COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 301

ON THE ASYMPTOTIC PROPERTIES OF CERTAIN TWO-STEP PROCEDURES COMMONLY USED IN THE ESTIMATION OF DISTRIBUTED LAG MODELS

David M. Grether and G.S. Maddala

October 8, 1970

ON THE ASYMPTOTIC PROPERTIES OF CERTAIN TWO-STEP PROCEDURES COMMONLY USED IN THE ESTIMATION OF DISTRIBUTED LAG MODELS*

by

David M. Grether and G.S. Maddala

I. Introduction

It is by now well known that in the presence of lagged dependent variables and serially correlated errors certain two-step procedures are not asymptotically as efficient as the method of Maximum Likelihood (ML). (See Amemiya and Fuller [1], Maddala [4], Dhrymes [2].) In the present paper we provide a convenient classification of the several two-step procedures that are currently in use, present some guidelines for choice between the different two-step estimators and show that the two-step procedures are inefficient not only when compared with the ML method but under some conditions when compared with the first step of the two-step procedure. The plan of the paper is as follows. Section II provides the classification of the two-step procedures. In Section III we compare the asymptotic covariance matrix of the two-step estimators with that of the ML estimators and the first step estimators in the two-step procedures. Section IV presents the conclusions of the paper.

*This research has been financed partly by the MSSB Workshop on Lags in Economic Behavior held at the University of Chicago in 1970 and partly by grants from the National Science Foundation and the Ford Foundation to the Cowles Foundation for Research in Economics at Yale University.

II. A Classification of Two-Step Procedures

Consider the model

$$y_t = \alpha y_{t-1} + \beta x_t + u_t \tag{1}$$

$$u_t = \rho u_{t-1} + e_t \quad |\rho| < 1 \tag{2}$$

where $\{e_t\}$ are serially independent. There are two commonly used two-step procedures.

Method 1: Write (1) and (2) as

$$y_t = \alpha y_{t-1} + \beta x_t + \rho u_{t-1} + e_t \tag{3}$$

In the first step we obtain consistent estimates $\hat{\alpha}$, $\hat{\beta}$ of the parameters α, β in (1)--say by the instrumental variable method. From these compute the residuals

$$\hat{u}_t = y_t - \hat{\alpha} y_{t-1} - \hat{\beta} x_t .$$

Now substitute $\hat{u}_{t-1}$ for u_{t-1} in equation (3) and estimate equation (3) by ordinary least squares. This method of using estimates of residuals as regressors is often used e.g. by Taylor and Wilson's Three Pass Least Squares [5], and Gupta [3], for the estimation of distributed lag models.

Method 2: In the first step we obtain, as before, consistent estimates $\hat{\alpha}$, $\hat{\beta}$ of the parameters α, β in (1). Then from the computed residuals we obtain a consistent estimate of the covariance matrix of the disturbances u_t, and estimate (1) by generalized least squares. This is the procedure discussed by Wallis [6].

Methods 1 and 2 differ in the way they treat the computed residuals --one uses them as regressors and the other uses them in obtaining an estimate of the covariance matrix. We will show later that the former procedure is not a desirable one.

An alternative two-step procedure, discussed by Dhrymes [2] is the following. Consider the distributed lag model

$$y_t = \frac{\alpha}{1 - \lambda L} x_t + e_t \tag{4}$$

where L is the lag operator defined by $L^k x_t \equiv x_{t-k}$. Define

$$\bar{y}_t = \frac{\alpha}{1 - \lambda L} x_t \tag{5}$$

we can write (4) as

$$y_t = \alpha x_t + \lambda \bar{y}_{t-1} + e_t \tag{6}$$

The two-step procedure is: obtain initial consistent estimates $\hat{\alpha}$, $\hat{\lambda}$ of α, λ. Compute

$$\hat{y}_t = \hat{\alpha} \sum_{i=0}^{t-1} \hat{\lambda}^i x_{t-i}$$

Substitute $\hat{y}_{t-1}$ for $\bar{y}_{t-1}$ in (6) and estimate (6) by ordinary least squares.

