A MONTE CARLO STUDY OF MEASUREMENT ERROR BIAS*

J.E.H. Davidson

NUMBER 80

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

A MONTE CARLO STUDY OF MEASUREMENT ERROR BIAS*

J.E.H. Davidson

NUMBER 80

November, 1975

* I would like to thank Dr. David Hendry for suggesting this work and for many discussions, and also Prof. J. D. Sargan for comments. I retain responsibility for the remaining errors. I should also like to thank the S.S.R.C. Project on Econometric Methodology for financial support.

Substituting in the least squares formula,

$$\hat{\beta} = (X'X)^{-1}X'y$$

yields, assuming zero expectations of the errors,

$$\text{plim}(\hat{\beta} - \beta) = (\bar{M} + \Omega)^{-1}(\omega^* - \Omega\beta) \qquad (2)$$

where $\bar{M} = \lim_{T \to \infty}\left(\frac{\bar{X}'\bar{X}}{T}\right)$

$$\Omega = \text{plim}\left(\frac{V'V}{T}\right)$$

$$\omega^* = \text{plim}\left(\frac{V'(\varepsilon + v)}{T}\right).$$

(ii) Direct simulation experiments were performed using the programme package NAIVE [5]. The programme generates time series of "exogenous variables" using normally distributed random numbers and a simple first-order autoregressive scheme from a starting value of zero. The endogenous variable is then obtained using input values for the structural parameters, and an additive normal random error. Rectangularly distributed "measurement errors" can be added to the variables. In effect, the procedure is then to feed the generated sample into a standard regression programme - as for GIVE [4] - , record the deviation of the parameter estimates from the known true values, and replicate this procedure a large number of times. The programme prints the mean of the recorded deviations, (in addition to their sampling standard error, the means of the usual summary statistics and other information), and this mean is taken as an estimate of the bias in the estimator. 100 replications were used in practice, which should be sufficient from the point of view of estimate precision.

The first set of experiments performed were designed to explore the response of the mean bias to various points in the parameter spaces of two simple models, and compare it at each point with the value for the inconsistency obtained by substituting the parameter values into the appropriate specialization of (2).

Computational cost strictly limits the number of experiments which can be conducted using direct simulation. Experimental design is therefore of great importance, since we seek the widest possible investigation of the parameter space without excessive confounding of the effects in a small number of experiments. Unfortunately, the number of parameters to be varied is large even for small models. Of the two models investigated one, the simplest possible two-variable model involves six parameters (including sample size); the second, a three variable dynamic model, involves eight. Even allowing only two values for each parameter, a full factorial experiment with six parameters involves sixty-four experiments. In the event, fully confounded fractional-factorial designs were employed, with randomization of the parameter combinations within the constraints set by the confounding requirement. (For discussion see Cochran and Cox [2], Ch. 6.)

---

* There is clearly scope here for the use of variance reduction techniques such as control variates; [6], [9]. Adaption of the programme for this purpose would be relatively simple, and should be considered for a more extensive investigation.

Following Hendry and Harrison, the chosen method of summarizing the information obtained in the simulations is to estimate the response surface of the mean bias, assuming approximate linearity. Where $\hat{\alpha}_T$ is the least squares estimator from a sample of size T of some slope coefficient α, we adopt the notation

$$I(\hat{\alpha}) = \underset{T\to\infty}{\text{plim}} \; (\hat{\alpha}_T - \alpha),$$

and also write $\underline{\theta}$ for the K-vector of all the parameters of the model, $\underline{\theta} = (\ldots., \alpha, \ldots.)$. Then by hypothesis,

$$E(\hat{\alpha}_T - \alpha) = \phi\left(I(\hat{\alpha}), \frac{I(\hat{\alpha})}{T}, \frac{\theta}{T}\right) \qquad (3)$$

is approximately linear in the arguments, provided that at least part of the non-linearity in the asymptotically vanishing terms can be accounted for by the term $I(\hat{\alpha})/T$. With an appropriate heteroscedasticity transformation, write

$$\sqrt{T}\, E(\hat{\alpha}_T - \alpha) = a_o\sqrt{T} + a_1\sqrt{T}\, I(\hat{\alpha}) + a_2 \frac{I(\hat{\alpha})}{\sqrt{T}} + \sum_{i=1}^{K} a_{2+i} \frac{\theta i}{\sqrt{T}} + \sqrt{T}\, U. \qquad (4)$$

where we expect that $a_1 = 1$.

Substituting for $E(\hat{\alpha}_T - \alpha)$ the mean bias $\bar{B}_j(\hat{\alpha}_T)$ obtained in simulation experiments with sample parameters $\underline{\theta}_j$, and sample size T_j, $j = 1, \ldots, m$, (4) can be estimated by least squares. Part of the stochastic term will then represent simulation sampling errors in the mean bias, but provided that sufficient replications have been performed to ensure reasonable approximation to the true bias, then the goodness of fit of the equation will reflect mainly the adequacy of the linear approximation. The t statistics can be used to test the significance of finite-sample effects associated with any given parameter. If such effects prove to be small, we have evidence that the inconsistency will be an adequate guide to finite sample behaviour.

(iii) The two models employed were as follows; in each case, Greek letters represent the true variables, and Roman letters the corresponding empirical variables.

(A)
$$\begin{aligned} \xi_t &= \alpha\,\chi_t + \varepsilon_t \\ y_t &= \xi_t \\ x_t &= \chi_t + v_t \end{aligned} \qquad (5)$$

(B)
$$\begin{aligned} \xi_t &= \beta_1\,\chi_t + \beta_2\,\xi_{t-1} + \varepsilon_t \\ y_t &= \xi_t + u_t \\ x_t &= \chi_t + v_t \qquad |\beta_2| < 1 \end{aligned} \qquad (6)$$

In each case, the independent variable χ_t was considered to be generated by

$$\chi_t = \gamma\,\chi_{t-1} + \eta_t,\ |\gamma| < 1 \qquad (7)$$

The error terms, $\varepsilon_t \sim N(0,\ \sigma_\varepsilon^2)$, $\eta_t \sim N(0,\sigma_\eta^2)$, $u_t \sim R(0,\sigma_u^2)$, $v_t \sim R(0,\ \sigma_v^2)$ are all serially and mutually independent.

