COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 400

A CONTROL VARIABLE INVESTIGATION OF THE PROPERTIES OF AUTOREGRESSIVE INSTRUMENTAL VARIABLES ESTIMATORS FOR DYNAMIC SYSTEMS

David F. Hendry and Frank Srba

August 19, 1975

A CONTROL VARIABLE INVESTIGATION OF THE PROPERTIES OF AUTOREGRESSIVE INSTRUMENTAL VARIABLES ESTIMATORS FOR DYNAMIC SYSTEMS*.

David F. Hendry and Frank Srba

The London School of Economics

June 1975.

* Financed by a Grant from the Social Science Research Council. We are very grateful to Robin W. Harrison for the preliminary work in writing the direct simulation programme NAIVE from which this study is developed. Completed while the first author was visiting at the Cowles Foundation, Yale University, where his work is supported by granted from the National Science and Ford Foundations.

A. Introduction

A large proportion of empirical econometric studies involve dynamic simultaneous equations systems and although there is often reasonable *a priori* economic theory to guide the specification, there is generally little or no prior information about either the dynamic structure or the autocorrelation properties of the error term. In practice, therefore, lag relationships are determined empirically using conventional econometric estimators such as Ordinary Least Squares (denoted OLS) or Instrumental Variables (IV) and the serious mis-specification biases which can result from attempting this in the presence of autocorrelation are now well established asymptotically (Hendry (1975)) and have been confirmed for finite samples (Hendry and Harrison (1974)). Further, many estimated models *do* exhibit autocorrelated residuals (Hickman (1972)) and as their one-period ahead forecasts are frequently no better than simple autoregressive extrapolations (Cooper (1972)) we have considerable evidence that the dynamics are not being appropriately "identified" in the sense of Box and Jenkins (1970).

Generalisations of OLS and IV which allow for autoregressive errors have, however, been available for some time (Sargan (1959), (1964)) and we denote these ALS and AIV (see section C). Their asymptotic properties in correctly specified situations are well-known (see Hendry (1974b) for a survey of the ALS and the AIV families of estimators) and an empirical application of all four estimators noted above to a small aggregate demand system demonstrated that the last two could yield valuable information about dynamic specification (Hendry (1974a)).

This provides a strong case for studying the finite sample behaviour of ALS and AIV in such models. However, the estimators require iterative solution of non-linear equations, effectively precluding a purely analytic derivation as well as making a straight simulation study very expensive. Thus we have adopted a Control Variable approach(Mikhail (1972)) solving part of the problem analytically via an asymptotic expansion, and the remainder by computer experiments. Additionally this clarifies the relationship between the numerical and the statistical properties of the estimators (compare Hendry (1974b)).

The Data Generation Process, the Monte Carlo methodology and the experimental design and analysis are as in Hendry and Harrison (1974) (denoted H-H henceforth)[where possible, in equivalent replications we have re-used their random numbers to enhance the accuracy of inter-estimator comparisons] and we reference that study for a detailed discussion of those aspects, although sections B, D and E summarise the main points for convenience. Section F reports our results.

B. Data Generation

The model used is defined by

(1) $$y_t = bY_t + cw_{1t} + dy_{t-1} + u_t \; ; \qquad u_t = ru_{t-1} + \varepsilon_t \qquad |r| < 1$$

(2) $$Y_t = ay_t + \sum_{i=1}^{4} f_i w_{it} + e_t \; ; \qquad \left|\frac{d}{1-ab}\right| < 1$$

(3) $$w_{it} = \lambda_i w_{it-1} + v_{it} \; ; \qquad |\lambda_i| < 1$$

The errors are distributed according to

(4) $$(\varepsilon_t, e_t) \sim NI(0, \Sigma) \qquad \text{where} \qquad \Sigma = (\sigma_{ij}) \qquad \text{and}$$

(5) $$v_{it} \sim NI(0, \omega_{ii}) \qquad \text{with}$$

(6) $$\mathcal{E}(v_{it}\varepsilon_s) = \mathcal{E}(v_{it}e_s) = \mathcal{E}(v_{it}v_{js}) = 0 \qquad \forall i, j, t, s. \; (it \neq js)$$

Uniform random numbers (ξ_j) were generated by a multiplicative congruential method and for (ε_t, e_t) converted to a bivariate normal distribution using the

Box-Muller procedure (see Hammersley and Handscomb (1964)); the v_{it} were made approximately normal using sums of twelve ξ_j. The first 49 data values were discarded and the next T constituted the sample for one replication; note that the w_{it} were usually stochastic (i.e. varied between replications).

The estimators were applied to (1), all of its parameters being over-identified in the econometric sense (see Sargan (1972)) and we concentrated on evaluating their first two moments. We investigated the existence of these as in H-H although a recent paper by Sargan (1975) shows that the non-existence of moments does not automatically invalidate a Monte-Carlo study as the results may estimate the finite moments of a distribution function which approximates the actual small sample distribution. In any event, we did achieve quite large efficiency gains from our Control Variables (compare Sargan (1974)).

From (1) - (6), the population moments can be derived as functions of the parameters (see Appendix A for those additional to H-H) and hence evaluated numerically for given parameter values.

C. The Econometric Estimators

Equation (1) can be written in standard regression notation as

(7) $$y = X\beta + u \qquad \text{where } \beta' = (b \; c \; d), \text{ and}$$

$$X' = (x_2 \; \ldots \; x_T) \qquad \text{with } x_t' = (Y_t \; w_{1t} \; y_{t-1}). \quad \text{Let}$$

$$Q' = (q_2 \; \ldots \; q_T) \quad \text{where} \quad q_t' = (w_{1t} y_{t-1} w_{2t} w_{3t} w_{4t} x_{t-1}') \text{ and define}$$

(8) $$y^+ = y - ry_1 , \qquad X^+ = X - rX_1 \quad \text{etc.} \quad \text{where (e.g.)}$$

X_1 denotes X lagged one period. Note that

(9) $$X_1' = DQ' \qquad \text{where} \quad D = [0 : I] \quad \text{and}$$

(10) $y^+ = X^+\beta + \varepsilon.$

The AIV estimator of (β, r) denoted $\tilde{}$ is obtained by minimising the quadratic form

