\ (\sigma^2) & 0 & 0 & N/2\sigma^4 \\ & & & \end{array} \qquad (3.13)$$

where $R_i(\lambda) = X_{i-1} + 2\lambda X_{i-2} + \cdots + (i-1)\lambda^{i-2}X_1$, and $R_i^*(\lambda) = R_i(\lambda) - \rho_1 R_i(\lambda) - \rho_2 R_{i-2}(\lambda)$. Notice that the matrix is block diagonal, which shows that the estimator of σ^2 is asymptotically independent of the estimators of β and λ, as previously claimed. The covariance matrix of β and λ can be found by inverting the upper left hand block.

We now proceed to the case in which ρ_1 and ρ_2 are not known and must also be estimated by maximum likelihood. The information matrix is in this case the following:

$$
\begin{array}{c|ccccc|}
(\beta) & \frac{1}{\sigma^2}\Sigma W_i^*(\lambda)^2 & \frac{1}{\sigma^2}\Sigma W_i^*(\lambda)R_i^*(\lambda) & 0 & 0 & 0 \\
(\lambda) & \frac{1}{\sigma^2}\Sigma W_i^*(\lambda)R_i^*(\lambda) & \frac{1}{\sigma^2}\Sigma R_i^*(\lambda)^2 & 0 & 0 & 0 \\
(\rho_1) & 0 & 0 & \frac{N\sigma^2}{\sigma^2} & \frac{N\sigma_u^2}{\sigma^2}\frac{\rho_1}{1-\rho_2} & 0 \\
(\rho_2) & 0 & 0 & \frac{N\sigma_u^2}{\sigma^2}\frac{\rho_2}{1-\rho_2} & \frac{N\sigma_u^2}{\sigma^2} & 0 \\
(\sigma^2) & 0 & 0 & 0 & 0 & \frac{N}{2\sigma^4} \\
\end{array}
\qquad (3.14)
$$

The matrix is again block diagonal, and the blocks corresponding to β and λ, and to σ^2, are exactly as they were in the case in which ρ_1 and ρ_2 were known. Hence the estimates of β, λ, and σ^2 have the same asymptotic variances as if ρ_1 and ρ_2 were known, so that the maximum likelihood estimators $\hat{\beta}$, $\hat{\lambda}$, and $\hat{\sigma}^2$ are still asymptotically efficient in the case where ρ_1 and ρ_2 are unknown.

The fact that the maximum likelihood estimators $\hat{\beta}$, $\hat{\lambda}$, and $\hat{\sigma}^2$ are asymptotically just as efficient as they would be if ρ_1 and ρ_2 were known has important implications for estimation. Consider first the case of no autocorrelation of any order, and consider the following three estimation procedures:

A. Estimation based on the assumption of no autocorrelation; i.e., ρ_1 and ρ_2 are set *a priori* to zero.

B. Estimation based on the assumption of first order autocorrelation; i.e., ρ_2 is set *a priori* to zero but ρ_1 is estimated.

C. Estimation based on the assumption of second order autocorrelation; i.e., ρ_1 and ρ_2 are both estimated.

(These clearly correspond to the procedures OLS, GLS1 and GLS2 in last chapter's linear model.) Then it is clear from the discussion above that in terms of the asymptotic efficiency of the estimates of β, λ, and σ^2, all three estimation procedures are identical. If the autocorrelation in the sample were of first-order form, then procedure (A) would yield asymptotically inefficient results, while procedures (B) and (C) would be equally asymptotically efficient. On the other hand, if the autocorrelation in the sample were second-order, then procedures (A) and (B) would yield inefficient results, while (C) would be asymptotically efficient. In other words, there is no loss in asymptotic efficiency in assuming a higher order of autocorrelation than is in fact present. On the other hand, assuming a lower order of autocorrelation than is in fact present leads to asymptotically inefficient results. (Note the close analogy to the asymptotic results of the last chapter.)

Finally, it should be noted that the asymptotic efficiency of maximum likelihood estimation is purchased

at a cost--first, because of the large computational burden, and, second, because of the possibility of inefficiency if too low an order of autocorrelation is specified. The second problem can be alleviated by choosing a very high order of autocorrelation, but this will only make the computational burden correspondingly larger. There is, therefore, perhaps something to be said for an estimation procedure such as Liviatan's,[12] which is not only simple computationally but which also yields results that are at least consistent regardless fo the covariance matrix of the disturbances. Of course, Liviatan's estimators are not as efficient as maximum likelihood;[13] if one's criterion for choosing an estimator is asymptotic efficiency, then maximum likelihood cannot be beaten.

3.3 Small Sample Properties of the Maximum Likelihood Estimators

The conclusion of the last section is that from the standpoint of asymptotic efficiency one should always assume a more general type of autocorrelation (e.g., second-order rather than first-order); there is nothing to lose by doing so,[14] and there is possibly much to gain.

---

[12] [34].

[13] For an evaluation of the difference in efficiency, see [2].

[14] Unless one is concerned with computational costs.

In small samples, however, this may not be true, if only because one loses degrees of freedom by estimating redundant ρ's. In other words, in small samples one must presumably strike some sort of balance between the possibility of assuming too low an order of autocorrelation (and hence possibly getting badly biased and inefficient results) and the possibility of assuming too high an order of autocorrelation (and hence possibly losing efficiency). It would therefore be desirable to know how large a loss there is in small sample efficiency when too high an order of autocorrelation is specified, as well as how bad the results are which one gets by assuming too low an order of autocorrelation. Unfortunately, the small sample properties of the estimators in a distributed lag model have not proven amenable to analytical solution, so that questions of this type must at present be investigated by Monte Carlo methods. This section will describe a Monte Carlo experiment which was performed in an attempt to discover the small sample properties of maximum likelihood estimators which assume no autocorrelation, first-order autocorrelation, and second-order autocorrelation, when the true type of autocorrelation in the sample is in turn each of the three types above. Note that this is essentially the same question that the experiment of the last chapter attempted to answer for a linear model.

The model considered was of the form of equation (3.4). The values chosen for the parameters were $\beta = 1$, $\lambda = 0.5$ and $\eta_0 = 0$, so that the structure of the model was

$$y_i = W_i(0.5) + u_i, \quad i = 1,\ldots,N \tag{3.16}$$

Three values of N were considered--20, 50, and 100. The sequence $X_1,\ldots,X_N$ was in each case constructed by reading independent N(0,1) deviates $\xi_1,\ldots,\xi_N$ off a "random" number generator and then transforming them as follows:

$$X_1 = \xi_1$$

$$X_i = \tau X_{i-1} + \sqrt{1 - \tau^2}\,\xi_i, \quad i = 2,3,\ldots,N \tag{3.17}$$

Each X_i thus has variance one, and the correlation between successive X's is τ. Two values of τ were used, 0.25 and 0.75.

Three types of disturbances were simulated--classical, first-order Markov, and second-order Markov. Classical disturbances were simulated by simply taking independent N(0,1) deviates and leaving them untransformed. Second-order distrubances were simulated by taking independent N(0,1) deviates $\xi_1,\ldots,\xi_N$ and transforming them as follows:

$$u_1 = \sigma_u \varepsilon_1$$

$$u_2 = [\rho_1/(1-\rho_2)]\, u_1 + \sigma_u\, [1-\rho_1^2/(1-\rho_2)^2]^{\frac{1}{2}}\, \varepsilon_2$$

$$u_i = \rho_1 u_{i-1} + \rho_2 u_{i-2} + \varepsilon_i, \quad i = 3,4,\ldots,N \qquad (3.18)$$

First-order disturbances were simulated in the same way except that in this case $\rho_2 = 0$.

Given the values of ρ_1, ρ_2, τ and N, then, u_i and X_i were generated as above, $W_i(\lambda)$ and λ^i were formed, y_i was generated, and the parameters were estimated by maximum likelihood, assuming in turn no autocorrelation, first-order autocorrelation, and second-order autocorrelation. For each specification of the model, this procedure was repeated a number of times, and the observed means, variances, and mean square errors of the parameters were calculated. The number of repetitions was 100 when N = 20 and 50 when N = 50 or N = 100.[15]

This study will concern itself only with the resulting estimates of β, λ, and σ^2.

---

[15]The smaller number of repetitions was used at the larger sample sizes to economize somewhat on computer time. A possible justification is that since the estimators should be better behaved at the larger sample sizes (at least those which are consistent), fewer observations should be required to characterize their distribution.