We first derive the inconsistencies for each model, the appropriate specializations of (2).

For A
$$\text{plim}\ (\hat{\alpha} - \alpha) = \frac{-\alpha\ \sigma_v^2}{m_{\chi\chi} + \sigma_v^2} \qquad (8)$$

where $m_{\chi\chi}$ is the population second moment of χ_t,

$$m_{\chi\chi} = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T} \chi_t^2\right).$$

For B

$$\text{plim}\ (\hat{\beta}_1 - \beta_1) = \frac{1}{\Delta}\left(-\ \beta_1\sigma_v^2(m_{\xi\xi} + \sigma_u^2) + \beta_2\sigma_u^2 m_{\chi\xi_{-1}}\right) \tag{9}$$

$$\text{plim}\ (\hat{\beta}_2 - \beta_2) = \frac{1}{\Delta}\left(-\ \beta_2\sigma_u^2(m_{\chi\chi} + \sigma_v^2) + \beta_1\sigma_v^2 m_{\chi\xi_{-1}}\right) \tag{10}$$

where $\Delta = (m_{\chi\chi} + \sigma_v^2)(m_{\xi\xi} + \sigma_u^2) - m^2_{\chi\xi_{-1}}$

and $m_{\xi\xi} = \text{plim}\ \frac{1}{T}\sum_{t=1}^{T}\xi_t^2$

$$m_{\chi\xi_{-1}} = \text{plim}\ \frac{1}{T}\sum_{t=1}^{T}\chi_t\ \xi_{t-1}$$

These results are derived in the Appendix, where it is also shown that

$$\text{(i)}\quad m_{\chi\chi} = \frac{\sigma_\eta^2}{1-\gamma^2}\ ,$$

$$\text{(ii)}\quad m_{\chi\xi_{-1}} = \frac{\beta_1\gamma}{1-\beta_2\gamma}\, m_{\chi\chi} \tag{11}$$

$$\text{(iii)}\quad m_{\xi\xi} = \frac{1}{1-\beta_2^2}\left(\frac{1+\beta_2\gamma}{1-\beta_2\gamma}\,\beta_1^2\, m_{\chi\chi} + \sigma_\varepsilon^2\right)$$

The formulae were used to write a routine to compute the inconsistencies for each sample point in the parameter space.

The approach to the design problem was slightly different for each model. Consider A first. σ_η^2 was set a a constant level of unity for all experiments, leaving five parameters to be varied, α, γ, σ_ε^2, σ_v^2 and T. To sample the parameter space fairly extensively, three values each of α, γ, σ_v^2 and T were chosen, and two values of σ_ε^2 :

$\alpha = .8, \ .2, \ -.5$

$\gamma = .5, \ .7, \ .9$

$\sigma_v^2 = 0, 1, 2$

$T = 20, 50, 80$

$\sigma_\varepsilon^2 = .1, \ 1$

A full-factorial design would give $3^4 \times 2 = 162$ parameter points. Actually employed was a $\frac{1}{9}$ factorial design, giving 18 experiments. This means in practice that out of a defined population of 162 parameter points, 18 were sampled by a restricted randomization, whereby each parameter value appears in equal proportion, and is combined with each of the values of the other parameters equi-proportionally. This ensures the most extensive possible plotting of the response surface for the given number of experiments.

The simulations were then carried out using the chosen parameter values, the first 30 observations being discarded (to minimize the influence of initial conditions), and the slope coefficients estimated using the next T observations, by Ordinary Least Squares only for these experiments. Each simulation was replicated 100 times.

The resulting mean biasses are shown in Table 1. For prediction purposes, four additional experiments were conducted with parameters chosen arbitrarily, but such that for two, the parameters lie wholly within the range of the main set of experiments, and for two, some parameters lie outside.

For model B, σ_η^2 is also set to unity for all experiments, and the parameters to be varied are β_1, β_2, γ , σ_v^2, σ_u^2, σ_ε^2 and T. This is a rather large number to be varied over a small number of experiments, and with the aim of making the design more manageable, it was decided to reduce the number of independent parameters by one, linking β_1 and β_2 by the relation,

$$\frac{\beta_1}{1 - \beta_2} = 1 \qquad \text{(i.e. } \beta_1 + \beta_2 = 1\text{)}.$$

This corresponds, when the difference equation of (6) is given the familiar interpretation of a Koyck transformation of a geometric lag scheme, to specifying a population of distributed lag equations with a constant long-run coefficient of unity, but varying mean lag. This restriction does, of course, reduce the possible information to be gained from the simulations, but with corresponding gains in the intensity of sampling of the overall parameter space for a given number of experiments. (An ad hoc justification for the restriction might be that, as in the case of the consumption function, the long run coefficient of the model of the form (6) is often thought to be approximately known a priori, and the mean lag is the unknown quantity of principal interest).

The number of independent parameters is thereby reduced to six, and it was decided, in contrast to A, to set each parameter at two values, and employ a 1/4-factorial design giving 16 experiments. The chose values were

$$(\beta_1, \beta_2) = (.4, .6), (.8, .2);$$
$$\gamma = .9, .99;$$
$$\sigma_v^2 = 0, 10,$$
$$\sigma_u^2 = 0, 10;$$
$$\sigma_\varepsilon^2 = 1, 10;$$
$$T = 30, 65.$$

The simulations were conducted in basically the same manner as for model A, but with an important variation. For B, the series for χ_t was generated only once for each experiment, the same set of numbers being used for each replication, whereas for A, the exogenous series was generated afresh for each replication.

The significance of this variation will be discussed in due course.