(11) $g(M) = \varepsilon' M \varepsilon$ where $M = Q(Q'Q)^{-1}Q'$

as a function of β and r (see Sargan (1964)). The IV or Two-stage Least Squares estimator (TSLS) minimises $u'Mu$ as a function of β only and with X_1 usually excluded from Q. The first order conditions for AIV are $\frac{\partial g}{\partial \beta} = 0$ and $\frac{\partial g}{\partial r} = 0$ which yield

(12) $\tilde{\beta} = (\tilde{X}^{+\prime} M \tilde{X}^{+})^{-1}(\tilde{X}^{+\prime} M \tilde{y}^{+})$ where $\tilde{X}^+ = X - \tilde{r}X_1$ and

(13) $\tilde{r} = (\tilde{u}_1'\tilde{u}_1)^{-1}(\tilde{u}_1'\tilde{u})$ where $\tilde{u} = y - X\tilde{\beta}$

to be solved jointly for $(\tilde{\beta}, \tilde{r})$ which minimise g(M).

In fact, ALS (denoted $\hat{}$) is now the special case in which the instrumental variables are just the regressors $(Q = X)$ which is <u>equivalent</u> to minimising

$$g(I) = \varepsilon'\varepsilon$$

as a function of (β, r) so that

(14) $\hat{\beta} = (\hat{X}^{+\prime}\hat{X}^{+})^{-1}(\hat{X}^{+\prime}\hat{y}^{+})$

and $\hat{r}$ is equivalent to (13) using $\hat{u}$ in place of $\bar{u}$. (Compare the corresponding Moving-Average Error model in Box and Jenkins (1970)).

In both cases iterative solution is involved, and to check for multiple optima and mimic the operational characteristics of these estimators, stepwise optimisation was used from the best local outcome determined by a grid search of g over $r \epsilon (-.9, -.8, \ldots, .8, .9)$. This is fast for this specific problem and has only failed to converge twice in 8000 trials.

An alternative numerical procedure often proposed for non-linear least squares problems is the Gauss-Newton algorithm (see e.g. Dixon (1972)) which for AIV requires iterating on

$$(15)\qquad \begin{pmatrix} \tilde{\beta} \\ \tilde{r} \end{pmatrix} = \begin{pmatrix} \tilde{X}^{+\prime} M\tilde{X}^{+} & \tilde{X}^{+\prime}\tilde{u}_1 \\ \tilde{u}_1' \tilde{X}^{+} & \tilde{u}_1' \tilde{u}_1 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{X}^{+\prime} M\tilde{y}^{+} \\ \tilde{u}_1' \tilde{u} \end{pmatrix}$$

and if $\tilde{X}^{+\prime}\tilde{u}_1 \neq 0$ this will be more efficient computationally than the successive substitution approach in (12)+(13). (Compare equations (22) and (23) below). However, from the optimum grid value the reduction in computer time using (15) would be small.

We have for AIV that

$$(16)\qquad \sqrt{T}\begin{pmatrix} \tilde{\beta} - \beta \\ \tilde{r} - r \end{pmatrix} \underset{A}{\sim} N\left(0,\ \sigma_{11}\begin{pmatrix} H & h \\ h' & \sigma_{11}/(1-r^2) \end{pmatrix}^{-1}\right)$$

where

$$(17)\qquad H = (PM_aP' - r(M_b + M_b') + r^2M_c), \qquad P = \text{plim } X'Q(Q'Q)^{-1}$$

$$M_a = \text{plim } T^{-1}Q'Q, \quad M_b = \text{plim } T^{-1}X'X_1 \quad \text{and} \quad M_c = \text{plim } T^{-1}X'X \quad \text{with}$$

$$(18)\qquad h = \text{plim } T^{-1}X^{+\prime}u_1 = (P - rD)\text{ plim } T^{-1}Q'u_1.$$

For the above choice of Q, this is the same asymptotic distribution as the Autoregressive generalisation of Limited Information Maximum Likelihood (see Sargan (1959)). $\tilde{\beta}$ and $\tilde{r}$ are independently distributed only if all the instruments are strictly exogenous (see Amemiya (1966)); note that $\text{plim } T^{-1}g(M) = 0$.

Similarly, if $a = \sigma_{12} = 0$ then ALS is consistent and asymptotically efficient, having the same form of variance matrix as (16) but with PM_aP' replaced by M_c;

also plim $T^{-1}g(I) = \sigma_{11}$. The small sample properties of ALS in reduced form dynamic equations have previously been investigated by direct simulation methods (see Malinvaud (1966)) and by Antithetic Variates (see Hendry and Trivedi (1972)).

If either a or σ_{12} is non-zero ALS ceases to be consistent and as plim $\hat{\beta} = \beta_p$ is a non-linear function of plim $\hat{r} = r_p$ an explicit formula for the inconsistency cannot be derived. Nevertheless, since we can calculate the population moments, by iteratively minimising plim $T^{-1}g(I)$ we can obtain the numerical value of (β_p, r_p) in every experiment. This can be used as a statistical control to explain inter-experiment variations in the central tendency (see section F) and to construct a "crude" control variable (see (30)).

D. Control Variables

The principles underlying the application of a Control Variable (denoted CV) (see Mikhail (1972)) are well illustrated by the case of ALS with fixed regressors ($d = a = \sigma_{12} = 0$ with the w_{jt} held fixed within experiments). When r is known (as it is to the experimenter) ALS is approximately Aitken's Generalised Least Squares estimator (AGLS):

$$(19) \qquad \beta^* = (X^{+'}X^{+})^{-1}(X^{+'}y^{+}) = \beta + (X^{+'}X^{+})^{-1}X^{+'}\varepsilon$$

which is exactly distributed as $N(\beta, \sigma_{11}(X^{+'}X^{+})^{-1})$ (ALS for estimated r is sometimes compared with this as a baseline to evaluate average inter-experiment outcomes). However, β^* is computable at every replication and can be used to control intra- experiment variation. Since $\hat{\beta}$ and β^* are asymptotically equivalent, $(\hat{\beta} - \beta^*)$ is $O(\frac{1}{T})$, and as

$$(20) \qquad (\hat{\beta} - \beta) \equiv (\beta^* - \beta) + (\hat{\beta} - \beta^*)$$

where the distribution of $(\beta^* - \beta)$ is known, we can estimate that of $(\hat{\beta} - \beta)$ more accurately indirectly via the right-hand decomposition than by direct simulation. Thus, for example, to estimate the bias $\mathcal{E}(\hat{\beta} - \beta)$ one could use the mean of the direct simulation outcomes $(\bar{\beta} - \beta)$ (say); but by calculating β^* at every replication, since from (20)

$$(21) \qquad \mathcal{E}(\hat{\beta} - \beta) \equiv \mathcal{E}(\beta^* - \beta) + \mathcal{E}(\hat{\beta} - \beta^*) = \mathcal{E}(\hat{\beta} - \beta^*)$$