3.4 Results of the Experiment

Tables 3.1 and 3.2 contain the results with $N = 20$, with $\tau = 0.25$ and 0.75 respectively. Consider first the case of no autocorrelation; the results are given in the upper left hand corner of Tables 3.1 and 3.2. (The figures under the headings 0, 1, and 2 are the results obtained by the estimation procedures assuming no autocorrelation, first-order autocorrelation, and second-order autocorrelation, respectively. That is, they correspond to last chapter's headings OLS, GLS1, and GLS2.) Note that the estimates of β and λ obtained by each of the estimation procedures are almost identical. If one loses small sample efficiency in the estimation of β and λ by assuming autocorrelation when none is in fact present, the loss is apparently very small, even for a sample as small as 20. The estimates of σ^2 are all significantly biased downward, and in the estimation of σ^2 one does seem to lose somewhat by assuming too high a degree of autocorrelation, in terms of the mean square error of the estimate. However, this is apparently due to an increase in the bias rather than to a larger variance, and the difference, while statistically significant, is actually numerically quite small. Finally, the results were rather insensitive to τ, the correlation between successive X_i.

TABLE 3.1.--N = 20 τ = 0.25.

		$\rho_1 = 0$ $\rho_2 = 0$			$\rho_1 = .4$ $\rho_2 = 0$			$\rho_1 = .8$ $\rho_2 = 0$		
		0	1	2	0	1	2	0	1	2
β	MEAN	1.006	1.010	1.016	0.979	0.982	0.989	0.761	1.034	1.036
	VAR	0.054	0.053	0.052	0.084	0.067	0.066	0.221	0.054	0.057
	MSE	0.054	0.053	0.052	0.088	0.067	0.066	0.277	0.056	0.059
λ	MEAN	0.478	0.466	0.459	0.491	0.473	0.471	0.637	0.507	0.513
	VAR	0.035	0.038	0.040	0.050	0.051	0.055	0.111	0.058	0.060
	MSE	0.035	0.039	0.041	0.050	0.052	0.055	0.130	0.058	0.060
σ^2	MEAN	0.873	0.854	0.835	0.911	0.753	0.736	1.467	0.794	0.782
	VAR	0.088	0.089	0.085	0.088	0.057	0.054	0.683	0.103	0.101
	MSE	0.104	0.111	0.112	0.096	0.118	0.124	0.901	0.145	0.148
		$\rho_1 = 0$ $\rho_2 = .4$			$\rho_1 = 0$ $\rho_2 = .8$			$\rho_1 = .4$ $\rho_2 = .4$		
		0	1	2	0	1	2	0	1	2
β	MEAN	0.996	1.016	1.019	0.804	0.813	1.028	0.785	0.982	1.027
	VAR	0.067	0.060	0.045	0.179	0.173	0.040	0.230	0.083	0.056
	MSE	0.067	0.061	0.045	0.218	0.208	0.041	0.276	0.083	0.057
λ	MEAN	0.475	0.460	0.461	0.574	0.571	0.486	0.649	0.539	0.516
	VAR	0.048	0.047	0.040	0.116	0.111	0.036	0.080	0.064	0.048
	MSE	0.049	0.048	0.042	0.121	0.117	0.036	0.101	0.066	0.048
σ^2	MEAN	0.984	0.941	0.764	2.178	2.128	0.778	1.215	0.848	0.718
	VAR	0.135	0.138	0.069	1.334	1.430	0.088	0.359	0.118	0.053
	MSE	0.135	0.141	0.124	2.722	2.702	0.137	0.406	0.141	0.132

TABLE 3.2.--N = 20 τ = 0.75.

		$\rho_1 = 0$ $\rho_2 = 0$			$\rho_1 = .4$ $\rho_2 = 0$			$\rho_1 = .8$ $\rho_2 = 0$		
		0	1	2	0	1	2	0	1	2
β	MEAN	0.994	0.994	0.995	0.979	0.972	0.961	0.924	0.990	0.995
	VAR	0.036	0.036	0.037	0.073	0.078	0.077	0.139	0.108	0.100
	MSE	0.036	0.036	0.037	0.073	0.078	0.079	0.145	0.108	0.100
λ	MEAN	0.500	0.500	0.502	0.507	0.512	0.521	0.511	0.493	0.494
	VAR	0.011	0.011	0.011	0.025	0.027	0.026	0.062	0.042	0.040
	MSE	0.011	0.011	0.011	0.025	0.027	0.026	0.062	0.042	0.040
σ^2	MEAN	0.843	0.827	0.818	0.945	0.830	0.821	1.224	0.684	0.656
	VAR	0.092	0.092	0.092	0.153	0.108	0.106	0.571	0.057	0.054
	MSE	0.119	0.121	0.125	0.156	0.137	0.138	0.619	0.156	0.172
		$\rho_1 = 0$ $\rho_2 = .4$			$\rho_1 = 0$ $\rho_2 = .8$			$\rho_1 = .4$ $\rho_2 = .4$		
		0	1	2	0	1	2	0	1	2
β	MEAN	0.971	0.969	0.977	0.948	0.959	1.010	0.745	0.902	0.983
	VAR	0.052	0.052	0.040	0.078	0.074	0.032	0.198	0.190	0.176
	MSE	0.053	0.053	0.040	0.081	0.076	0.032	0.264	0.200	0.177
λ	MEAN	0.512	0.512	0.505	0.508	0.498	0.475	0.643	0.525	0.514
	VAR	0.017	0.017	0.017	0.036	0.035	0.014	0.074	0.070	0.076
	MSE	0.017	0.017	0.017	0.036	0.035	0.015	0.095	0.071	0.077
σ^2	MEAN	1.000	0.990	0.829	1.823	1.764	0.704	1.067	0.853	0.744
	VAR	0.152	0.144	0.103	0.843	0.813	0.059	0.400	0.185	0.093
	MSE	0.152	0.144	0.132	1.521	1.396	0.146	0.405	0.207	0.158

Consider next the specification of first-order autocorrelation. Two values of ρ_1 were considered, 0.4 and 0.8. The results are given in the two upper right hand columns of Tables 1 and 2. Notice first that in all cases the estimators assuming first-order autocorrelation and those assuming second-order autocorrelation are extremely close together. Once again there is practically no real loss in assuming too high a degree of autocorrelation; the asymptotic equivalence of these two procedures apparently comes through quite well even at sample size 20. Next consider the estimators of β and λ only, and compare the estimators which assume no autocorrelation with those which assume either first- or second-order autocorrelation. With $\rho_1 = 0.4$ the results are somewhat mixed, but with $\rho_1 = 0.8$ the inferiority of assuming $\rho_1 = 0$ is apparent. The estimates assuming no autocorrelation are clearly biased and inefficient. Finally, consider the estimators of σ^2. The estimators assuming no autocorrelation have a considerably larger variance than those which correctly assume autocorrelation to be present, expecially when ρ_1 is large. Also note that when autocorrelation is present, estimation assuming no autocorrelation leads to an overestimate of σ^2.[16] All of these conclusions are again true for both values of τ.

---

[16]At least it leads to higher estimates than if autocorrelation were in fact not present. Since σ^2 was somewhat underestimated in the absence of autocorrelation,

The last specification considered is that of second-order autocorrelation. The results are given in the bottom half of Tables 1 and 2. Three cases are considered: $\rho_1 = 0$, $\rho_2 = 0.4$; $\rho_1 = 0$, $\rho_2 = 0.8$; and $\rho_1 = \rho_2 = 0.4$. As could be expected, estimation assuming second-order autocorrelation leads to the best (in the sense of mean square error) estimates of β, λ, and σ^2. (It does not always lead to the least biased estimates of σ^2; this will be discussed further in the next section.) Estimation assuming first-order autocorrelation seems somewhat better than estimation assuming no autocorrelation, especially in the case in which ρ_1 is non-zero. Also it is once again true that assuming too low an order of autocorrelation leads to an overestimate of σ^2. Finally, the results are again rather insensitive to the value of τ.

This completes the experiment for the case N = 20. The salient result is that the known asymptotic properties of the estimators seemed to hold fairly well even for such a small sample. In particular, understating the order of autocorrelation leads to very bad results, while overstating the order of autocorrelation is essentially costless. Apparently the only serious problem with

---

the effect of increasing ρ_1 is to first push the mean of $\hat{\sigma}^2$ toward the true value before actually leading to an overestimate.

assuming too high an order of autocorrelation is that the resulting estimate of σ^2 is apt to be more downward biased than if the correct order of autocorrelation were assumed.