The simulation results and computed inconsistencies for B are shown in Table 2.

The bias response function for model A was estimated as,

$$\sqrt{T}_j\ \bar{B}_j(\hat{\alpha}) = a_o \sqrt{T}_j + a_1 \sqrt{T}_j\ I_j(\hat{\alpha}) + a_2 \frac{I_j(\hat{\alpha})}{\sqrt{T}_j} + a_3 \frac{\alpha_j}{\sqrt{T}_j} + a_4 \frac{\gamma_j}{\sqrt{T}_j} + a_5 \frac{(\sigma_v^2)_j}{\sqrt{T}_j} + a_6 \frac{(\sigma_\varepsilon^2)_j}{\sqrt{T}_j} + a_7 \frac{1}{\sqrt{T}_j}, \qquad j = 1, \ldots, m \tag{12}$$

where m is the number of experiments.

For model B,

$$\sqrt{T}_j\ \bar{B}_j(\hat{\beta}_i) = b_{i1} \sqrt{T}_j\ I_j(\hat{\beta}_i) + b_{i2} \frac{I_j(\hat{\beta}_i)}{\sqrt{T}_j} + b_{i3} \frac{\beta_{1j}}{\sqrt{T}_j} + b_{i4} \frac{\gamma_j}{\sqrt{T}_j} + b_{i5} \frac{(\sigma_v^2)_j}{\sqrt{T}_j} + b_{i6} \frac{(\sigma_u^2)_j}{\sqrt{T}_j} + b_{i7} \frac{(\sigma_v^2)_j (\sigma_u^2)_j}{\sqrt{T}_j} + b_{i8} \frac{(\sigma_\varepsilon^2)_j}{\sqrt{T}_j} + b_{i9} \frac{1}{\sqrt{T}_j} \qquad j = 1, \ldots, m \quad i = 1, 2 \tag{13}$$

Note that it was decided to include one second-order interaction term, $\sigma_v^2 \cdot \sigma_u^2$, and also that because of the dependence between β_1 and β_2, only β_1 is included as a regressor. Also, because T is set at only two levels, one of the variables $\sqrt{T}_j$ and $1/\sqrt{T}_j$ is redundant. (The former corresponds to the intercept in the regression, the latter to the dummy parameter, 1, in the model). The intercept term is therefore dropped.

TABLE I

Inconsistencies and Simulation Mean Biasses for Model A.

	γ	α	σ_v^2	σ_ε^2	T	plim $(\hat{\alpha}-\alpha)$	$(\widehat{\alpha-\alpha})$ *
1	·7	-·5	1	·1	20	·169	·195 (·009)
2	·7	-·5	2	1	80	·252	·263 (·008)
3	·7	·2	0	·1	80	0	·002 (·003)
4	·7	·2	2	1	50	-·101	-·107 (·008)
5	·7	·8	0	1	20	0	·018 (·020)
6	·7	·8	1	·1	50	-·270	-·303 (·007)
7	·9	-·5	0	·1	50	0	-·007 (·003)
8	·9	-·5	2	1	20	·138	·230 (·016)
9	·9	·2	1	·1	80	-·032	-·043 (·002)
10	·9	·2	0	1	20	0	-·026 (·018)
11	·9	·8	1	1	50	-·128	-·201 (·009)
12	·9	·8	2	·1	80	-·220	-·284 (·008)
13	·5	-·5	0	·1	80	0	-·003 (·003)
14	·5	-·5	1	1	50	·214	·217 (·010)
15	·5	·2	2	·1	50	-·120	-·118 (·003)
16	·5	·2	1	1	20	-·086	-·092 (·015)
17	·5	·8	0	1	80	0	-·025 (·010)
18	·5	·8	2	·1	20	-·480	-·515 (·011)
19	·7	·5	1	·1	30	-·217	-·245 (·007)
20	·95	-·2	·5	·1	70	·009	·021 (·002)
21	·4	-·8	4	·1	100	·617	·617 (·004)
22	·9	1·5	3	·1	40	-·545	-·737 (·189)

* Simulation standard errors in brackets.

TABLE II

Inconsistencies and Simulation Mean Biasses for Model B.

	γ	β_1	β_2	σ_v^2	σ_u^2	σ_ε^2	T	$\text{plim}(\hat{\beta}_1-\beta_1)$	$\overline{(\hat{\beta}_1-\beta_1)}$ *	$\text{plim}(\hat{\beta}_2-\beta_2)$	$\overline{(\hat{\beta}_2-\beta_2)}$ *
1	.9	.8	.2	0	0	10	30	.0	.020 (.024)	.0	−.022 (.013)
2	.9	.4	.6	10	0	10	30	−.278	−.293 (.010)	.057	.028 (.010)
3	.9	.4	.6	0	10	1	30	.369	.412 (.021)	−.472	−.517 (.013)
4	.9	.8	.2	10	10	1	30	.534	−.614 (.012)	.028	.015 (.013)
5	.9	.8	.2	0	0	1	65	.0	.019 (.007)	.0	−.016 (.006)
6	.9	.4	.6	10	0	1	65	−.323	−.345 (.002)	.223	.241 (.003)
7	.9	.4	.6	0	10	10	65	.176	.216 (.020)	−.225	−.248 (.009)
8	.9	.8	.2	10	10	10	65	−.531	−.571 (.009)	.017	.032 (.008)
9	.99	.4	.6	0	0	1	30	.0	.044 (.009)	.0	−.057 (.009)
10	.99	.8	.2	10	0	1	30	−.665	−.701 (.005)	.646	.662 (.005)
11	.99	.8	.2	0	10	10	30	.093	.139 (.020)	−.094	−.153 (.013)
12	.99	.4	.6	10	10	10	30	−.003	−.198 (.013)	−.079	−.200 (.012)
13	.99	.4	.6	0	0	10	65	.0	.045 (.008)	.0	−.047 (.007)
14	.99	.8	.2	10	0	10	65	−.412	−.510 (.006)	.338	.321 (.006)
15	.99	.8	.2	0	10	1	65	.164	.177 (.010)	−.166	−.179 (.008)
16	.99	.4	.6	10	10	1	65	.040	−.120 (.007)	−.131	−.204 (.008)

* Simulation standard errors in brackets.