$(\bar{\beta} - \bar{\beta}^*)$ (denoted the pooled estimator below) is a much more efficient estimator (note that $(\hat{\beta} - \beta^*)$ depends only on $(\hat{r} - r)$ and $\hat{r}$ is an asymptotically efficient estimator of r).

More generally, our principle is to construct a (simulation) estimator with a known finite sample distribution which is the same as the asymptotic distribution of the econometric estimator and simulate the difference between these. Since this (inter alia) provides general and analytic expressions for the distribution of $\hat{\beta}$ to $O(\frac{1}{T})$ whereas experiments are specific and imprecise, unless very large numbers of replications are used simulation estimated moments (if they exist) may have larger Mean Square Errors than those based on the approximate distribution. Further, this is increasingly important as T increases (especially for iterative estimators) and hence our approach highlights the need to ask "How small is 'small'?" to justify resort to Monte Carlo experiment in place of the conventional "How large is 'large'?" to justify asymptotic results. (Compare Goldfeld and Quandt (1972)). Of course, since the Control Variable approach combines both aspects, we achieve more precise results than either in isolation.

The asymptotic distribution of AIV is given in (16) and hence we can easily obtain the required CV using

$$(22)\qquad \begin{pmatrix} \beta^o - \beta \\ r^o - r \end{pmatrix} = T^{-1} \begin{pmatrix} H & h \\ h' & \sigma_{11}/(1-r^2) \end{pmatrix}^{-1} \begin{pmatrix} (P-rD)\ Q'\varepsilon \\ u_1'\ \varepsilon \end{pmatrix}$$

Note that this has the same form as the Gauss-Newton algorithm (15). One might, however, have attempted to base the CV on a generalisation of (19) leading to

$$(23)\qquad (\beta^* - \beta) = H^{-1}(P - rD)\,\frac{Q'\varepsilon}{T} \quad \text{and (say)}$$

$$(24)\qquad (r^* - r) = \left(\frac{1 - r^2}{\sigma_{11}}\right)\left(\frac{u_1'\varepsilon}{T}\right)$$

which have the same forms as the successive-substitution equations (12) and (13). It is interesting that (β^*, r^*) is statistically inefficient relative to (β^o, r^o) unless $h = 0$ (as it does if X is fixed, confirming the efficiency of (19)), whereas the formulation which induces an efficient optimisation algorithm also induces an efficient simulation one. It must be stressed that (β^o, r^o) is non-iterative and trivial to calculate as only $Q'\varepsilon$ and $u_1'\varepsilon$ vary between replications and these are calculated in any case for (12) and (13)(in deviation form).

When it is valid to apply ALS to (1) $(a = \sigma_{11} = 0)$ the efficient CV is just the appropriate specialisation of (22):

$$(25)\qquad \begin{pmatrix} \beta^\dagger - \beta \\ r^\dagger - r \end{pmatrix} = \begin{pmatrix} [(1+r^2)M_c - r(M_b+M_b')] & h \\ h' & \sigma_{11}/(1-r^2) \end{pmatrix}^{-1} \begin{pmatrix} \frac{X^{+'}\varepsilon}{T} \\ \frac{u_1'\varepsilon}{T} \end{pmatrix}$$

More generally when ALS is inconsistent (if a or σ_{12} is non-zero when $b \neq 0$) (25) is not an appropriate CV. We have been unable to derive its asymptotic distribution in this case to develop a fully efficient CV, but we did construct an unbiased estimator of the inconsistency to investigate the accuracy gains from CV's based on correlations only (see Hammersley and Hardscomb (1964)) i.e. not justified by asymptotic theory. The ALS equivalent of (15) is

$$(26)\quad \begin{pmatrix} \hat{\beta}-\beta \\ \hat{r}-r \end{pmatrix} = \begin{pmatrix} \hat{X}^{+\prime}\hat{X}^{+} & \hat{X}^{+\prime}\hat{u}_1 \\ \hat{u}_1'\hat{X}^{+} & \hat{u}_1'\hat{u}_1 \end{pmatrix}^{-1} \left[\begin{pmatrix} \hat{X}^{+\prime}\varepsilon \\ \hat{u}_1'\varepsilon \end{pmatrix} + (r-\hat{r}) \begin{pmatrix} \hat{X}^{+\prime}X_1 \\ \hat{u}_1' X_1 \end{pmatrix} (\hat{\beta}-\beta) \right] = \Phi^{-1}(\phi_1+\phi_2)$$

where (e.g.) $\phi_1 = T^{-1}\begin{pmatrix} \hat{X}^{+\prime}\varepsilon \\ \hat{u}_1'\varepsilon \end{pmatrix}$. Let plim $\Phi = \Phi_p$, plim $\phi_i = \phi_{ip}$, then

$$(27)\qquad \text{plim}\begin{pmatrix} \hat{\beta}-\beta \\ \hat{r}-r \end{pmatrix} = \begin{pmatrix} \beta_p-\beta \\ r_p-r \end{pmatrix} = \Phi_p^{-1}(\phi_{1p}+\phi_{2p}),$$

where

$$(28)\qquad \Phi_p = \begin{pmatrix} [(1+r_p^2)M_c - r_p(M_b+M_b')] & \theta \\ \theta' & \sigma_p^2/(1-r_p^2) \end{pmatrix}$$

$$\theta = \text{plim}\, T^{-1}(X - r_pX_1)'(y_1 - X_1\beta_p), \quad \sigma_p^2 = \text{plim}\, \hat{\sigma}^2 \text{ and } \hat{\sigma}^2 = T^{-1}\hat{\varepsilon}'\hat{\varepsilon}.$$