The procedure is similar to Method 1 described above inasmuch as an estimated value of one of the regressors is substituted for the regressor.[1]

---

[1] This is also similar to the procedure followed in two-stage least squares.

In all these problems, there are two questions that arise. The first is the question: when is the two-step procedure inefficient compared with the ML method? The second is the question of whether the two-step procedure iterated further produces the ML estimates.

The first question which is discussed by Maddala [4] is easily answered by looking at the information matrix.[2] Suppose the model involves two sets of parameters θ_1 and θ_2, where θ_1 is the set of parameters of interest and θ_2 are the 'nuisance' parameters. In the model given by equations (1) and (2) θ_1 consists of α, β and θ_2 consists of ρ. Let V_1 be the asymptotic covariance matrix of the ML estimates of θ_1 when the 'nuisance' parameters θ_2 are known. Let V_2 be the corresponding covariance matrix when the parameters θ_2 are not known but θ_1 and θ_2 are jointly estimated by the method of ML. Finally, let V_3 be the covariance matrix of a 2-step estimator of θ_1 based on a consistent estimate of θ_2. If the information matrix is block-diagonal i.e. the ML estimates of θ_1 and θ_2 are uncorrelated, then $V_1 = V_2 = V_3$. Otherwise $V_1 < V_2 < V_3$ (the inequality being used in the sense that $V_1 - V_2$ and $V_2 - V_3$ are negative semi-definite).

Consider, for example, the following model:

$$y_t = \frac{\alpha}{1-\lambda L} x_t + \frac{1}{1-\rho L} e_t \tag{7}$$

This is a distributed lag model with serially correlated errors. However,

---

[2] The information matrix is the expected value of the matrix of second partial derivatives of the logarithm of the likelihood function were a negative sign.

one can easily verify that the information matrix is block-diagonal--the ML estimates of (α, λ) and ρ are uncorrelated. Hence a two-step procedure based on a consistent estimate of ρ is asymptotically as efficient as the ML estimate with ρ estimated or with ρ known. On the other hand, in the case of the model given by equations (1) and (2) this is not the case. Considering the fact that we can write equation (7) as:

$$y_t = \alpha x_t + \lambda y_{t-1} + \frac{1 - \lambda L}{1 - \rho L} e_t$$

and comparing it with equations (1) and (2) which are: $y_t = \alpha y_{t-1} + \beta x_t + \frac{1}{1 - \rho L} e_t$, it might appear that the problems are the same--indeed that the model given by (7) is more complicated than the model given by equations (1) and (2). This is not the case. Thus, in any problem, to appraise the properties of two-step procedures, it is first desirable to check what the information matrix looks like.

As to the second question- whether the two-step procedures iterated produce the ML estimates, this can also be checked from the equations to be solved for obtaining the ML estimates of the parameters. Since this will automatically be done in the course of evaluating the information matrix, there is no additional burden involved. For the model given by equations (1) and (2) it is easy to check that Method 2, iterated beyond the second step produces ML estimates whereas Method 1 does not.

As explained above, the two-step procedures differ in the way in which the information about the serial correlation is used. It should be emphasized as brought out in the discussion above that the relative merits of two-step procedures and ML depend crucially on the model being estimated.

III. Asymptotic Efficiency of the Two-Step Estimators

First we will consider the case where the calculated residuals are used as regressors. This procedure is obviously inefficient even in the case where there are no lagged dependent variables.

Consider the model

$$y_t = \beta x_t + u_t , \quad u_t = \rho u_{t-1} + e_t \quad |\rho| < 1$$

where e_t are serially independent with variance σ_e^2. These equations imply $y_t = \beta x_t + \rho u_{t-1} + e_t$. To simplify the algebra we will assume that the x_t series is first order autoregressive i.e. $x_t = \lambda x_{t-1} + w_t$ where w_t are serially independent. Let σ_x^2 be the variance of x.