Table 3

Bias Regression for Model A (18 observations)

$$\sqrt{T}\,\bar{B}(\hat{\alpha}) = \underset{(2.34)}{-.028\sqrt{T}} + \underset{(13.15)}{1.100\sqrt{T}\,I(\hat{\alpha})} + \underset{(.052)}{.178\,I(\hat{\alpha})/\sqrt{T}} - \underset{(.658)}{.650\,\alpha/\sqrt{T}}$$

$$- \underset{(.505)}{.915\,\gamma/\sqrt{T}} + \underset{(.517)}{.189\,\sigma_v^2/\sqrt{T}} - \underset{(.332)}{.262\,\sigma_\varepsilon^2/\sqrt{T}} + \underset{(1.07)}{1.58/\sqrt{T}}$$

$R^2 = .990$ $s = .171$ $d = 1.37$

t values in brackets.

$\chi^2(4) = 31.84$ (forecast stability test)

Predictions (see Table 1, 19-22)

Equation	(19)	(20)	(21)	(22)
$\bar{B}(\hat{\alpha})-\hat{B}(\hat{\alpha})$	-.007	.026	-.059	-.116
$\bar{B}(\hat{\alpha})-I(\hat{\alpha})$	.028	.020	.000	-.192

$\hat{B}(\hat{\alpha}) = \frac{1}{\sqrt{T}}$ times the prediction of the bias regression

---

Table 4

Bias Regressions for Model B (32 observations)

$$\sqrt{T}\,\bar{B}(\hat{\beta}_1) = \underset{(10.55)}{1.07\sqrt{T}\,I(\hat{\beta}_1)} - \underset{(2.69)}{14.63\,I(\hat{\beta}_1)/\sqrt{T}} - \underset{(1.89)}{6.85\,\beta_1/\sqrt{T}} - \underset{(1.60)}{16.65\,\gamma/\sqrt{T}}$$

$$- \underset{(3.58)}{.821\,\sigma_v^2/\sqrt{T}} + \underset{(1.66)}{.263\,\sigma_u^2/\sqrt{T}} - \underset{(1.98)}{.036(\sigma_u^2\sigma_v^2)/\sqrt{T}} + \underset{(.782)}{.085\,\sigma_\varepsilon^2/\sqrt{T}} + \underset{(2.07)}{20.70/\sqrt{T}}$$

$R^2 = .975$ $s = .406$ $d = 2.15$

$$\sqrt{T}\,\bar{B}(\hat{\beta}_2) = \underset{(13.55)}{1.07\sqrt{T}\,I(\hat{\beta}_2)} + \underset{(1.05)}{3.23\,I(\hat{\beta}_2)/\sqrt{T}} + \underset{(.244)}{.477\,\beta_1/\sqrt{T}} - \underset{(3.47)}{23.98\,\gamma/\sqrt{T}} - \underset{(.533)}{.066\,\sigma_v^2/\sqrt{T}}$$

$$+ \underset{(1.51)}{.156\sigma_u^2/\sqrt{T}} - \underset{(1.25)}{.012(\sigma_u^2\sigma_v^2)/\sqrt{T}} - \underset{(1.06)}{.053\,\sigma_\varepsilon^2/\sqrt{T}} + \underset{(2.91)}{21.19/\sqrt{T}}$$

$R^2 = .990$ $s = .190$ $d = 1.49$

The regression estimates are presented in Tables 3 and 4, and Table 3 shows in addition the prediction errors for model A, together with the corresponding difference between mean bias and inconsistency.

A methodological problem emerged in the course of the experiments which must be borne in mind when interpreting the results. The first experiments were in fact performed on model B, and atypical discrepancies between mean bias and inconsistency were noted in one or two of the experiments. Estimation of the response surface produced values of b_{11} and b_{21} significantly different from unity, strongly suggesting a misspecification. After careful checks, it was decided that the cause lay with the fact that the NAIVE programme holds the exogenous variables fixed in repeated samples. When the autoregressive coefficient, γ in (7), is close to 1, the estimated second moment of the series for samples of moderate size is likely to be quite a long way from the corresponding probability limit - a short realization of the series being, in fact, close in appearance to a random walk.

Therefore, the variance of the estimated bias was not, in fact, being reduced by replication (since the same random effect was appearing repeatedly); comparison with the inconsistency computed from the asymptotic moment of the series must have a strictly qualified interpretation.

To test this conjecture, the simulations were run a second time using a new set of random numbers, and some of the mean biasses obtained were found to differ significantly from the previous estimates, (using two standard errors as a criterion). The response surface was re-estimated from the two sets of results combined, and it is these regression estimates which are shown in table 4. The coefficients of $\sqrt{T}\, I(\hat{\beta}_1)$ and $\sqrt{T}\, I(\hat{\beta}_2)$ are now both found to lie within one standard error of unity. Since there seemed to be no point in reproducing both sets of results, the mean biasses shown in table 2 are

are simple averages of the two figures, and the quoted standard errors, $\sqrt{2}$ times the average standard error.

In the light of this experience, the simulations for model A were run with the exogenous variable generated separately for each replication. This does, of course, considerably increase the computational cost. A lower range of values for the autoregression coefficient were chosen also, since the divergence of the small-sample variance estimates from the asymptotic limits is thereby reduced.

Bearing these problems in mind, we pass now to the general interpretation of the results. We see that for model A, none of the finite-sample effects appear to be significant at the 5% level, although the intercept term is significant, and the fit of the equation is good. It will be observed that in two cases - which happen to correspond to the smaller sample sizes - the predictions from the regression are better than the inconsistencies as approximations to the mean bias, while for the other two cases, they are worse. Note that while the estimate of the coefficient of $\sqrt{T}\ I(\hat{\alpha})$ is within two standard errors of unity, it is nevertheless rather large. Disappointingly, the evidence suggests that there is some finite sample effect at work which is not caught by the chosen specification of the response function.