Further, (β_p, r_p) and all the population moments are known. Note that

$$(29)\qquad \phi_{1p} = \begin{pmatrix} (\sigma_{12}+a\sigma_{11})/(1-ab) \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \phi_{2p} = (r-r_p)\begin{pmatrix} (M_b - r_pM_c) \\ \text{plim}\, T^{-1}(y_1'X_1 - \beta_p'M_c) \end{pmatrix}(\beta_p-\beta)$$

Thus one unbiased estimator of the inconsistency is given by

$$(30)\qquad T^{-1}\Phi_p^{-1}\left[\begin{pmatrix} (X - r_pX_1)'\varepsilon \\ (y_1 - X_1\beta_p)'\varepsilon \end{pmatrix} + T\phi_{2p}\right]$$

Additionally, $\sigma_p^2\Phi_p^{-1}$ provides the plims of the estimated variances of ALS. Overall, however, this seems precisely the type of problem for which straight Monte Carlo methods may have a role.

E. Design and Analysis of Experiments

Each design is briefly described in the relevant results section, but they are all variants of "factorial" forms. The analyses primarily relate to the first two moments of the distributions and to the efficiency gains achieved by estimating the central tendency using control variables (defined as the ratio of the simulation variance of the pooled estimator as in (21) to that of the mean direct simulation outcome). The general factor response surfaces adopted to explain variations in the first moments were of the form

$$(31)\qquad (\overline{\mathcal{E}(\hat{\beta}_i)} - \beta) = \mathfrak{J}_i(\beta_{ip}, b, r, c, d, T, \sigma_{11}, \sigma_{12})$$

where $\hat{\beta}_i$ is the econometric estimator with plim $\hat{\beta}_i = \beta_{ip}$, $\overline{\mathcal{E}(\hat{\beta}_i)}$ is the simulation estimate of $\mathcal{E}(\hat{\beta}_i)$. For the second moments we considered

$$(32)\qquad \ln sD_i = \delta_1 \ln ASE_i + \delta_2/T$$

where ASE_i is the asymptotic standard error of $\hat{\beta}_i$ and SD_i is the simulation estimate of the sampling standard error (see Naylor (1971 ch.7)). (Note that an identical functional relationship holds for the equivalent variances). A similar regression to (32) was used for the estimated standard errors of the $\hat{\beta}_i$ - denoted ESE_i - the square roots of the relevant conventional variance formulae (such as $\hat{\sigma}_{11}(\hat{X}^{+'}\hat{X}^{+})^{-1}$ for fixed regressor ALS estimates). For a given specification of $\mathfrak{J}_i$ the residual variances of such regressions (s^2) are due to errors in estimating the regressand (s_1^2) and inappropriate functional forms (s_2^2) and if we assume that these are independent we can measure their importance by comparing the direct (d) and the pooled estimates (e). Thus $s_d^2 = s_{1d}^2 + s_{2d}^2$ and $s_e^2 = s_{1e}^2 + s_{2d}^2$ as $s_{2d}^2 = s_{2e}^2$; however, from the within replication simulation results we can estimate that $s_{1d}^2 \approx Ks_{1e}^2$ and hence $s_{2d}^2 = (Ks_e^2 - s_d^2)/(K-1)$ while

$s_{1e}^2 = s_e^2 - s_{2d}^2$. The relative magnitudes of these indicate whether further resources should be devoted to more replications to reduce s_{1e}^2 or better formulation of the approximation to η_i. Generally, we chose η_i such that the dependent variable would have an approximately constant variance across experiments (see e.g. Nicholls, Pagan and Terrell (1975)). Note that there will usually be considerable testable prior theory about the signs and magnitudes of these regression coefficients (e.g. $\delta_1 = 1$). The overall validity of the fitted regressions can also be evaluated by retaining a small number (n) of randomly chosen experimental outcomes and testing the regression predictions (F_i) against these observations (O_i). This is done below using the result that $\sum_{i=1}^{n} (O_i - F_i)^2/s^2 \underset{A}{\sim} \chi_n^2$.

F. Results

I. Firstly we considered the ALS/AGLS combination with fixed regressors for an equation of the form

$$(33) \qquad y_t = \sum_{j=2}^{4} \beta_j w_{jt} + u_t, \qquad u_t = r u_{t-1} + \varepsilon_t$$

with $\beta_j = 1 \ \forall j$, $\sigma_{11} = 1$ (this is now just a scale effect on AGLS variances leaving the $\hat{\beta}_j$ unaffected). The w_{jt} were generated prior to each experiment using $(\lambda_2, \lambda_3, \lambda_4) = (.8, .4, .2)$ and $(\omega_{22}, \omega_{33}, \omega_{44}) = (.36, .84, \frac{1}{12})$, and held fixed between replications. We chose T = (20, 40, 60, 80) (one observation was "lost" to allow the estimation of r) and r = ($\pm$.9, $\pm$.5) and we used a complete factorial design of 16 experiments. The CV for $\hat{\beta}$ was (19) and that for $\hat{r}$ was (24); each experiment was replicated 100 times. The results for 2 (randomly chosen) experiments are recorded in Table I for $\hat{\beta}_3$ and $\hat{r}$. In all experiments the biases estimated from the pooled estimates for all the $\hat{\beta}_j$ were small and insignificant despite very dramatic efficiency gains (see Table II). The Direct Simulation estimate of $\mathcal{E}(\hat{r})[\bar{r}]$ revealed significant bias and regression analysis of 14 experiments yielded $(\ell = 1/\sqrt{T}(1-r^2))$

$$(34) \qquad T\ell(\bar{r} - r) = \underset{(.15)}{-.17\ell} - \underset{(.17)}{1.27r\ell} \qquad s = .17 \quad R^2 = .82 \quad \chi^2_2 = 0.3$$

Comparably, for the pooled estimate of $\mathcal{E}(\hat{r})$ we obtained

$$(35) \qquad T\ell(\bar{r} - \bar{r}^*) = \underset{(.10)}{-.25\ell} - \underset{(.12)}{1.35r\ell} \qquad s = .11 \quad R^2 = .92 \quad \chi^2_2 = 1.6$$

implying that

$$\mathcal{E}(\hat{r} - r) \simeq -\left(\frac{1+5.4r}{4T}\right) \qquad \text{(compare Shenton and Johnson (1965)).}$$

From Table II the variance of the estimate of $\mathcal{E}(\hat{r})$ from direct simulation is roughly four times that of the pooled estimate and hence the remaining gains in reducing s^2 in (35) by more accurate simulation are of the order of .005 as against .008 due to functional mis-specification.