As a further check on the rate of convergence of the results to the known asymptotic results, the experiment was partially duplicated with N = 50. Since the value of τ did not seem to materially affect the results in the previous case, only one value (0.75) was used at this sample size. The results are given in Table 3.3. The asymptotic properties of the estimators are in this case an extremely good guide to the observed properties. Under the null hypothesis the estimation procedures based on all three assumptions gave essentially identical results, and the downward bias of σ^2 has largely disappeared. In the case of first-order autocorrelation the estimation procedures which assumed first- and second-order autocorrelation gave almost identical results, while estimation based on the assumption of no autocorrelation gave substantially worse results. In the two cases of second-order autocorrelation, estimation based on second-order autocorrelation gave the best results, while estimation based on the assumption of first-order autocorrelation generally gave better results than estimation based on the assumption of no autocorrelation. These are

TABLE 3.3.--N = 50 τ = 0.75.

		$\rho_1 = 0$ $\rho_2 = 0$			$\rho_1 = .4$ $\rho_2 = 0$		
		0	1	2	0	1	2
β	MEAN	1.001	1.000	1.000	1.043	1.059	1.060
	VAR	0.026	0.026	0.027	0.032	0.025	0.026
	MSE	0.026	0.026	0.027	0.034	0.029	0.029
λ	MEAN	0.497	0.497	0.497	0.470	0.463	0.464
	VAR	0.0086	0.0086	0.0088	0.0101	0.0093	0.0095
	MSE	0.0086	0.0086	0.0088	0.0110	0.0106	0.0109
σ^2	MEAN	0.948	0.946	0.940	1.106	0.963	0.957
	VAR	0.030	0.030	0.030	0.062	0.042	0.043
	MSE	0.032	0.033	0.033	0.073	0.043	0.045

		$\rho_1 = 0$ $\rho_2 = .4$			$\rho_1 = .4$ $\rho_2 = .4$		
		0	1	2	0	1	2
β	MEAN	0.988	0.987	0.988	0.898	0.974	1.004
	VAR	0.032	0.032	0.028	0.128	0.050	0.043
	MSE	0.033	0.032	0.028	0.139	0.051	0.043
λ	MEAN	0.504	0.505	0.500	0.541	0.492	0.483
	VAR	0.0106	0.0102	0.0081	0.048	0.029	0.018
	MSE	0.0106	0.0102	0.0081	0.050	0.029	0.018
σ^2	MEAN	1.004	1.001	0.869	1.536	0.987	0.831
	VAR	0.057	0.056	0.037	0.348	0.076	0.023
	MSE	0.057	0.056	0.054	0.636	0.076	0.058

essentially the same results as were obtained with N = 20; the difference is that they hold more distinctly in this case.

The last specification considered is the null hypothesis with N = 100. The results are given in Table 3.4. As would be expected, all three procedures give essentially the same results. This was done primarily as a check of whether the calculations were performed correctly, and they do seem to have been.

3.5 Comments and Summary

In this chapter we have developed the maximum likelihood estimators of a distributed lag model when the disturbances follow a second-order process. It has also been pointed out that there is no loss in the asymptotic efficiency of maximum likelihood estimation when one assumes an order of autocorrelation higher than the true order, while assuming too low an order leads to asymptotically inefficient results. These results parallel those shown in Chapter II in the context of a linear model, and the implication is the same--asymptotically, one should always assume a very general type of autocorrelation.

A Monte Carlo experiment was then performed to see how well these asymptotic results hold in small samples. The experiment suggested that, even with samples

TABLE 3.4.--N = 100 $\tau = 0.75$.

		$\rho_1 = 0$ $\rho_2 = 0$		
		0	1	2
β	MEAN	1.0003	1.0005	1.0000
	VAR	0.0092	0.0093	0.0092
	MSE	0.0092	0.0093	0.0092
λ	MEAN	0.4991	0.4990	0.4991
	VAR	0.0028	0.0028	0.0028
	MSE	0.0028	0.0028	0.0028
σ^2	MEAN	0.9668	0.9660	0.9632
	VAR	0.0230	0.0229	0.0230
	MSE	0.0240	0.0241	0.0243

as small as 20, the asymptotic properties of the estimators provide a rather good guide to the small sample properties. That is, even in small samples there is very little loss in assuming too high an order of autocorrelation, but there is a considerable loss in assuming too low an order. This is essentially the same result obtained by the Monte Carlo experiment of the last chapter, at least as far as the regression coefficient β is concerned.

Recall, however, that in the last chapter it was found that GLS did not necessarily lead to "better" estimators of σ_u^2, even when autocorrelation was actually present. This seemed puzzling, and it was conjectured at that time that this might have occurred either because $\hat{\sigma}_u^2$ was considered rather than σ^2, or because the criterion for choosing the "best" estimator of σ_u^2 was just the size of the bias. It now seems probable that the latter reason was the cause of the problem. After all, the experiment of this chapter considered $\hat{\sigma}^2$, not $\hat{\sigma}_u^2$, and it was still common for an estimation procedure which assumed too low an order of autocorrelation to give the least biased estimator of σ^2. However, estimation procedures which assumed the <u>true</u> order did give the estimators of σ^2 which typically had the smallest variance and mean square error. Hence it appears that whether or not estimation which takes autocorrelation into account (when it is present) gives "better" estimates of the variance (either

σ^2 or σ_u^2) depends on whether one considers the "best" estimators to be those with the smallest bias or those with the smallest mean square error.

CHAPTER IV

TESTING FOR SECOND-ORDER AUTOCORRELATION: A GENERALIZATION OF THE DURBIN-WATSON TEST

4.1 Introduction

Because GLS procedures are computationally more complicated than OLS and because there does seem to be some loss in small sample efficiency involved in using GLS under the FIC, it is desirable to be able to test the null hypothesis of the independence of the disturbances against the alternative of autocorrelation. In the context of a linear model, the test which has been most often used to test for autocorrelation is the Durbin-Watson test,[1] for which the test statistic is the ratio

$$d_1 = \frac{\sum_{2}^{N} (\tilde{u}_i - \tilde{u}_{i-1})^3}{\sum_{1}^{N} \tilde{u}_i^2} \qquad (4.1)$$

the $\tilde{u}_i$ being the residuals from the OLS regression of y

[1] [14] and [15].

on X.[2] This test was specifically designed to detect autocorrelation in the form of a first-order scheme. In fact, the statistic d_1 can be written

$$d_1 \doteq 2(1 - \hat{\rho}_1), \tag{4.2}$$

where $\hat{\rho}_1$ is the C-O estimator of ρ_1 given by equation (2.3). Clearly positive first-order autocorrelation ($\rho_1 > 0$) will tend to lead to small values of d_1; negative first-order autocorrelation will tend to lead to large values. It should also be clear that this test may not be very effective in detecting types of autocorrelation other than first-order; an obvious example of a type of autocorrelation to which it would be insensitive would be a second-order scheme with $\rho_1 = 0$.

In order to test for second-order autocorrelation, we will propose the second-order Durbin-Watson test, to be based on the test statistic

$$d_2 = \frac{\sum_{2}^{N} (\tilde{u}_i - \tilde{u}_{i-1})^2 + \sum_{3}^{N} (\tilde{u}_i - \tilde{u}_{i-2})^2}{\sum_{1}^{N} \tilde{u}_i^2} \tag{4.3}$$

---

[2]The Durbin-Watson test is not applicable to a distributed lag model such as the one presented in the last chapter. A test which is asymptotically valid in such a model has been recently suggested in [13].

This test should be able to detect first- and second-order autocorrelation. It should also be able to detect more different types of autocorrelation than the ordinary first-order test, since the second-order scheme on which it is based should be able to approximate more different types of autocorrelation than can a first-order scheme. In the next chapter we will discuss the power of both tests against various alternatives; the remainder of this chapter will be devoted to the consideration of the distribution of the above statistic under the null hypothesis of the FIC.