In model B, however, a number of 5%-significant finite-sample effects appear. The high significance of $\gamma/\sqrt{T}$ alone is a paradoxical feature of the regression for β_2, and the most reasonable explanation would seem to be that this is connected with the data generation problem. As in model A, both equations have the coefficients of the inconsistency term on the high side of one.

A fundamental difficulty with this technique of response surface estimation is related to the problem with the generation of the exogenous variables already described. One of the sources of deviation between the inconsistency and the simulation bias must be the bias in the estimate of the asymptotic variance of x_t from a finite sample.

If the added measurement errors are ignored for simplicity, then expanding the autoregression gives

$$x_t = \gamma\, x_o + \sum_{j=0}^{t-1} \gamma^j \eta_{t-j}$$

NAIVE sets the starting value x_o equal to zero. In this case, it is easily shown that

$$E\left[\frac{1}{T}\sum_{t=1}^{T} x_t^2\right] = \sigma_\eta^2 \left[1 + \frac{T-1}{T}\gamma^2 + \frac{T-2}{T}\gamma^4 + \ldots + \frac{1}{T}\gamma^{2(T-1)}\right]$$

$$= \frac{\sigma_\eta^2}{1-\gamma^2}\left[1 - \frac{\gamma^2(1-\gamma^{2T})}{T(1-\gamma^2)}\right] < \frac{\sigma_\eta^2}{1-\gamma^2} \quad \text{when } T < \infty$$

The downward bias increases inversely with T, and directly with $|\gamma|$.

The problem is that this source of bias is of no interest in the present study, since the use of the autoregregressive scheme for the generation of x_t represents no more than computational convenience. In this case, our interest in asymptotic formulae depends mainly on the fact that moments can be estimated from sample data and substituted into them.

It seems reasonable to suppose that this source of bias is an important component of the deviation of mean bias from inconsistency. Inspection of Table 1, for example, reveals a tendency for larger upward deviations of the mean bias for the experiments for which $\gamma = .9$, than for others. (It is frankly puzzling that the effect is not revealed by the bias regression, but the experimental sample is of course rather small.)

A simple method of checking the importance of the effect would be to replace the asymptotic formula $(\sigma_\eta^2/(1 - \gamma^2))$ by the corresponding estimated moment of the generated series (and revert to holding the latter constant at each replication). Sampling variability in the exogenous variable is thereby eliminated altogether from the experiments. This is unfortunately not possible with the experiments presented here, but the programme has since been updated so that the sample covariance matrix appears in the output.

(iv) Measurement error in dynamic models may mean that the residual errors on the equation are autocorrelated, as is seen by writing (6) with the observed variables substituted into the equation,

$$y_t = \beta_1 x_t + \beta_2 y_{t-1} + \omega_t$$

$$\text{where} \quad \omega_t = \varepsilon_t - \beta_1 v_t + (u_t - \beta_2 u_{t-1}) \qquad (14)$$

The presence of the moving average component, $u_t - \beta_2 u_{t-1}$, suggests that some reduction in bias may be possible by the use of Autoregressive Least Squares. We may appeal to Hendry and Trivedi's [7] finding that these methods are fairly robust against errors of specification (i.e., moving average error processes are approximated by autoregressive schemes); but obviously, much depends on the relative dominance of the moving average component, since the autoregressive transformation employed in ALS will convert the serially independent component into a moving average. This is the kind of question which simulation methods are well suited to assess. ALS estimates were computed in the course of the simulations for model B, and in Table 5, the OLS and ALS biasses are presented for comparison for the eight experiments in which y_t contained measurement error, both with and without errors in x_t also.

TABLE 5

$\sigma_u^2 = 10$

							$\overline{(\hat{\beta}_1 - \beta_1)}$		$\overline{(\hat{\beta}_2 - \beta_2)}$	
	γ	β_1	β_2	σ_v^2	σ_ε^2	T	OLS	ALS	OLS	ALS
3	.9	.4	.6	0	1	30	.418 (.036)	.391 (.041)	−.513 (.020)	−.443 (.029)
4	.9	.8	.2	10	1	30	−.677 (.015)	−.695 (.018)	.020 (.020)	−.021 (.039)
7	.9	.4	.6	0	10	65	.247 (.036)	.269 (.049)	−.264 (.013)	−.301 (.033)
8	.9	.8	.2	10	10	65	−.646 (.010)	−.646 (.011)	.033 (.012)	.107 (.030)
11	.99	.8	.2	0	10	30	.134 (.024)	.153 (.025)	−.139 (.018)	−.165 (.025)
12	.99	.4	.6	10	10	30	−.150 (.014)	−.177 (.017)	−.166 (.016)	−.207 (.037)
15	.99	.8	.2	0	1	65	.177 (.014)	.133 (.021)	−.174 (.012)	−.128 (.021)
16	.99	.4	.6	10	1	65	−.219 (.010)	−.250 (.011)	−.214 (.012)	−.117 (.033)

The quoted simulation standard errors provide a guide to the relative efficiency of the methods. (Note, these statistics are $1/\sqrt{100}$ times the standard deviation of the estimates over replications).

It will be observed that while the bias is reduced in both coefficients when the equation errors are small and there is no measurement error in x_t (experiments 3 and 15), and is reduced in $\hat{\beta}_2$ even with errors in x_t, (4 and 16), it is liable to be made considerably worse when the errors have a large independent component. Since it will rarely be possible to know very much about the relative magnitudes of the error components, these results cannot provide reassurance that ALS will provide superior estimates. The problem of a mixed autocorrelated and independent error term will also be found when the <u>equation</u> errors are autocorrelated and the measurement errors independent – this situation will also be discussed in the course of the following section.