Next for the pooled estimator of $sD(\hat{r}) = \sqrt{\mathcal{E}(\hat{r}-r)^2}$ we found

$$(36) \qquad \ln \overline{sD(\hat{r})} = \underset{(0.4)}{1.00\ln ASE(\hat{r})} + \underset{(2.8)}{3.2/T} \qquad s = .21 \quad R^2 = .85 \quad \chi^2_2 = 0.6$$

where $ASE(\hat{r})$ is the asymptotic standard error (also equal to $\sqrt{\mathcal{E}(r^*-r)^2}$) strongly confirming the relationship between the actual finite sample variance and its asymptotic value. Finally, for the mean direct simulation estimated Standard Errors $ESE(\hat{r}) = \sqrt{\hat{\sigma}_{11}/\hat{u}_1'\hat{u}_1}$:

$$(37) \qquad \ln \overline{ESE(\hat{r})} = \underset{(.01)}{0.99\ln ASE(\hat{r})} + \underset{(0.9)}{3.7/T} \qquad s = .06 \quad R^2 = .98 \quad \chi^2_2 = 0.6$$

again affirming the value of asymptotic theory in small sample situations.

TABLE I

Experiments (T, r) = (39, -.5) and (79, .9)

	$\bar{B}$	B*	F_1	sD	F_2	ESE	F_3	ASE
β_3	-.019	.000	-	.097	-	.096	-	.098
	(.010)	(.001)				(.001)		
r	.008	-.009	.011	.152	.150	.146	.154	.139
	(.012)	(.005)				(.001)		
β_3	.001	.000	-	.117	-	.117	-	.117
	(.011)	(.0003)				(.001)		
r	-.013	-.019	-.018	.043	.051	.052	.052	.049
	(.005)	(.003)				(.001)		

Legend

$\bar{B}$ is the mean direct simulation estimated bias, B* the pooled estimated bias, F_1 the prediction from the regression (35), sD the pooled estimated sampling standard error, F_2 the prediction from (36), ESE the estimated standard error, F_3 the prediction from (37) and ASE is the asymptotic standard error. Simulation Standard Errors in Parentheses.

TABLE II

Average Efficiency Gains in Estimating Biases

	T ≤ 39		T ≥ 59	
	β_3	r	β_3	r
|r|=0.5	1507	2	1218	3
|r|=0.9	89	4	95	8

For fixed regressor models estimated by ALS, however, the alternative simulation variance reduction method of Antithetic Variates has <u>infinite efficiency</u> for estimating the bias in $\hat{\beta}$ when ε is symmetrically distributed (compare Kakwani (1967)). Since $\hat{r}$ is the same for ε and $-\varepsilon$, the latter produces only a sign change in $\hat{\beta}$ and the average of two replication is thus identically zero, verifying the CV results on bias for $\hat{\beta}$.

II. Next we considered ALS for the dynamic model

(38) $\quad y_t = bY_t + dy_{t-1} + u_t \qquad\qquad u_t = ru_{t-1} + \varepsilon_t$

(39) $\quad Y_t = w_{1t} + e_t \qquad\qquad \varepsilon_t, e_t \sim NI(0, 1)$ and $\sigma_{12} = 0$

(40) $\quad w_{1t} = .8w_{1t-1} + v_t \qquad\qquad v_t \sim NI(0, 4)$

In this situation, ALS is consistent and asymptotically efficient, whereas OLS is inconsistent; we used the CV (25) for the former and the CV developed in H-H for OLS. We chose a Graeco-Latin square design of 16 experiments (see Cochran and Cox (1957 p. 146)) generated by T = (20, 40, 60, 80), b = ($\pm$1, $\pm$10), d = ($\pm$.8, $\pm$.4) and r = ($\pm$.75, $\pm$4) with 100 replications. The results for 2 of these experiments are recorded in Table III and the efficiency gains in Table 1V.

Checking the efficiency of CV's of the form of (23) and (24) revealed them to be about 90% efficient relative to (25) for this model. It is clear from Table IV that the efficiency gains are large (especially for $T \geq 59$) and hence the distributions of OLS and ALS are close to their asymptotic approximations.

As before, we used regression analysis to summarise the experimental outcomes of the remaining 14 experiments for ALS and the relationships between

TABLE III

Experiments (T, b, d, r) = (19, -10, -.8, -.4) and (79, 1, .4, .75)

	OLS			ALS							
	I	$\bar{B}$	B^{**}	$\bar{B}$	B^{*}	F_1	sD	F_2	ESE	F_3	ASE
b	-.004	-.009	-.004	-.001	-.003	-.014	.100	.096	.093	.089	.081
		(.010)	(.006)	(.011)	(.006)				(.002)		
d	-.001	.000	.000	.000	.000	/	.016	.014	.013	.014	.012
		(.002)	(.001)	(.002)	(.001)				(.001)		
r	/	/	/	.085	.061	.067	.248	.247	.233	.236	.210
				(.025)	(.015)				(.003)		
b	-.148	-.127	.017	.007	.001	.000	.047	.048	.049	.049	.048
		(.008)	(.003)	(.005)	(.001)				(.001)		
d	.137	.138	-.007	-.003	.002	/	.042	.042	.045	.042	.040
		(.005)	(.002)	(.004)	(.002)				(.001)		
r	/	/	/	-.012	-.024	-.032	.074	.077	.083	.077	.075
				(.008)	(.004)				(.001)		

Legend as Table I; I denotes inconsistency, and B^{**} the deviation of the pooled estimate of the bias from I.