4.2 Calculation of Significance Points

The statistics d_1 and d_2 can each be written in matrix form as

$$d_i = \frac{\tilde{u}'A_i\tilde{u}}{\tilde{u}'\tilde{u}}, \quad i = 1,2; \tag{4.4}$$

where $\tilde{u}$ is the vector of least squares residuals and A_1 and A_2 are NxN matrices defined as follows:

$$A_1 = \begin{vmatrix} 1 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & \dots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & -1 & 1 \end{vmatrix}$$

$$A_2 = \begin{vmatrix} 2 & -1 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 4 & -1 & -1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 4 & -1 & -1 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & -1 & 4 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & -1 & -1 & 2 \end{vmatrix}$$

(4.5)

This is useful since a number of results have been established in the literature for the distribution of a test statistic $d = \frac{\tilde{u}'A\tilde{u}}{\tilde{u}'\tilde{u}}$, where A is any real, non-singular, symmetric, positive definite matrix. In particular, define

$$M = I - X(X'X)^{-1}X', \tag{4.6}$$

and let

$$Z = MA \tag{4.7}$$

Then Z has N-K real positive characteristic roots (N and K being the dimensions of the regressor matrix X) and K zero roots; number the positive roots, in increasing order, $\pi_1, \pi_2, \ldots, \pi_{N-K}$. Durbin and Watson have shown[3] that d can be rewritten as follows:

$$d = \frac{\sum_{1}^{N-K} v_i^2 \pi_i}{\sum_{1}^{N-K} v_i^2}, \tag{4.8}$$

where the v_i are independent N(0,1) variables. Now, following Koerts and Abrahamse,[4] one can note that $P(d < d^*) =$

$$P\left[\sum_{1}^{N-K} \pi_i v_i^2 < d^* \sum_{1}^{N-K} v_i^2\right] = P\left[\sum_{1}^{N-K} \eta_i v_i^2 < 0\right], \text{ where}$$

where $\eta_i = \pi_i - d^*$. Using a result from Imhof,[5] Koerts and Abrahamse note that

---

[3] [14].

[4] [30].

[5] [25].

$$P\left[\sum_{1}^{N-K} \eta_i v_i^2 < 0\right] = \tfrac{1}{2} - \frac{1}{\pi}\int_0^{\infty} \frac{\sin\left[\tfrac{1}{2}\sum_{1}^{N-K} \arctan(\eta_i r)\right]}{r \prod_{1}^{N-K} (1+\eta_i^2 r^2)^{\frac{1}{4}}}\, dr. \tag{4.9}$$

Numerical integration is feasible since Imhof has provided the limit of the integrand as $r \to 0$; it is $\tfrac{1}{2}\sum_{1}^{N-K} \eta_i$. He also provides a bound for the truncation error caused by integrating over the finite range [0,R]; to hold the truncation error to ε one must take R equal to

$$R = \left[\pi \left(\frac{N-K}{2}\right) \varepsilon \prod_{1}^{N-K} |\eta_i|^{\frac{1}{2}}\right]^{\frac{-2}{N-K}}. \tag{4.10}$$

Thus the exact probability that d lies below any value d* can be calculated by numerical integration, even though the form of the distribution of d is not known. This procedure was developed by Koerts and Abrahamse[6] for the statistic d_1; it clearly can also be applied to d_2. All that need be done is to insert the proper A_i in (4.7); from there on the procedure is the same in each case.

---

[6] See [30] or [31]. The same procedure was subsequently but independently developed in [41].

4.3 Approximations

The exact procedure of the preceding section has the drawback of being rather difficult computationally, so that one might sometimes wish to resort to an approximation procedure so as to save computational effort. Henshaw[7] and Durbin and Watson[8] have given fairly comprehensive reviews of the available procedures, so that only a few brief comments need be made here.

Given that d has been written as in (4.8), the moments of d are readily computed. In particular, as noted by Durbin and Watson,[9]

$$E(d) = \frac{1}{N-K} \sum_{1}^{N-K} \pi_i \equiv \bar{\pi} \tag{4.11}$$

and

$$\mathrm{Var}(d) = \frac{2 \sum_{1}^{N-K} (\pi_i - \bar{\pi})^2}{(N-K)(N-K+2)}. \tag{4.12}$$

Now it has been proven that the distribution of d is asymptotically normal,[10] but it is not known how good a fit the normal distribution would provide in small samples. In fact, there is some limited evidence to suggest that the beta distribution may provide a better

---

[7] [22].

[8] [16].

[9] [15].

[10] [4].

approximation to the distribution of d,[11] and the beta distribution has generally been used to approximate the distribution of d. Since it is clear from (4.8) that the possible range of d is $[\pi_1, \pi_{N-K}]$, and since the beta distribution over a given range is a two parameter distribution, it is possible to fit a beta distribution having the same mean, variance and range as the true distribution of d. This is essentially the procedure of Henshaw; however, he goes to rather great lengths to avoid computing the eigenvalues of Z. Where direct eigenvalue calculation is possible, the following procedure is somwhat simpler, at least conceptually.

Having calculated Z and its roots, calculate E(d) and Var(d) from (4.11) and (4.12). Normalize d to the range [0,1] by defining

$$x = \frac{d - \pi_1}{\pi_{N-K} - \pi_1} \tag{4.13}$$

Then clearly

$$E(x) = \frac{E(d) - \pi_1}{\pi_{N-K} - \pi_1} \tag{4.14}$$

and

---

[11] For some evidence see [43] or [5].

$$\text{Var}(x) = \frac{\text{Var}(d)}{(\pi_{N-K} - \pi_1)^2} \tag{4.15}$$

Assume that x is a beta variable; clearly it has range [0,1]. It is well known that such a variable with density

$$\frac{1}{\beta(p,q)} \; x^{p-1}(1 - x)^{q-1} \tag{4.16}$$

has

$$E(x) = \frac{p}{p+q} \tag{4.17}$$

and

$$\text{Var}(x) = \frac{pq}{(p+q)^2(p+q+1)} \cdot \tag{4.18}$$

Since E(x) and Var(x) are known from (4.14) and (4.15), (4.17) and (4.18) can be solved for p and q. The results can be stated as follows:

$$h = \frac{E(x)}{1 - E(x)}$$

$$q = \frac{h}{\text{Var}(x)(1 + h)^3} - \frac{1}{1 + h}$$

$$p = qh \tag{4.19}$$

Given q and p, one can find the critical values of x in any table of the incomplete beta function[12] and get the corresponding critical value of d from the relation

$$d_\alpha = \pi_1 + x_\alpha(\pi_{N-K} - \pi_1). \tag{4.20}$$

This procedure fits a beta variable of the same mean and variance as d into the exact range of d; the only element of approximation is the use of the beta distribution. Durbin and Watson[13] and Theil and Nagar[14] have also proposed beta approximations, but each makes approximations about the mean, variance and range of d that are replaced here by exact results.

Finally, this procedure also clearly applies to each of the d_i defined above. Once again all that need be done is to insert the proper A_i into (4.7) and from there on the procedure is identical in each case.

4.4 The Bounds Test

Because of the substantial computational burdens involved in the procedures of either of the last two sections, it would clearly be desirable to avoid them as often as possible. Durbin and Watson have provided a partial solution in the case of the first-order test, by

---

[12] For example, [39].

[13] [15].

[14] [48].

tabulating the critical points of statistics d_L and d_U whose critical points bound the critical points of d_1.[15] In this section we will provide similar bounds for the distributions of the higher order tests defined above.

Again consider the statistic $d = \frac{\tilde{u}'A\tilde{u}}{\tilde{u}'\tilde{u}}$, A being one of the A_i in (4.5). Then A has N-1 positive characteristic roots; number them in increasing order $\lambda_1,\lambda_2,\ldots,\lambda_{N-1}$. Recall that $\pi_1,\ldots,\pi_{N-K}$ are the positive roots, in increasing order, of Z = MA. Then the basis for the present procedure is the fact, proved by Durbin and Watson[16] that

$$\lambda_i \leq \pi_i \leq \lambda_{i+K'}, \quad i = 1,2,\ldots,N-K \tag{4.21}$$

where K' = K-1 = the number of regressors not including the constant term (which must be present).

It is then natural to define the variables d_L and d_U as follows:

$$d_L = \frac{\sum_1^{N-K} \lambda_i v_i^2}{\sum_1^{N-K} v_i^2}, \quad d_U = \frac{\sum_1^{N-K} \lambda_{i+K'} v_i^2}{\sum_1^{N-K} v_i^2} \tag{4.22}$$

---

[15] [15].

[16] [14].