(v) Inspection of Table 2 will show that measurement errors can affect multivariate regression estimates in complex ways. Consider for example the three-variable model as in (6) but with dynamics unspecified.

$$y_t = \beta_1 \chi_{1t} + \beta_2 \chi_{2t} + \varepsilon_t$$

$$x_{1t} = \chi_{1t} + v_{1t} \qquad (15)$$

$$x_{2t} = \chi_{2t} + v_{2t}$$

with all errors independent, the inconsistency in $\hat{\beta}_1$, by analogy with (9), is

$$\text{plim}\,(\hat{\beta}_1 - \beta_1) = \frac{1}{\Delta}\left(-\beta_1 \sigma^2_{v_1}\left(m_{\chi_2\chi_2} + \sigma^2_{v_2}\right) + \beta_2 \sigma^2_{v}\, m_{\chi_1\chi_2}\right) \qquad (16)$$

and similarly for $\hat{\beta}_2$.

It is apparent that if χ_{1t} and χ_{2t} are negatively correlated, and β_1 and β_2 have the same sign - or vice versa - then the errors in x_{1t} and x_{2t} will have a reinforcing effect on the bias in each; whereas, if χ_{1t} and χ_{2t} are positively correlated, and β_1 and β_2 have the same sign (or vice versa) then the errors in x_{1t} and x_{2t} have, asymptotically at least, a counterbalancing effect, and the bias in $\hat{\beta}_1$, say, is smaller, the larger (within limits) the variance of the errors in x_{2t}. (The second of these effects is to be observed in the results for model B.) It is clearly useful to have such information about the expected behaviour of our models, but adding more variables to the equation, and dropping the assumption of independence of the errors would rapidly make it very difficult to determine such effects analytically.

A case of special interest is that where estimation is performed on the first (or higher order) difference of the variables. Differencing is a common procedure in econometrics, being used for instance to obtain a finite approximation to the rate of change of the variable, when this quantity is specified in the theoretical model to be estimated; wage-price models are well-known examples, e.g. Lipsey & Parkin [8]. Or differencing may be used simply to transform a non-stationary series into a stationary one, as in certain forecasting models (see Box & Jenkins, [1]). Yet it may be the case that even if the level magnitudes of the data are measured sufficiently accurately from the estimation point of view, the difference magnitudes are not, particularly when the true variables are "slowly changing" - i.e. exhibit strong first-order autocorrelation - but the measurement errors are serially independent.

The last set of experiments are designed to throw light on the behaviour of least squares estimators with simple models in the differences of variables - either correctly or incorrectly specified - in the presence of measurement errors of such a size as to constitute a serious source of bias in the differences. Also considered is the effect of autocorrelation in the measurement errors.

Two models were used to generate the data,

$$\text{I.} \quad \xi_t = .5\, \chi_t + \varepsilon_{1t} \tag{17}$$

and

$$\text{II.} \quad \Delta\xi_t = .5\, \Delta\chi_t + \varepsilon_{2t} \tag{18}$$

where $\Delta\xi_t = \xi_t - \xi_{t-1}$

$\Delta\chi_t = \chi_t - \chi_{t-1}$

and $\varepsilon_{1t}, \varepsilon_{2t} \sim NI(0, .25)$

In each case, χ_t is generated by

$$\chi_t = z_{1t} + z_{2t} + \eta_{3t}$$

where

$$\begin{aligned} z_{1t} &= .99\, z_{1t-1} + \eta_{1t} \\ z_{2t} &= .5\, z_{2t-1} + \eta_{2t} \\ \eta_{it} &\sim NI(0, 1), \quad i = 1, 2, 3 \end{aligned} \tag{19}$$

(note, the reason for the rather elaborate specification of (19) was to make z_{1t} and z_{2t} available as instruments for the IV estimation of the equations. However, problems related to ensuring that valid instruments were used in the programme were not fully resolved, and these results are not presented.)

We also have,

$$y_t = \xi_t$$
$$x_t = \chi_t + v_t,$$

and there are two specification for the measurement error,

(a) $v_t \sim R(0, 1)$

(b) $v_t = .5\, v_{t-1} + \zeta_t, \quad \zeta_t \sim R(0, .75)$

giving four model specifications, Ia, Ib, IIa, and IIb.

From samples of size 40, the equations estimated from models Ia and Ib were

(i) $y_t = \alpha_{11}\, x_t + u_{1t}$

(ii) $\Delta y_t = \alpha_{21}\, \Delta x_t + u_{2t}$,

and from models IIa and IIb,

(iii) $\Delta y_t = \alpha_{31}\, \Delta x_t + u_{3t}$

(iv) $y_t = \alpha_{41}\, x_t + \alpha_{42}\, y_{t-1} + \alpha_{43}\, x_{t-1} + u_{4t}$

(v) $y_t = \alpha_{51}\, x_t + u_{5t}$.

Note that equations (ii) and (v) represent incorrect specifications, (i) and (iii) represent correct specifications, and (iv) is the unrestricted version of (iii), assuming only a general dynamic specification for the model - as generated α_{42} equals 1, and α_{43} equals -.5.