The variance of the ordinary least squares (OLS) estimate of β in this case is $\frac{\sigma_e^2}{\sigma_x^2}\left[\frac{1+\lambda\rho}{(1-\lambda\rho)(1-\rho^2)}\right]$. The variance of the ML estimate of β is $\frac{\sigma_e^2}{\sigma_x^2}\left[\frac{1}{1-2\rho\lambda+\rho^2}\right]$.

Consider next the following two-step procedure. From the OLS estimate $\hat{\beta}$ of β, obtain the calculated residuals $\hat{u}_t = y_t - \hat{\beta}x_t$. In the second step estimate β from a regression of y_t on x_t and $\hat{u}_{t-1}$. Writing the model as:

$$y_t = \beta x_t + \rho\hat{u}_{t-1} + \rho(u_{t-1} - \hat{u}_{t-1}) + e_t$$

we find that the variance of the two-step estimator of β is: (see Appendix for details)

$$\frac{\sigma_e^2}{\sigma_x^2}\left(\frac{1+\rho\lambda}{1-\rho\lambda}\right)\left[1+\frac{\rho^2\lambda^2}{1-\rho^2}\right]$$

It can be easily verified that the two-step estimator of β is more efficient than the OLS estimator of β but is less efficient than the ML estimator. Specifically,

$$\frac{V(\text{2 step estimator})}{V(\text{OLS estimator})} = 1 - \rho^2(1 - \lambda^2)$$

and $$\frac{V(\text{2 step estimator})}{V(\text{ML estimator})} = [1 - \rho^2(1 - \lambda^2)]\left[1 + \frac{2\rho^2(1 - \lambda^2)}{(1 - \rho\lambda)(1 - \rho^2)}\right]$$

For $\lambda = \rho = \cdot 7$ the last ratio is $2{\cdot}25$ and for $\lambda = \rho = \cdot 9$ it is $7{\cdot}2$. Thus for higher (positive) values of λ and ρ which is likely to be the case with observed economic time-series, this two-step procedure is considerably less efficient than an alternative two-step procedure.

In this example it is well known that if we used the residuals to get a consistent estimate $\hat{\rho}$ of ρ and used a quasi-first difference transformation of y and x based on $\hat{\rho}$, the resulting estimate of β is asymptotically as efficient as the ML estimate. Thus using the estimated residuals as regressors is an inefficient way of taking account of the serial correlation in the residuals. Similar conclusions hold if we have lagged dependent variables. Since the algebra gets very complicated, we will consider the case where the x_t are serially independent. If they are serially correlated, we would expect the two-step procedures to perform worse.

Consider the model given by equations (1) and (2)

$$y_t = \alpha y_{t-1} + \beta x_t + u_t \tag{1}$$

$$u_t = \rho u_{t-1} + e_t \;, \; |\rho| < 1 \tag{2}$$

where e_t are serially independent with a common variance σ_e^2. Let σ_x^2 be the variance of x_t and $\lambda = \sigma_e^2/\sigma_x^2$. If x_t and x_{t-1} are used as instrumental variables to obtain consistent estimates of α and β in (1), the covariance matrix of the instrumental variable (IV) estimator is

$$V_{IV} = \frac{\lambda}{1-\rho^2}\begin{bmatrix} \frac{1}{\beta^2} & \frac{\rho}{\beta} \\ \frac{\rho}{\beta} & 1 \end{bmatrix}$$

The covariance matrix of the ML estimator of (α, β, ρ) is

$$V_{ML} = \begin{bmatrix} \frac{1}{1-\alpha^2} + \frac{\beta^2}{\lambda}\,\frac{1-2\alpha\rho+\rho^2}{1-\alpha^2} & -\frac{\beta\rho}{\lambda} & \frac{1}{1-\alpha\rho} \\ -\frac{\beta\rho}{\lambda} & \frac{1+\rho^2}{\lambda} & 0 \\ \frac{1}{1-\alpha\rho} & 0 & \frac{1}{1-\rho^2} \end{bmatrix}^{-1}$$