Comparing these with (4.8), it is evident that the distribution of d is bounded by the distribution of d_L and d_U. Now, given the matrix A, the significance points of d_L and d_U can be calculated by the methods of section 4.2 and tabulated. Note that they depend on the matrix X only in that they depend on N and K'.

The critical points of $(d_1)_L$ and $(d_1)_U$ have been tabulated in Durbin and Watson.[17] Tables 4.1 - 4.3 of this paper contain the significance points of $(d_2)_L$ and $(d_2)_U$. To use the tables, simply compute the value of d and compare it to the critical points of d_L and d_U for the given values of N and K' and the desired alpha level. If d is less than the alpha level critical point of d_L, the null hypothesis is rejected at that alpha level. If d is greater than the critical point of d_U, the null hypothesis is accepted at that alpha level. If d falls between the critical points of d_L and d_U, the test is inconclusive; the critical point of d itself can then be calculated by the methods of sections 4.2 or 4.3. This procedure again applies to all of the d_i defined in this paper.

---

[17] [14].

TABLE 4.1.--0.01 level critical values of $(d_2)_L$ and $(d_2)_U$.

		K' = 1		K' = 2		K' = 3		K' = 4		K' = 5	
	15	1.99	2.61	1.80	3.11	1.63	3.59	--	--	--	--
	20	2.26	2.70	2.13	3.03	1.98	3.41	--	--	--	--
	25	2.45	2.79	2.35	3.04	2.24	3.30	2.10	3.59	--	--
	30	2.59	2.87	2.50	3.07	2.41	3.27	2.31	3.50	--	--
	35	2.70	2.94	2.63	3.10	2.54	3.27	2.46	3.45	2.38	3.65
	40	2.79	2.99	2.73	3.13	2.66	3.28	2.59	3.44	2.52	3.59
N	45	2.88	3.04	2.82	3.16	2.74	3.28	2.69	3.43	2.63	3.56
	50	2.95	3.08	2.88	3.19	2.82	3.31	2.77	3.43	2.72	3.54
	60	3.04	3.15	2.99	3.25	2.94	3.34	2.90	3.42	2.86	3.52
	70	3.12	3.22	3.07	3.30	3.03	3.38	2.99	3.44	2.96	3.53
	80	3.17	3.27	3.13	3.33	3.10	3.40	3.06	3.46	3.03	3.54
	90	3.21	3.30	3.18	3.35	3.15	3.42	3.12	3.48	3.09	3.55
	100	3.25	3.32	3.22	3.37	3.20	3.43	3.17	3.49	3.15	3.56

TABLE 4.2.--0.05 level critical values of $(d_2)_L$ and $(d_2)_U$.

		K' = 1		K' = 2		K' = 3		K' = 4		K' = 5	
	15	2.52	3.10	2.33	3.53	2.12	3.92	--	--	--	--
	20	2.74	3.15	2.59	3.45	2.45	3.76	--	--	--	--
	25	2.88	3.20	2.77	3.43	2.65	3.68	2.53	3.91	--	--
	30	2.99	3.25	2.90	3.43	2.80	3.63	2.70	3.83	--	--
	35	3.07	3.29	2.99	3.45	2.91	3.61	2.83	3.77	2.75	3.94
	40	3.14	3.33	3.07	3.46	3.00	3.60	2.93	3.74	2.86	3.88
N	45	3.19	3.36	3.13	3.47	3.07	3.59	3.01	3.72	2.95	3.84
	50	3.24	3.39	3.18	3.49	3.12	3.59	3.07	3.70	3.03	3.82
	60	3.31	3.44	3.26	3.52	3.21	3.60	3.17	3.69	3.13	3.78
	70	3.36	3.46	3.32	3.53	3.28	3.61	3.24	3.68	3.21	3.76
	80	3.39	3.48	3.36	3.54	3.33	3.61	3.30	3.68	3.27	3.76
	90	3.42	3.50	3.40	3.55	3.37	3.61	3.34	3.68	3.32	3.75
	100	3.45	3.51	3.42	3.56	3.40	3.61	3.38	3.67	3.36	3.74

TABLE 4.3.--0.10 level critical values of $(d_2)_L$ and $(d_2)_U$.

		K' = 1		K' = 2		K' = 3		K' = 4		K' = 5	
	15	2.80	3.36	2.60	3.74	2.40	4.09	--	--	--	--
	20	2.98	3.38	2.84	3.67	2.69	3.95	--	--	--	--
	25	3.11	3.42	2.99	3.64	2.88	3.86	2.76	4.08	--	--
	30	3.20	3.45	3.10	3.62	3.01	3.81	2.91	3.99	--	--
	35	3.26	3.48	3.19	3.63	3.11	3.78	3.03	3.93	2.94	4.09
	40	3.31	3.50	3.25	3.63	3.18	3.76	3.11	3.89	3.04	4.03
N	45	3.35	3.52	3.30	3.63	3.24	3.75	3.18	3.87	3.12	3.99
	50	3.39	3.54	3.33	3.64	3.26	3.74	3.23	3.85	3.18	3.96
	60	3.44	3.58	3.40	3.66	3.35	3.74	3.31	3.82	3.28	3.91
	70	3.48	3.58	3.45	3.65	3.41	3.73	3.37	3.81	3.34	3.88
	80	3.51	3.59	3.48	3.65	3.45	3.72	3.42	3.80	3.39	3.87
	90	3.53	3.60	3.51	3.65	3.48	3.71	3.46	3.78	3.43	3.85
	100	3.55	3.61	3.53	3.65	3.51	3.71	3.49	3.77	3.47	3.83

CHAPTER V

THE POWER OF THE GENERALIZED DURBIN-WATSON TEST

5.1 Analytical Results

It is well known[1] that the power of tests of the Durbin-Watson type depends not only on the alpha level and the alternative Ω matrix, but also on the regressor matrix X. That is, the power of the test can be calculated exactly, but this requires knowledge of α, Ω, and X. Unfortunately, the fact that the power function is X-dependent makes an analytical comparison of the power of the tests rather intractable. It should not be surprising, therefore, that even for the case of the first-order Durbin-Watson test there has been practically no analytical work done comparing its power to that of other tests. Rather, resort has usually been made to Monte Carlo methods.

One exception is the recent work by Durbin and Watson[2] in which it is shown that if the columns of X are linear combinations of the eigenvectors of the matrix A_1

---

[1] For example, see [30] or [31].

[2] [16].

(the matrix such that $d_1 = \frac{\tilde{u}'A_1\tilde{u}}{\tilde{u}'\tilde{u}}$), then the first-order Durbin-Watson test is a uniformly most powerful invariant test[3] against the alternative hypothesis of first-order autocorrelation. For arbitrary X matrices the first-order Durbin-Watson test is a locally most powerful invariant test in some neighborhood of the null hypothesis.

It would clearly be nice to be able to make analogous statements about the powers of the second-order Durbin-Watson tests when the alternative hypothesis is second-order autocorrelation. Unfortunately this is not possible. Durbin and Watson's proof for the case of first-order autocorrelation[4] begins by considering the log likelihood function generated by the first-order Markov normal variables $u_1,\ldots,u_N$, conditional on u_0; this is proportional to the following expression:

$$L_1^* = (1 + p_1^2) \sum_1^N u_i^2 - \rho_1^2(u_1^2 + u_N^2) - 2\rho_1 \sum_2^N u_i u_{i-1}. \tag{5.1}$$

For this they substitute the closely similar expression

$$L_1^{**} = (1 + \rho_1^2) \sum_1^N u_i^2 - \rho_1(u_1^2 + u_N^2) - 2\rho_1 \sum_2^N u_i u_{i-1}. \tag{5.2}$$

---

[3] For a discussion of invariant tests see [33], Chapter 5.

[4] [16].