The main result for the simulations using OLS and ALS are shown in Table 6. The mean biasses and root mean squared errors for each model and each estimator are shown,together with the mean ALS estimates of the autoregression coefficient, ρ. The rows of the table labelled ASE and SSD show, respectively, the means

TABLE 6

		Ia		IIa					Ib		IIb				
		(i)	(ii)	(iii)	(iv)			(v)	(i)	(ii)	(iii)	(iv)			(v)
	Variable	x_t	Δx_t	Δx_t	x_t	y_{t-1}	x_{t-1}	x_t	x_t	Δx_t	Δx_t	x_t	y_{t-1}	x_{t-1}	x_t
OLS	Bias	−.032 (.004)	−.159 (.007)	−.157 (.006)	−.157 (.006)	−.135 (.017)	.216 (.008)	−.053 (.046)	−.029 (.004)	−.090 (.007)	−.098 (.005)	−.088 (.005)	−.086 (.008)	.127 (.006)	.001 (.054)
	RMSE	.049	.17[illegible]	.166	.168	.215	.231	.458	.051	.113	.109	.101	.121	.142	.537
	ASE	.026	.060	.051	.056	.075	.065	.086	.025	.059	.047	.053	.064	.061	.096
	SSD	.036	.070	.055	.059	.167	.083	.455	.043	.068	.046	.050	.085	.063	.537
	s^2	.494	.841	.611	.559			5.73	.465	.695	.445	.448			6.48
	D.W.	1	58	12	1			92	14	35	0	1			96
ALS	Bias	−.034 (.004)	−.140 (.006)	−.147 (.006)	−.152 (.006)	−.242 (.042)	.251 (.017)	−.164 (.006)	−.034 (.005)	−.095 (.006)	−.101 (.005)	−.089 (.005)	−.237 (.039)	.185 (.016)	−.096 (.006)
	RMSE	.053	.151	.157	.163	.484	.301	.175	.058	.113	.111	.103	.454	.245	.112
	ASE	.026	.057	.053	.057	.085	.070	.050	.029	.055	.048	.053	.085	.068	.047
	SSD	.041	.057	.057	.057	.419	.166	.060	.047	.060	.047	.051	.388	.160	.057
	ρ	−.022	−.466	−.257	−.070			.927	.196	−.417	−.103	.063			.943
	s^2	.489	.663	.572	.536			.585	.446	.572	.444	.441			.452
	χ^2	11	87	32	25			100	22	73	6	15			100

of the asymptotic standard errors of the coefficients computed in the usual way at each replication, and the actual sampling standard deviations of the biasses. Comparison of these two quantities gives an indication of the bias in the former, the conventional measure of estimate precision. The rows labelled "D.W." and "χ^2" refer to the Durbin Watson statistic, and the χ^2 test of the hypothesis $\rho = 0$; the quoted figures are not (obviously) mean values of the statistics, but the number of times out of the 100 replications that the respective null-hypotheses were rejected.

A detailed interpretation of the results is best left to the interested reader; the interaction of effects is complex, and the question of extrapolation from the experimental framework must be a matter of rather subjective judgement. It is more appropriate in exploratory experiments of this nature to draw attention to the main suggestive features than to attempt definite conclusions.

Although incidental to the main motivation of the experiments, results of considerable interest are the low power of the Durbin Watson test, judged against the χ^2 test - and of the χ^2 test itself in several cases; and also, the incidence of downward bias in the estimated standard errors of the coefficients, sometimes severe.

Equation (i), the correct specification of the "levels" model, provides a baseline with which the effect of measurement errors of the given variance under various model specifications can be compared. (Note, the serially independent and the autocorrelated measurement errors are specified so as to have an equal asymptotic variance.) The measurement errors produce a level of bias of the order of 6%. The effect of differencing on the bias is similar whether the difference specification is correct (equation (iii)) or incorrect (equation (ii)). For independent measurement errors (specification (a)), the bias rises to over 30%; with autocorrelated measurement errors, the increase is less severe, but still of the order of 20%.

Considering equation (iv), where the difference restrictions are valid but not imposed, it should be noted that the biasses in the coefficients of x_t and x_{t-1} are asymptotically equal in magnitude, although opposite in sign. Significant small-sample effects appear to be present, reinforcing the evidence of the results on model B, discussed in section (iii) above.

Equation (v) was run in order to test the effect of mis-specifying a "differences" model to avoid bias, by estimating in the levels. On a mean squared error criterion, it is clear that this equation performs extremely badly, and it appears unlikely that the reduction in bias could compensate for lack of efficiency.

The "differences" model is of course a particular example of the "levels" model with autoregressive errors, the error process having roots on the unit circle. The last result draws attention to a problem pointed out by Grether and Maddala, [3] p. 257 (and alluded to here in section (iv)); that the use of the quasi-difference transformation to eliminate serial correlation of the errors, as in ALS, can result in an unforeseen increase in measurement error bias. An interesting field of further experimentation would be to study this bias/efficiency trade-off in greater detail; the results for equation (v) seem to suggest that correct specification is desirable, but this preference might not hold at other points in the parameter space.

APPENDIX

Model A:

From (5),

$$\hat{\alpha} = \alpha + \frac{\sum_{t=1}^{T} x_t \omega_t}{\sum_{t=1}^{T} x_t} \qquad \text{where } \omega_t = \varepsilon_t - \alpha v_t.$$

Hence,

$$\text{plim } (\hat{\alpha} - \alpha) = \frac{\text{plim } \frac{1}{T} \sum_{t=1}^{T} (\chi_t + v_t)(\varepsilon_t - \alpha v_t)}{\text{plim } \frac{1}{T} \sum_{t=1}^{T} (\chi_t + v_t)^2}$$

$$= \frac{-\alpha\sigma_v^2}{m_{\chi\chi} + \sigma_v^2} \qquad (8)$$

Model B:

From (6),

$$\hat{\beta}_1 = \beta_1 + \frac{\left(\sum_{t=1}^{T} x_t\omega_t\right)\left(\sum_{t=1}^{T} y_{t-1}^2\right) - \left(\sum_{t=1}^{T} y_{t-1}\omega_t\right)\left(\sum_{t=1}^{T} x_t y_{t-1}\right)}{\left(\sum_{t=1}^{T} x_t^2\right)\left(\sum_{t=1}^{T} y_{t-1}^2\right) - \left(\sum_{t=1}^{T} x_t y_{t-1}\right)^2}$$

where $\omega_t = \varepsilon_t + u_t - \beta_1 v_t - \beta_2 u_{t-1}$

$$\text{plim } \frac{1}{T} \sum_{t=1}^{T} x_t\omega_t = \text{plim } \frac{1}{T} \sum_{t=1}^{T} (\chi_t + v_t)(\varepsilon_t + u_t - \beta_1 v_t - \beta_2 u_{t-1}) = -\beta_1\sigma_v^2$$

$$\text{plim } \frac{1}{T} \sum_{t=1}^{T} y_{t-1}\omega_t = \text{plim } \frac{1}{T} \sum_{t=1}^{T} (\xi_{t-1} + u_{t-1})(\varepsilon_t + u_t - \beta_1 v_t - \beta_2 u_{t-1}) = -\beta_2\sigma_u^2$$