Consider now the following two-step procedure: From the consistent estimates $(\hat{\alpha}, \hat{\beta})$ of (α, β) compute the residuals $\hat{u}_t = y_t - \hat{\alpha} y_{t-1} - \hat{\beta} x_t$. Next regress y_t on y_{t-1}, x_t, and $\hat{u}_{t-1}$. The covariance matrix of this two-step estimator of (α, β) is: (see Appendix for details)

$$V_{2\ \text{step}} = \lambda\sigma_x^2 A + \rho^2 ABV_{IN}B'A + \rho ABDC'A + \rho ACDB'A$$

where

$$A^{-1} = \sigma_x^2 \begin{bmatrix} \frac{\beta^2}{1-\alpha^2} + \frac{\lambda(1+\lambda\rho)}{(1-\alpha\rho)(1-\alpha^2)(1-\rho^2)} & 0 & \frac{\lambda}{(1-\alpha\rho)(1-\rho^2)} \\ 0 & 1 & 0 \\ \frac{\lambda}{(1-\alpha\rho)(1-\rho^2)} & 0 & \frac{\lambda}{1-\rho^2} \end{bmatrix}$$

$$B = \sigma_x^2 \begin{bmatrix} \frac{\alpha\beta^2}{1-\alpha^2} + \frac{\lambda(\rho+\alpha)}{(1-\alpha^2)(1-\rho^2)(1-\alpha\rho)} & \beta \\ 0 & 0 \\ \frac{\lambda\rho}{(1-\alpha\rho)(1-\rho^2)} & 0 \end{bmatrix}$$

$$C = \lambda\sigma_x^4 \begin{bmatrix} \beta & 0 \\ \rho & 1 \\ 0 & 0 \end{bmatrix} \qquad D = \frac{1}{\sigma_x^2} \begin{bmatrix} \frac{1}{\beta} & 0 \\ 0 & 1 \end{bmatrix}$$

It is not possible to show that the two-step estimator is necessarily more efficient than the IV estimator. The variance of the two-step estimator of β is λ, whereas the variance of the IV estimator of β is $\frac{\lambda}{\beta^2(1-\rho^2)}$. Thus the comparison depends on $\beta^2(1-\rho^2)$. As for the variance of α, since the expressions are very complicated, we have computed the magnitudes for selected values of the parameters to show that things can go either way. Table I gives the variance of the two-step estimates of α as compared with those of the IV estimator and the ML estimator.

TABLE I

Variance of $\hat{\alpha}$

($\beta = 1$, $\sigma_e^2 = 1$)

	α	ρ	V(2 step)	V_{IV}	V_{ML}
$\lambda = .1$	.9	.95	1.14	1.02	.166
	.95	.9	.475	.526	.143
	.05	.1	.102	.101	.101
	.1	.05	.100	.100	.100
	.1	.9	.626	.526	.078
	.9	.1	.021	.101	.021
	.48	.53	.166	.139	.127
	.53	.48	.144	.130	.122
	.5	.5	.154	.133	.125
$\lambda = 1.0$	.9	.95	11.4	10.2	.859
	.95	.9	4.73	5.26	.556
	.05	.1	1.03	1.01	1.00
	.1	.05	1.00	1.00	.997
	.1	.9	7.28	5.26	.505
	.9	.1	.125	1.01	.119
	.48	.53	1.77	1.39	1.27
	.53	.48	1.42	1.30	1.22
	.5	.5	1.58	1.33	1.25

A glance at the table indicates that the variance of the two-step estimator of α is higher than the variance of the IV estimator of α for the values of $\rho > \alpha$. Both the estimators are considerably less efficient than the ML estimator for high values of α and ρ. Even if α is large, if ρ is small, the 2-step estimator of α is almost as efficient as the ML estimator and there is considerable gain in efficiency as compared with the IV estimator. The results do not depend too heavily on the magnitude of the noise-signal ratio λ.