TABLE IV

Average Efficiency Gains

	$T \leq 39$		$T \geq 59$	
	OLS	ALS	OLS	ALS
b	8	6	13	14
d	7	5	10	11
r	/	4	/	9

the estimated finite sample moments and those derived from the asymptotic expressions. Despite the substantial efficiency gains, very few of the estimated biases in the ALS estimates of b and d were significant and no important determinants of the latter could be established. For b and r we obtained ($\ell = 1/\sqrt{T}\ ASE(\hat{b})$ for b and $1/\sqrt{T}\ ASE(\hat{r})$ for r)

(41) $T\ell(\overline{\mathcal{E}(\hat{b}-b)}) = \underset{(.06)}{.13r\ell} + \underset{(.09)}{.17dr\ell} - \underset{(.01)}{.02bd\ell} - \underset{(.06)}{.07d\ell} - \underset{(.04)}{.07\ell}$ $\quad s = .35$, $R^2 = .50$, $\chi^2_2 = 5.8$

(42) $T\ell(\overline{\mathcal{E}\hat{r}-r}) = \underset{(.01)}{.03b\ell} + \underset{(.15)}{.92d\ell} - \underset{(.14)}{2.10r\ell} - \underset{(.09)}{.55\ell}$ $\quad s = .45$, $R^2 = .96$, $\chi^2_2 = 4.8$

so that (very roughly) we now have $\mathcal{E}(\hat{r}-r) \simeq -\frac{(1+4r-2d)}{2T}$ (compare (35): if $d = 0$, $T = 50$ and $r = \frac{1}{2}$, (35) yields a bias of $-.02$ whereas (42) yields $-.03$). Table V records the results of the regressions for sD and ESE of the form (32). As in the static case, the asymptotic result is very accurate; none of the coefficients of ASE differ significantly from unity, $\frac{1}{T}$ adds little to the explanation and R^2 is very high for an experimental context.

TABLE V

Regression Results for sD and ESE.

	sD				ESE			
Parameter	δ_1	δ_2	s	R^2/χ^2_2	δ_1	δ_2	s	R^2/χ^2_2
b	1.04 (.02)	1.11 (.38)	.08	.96 0.3	1.01 (.02)	.55 (.41)	.08	.94 0.2
d	1.00 (.01)	0.44 (.31)	.08	.99 4.2	1.03 (.01)	1.11 (0.41)	.10	.99 1.6
r	1.03 (.05)	0.89 (0.66)	.14	.83 0.1	1.01 (.04)	.59 (0.47)	.10	.91 0.5

III. The final study reproduced the two series of 40 experiments in H-H section 6.3 applying ALS and AIV to the same data. The 100 replications, direct simulation group allowed us to study the relationship between the inconsistency and the finite sample bias for an iterative estimator (ALS) while the 25 replications, control variable group investigated the efficiency of the CV for AIV and hence the accuracy of the asymptotic approximation. Corresponding to equation (50) and Table 4 in H-H we obtained for ALS biases the results shown in Table VI.

TABLE VI

ALS Inconsistency Regressions

	R^2	s_1	s_2	s_3	s_4	m	χ^2_4
$\hat{d}$	0.76	0.63	0.67	0.95	0.85	6	2.8
$\hat{b}$	0.99	0.19	0.20	0.30	1.29	7	7.7
$\hat{c}$	0.61	0.51	0.52	0.55	0.64	4	2.2
$\hat{r}$	0.76	0.70	0.73	1.10	0.94	6	5.0

1, 2, 3, 4 subscripts on s respectively denote: the final regression (with m regressors; all regressors; only I; all but I.

The coefficient of the inconsistency was never significantly different from unity but the s_3 values reveal it to have much lower "explanatory power" than for OLS and TSLS. This is primarily due to a serious problem of multiple optima (see section G) and in a similar series of 40 experiments where the v_{it} had much larger variances and there were no multiple optima, I and B were substantially more highly correlated.

Let PE denote the plim of the estimated standard error (ESE) then we considered regressions of the form

(43) $\ln ESE_{ij} = \delta_1 \ln PE_{ij} + \delta_2/T_j$ $(i = 1, \ldots, 4, \ j = 1, \ldots, 36)$,

(PE is available from $\sigma_p^2 \phi_p^{-1}$). The results are recorded in Table VII (for $\hat{\sigma}^2$ the first regressor is σ_p^2)

TABLE VII

ALS Regressions of form (43)

	δ_1	δ_2	R^2	χ_4^2
$\hat{d}$	1.00 (0.011)	4.08 (0.68)	0.983	5.9
$\hat{b}$	1.00 (0.004)	2.44 (0.30)	0.996	1.8
$\hat{c}$	0.99 (0.021)	3.24 (0.88)	0.961	3.1
$\hat{r}$	0.99 (0.019)	4.80 (0.94)	0.919	1.1
$\hat{\sigma}^2$	1.00 (0.004)	0.09 (0.14)	0.999	4.7

It is obvious how accurate asymptotic theory is for these "second moments" even in this very complicated case. The detailed results for the four experiments retained for "forecast" tests are shown in Table VIII (based on 25 replications). This also records the outcomes of the CV experiments for AIV and it is noteworthy that this complete set of 40 experiments computing both ALS (controlled via (30)) and AIV (controlled via (25))required only 56 seconds CPU time on the CDC 7600.* The efficiency gains of the CV in determining the bias in AIV are recorded in Table IX and demonstrate that the correlation between the actual distribution of AIV and its asymptotic approximation for samples of 55 or more observations is high (the CV for $\tilde{\sigma}^2$ was simply $(T^{-1}\varepsilon'\varepsilon - \sigma^2)$). We had relatively little success in establishing the determinants of the observed biases in AIV, using regression analysis (again because of the

* Of this, the calculations for both ALS and AIV took 19 seconds, data generation and processing 37 seconds.

multiple optima problem) but summary statistics for these are reported in Table X. Results for regressions like (43) are recorded in Table XI together with those for sD (note that PE = ASE for AIV). Again we see that knowing the asymptotic result almost obviates the need to do the Monte Carlo!