Hence,

$$\text{plim}\ (\hat{\beta}_1 - \beta_1) = \frac{-\beta_1\sigma_v^2(m_{\xi\xi} + \sigma_u^2) + \beta_2\sigma_u^2 m_{\chi\xi_{-1}}}{(m_{\chi\chi}+\sigma_v^2)(m_{\xi\xi}+\sigma_u^2) - m_{\chi\xi_{-1}}^2} \qquad (9)$$

and plim $(\hat{\beta}_2 - \beta_2)$ is derived analogously.

Moment formulae (11)

(i) $$\chi_t^2 = (\gamma\chi_{t-1} + \eta_t)^2$$

$$= \gamma^2\chi_{t-1}^2 + \eta_t^2 + 2\gamma\chi_{t-1}\eta_t$$

$\bar{E}(\chi_t^2) = \bar{E}(\chi_{t-1}^2) = m_{\chi\chi}$, where $\bar{E}$ denotes the asymptotic expectation

Thus, $$m_{\chi\chi} = \gamma^2 m_{\chi\chi} + \sigma_\eta^2$$

$$\Rightarrow m_{\chi\chi} = \frac{\sigma_\eta^2}{1-\gamma^2}$$

(ii) $$\chi_t\xi_{t-1} = \gamma\chi_{t-1}\xi_{t-1} + \eta_t\xi_{t-1}$$

$$= \gamma\beta_1\chi_{t-1}^2 + \gamma\beta_2\chi_{t-1}\xi_{t-2} + (\gamma\chi_{t-1}\varepsilon_{t-1} + \eta_t\xi_{t-1})$$

$$\bar{E}(\chi_t\xi_{t-1}) = \bar{E}(\chi_{t-1}\xi_{t-2}) = m_{\chi\xi_{-1}}$$

$$\Rightarrow m_{\chi\xi_{-1}} = \gamma\beta_1 m_{\chi\chi} + \gamma\beta_2 m_{\chi\xi_{-1}} + 0$$

$$= \frac{\gamma\beta_1}{1-\gamma\beta_2} m_{\chi\chi}$$

(iii) $$\xi_t^2 = (\beta_1\chi_t + \beta_2\xi_{t-1} + \varepsilon_t)^2$$

$$= \beta_1^2\chi_t^2 + \beta_2^2\xi_{t-1}^2 + \varepsilon_t^2 + 2\beta_1\beta_2\chi_t\xi_{t-1} + 2(\beta_1\chi_t + \beta_2\xi_{t-1})\varepsilon_t$$

$$\bar{E}(\xi_t^2) = \bar{E}(\xi_{t-1}^2) = m_{\xi\xi},$$

$$\Rightarrow \quad m_{\xi\xi} = \beta_1^2 m_{\chi\chi} + \beta_2^2 m_{\xi\xi} + \sigma_\varepsilon^2 + 2\beta_1\beta_2 m_{\chi\xi_{-1}}$$

$$= \frac{1}{1-\beta_2^2}\left[\left[\beta_1^2 + 2\beta_1\beta_2 \frac{\gamma\beta_1}{1-\gamma\beta_2}\right] m_{\chi\chi} + \sigma_\varepsilon^2\right],$$

using the previous result to substitute for $m_{\chi\xi_{-1}}$;

$$m_{\xi\xi} = \frac{1}{1-\beta_2^2}\left[\frac{\beta_1^2(1 - \gamma\beta_2 + 2\gamma\beta_2)}{1 - \gamma\beta_2} m_{\chi\chi} + \sigma_\varepsilon^2\right]$$

$$= \frac{1}{1-\beta_2^2}\left[\frac{1 + \beta_2\gamma}{1 - \beta_2\gamma} \beta_1^2 m_{\chi\chi} + \sigma_\varepsilon^2\right]$$

REFERENCES

[1] Box, G.E.P., and Jenkins, G.M., (1970), <u>Time Series Analysis, Forecasting and Control</u>, Holden Day.

[2] Cochran, W.G., and Cox, G.M. (1957), <u>Experimental Designs</u>, (2nd Ed.), John Wiley.

[3] Grether, D.M., and Maddala, G.S., (1973), "Errors in Variables and Serially Correlated Disturbances in Distributed Lag Models", <u>Econometrica</u>, Vol. 41, No. 2.

[4] Hendry, D.F., (1973) "Users' Manual for GIVE: General Instrumental Variable Estimation of Linear Equations with Lagged Dependent Variables and First Order Autoregressive Errors", mimeo, London School of Economics.

[5] ___________, (1974) "Users' Manual for NAIVE (Numerical Analysis of Instrumental Variable Estimators): Monte Carlo Analysis of GIVE estimators", mimeo, London School of Economics.

[6] ___________, and Harrison, R.W. (1974) "Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two Stage Least Squares", <u>Journal of Econometrics</u>, Vol. 2, No. 2.

[7] ___________, and Trivedi, P.K., (1972), "Maximum Likelihood Estimation of Difference Equations with Moving Average Errors: A Simulation Study", <u>Review of Economic Studies</u>, Vol. XXXIX (2).

[8] Lipsey, R.G., and Parkin, J.M. (1970), "Incomes Policy: A Re-appraisal", <u>Economica</u>, Vol. XXXVII, 145.

[9] Mikhail, W.M., (1972), "Simulating the Small Sample Properties of Econometric Estimators", Journal of the American Statistical Association, 67, 620-624.