Suppose that instead of using $\hat{u}_{t-1}$ as a regressor, we obtain a consistent estimate of the covariance matrix of the disturbances and estimate the parameters in equation (1) by a generalized least squares (GLS) procedure based on the estimated covariance matrix. This is the two-step procedure discussed in Wallis [6]. The expression for the covariance matrix of the two step estimator of (α, β) is[3]

$$V_{2\ \text{step}} = V_A + V_A Q_1 V_A + V_A Q V_{IV} Q V_A$$

where
$$V_A = \begin{bmatrix} \frac{1}{1-\alpha^2} + \frac{\beta^2}{\lambda}\left(\frac{1-2\alpha\rho+\rho^2}{1-\alpha^2}\right) & -\frac{\beta\rho}{\lambda} \\ -\frac{\beta\rho}{\lambda} & \frac{(1+\rho^2)}{\lambda} \end{bmatrix}^{-1}$$

---

[3]There is an error in the expression given by Wallis [6]. The second term in his expression is $2V_A Q V_A$. The error consists in the definition of Q and in the omission of a term in expression (23), p. 567. When that is corrected we get the expression given here. Our expression also checks with the one given in Dhrymes [2].

$$Q_1 = \begin{bmatrix} \frac{1 + 3\rho^2}{(1 - \alpha\rho)^2} & 0 \\ 0 & 0 \end{bmatrix} \qquad Q = \begin{bmatrix} \frac{1 + \rho^2}{(1 - \alpha\rho)^2} & 0 \\ 0 & 0 \end{bmatrix}$$

V_A is the covariance matrix of the ML estimators of (α, β) when ρ is known, and $\lambda = \sigma_e^2/\sigma_x^2$.

Considering the fact that V_A and V_{IV} have an expression σ_e^2/σ_x^2 and that Q_1 and Q are matrices of constants, it is easy to see that if σ_e^2/σ_x^2 is high, then the two-step estimator could be considerably less efficient than even the IV estimator. On the other hand if σ_e^2/σ_x^2 is small, unless $\rho\alpha$ is large, there could be a considerable gain in efficiency as compared with the IV estimator. Table II presents a comparison of the variances of α for the two-step, IV, ML and the Aitken (ρ known) estimators. Table III shows the same information on the estimators for β .

The tables illustrate the above remarks: for ρ and α large the two step procedure is worse than the IV estimates, but when $\rho\alpha$ is not large there can be a substantial gain in efficiency using the 2 step procedures. Note, also that the variances are quite sensitive to variation in the noise-signal ratio λ . For this model the information matrix is not block diagonal, so the estimates with ρ known (V_A) have smaller variances than the ML estimates which in turn have smaller variances than the two-step estimators. For many values of the parameters the IV estimates are nearly fully efficient. Since the x_t were assumed to be uncorrelated using x_{t-1} as an instrument brings in an independent piece of information, if x_t is auto-correlated we would expect the IV estimates to be less efficient.

TABLE II

Variance of $\hat{\alpha}$
($\beta = 1$)

	α	ρ	V(2 step)	V_{IV}	V_A	V_{ML}
$\lambda = .1$	.9	.95	75.7	1.02	.094	.166
	.95	.9	15.9	.526	.062	.143
	.05	.1	.101	.101	.092	.101
	.1	.05	.100	.100	.091	.100
	.1	.9	.116	.526	.077	.078
	.9	.1	.021	.101	.020	.021
	.48	.53	.157	.139	.109	.127
	.53	.48	.145	.130	.105	.122
	.5	.5	.150	.133	.107	.125
$\lambda = 1.0$	.9	.95	2500.	10.2	.172	.859
	.95	.9	33.4	5.26	.092	.556
	.05	.1	1.03	1.01	.501	1.00
	.1	.05	1.00	1.00	.497	.997
	.1	.9	6.45	5.26	.453	.505
	.9	.1	.134	1.01	.104	.119
	.48	.53	2.94	1.39	.480	1.27
	.53	.48	2.38	1.30	.453	1.22
	.5	.5	2.60	1.33	.469	1.25