Then the key result is the following:

$$L_1^{**} = u'[(1 - \rho_1)^2 I + \rho_1 A_1]u, \tag{5.3}$$

where I is the identity matrix and A_1 is as previously defined. Now in the case of second-order autocorrelation the log likelihood function is proportional to L_2^* defined as follows:

$$L_2^* = (1 + \rho_1^2 + \rho_2^2) \sum_1^N u_i^2 - 2\rho_1 \sum_2^N u_i u_{i-1} - 2\rho_2 \sum_3^N u_i u_{i-2}$$

$$+ \rho_1\rho_2 \sum_2^{N-1} u_i u_{i-1} - (\rho_1^2 + \rho_2^2)(u_1^2 + u_N^2)$$

$$- \rho_2^2(u_2^2 + u_{N-1}^2). \tag{5.4}$$

Unfortunately one can <u>not</u> find suitable constants c_1 and c_2 such that the following is true:

$$L_2^* \triangleq u'(c_1 I + c_2 A_2)u; \tag{5.5}$$

this becomes evident immediately if one expands the above expression. Hence the line of reasoning of the Durbin-Watson proof breaks down in this case, so that their result can unfortunately not be generalized to the second-order case. Another way to see this is to note that the following <u>is</u> true:

$$L_2^* \doteq u'[(1 - \rho_1 - \rho_2)^2 I + (\rho_1 - \frac{\rho_1\rho_2}{2})A_1 + \rho_2(A_2 - A_1)]u.$$

(5.6)

Following the development of Durbin and Watson, one can now proceed to derive the locally most powerful invariant test, whose rejection region turns out to be defined as follows:

$$(\rho_1 - \frac{\rho_1\rho_2}{2}) \frac{\tilde{u}'A_1\tilde{u}}{\tilde{u}'\tilde{u}} + \rho_2 \frac{\tilde{u}'(A_2 - A_1)\tilde{u}}{\tilde{u}'\tilde{u}} < d^* \quad (d^* \text{ some constant}).$$

(5.7)

Recalling that $d_2 = \frac{\tilde{u}'A_2\tilde{u}}{\tilde{u}'\tilde{u}}$, this clearly is not of the form $d_2 < d^*$, so that the second-order Durbin-Watson test is not locally most powerful invariant against the alternative of second-order autocorrelation. Indeed, it is easy to see that there does not exist a matrix A not containing the parameters ρ_1 and ρ_2 such that the test based on the statistic $\frac{\tilde{u}'A\tilde{u}}{\tilde{u}'\tilde{u}}$ is locally most powerful invariant in this case.

To summarize these results, there is some theoretical reason to believe that the first-order Durbin-Watson test should be more powerful than the second-order test in the case of first-order autocorrelation, at least

for small values of ρ_1. At present it is not possible to make an analytical comparison of the tests in the cases of autocorrelation of higher orders. From a purely intuitive point of view, however, it would seem reasonable to expect that the first- and second-order Durbin-Watson tests would be most appropriate in testing for autocorrelation of first- and second-order, respectively. The next section will describe a Monte Carlo experiment which was performed in an attempt to test this admittedly intuitive hypothesis.

5.2 A Monte Carlo Comparison of the Tests

The purpose of this section is to compare by Monte Carlo methods the performance of the first- and second-order Durbin-Watson tests in detecting first- and second-order autocorrelation. Such a comparison could be made in at least two different ways. The first would be, given a regressor matrix X and disturbance term covariance matrix Ω, simply to calculate the power of each of the tests exactly by the methods of Koerts and Abrahamse.[5] The second possible procedure would be to generate random deviates having covariance matrix Ω, to use them to generate observations on the dependent variable y, and to regress y on X and actually carry out the tests. By repeating this procedure many times with

[5] [30].

independent sets of random deviates one should get a good idea of the distribution of the test statistics. If one is simply interested in the power of the various tests, the first procedure would of course be simpler. The second procedure, however, should more readily reveal other features of the distribution of the test statistics (e.g., mean, variance, shape, etc.) which would be extremely difficult to calculate by Koerts-Abrahamse methods. This Monte Carlo experiment was therefore run along the lines of the second procedure outlined above.

In performing the experiment, three different X matrices were used, of dimension 20 x 6, 35 x 4, and 50 x 2.[6] Each consisted of a constant term plus K-1 columns of random digits taken from a listing compiled by the Rand Corporation.[7] For each X matrix, each test was set at the 0.01, 0.05 and 0.10 alpha levels by means of the beta approximation described in the last chapter. The procedure was then to run 1000 independent trials under each specification of Ω, each time constructing observations on y, to calculate the observed

---

[6]Note that different numbers of regressors were used at each sample size, so that the results are somewhat more general than if the same type of X matrix had been used in each case. The price paid for this added generality is that it is now not valid to compare the powers of the tests as sample size increases; this did not seem to be a significant cost. Without known asymptotic properties, it is not terribly informative to watch the behavior of the tests as sample size increases.

[7][42].

mean and standard deviation of each test statistic, and to count the number of rejections by each test at each alpha level.

Specifically, all elements of the regression coefficient vector were taken equal to one,[8] so that, given u, y was obtained by

$$y_i = \sum_j X_{ij} + u_i, \quad i = 1,2,\ldots,N. \tag{5.8}$$

Observations on u were created by suitable transformations of random N(0,1) deviates taken from the Rand list, as indicated below.

The first specification considered is the null hypothesis of the full ideal conditions (FIC). Here the random deviates were simply left untransformed. Table 5.1 gives the number of rejections by each test in 1000 independent trials. The number of rejections should average 10, 50 and 100 with standard deviations of 3.2, 6.9, and 9.5. An asterisk indicates that the number of rejections obtained was significantly different from the expected number at the 5% level. It will be noted that the beta approximation performed quite well; it was therefore judged good enough to be used in the rest of the experiment as well. Table 5.2 compares the actual

---

[8] This was done purely for convenience, and does not in any case affect the distribution of the residuals, so that it makes no difference.

TABLE 5.1.--Number of rejections per 1,000 trials under the null hypothesis of no autocorrelation.

		Alpha-Level		
		0.01	0.05	0.10
N = 20	d_1	15	52	112
	d_2	20*	61	113
N = 35	d_1	18*	81*	145*
	d_2	17	60	113
N = 50	d_1	15	67*	108
	d_2	12	46	95

TABLE 5.2.--Means and standard deviations of d_1 and d_2 under the null hypothesis.

		Mean		Standard Deviation	
		Actual	Predicted	Actual	Predicted
N = 20	d_1	2.130	2.141	.4365	.4308
	d_2	3.953	3.973	.5788	.5800
N = 35	d_1	1.955	1.998	.3165	.3167
	d_2	3.982	3.995	.4585	.4502
N = 50	d_1	1.986	1.986	.2778	.2769
	d_2	3.948	3.950	.3841	.3901

mean and standard deviation of the statistics with the "predicted" mean and standard deviation used in calculating the beta approximation. The actual values are in all cases extremely close to the theoretical correct values, as would be expected in a sample of 1000 trials.

The next specification considered is that of first-order autocorrelation. For each trial at each value of ρ_1, the procedure is to take random N(0,1) deviates $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ and to transform them as follows:

$$u_1 = \varepsilon_1$$

$$u_j = \rho_1 u_{j-1} + (1 - \rho_1^2)^{\frac{1}{2}} \varepsilon_j, \quad j = 2,3,\ldots,N \qquad (5.9)$$

Values of ρ_1 considered were 0.2, 0.4, 0.6, and 0.8. Table 5.3 gives the number of rejections, as before, while Table 5.4 gives the means and standard deviations of the test statistics. In general, the results confirm our expectations. For any test at any alpha level and sample size, the expected value of the statistic decreases and the power of the test increases as ρ_1 increases. The effect on the standard deviation of the statistics is ambiguous. Also, for any alpha level and sample size, the first-order test is almost invariably most powerful. However, the difference in power is generally not very large.

TABLE 5.3.--Number of rejections under first-order autocorrelation.

		$\rho_1 = 0.2$			$\rho_1 = 0.4$			$\rho_1 = 0.6$			$\rho_1 = 0.8$		
Alpha-Level:		.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
N = 20	d_1	59	148	246	150	327	447	328	569	687	596	761	835
	d_2	56	139	206	148	291	402	324	539	637	595	739	825
N = 35	d_1	120	316	448	415	670	787	761	917	958	940	984	992
	d_2	89	212	307	297	511	628	670	829	891	921	980	988
N = 50	d_1	205	425	554	694	868	922	973	994	998	1000	1000	1000
	d_2	121	287	421	528	730	822	906	964	977	996	998	1000

TABLE 5.4.--Means and standard deviations of d_1 and d_2 under first-order autocorrelation.