Finally, to interrelate the results for the four estimators which we have investigated we report in Table XII their rankings by Root Mean Squared Error (RMSE) over the 40 experiments by sample size and the magnitude of r. The results are exactly as one would anticipate from theory: AIV is optimal in large samples with considerable autocorrelation; TSLS is optimal for large T, small $|r|$; OLS for small T, small $|r|$; and ALS for small T, large $|r|$. These results obtain despite the multiple optima problem and hence if T is over 50 we would strongly recommend using AIV (computing the grid to check for bimodality). For smaller values of T, ALS appears to perform reasonably well on average even in the presence of simultaneity. Overall, the reduced form R^2's for the two jointly determined equations in our system correspond more closely with the range of values observed for models in first differences rather than in levels (see Table XIII) - as seems appropriate for stationarity. Increasing the variances of the w_{it} by a factor of 10 produces R^2's matching "levels" models and in this state of the world the relative ranking of AIV is higher (see Table XIV) and the efficiency gains are substantially better (see (3) in Table IX).

TABLE VIII

Outcomes in Four Control Variate Experiments

	θ	ALS $\hat{F}_1$	ALS $\hat{B}$**	ALS $\bar{B}$	ALS I	ALS R	AIV $\tilde{F}_1$	AIV $\tilde{B}$*	AIV $\bar{B}$	AIV R
b	-.2	.67(.05)	.03(.06)	0.67(.05)	0.65	.02	.19(.06)	.34(.09)	.35(.06)	-.18
d	-.4	.58(.16)	-.32(.09)	0.85(.09)	1.12	.58	.38(.19)	.93(.13)	.90(.09)	-.47
c	1.	-.56(.13)	.34(.18)	-0.54(.18)	-0.85	.41	-.22(.17)	-.27(.31)	-.62(.22)	.46
r	.85	-.59(.18)	.30(.10)	-0.94(.11)	-1.17	.79	-.52(.19)	-.95(.13)	-.92(.12)	-.16
σ^2	4.	-1.46(.10)	-	-1.69(.18)	-1.55		-	-1.19(.27)	-	
b	-.8	.03(.03)	.00(.003)	.04(.010)	.03	.96	-.03(.03)	.00(.006)	.01(.013)	.94
d	.4	.02(.08)	.00(.005)	-.01(.015)	-.02	.92	.05(.10)	.00(.004)	.00(.017)	.93
c	1.	.01(.07)	.00(.016)	-.04(.050)	.00	.91	-.01(.09)	.00(.016)	-.04(.050)	.93
r	.55	-.02(.09)	-.03(.015)	.00(.024)	.01	.79	-.09(.10)	-.03(.015)	-.01(.024)	.82
σ^2	.25	.01(.05)	-	.01(.010)	.00		-	.00(.004)	-	
b	-.8	.45(.03)	-.02(.02)	.44(.03)	.45	.80	.06(.03)	.04(.02)	.09(.04)	.94
d	-.4	.16(.08)	.10(.05)	.12(.05)	.08	.57	.13(.10)	.07(.03)	-.04(.05)	.78
c	1.	-.10(.07)	-.02(.06)	-.12(.09)	-.13	.85	-.04(.09)	-.04(.05)	-.04(.09)	.89
r	0.	-.10(.09)	-.12(.05)	-.08(.05)	-.02	.61	-.12(.10)	-.09(.03)	-.05(.05)	.83
σ^2	4.	-.49(.05)	-	-.63(.12)	-.44		-	.09(.[illegible]1)	-	
b	.8	.39(.02)	.03(.01)	.31(.02)	.32	.80	.02(.03)	.00(.01)	-.02(.02)	.93
d	-.7	.06(.07)	.01(.01)	.04(.01)	.05	.79	-.02(.0[illegible])	.02(.01)	.00(.01)	.61
c	1.	-.13(.06)	-.03(.04)	-.09(.05)	-.10	.65	-.02(.0[illegible])	-.02(.02)	.02(.05)	.88
r	-.85	.12(.08)	.01(.01)	.10(.01)	.08	.94	.09(.0[illegible])	-.01(.02)	.03(.02)	.76
σ^2	4.	-1.10(.05)	-	-1.14(.08)	-1.08		-	0.15(.12)	-	

Legend as Table I; B** denotes deviation of the bias from the inconsistency (I); θ the true value of the parameter. The sample sizes are (15, 55, 55, 75). R denotes the correlation between the estimator and its CV.

TABLE IX

Efficiency Gains from AIV Control Variables.

		N	$\tilde{b}$	$\tilde{d}$	$\tilde{c}$	$\tilde{r}$	$\tilde{\sigma}^2$
(1)	$T \leq 35$	7	1.9	1.8	2.3	2.0	9.8
(2)	$T < 35$	13	2.8	2.7	2.7	2.1	10.6
(3)	$T \leq 35$	20	5.4	4.2	3.4	2.9	—
(1)	$T \geq 55$	4	4.0	2.9	3.7	2.6	7.4
(2)	$T \geq 55$	16	6.5	6.1	7.8	6.9	16.2
(3)	$T \geq 55$	20	11.5	9.2	10.4	7.3	—

(1),(2),(3) denote experiments with and without asymptotic multiple optima in w.. and the overall for 10w respectively, and N is the number of experiments in that category.

TABLE X

AIV Bias Regressions Summary

	DS			CV			
	R^2	s	χ^2_4	R^2	s	χ^2_4	
$\tilde{d}$	.46	.71	6.8	.45	.73	9.3	d, r
$\tilde{b}$	.75	.22	10.3	.78	.22	8.6	σ_{11}, br
$\tilde{c}$	.48	.69	1.2	.32	.64	0.1	d
$\tilde{r}$	.51	.70	10.7	.51	.72	7.2	b, r

DS denotes direct simulation based on 100 replications; CV was based on 25 replications. Reported statistics are for regression of $\sqrt{T}B^*$ on $(b, d, r, \sigma_{12}, \sigma_{11}, T, 1, dr, bd, br)/\sqrt{T}$ and those variables with significant values are recorded in the final column.