TABLE III

Variance of $\hat{\beta}$
($\beta = 1$)

	α	ρ	V(2 step)	V_{IV}	V_A	V_{ML}
$\lambda = .1$	.9	.95	18.9	1.02	.076	.094
	.95	.9	3.98	.526	.071	.091
	.05	.1	.100	.101	.100	.100
	.1	.05	.100	.100	.100	.100
	.1	.9	.084	.526	.074	.075
	.9	.1	.099	.101	.099	.099
	.48	.53	.105	.139	.097	.100
	.53	.48	.103	.130	.097	.100
	.5	.5	.104	.133	.097	.100
$\lambda = 1.0$	.9	.95	624.0	10.2	.569	.740
	.95	.9	83.1	5.26	.575	.690
	.05	.1	1.00	1.01	.995	1.00
	.1	.05	1.00	1.00	.999	1.00
	.1	.9	2.15	5.26	.664	.677
	.9	.1	.991	1.01	.991	.991
	.48	.53	1.28	1.39	.863	.997
	.53	.48	1.18	1.30	.882	.998
	.5	.5	1.22	1.33	.875	1.00

Finally, to complete our illustrations, we have investigated the following model considered by Dhrymes [2]:

$$y_t = \frac{\alpha}{1 - \lambda L} x_t + e_t$$

where e_t are serially independent.[4]

To simplify the algebra we will assume that the x_t series is serially uncorrelated. What we are interested in showing is that the 2-step estimator can be inefficient even when compared with the IV estimator.

If x_t and x_{t-1} are used as instrumental variables, the covariance matrix of the IV estimator of (α, λ) is

$$V_{IV} = \frac{\sigma_e^2}{\sigma_x^2} \begin{bmatrix} 1 + \lambda^2 & -\frac{\lambda}{\alpha} \\ -\frac{\lambda}{\alpha} & \frac{1 + \lambda^2}{\alpha^2} \end{bmatrix}$$

The covariance matrix of the ML estimator of (α, λ) is given by

$$V_{ML}^{-1} = \frac{\sigma_x^2}{\sigma_e^2(1 - \lambda^2)} \begin{bmatrix} 1 & \frac{\alpha\lambda}{1 - \lambda^2} \\ \frac{\alpha\lambda}{1 - \lambda^2} & \frac{\alpha^2(1 + \lambda^2)}{(1 - \lambda^2)^2} \end{bmatrix}$$

[4] Dhrymes considers the case where the errors are first order autocorrelated but this is an unnecessary complication and hence we have considered the case of serially uncorrelated errors. As mentioned earlier a look at the information matrix reveals that serial correlation is not a problem.

Consider now the two-step procedure described as follows: From the preliminary consistent estimates $(\hat{\alpha}, \hat{\lambda})$ of (α, λ), define

$$\hat{y}_{t-1} = \hat{\alpha} \sum_{i=0}^{t-1} \hat{\lambda}^i x_{t-i}$$

Regress y_t on x_t and $\hat{y}_{t-1}$ to obtain the second-step estimates of α and λ. The covariance matrix of this two-step estimator is derived in Chapter 5, Dhrymes [2]. It is given by

$$V_{2\ \text{step}} = \sigma_e^2[A - AQBG'A' - AGB'Q'A'] + AQV_{IV}QA'$$

where
$$A = \frac{1}{\sigma_x^2}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1-\lambda^2}{\alpha^2} \end{bmatrix}$$

$$Q = \frac{\alpha\lambda\sigma_x^2}{1-\lambda^2}\begin{bmatrix} 0 & 0 \\ 1 & \frac{\alpha\lambda}{1-\lambda^2} \end{bmatrix} \qquad BG' = \begin{bmatrix} 1 & 0 \\ -\frac{\lambda}{\alpha} & 1 \end{bmatrix}$$