		$\rho_1 = 0.2$		$\rho_1 = 0.4$		$\rho_1 = 0.6$		$\rho_1 = 0.8$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
N = 20	d_1	1.887	.4434	1.636	.4451	1.371	.4490	1.098	.4590
	d_2	3.700	.6434	3.354	.7112	2.914	.7831	2.389	.8573
N = 35	d_1	1.678	.3107	1.390	.3010	1.113	.2885	0.863	.2787
	d_2	3.648	.5168	.3202	.5622	2.657	.5863	2.036	.5827
N = 50	d_1	1.625	.2710	1.266	.2567	0.908	.2344	0.554	.1996
	d_2	3.533	.4567	2.983	.4949	2.302	.5200	1.496	.4999

The third specification considered is second-order autocorrelation. This is constructed by transforming the N(0,1) deviates ε_i, i = 1,2,...,N, as follows:

$$u_1 = \varepsilon_1$$

$$u_2 = \rho_1 u_1 + (1 - \rho_1^2)^{\frac{1}{2}} \varepsilon_2$$

$$u_j = \rho_1 u_{j-1} + \rho_2 u_{j-2} + a_j^{\frac{1}{2}} \varepsilon_j, \quad j = 3,4,...,N, \tag{5.10}$$

where

$$a_j = 1 - \rho_1^2 - \rho_2^2 - 2\rho_1^2\rho_2 \sum_{i=0}^{j-3} \rho_2^i. \tag{5.11}$$

Table 5.5 gives the number of rejections for the six cases in which both ρ_1 and ρ_2 are non-zero, while Table 5.6 gives the means and standard deviations. Note that for all alpha levels and all sample sizes, the second-order test is most powerful. This is true even when ρ_1 is large relative to ρ_2. Also note that the means of all the test statistics are lower, and the standard deviations higher, than under the null hypothesis.

Table 5.7 gives the number of rejections for the four cases in which $\rho_1 = 0$ but $\rho_2 \neq 0$, with Table 5.8 giving the means and standard deviations. As might be expected, the second-order test is most powerful. In fact, the first-order test shows very little power.

TABLE 5.5.--Number of rejections under second-order autocorrelation.

		$\rho_1 = 0.2$ $\rho_2 = 0.2$			$\rho_1 = 0.2$ $\rho_2 = 0.4$			$\rho_1 = 0.2$ $\rho_2 = 0.6$			$\rho_1 = 0.4$ $\rho_2 = 0.2$			$\rho_1 = 0.4$ $\rho_2 = 0.4$			$\rho_1 = 0.6$ $\rho_2 = 0.2$		
Alpha-Level:		.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
N = 20	d_1	86	178	280	114	216	308	148	254	337	214	391	519	284	458	552	450	634	711
	d_2	117	239	323	209	355	463	303	458	552	258	447	557	410	567	649	502	665	729
N = 35	d_1	180	344	468	228	376	485	271	399	482	492	706	806	567	737	830	817	930	957
	d_2	225	419	554	496	646	750	628	795	863	551	748	820	767	880	926	356	940	966
N = 50	d_1	302	512	620	423	611	669	551	668	723	785	902	950	869	944	960	982	996	1000
	d_2	422	638	742	717	837	901	882	951	979	821	923	961	961	988	994	970	994	998

TABLE 5.6.--Means and standard deviations of d_1 and d_2 under second-order autocorrelation.

		$\rho_1 = 0.2$ $\rho_2 = 0.2$		$\rho_1 = 0.2$ $\rho_2 = 0.4$		$\rho_1 = 0.2$ $\rho_2 = 0.6$		$\rho_1 = 0.4$ $\rho_2 = 0.2$		$\rho_1 = 0.4$ $\rho_2 = 0.4$		$\rho_1 = 0.6$ $\rho_2 = 0.2$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
N = 20	d_1	1.885	.4935	1.873	.5597	1.875	.6401	1.583	.5029	1.524	.5759	1.275	.5119
	d_2	3.462	.6912	3.214	.7356	2.989	.8252	3.065	.7886	2.777	.8761	2.576	.8829
N = 35	d_1	1.661	.3599	1.642	.4161	1.641	.4847	1.332	.3446	1.267	.3911	1.030	.3253
	d_2	3.317	.5587	2.990	.6036	2.688	.6469	2.817	.6094	2.430	.6515	2.223	.6258
N = 50	d_1	1.568	.3202	1.488	.3828	1.369	.4669	1.136	.2990	0.954	.3456	0.811	.2832
	d_2	3.117	.4968	2.675	.5616	2.196	.6426	2.464	.5520	1.878	.6072	1.676	.6302

TABLE 5.7.--Number of rejections under second-order autocorrelation ($\rho_1 = 0$).

		$\rho_1 = 0.2$			$\rho_2 = 0.4$			$\rho_2 = 0.6$			$\rho_2 = 0.8$		
Alpha-Level:		.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
N = 20	d_1	21	67	116	33	86	124	45	98	140	54	104	136
	d_2	43	115	182	87	183	271	134	256	360	174	312	408
N = 35	d_1	46	96	142	54	116	175	58	118	179	62	104	133
	d_2	68	175	256	153	332	448	230	461	547	325	528	641
N = 50	d_1	32	91	146	50	123	176	77	151	187	90	145	182
	d_2	82	207	338	267	484	626	492	681	785	587	748	843

TABLE 5.8.--Means and standard deviations of d_1 and d_2 under second-order autocorrelation ($\rho_1 \equiv 0$).

		$\rho_1 = 0.2$		$\rho_2 = 0.4$		$\rho_2 = 0.6$		$\rho_2 = 0.8$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
N = 20	d_1	2.174	.4807	2.216	.5351	2.273	.5987	2.387	.6620
	d_2	3.748	.5978	3.541	.6307	3.351	.6670	3.212	.6972
N = 35	d_1	2.031	.4088	2.054	.4299	2.228	.5307	2.371	.6293
	d_2	3.661	.5063	3.434	.5165	3.280	.5438	3.103	.5842
N = 50	d_1	2.002	.3285	2.032	.3929	2.096	.4830	2.260	.6207
	d_2	3.612	.4173	3.289	.4600	2.997	.4831	2.800	.6151

Looking at the table of means and standard deviations, one can see that in this case the mean of the first-order test statistic actually increases over its value under the FIC. It is only because the standard deviation also increases as ρ_2 increases that we obtain more rejections than under the null hypothesis. In fact, if the increase in the mean predominated over the increase in the standard deviation, one could actually get less rejections in this case than under the null hypothesis. Clearly the first-order test is not suitable when $\rho_1 = 0$ and $\rho_2 \neq 0$.

5.3 Summary

To summarize these results, some general patterns clearly appear. The first-order test appears to be most powerful for the case of first-order autocorrelation and the second-order test most powerful for second-order autocorrelation. Hence if a test is used which is of higher order than the true order of the autocorrelation scheme, some loss of power apparently results relative to the case in which the test of "correct" order is used. However, the Monte Carlo evidence suggests that this loss is fairly small. On the other hand, use of a test of "too-low" order also forfeits power, and this loss can be very substantial indeed; this is especially true if the lower order ρ's are small. All these results agree with the intuition expressed in section 5.1. Of course, it

should be clear that these results are dependent on the particular X matrices used in the experiment. Certain X matrices may (or may not) exist for which these conclusions would not hold.

To the extent that these conclusions are generally valid, however, they would seem to imply that one should perhaps be wary of using the ordinary Durbin-Watson test in the common case of testing for autocorrelation which is not known a priori to be of first-order form. Even if the autocorrelation in the sample should happen to be of first-order form, use of d_2 rather than d_1 would entail only a fairly small loss of power. On the other hand, cases do exist for which the use of d_1 rather than d_2 would entail an almost complete loss of power. To put the same point somewhat differently, the second-order Durbin-Watson test is more generally applicable than the first-order test, and it would appear to be useful in the general case of testing for autocorrelation of unknown form.

Finally, it should be noted that other tests have recently been proposed which are not tied to the idea of first-order autocorrelation. For example, Durbin[9] has proposed as test based on the cumulative periodogram of the residuals which may be useful in detecting autocorrelation of a general nature; an interesting topic

---

[9] [12].

for further research would be to compare the power of this test with the test proposed here.

CHAPTER VI

CONCLUDING REMARKS

As noted in Chapter I, autocorrelation can cause serious problems in econometric regression analysis. Econometricians have therefore developed testing procedures to test for its presence, and estimation procedures to alleviate the problems which it causes when it is present. These procedures must of necessity make some assumptions about what types of autocorrelation might be present. In particular, the testing and estimation procedures which have most commonly been used have been based on the assumption that the autocorrelation in the sample is of first-order form. If autocorrelation is present, but not of first-order form, one can only hope that a first-order scheme is in some sense a reasonable approximation to the true scheme. If it is not, the ordinary procedures may not be very appropriate.