After carrying out the necessary multiplications, we find that

$$V_{2\ \text{step}} - V_{IV} = \frac{\sigma_e^2}{\sigma_x^2}\begin{bmatrix} -\lambda^2 & \frac{\lambda^3}{\alpha(1-\lambda^2)} \\ \frac{\lambda^3}{\alpha(1-\lambda^2)} & \frac{\lambda^2 - 2\lambda^4 + 2\lambda^6 + \lambda^8}{\alpha^2(1-\lambda^2)^2} \end{bmatrix}$$

The first diagonal element is always negative and the second diagonal element is always positive (for $\lambda^2 < 1$) . Thus the 2-step estimator is not necessarily more efficient than the IV estimator (i.e. the first step of the two-step procedure).

IV. Conclusions

Many two-step procedures have during recent years crept into econometric literature. It is by now well known that though these two-step estimators are consistent, they are less efficient than the ML estimators. However, there are two important questions that remain unanswered. How much loss in efficiency is there in the use of the two-step procedure? Are the two-step estimators at least more efficient than the IV estimators? We have tried to answer these two questions in this paper.

As to the first question, we have given the relevant variances for selected values of the parameters for two commonly used two-step procedures in the case of the model given by equations (1) and (2). The indications are that for high values of ρ, there could be a considerable loss of efficiency. Since most economic time-series exhibit high positive serial correlation (ρ = .7 or .8) we do not advocate the use of 2-step procedures. Further, we have also shown that the variance of the 2-step estimator can in some cases be higher than the variance of the IV estimator and this is so for plausible values of the parameters. Thus, not only are the 2-step estimators less efficient than the ML estimators, but also less efficient than the IV estimators in some cases.

For models similar to the one given in equation (4) i.e. where the information matrix is block diagonal appropriate two step procedures may be as efficient as ML. However, if asymptotic efficiency is at all a guide, we would not recommend the use of the two-step procedures for models of the type given in equations (1) and (2). In these cases it is better to use the ML method. Since computational ease is not a powerful argument in favor of the 2-step procedures in the present age of high-speed computers, we do not see any case for the use of these procedures in econometric work.

APPENDIX

1) Variance of the two-step estimator of β in the model

$$y_t = \beta x_t + u_t$$

$$u_t = \rho u_{t-1} + e_t \qquad |\rho| < 1$$

Let $\hat{\beta}$ be the OLS estimator of β

$$\hat{u}_t = y_t - \hat{\beta} x_t$$

In the second step we regress y_t on x_t and $\hat{u}_{t-1}$. Since $y_t = \beta x_t + \rho\hat{u}_{t-1} + \rho(u_{t-1} - \hat{u}_{t-1}) + e_t$, if β^*, ρ^* are the 2-step estimators, we have

$$\sqrt{T}\begin{bmatrix} \beta^* - \beta \\ \rho^* - \rho \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\Sigma x_t^2 & \frac{1}{T}\Sigma x_t\hat{u}_{t-1} \\ \frac{1}{T}\Sigma x_t\hat{u}_{t-1} & \frac{1}{T}\Sigma\hat{u}_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{\Sigma x_t e_t}{\sqrt{T}} + \frac{\rho}{\sqrt{T}}\Sigma x_t(u_{t-1} - \hat{u}_{t-1}) \\ \frac{\Sigma\hat{u}_{t-1}e_t}{\sqrt{T}} + \frac{\rho}{\sqrt{T}}\Sigma\hat{u}_{t-1}(u_{t-1} - \hat{u}_{t-1}) \end{bmatrix}$$

The plim of the expression in the first bracket is $\begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}^{-1}$.

Hence $\sqrt{T}(\beta^* - \beta) \sim \frac{1}{\sigma_x^2}\frac{\Sigma x_t e_t}{\sqrt{T}} + \frac{\rho}{\sigma_x^2}\frac{\Sigma x_t(u_{t-1} - \hat{u}_{t-1})}{\sqrt{T}}$.

But $u_{t-1} - \hat{u}_{t-1} = (\hat{\beta} - \beta)x_{t-1}$, and if x_t are first-order autoregressive with autocorrelation coefficient λ, we have

$$\sqrt{T}(\beta^* - \beta) \sim \frac{1}{\sigma_x^2}\frac{\Sigma x_t e_t}{\sqrt{T}} + \sqrt{T}(\hat{\beta} - \beta)\rho\lambda$$

Hence $$V[\sqrt{T}(\beta^* - \beta)] = \frac{\sigma_e^2}{\sigma_x^2} + V(\hat{\rho}) \cdot \rho^2\lambda^2 + \frac{2\sigma_e^2}{\sigma_x^2} \cdot \frac{\rho\lambda}{1 - \rho\lambda}$$

$$= \frac{\sigma_e^2}{\sigma_x^2} \left[\frac{1 + \rho\lambda}{1 - \rho\lambda} \right] \left[1 + \frac{\rho^2\lambda^2}{1 - \rho^2} \right]$$

2) Variance of the two-step estimator of (α, β) in the model

$$y_t = \alpha y_{t-1} + \beta x_t + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

Let $\hat{\alpha}$, $\hat{\beta}$ be the IV estimates and

$$\hat{u}_t = y_t - \hat{\alpha} y_{t-1} - \hat{\beta} x_t$$

Since $y_t = \alpha y_{t-1} + \beta x_t + \rho \hat{u}_{t-1} + \rho(u_{t-1} - \hat{u}_{t-1}) + e_t$ defining

$$Z = (y_{-1}, \ x, \ \hat{u}_{-1})$$

$$Z_1 = (y_{-1}, \ x)$$

$$X = (x_{-1}, \ x)$$

$$\delta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

and δ^*, ρ^* as the two step estimators, we have

$$T \begin{pmatrix} \delta^* - \delta \\ \rho^* - \rho \end{pmatrix} \begin{pmatrix} \delta^* - \delta \\ \rho^* - \rho \end{pmatrix}' = T(Z'Z)^{-1} Z' \{ ee' + \rho^2 (u_{-1} - \hat{u}_{-1})' + \rho e (u_{-1} - \hat{u}_{-1})'$$
$$+ \rho (u_{-1} - \hat{u}_{-1}) e' \} Z (Z'Z)^{-1}$$

The first term is $\sigma_e^2 \left(\frac{Z'Z}{T}\right)^{-1}$. Since $u_{-1} - \hat{u}_{-1} = (Z_1)_{-1}(\hat{\delta} - \delta)$ the second term is

$$\rho^2\left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'Z_{1,-1}}{T}\right)V_{IV}\left(\frac{Z'_{1,-1}Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}$$

Noting that $(\hat{\delta} - \delta) = (X'Z_1)^{-1}x'u$ the third term can be written as:

$$\left(\frac{Z'Z}{T}\right)^{-1} \frac{Z'eu'x}{T} \left(\frac{x'Z_1}{T}\right)^{-1}\left(\frac{Z'_{1,-1}Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}$$

Evaluating these expressions separately, we get the result mentioned in the paper.

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

[2] Dhrymes, P.J. Distributed Lags: Problems of Formulation and Estimation (San Francisco: Holden-Day, 1970).

[3] Gupta, Y.P. "Least Squares Variant of the Dhrymes Two-Step Estimation Procedure of the Distributed Lag Model," International Economic Review, 1969, pp. 112-113.

[4] Maddala, G.S. "Generalized Least Squares with an Estimated Covariance Matrix," Econometrica (forthcoming).

[5] Taylor, L.D. and T.A. Wilson. "Three Pass Least Squares: A Method for Estimating Models with a Lagged Dependent Variable," Review of Economics and Statistics, 1964, pp. 329-346.

[6] Wallis, K.F. "Lagged Dependent Variables and Serially Corrleated Residuals--A Reappraisal of Three-Pass Least Squares," Review of Economics and Statistics, 1967, pp. 555-567.