Having argued that the usual first-order autocorrelation scheme is unduly restrictive, this study then proposed a generalization in the form of a second-order scheme. The common testing and estimation procedures were generalized to forms appropriate for this case. Finally, the new procedures were compared to the

original procedures in terms of their performance in the presence of various types of autocorrelation. It was typically found that the "best" procedures in each case were those which assumed the true order of autocorrelation. However, there was a fundamental asymmetry in that the losses involved in assuming too high an order of autocorrelation were generally rather small, while the losses involved in assuming too low an order were often quite serious. This would seem to imply that when one does not know a priori what type of autocorrelation is present, one should proceed under rather general assumptions about its form.

Of course, there must be some limit on how general a process one can assume and still get meaningful results. (After all, without some restrictions on Ω estimation is literally impossible.) This study does not claim to have discovered where that limit might lie. However, it does seem clear that the assumption of second-order autocorrelation lies well within the permitted range of generality. It would therefore seem that testing and estimation procedures based on the assumption of second-order autocorrelation might often be more appropriate than those which assume first-order autocorrelation, at least when one does not have a priori knowledge of the true form of the disturbance term covariance matrix.

REFERENCES

REFERENCES

1. Amemiya, T. "Specification Analysis in the Estimation of the Parameters of a Simultaneous Equation Model with Autocorrelated Residuals." Econometrica (1966), pp. 283-306.

2. Amemiya, T. and W. Fuller. "A Comparative Study of Alternative Estimators in a Distributed Lag Model." Econometrica (1967), pp. 509-529.

3. Anderson, R. L. "Distribution on the Serial Correlation Coefficient." Annals of Mathematical Statistics (1943), pp. 1-13.

4. Anderson, T. W. "On the Theory of Testing Serial Correlation." Skandinavisk Aktuarietidskrift (1948), pp. 88-116.

5. Anderson, R. L. and T. W. Anderson. "Distribution of the Circular Correlation Coefficient for Residuals from a Fitted Fourier Series." Annals of Mathematical Statistics (1950), pp. 59-81.

6. Anscombe, F. "Examination of Residuals." Proceedings of the Fourth Berkely Symposium on Mathematical Statistics and Probability, Vol. I, University of California Press (1961), pp. 1-36.

7. Cochrane, D. and G. H. Orcutt. "Application of Least-Squares Regression to Relationships Containing Autocorrelated Error Terms." Journal of the American Statistical Association (1949), pp. 32-61.

8. Dhrymes, P. "Efficient Estimation of Distributed Lags with Autocorrelated Errors." International Economic Review (1969), pp. 47-67.

9. Dhrymes, P. Distributed Lags: Problems of Formulation and Estimation (forthcoming).

10. Durbin, J. "The Fitting of Time-Series Models." Review of the International Statistical Institute (1960), pp. 233-243.

11. Durbin, J. "Estimation of Parameters in Time Series Regression Models." Journal of the Royal Statistical Society, Series B (1960), pp. 139-153.

12. Durbin, J. "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals." Biometrika (1969), pp. 1-15.

13. Durbin, J. "Testing for Serial Correlation in Least Squares Regression when Some of the Regressors are Lagged Dependent Variables." Econometrica (forthcoming).

14. Durbin, J. and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression I." Biometrika (1950), pp. 409-428.

15. Durbin, J. and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression II." Biometrika (1951), pp. 159-178.

16. Durbin, J. and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression III." Biometrika (forthcoming).

17. Goldberger, A. Econometric Theory. New York: Wiley (1964).

18. Grenander, U. "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance." Annals of Mathematical Statistics (1954), pp. 252-272.

19. Griliches, Z. "A Note on the Serial Correlation Bias in Estimates of Distributed Lags." Econometrica, (1961), pp. 65-73.

20. Griliches, Z. and P. Rao. "Small Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors." Journal of the American Statistical Association (1969), pp. 253-272.

21. Hart, B. and J. von Neumann. "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance." Annals of Mathematical Statistics (1942), pp. 207-214.

22. Henshaw, R. "Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances." Econometrica (1966), pp. 646-660.

23. Hildreth, C. "Asymptotic Distribution of Maximum Likelihood Estimators in a Linear Model with Autoregressive Disturbances." Annals of Mathematical Statistics (1969), pp. 583-594.

24. Hildreth, C. and J. R. Lu. "Demand Relations with Autocorrelated Disturbances." Agricultural Experiment Station Technical Bulletin No. 276, Michigan State University (1960).

25. Imhof, P. "Computing the Distribution of Quadratic Forms in Normal Variables." Biometrika (1961), pp. 419-426.

26. Johnston, J. Econometric Methods. New York: McGraw-Hill (1963).

27. Kadiyala, K. R. "Testing for the Independence of Regression Disturbances." Econometrica (1970), pp. 97-117.

28. Klein, L. "The Estimation of Distributed Lags." Econometrica (1958), pp. 553-565.

29. Koerts, J. "Some Further Notes on Disturbance Estimates in Regression Analysis." Journal of the American Statistical Association (1967), pp. 169-183.

30. Koerts, J. and A. P. J. Abrahamse. "On the Power of the BLUS Procedure." Journal of the American Statistical Association (1968), pp. 1227-1236.

31. Koerts, J. and A. P. J. Abrahamse. On the Theory and Application of the General Linear Model. Rotterdam: University of Rotterdam Press (1969).

32. Koyck, L. Distributed Lags and Investment Analysis. Amsterdam: North Holland Publishing Company (1955).

33. Lehman, E. Testing Statistical Hypotheses. New York: Wiley (1959).

34. Liviatan, N. "Consistent Estimation of Distributed Lags." International Economic Review (1963), pp. 44-52.

35. Lyttkens, E. "Standard Errors of Regression Coefficients by Autocorrelated Residuals." In H. Wold, Econometric Model Building: Essays on the Causal Chain Approach. Amsterdam: North Holland Publishing Company (1964).

36. Maddala, G. S. "Generalized Least Squares with an Estimated Variance-Covariance Matrix." Econometrica (forthcoming).

37. Malinvaud, E. Statistical Methods of Econometrics. Chicago: Rand-McNally (1966).

38. Neumann, J. von. "Distribution of the Ratio of the Mean-Square Successive Difference to the Variance." Annals of Mathematical Statistics (1941).

39. Pearson, K. Tables of the Incomplete Beta Function. Cambridge: Biometrika Office (1934).

40. Prais, S. J. and C. B. Winsten. "Trend Estimators and Serial Correlation." Cowles Commission Discussion Paper No. 383, Chicago (1954).

41. Press, S. and R. Brooks. "Testing for Serial Correlation in Regression (revised)." Center for Mathematical Studies in Business and Economics. University of Chicago, Report 6911 (1969).

42. RAND Corporation. One Million Random Digits and One Hundred Thousand Deviates. Santa Monica (1950).

43. Rubin, H. "On the Distribution of the Serial Correlation Coefficient." Annals of Mathematical Statistics (1945), pp. 211-215.

44. Sargan, J. D. "The Estimation of Economic Relationships using Instrumental Variables." Econometrica (1958), pp. 393-415.

45. Sargan, J. D. "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals." Econometrica (1961), pp. 414-426.

46. Theil, H. "Analysis of Disturbances in Regression Analysis." Journal of the American Statistical Association (1965), pp. 1067-1079.

47. Theil, H. "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances." Journal of the American Statistical Association (1968), pp. 242-253.

48. Theil, H. and A. Nagar. "Testing the Independence of Regression Disturbances." Journal of the American Statistical Association (1961), pp. 793-806.

49. Watson, G. S. "Serial Correlation in Regression Analysis." Biometrika (1955), pp. 327-342.

50. Watson, G. S. "Linear Least Squares Regression." Annals of Mathematical Statistics (1967), pp. 1679-1699.

51. Wickens, M. R. "The Consistency and Efficiency of Generalized Least Squares in Simultaneous Equations Systems with Autocorrelated Errors." Econometrica (1969), pp. 651-659.

52. Zellner, A. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." Journal of the American Statistical Association (1962), pp. 348-368.

53. Zeller, A. and M. Geisel. "Analysis of Distributed Lag Models with Applications to Consumption Function Estimation." Econometrica (forthcoming).