TABLE XI

AIV Regressions of form (43) for ESE and sD

	ESE				sD			
	δ_1	δ_2	R^2	χ^2_4	δ_1	δ_2	R^2	χ^2_4
$\tilde{d}$	1.00(.019)	2.97(1.12)	.955	0.8	0.96(.044)	-0.29(2.59)	.755	1.4
$\tilde{b}$	0.99(.006)	0.51(0.41)	.993	6.2	1.00(.012)	2.71(0.77)	.978	10.3
$\tilde{c}$	1.00(.014)	2.94(0.56)	.983	22.0	1.05(.026)	7.45(1.01)	.961	10.7
$\tilde{r}$	0.99(.021)	3.49(o.98)	.918	14.1	0.95(.057)	3.68(2.65)	.488	9.5

TABLE XII

Overall Ranking of Estimators by Ranking RMSE's in the 40 experiments.

	T ≤ 35					T ≥ 55				
	N	OLS	TSLS	ALS	AIV	N	OLS	TSLS	ALS	AIV
r ≥ .5	8	2½	4	1	2½	6	3	4	2	1
|r| <.5	6	1	2	3	4	6	3	1	4	2
r ≤ -.5	6	3	4	2	1	8	4	3	2	1

TABLE XIII

Reduced Form R^2's

	ω_{ii}		$10\omega_{ii}$	
	$R^2 > .95$	$R^2 < .75$	$R^2 > .95$	$R^2 < .75$
y	3	22	27	3
Y	1	24	33	0

TABLE XIV

Ranking by RMSE in the 40 Experiments with $10\omega_{ii}$

	$T \leq 35$				$T \geq 55$			
	OLS	TSLS	ALS	AIV	OLS	TSLS	ALS	AIV
$r \geq .5$	4	3	2	1	4	3	2	1
$\|r\| < .5$	2	1	4	3	3	1	4	2
$r \leq -.5$	3	4	2	1	3	4	2	1

G. Multiple Optima Problems

In an earlier study of ALS (see Hendry and Trivedi (1972)) the first author had encountered a considerable number of cases of multiple optima to the likelihood function in finite sample experiments. In particular, repeating one experiment of 40 trials from 3 different initial values (a consistent estimator of r, $\hat{r} = 0$ and $\hat{r} = r$) revealed 25 instances of convergence to totally different final values, at all of which the asymptotic variance matrix was positive definite. Further, $\hat{\sigma}^2$ was not necessarily smallest at the solutions obtained when commencing from $\hat{r} = r$. Clearly, statistics such as "Bias" or "Mean Square Error" as used above are unhelpful when the frequency distribution of the estimor is bimodal.

In the present study we computed the asymptotic function values plim $T^{-1}g$ on a grid over r and observed 11 experiments out of 40 (in set III) where this had two optima for AIV and of these, 9 also revealed 2 optima for ALS. The CV's performed rather badly in these experiments (see Table IX), and both estimators manifested very large standard deviations (quite unrelated to the asymptotic standard errors) suggesting a bimodal frequency distribution.

This problem appears to be very closely related to the "identification" of the parameters of dynamic models with autoregressive errors as follows. Let

(44) $$y_t = dy_{t-1} + u_t \qquad \text{where } u_t = ru_{t-1} + \varepsilon_t \text{ so that}$$

(45) $$y_t = (d+r)y_{t-1} - rd\, y_{t-2} + \varepsilon_t.$$

From such an equation d and r cannot be uniquely identified, and therefore $\Sigma\, \varepsilon_t^2$ has the same value at the two solutions (d, r) and (r, d). Thus g is bimodal as a function of r if $d \neq r$. Introducing an exogenous variable w_t removes the identification problem but need not remove the bimodality. Thus the transformed equation becomes

$$(46) \qquad y_t = (d+r)y_{t-1} - rd\, y_{t-2} + cw_t - rcw_{t-1} + \varepsilon_t$$

but (as is clear from continuity) for "small" Var(w), g remains bimodal. Asymptotically, the global minimum is certainly at (d, c, r) rather than (r, c*, d) but in finite samples, sampling fluctuations can lead to the latter yielding the minimum. To test this explanation, we increased the variances of all the w_{it} by a factor of 10 and reran the experiments using identical random numbers. The reduced form R^2's are reported in Table XIII. Now only one instance of asymptotic bimodality was found (experiment 7 for ALS - this being one of the four outcomes we were attempting to predict!) Further, the Hendry-Trivedi findings were also of this form in that the estimates of r in the two solutions corresponded roughly to r and to the coefficient of the one period lagged dependent variable.

We conclude that it is important to graph g to check for multiple optima and if such are located, to treat the estimates with caution even if they do correspond to the sample global optimum. The ALS, AIV results in section F correspond to "averaging" over states of the world in which this problem appeared to occur about 25 % of the time and (obviously) in unimodal situations their distributions would be much closer to the asymptotic approximations (although in 4 of the 11 cases AIV still had the smallest Mean Square error of the four estimators considered). This is borne out by Table XIV.

APPENDIX

The required population moments additional to those recorded in Appendix I of Hendry and Harrison (1974) are as follows: (Z_t denotes w_{it}).

(A1) $E(Z_t\ Y_{t-1}) = \Lambda_1 E(Z_t\ Y_t)$ (A2) $E(Y_t\ Z_{t-1}) = aE(y_t\ Z_{t-1})$

(A3) $E(y_t\ Z'_{t-1}) = \frac{1}{1-ab}(c'\Lambda_1 M_z + dE(y_t\ Z_t))$ (A4) $E(Z_t\ y_{t-2}) = \Lambda_1 E(Z_t\ y_{t-1})$

(A5) $E(Y_t\ Y_{t-1}) = \frac{1}{1-ab}(ac'\Lambda_1 E(Y_t\ Z_t) + adE(y_t\ Y_t) + f'\Lambda E(Y_t\ w_t) + arE(Y_t\ u_t))$

(A6) $E(y_t\ Y_t) = aE(y_t^2) + f'E(y_t\ w_t) + (b\sigma_{22}+\sigma_{12})/(1-ab)$

(A7) $E(y_t\ y_{t-1}) = bE(Y_t\ y_{t-1}) + c'E(Z_t\ y_{t-1}) + dE(y_t^2) + rE(y_t\ u_t)$

(A8) $E(y_t\ Y_{t-1}) = bE(Y_t\ Y_{t-1}) + c'E(Z_t\ Y_{t-1}) + dE(y_t\ Y_t) + rE(Y_t\ u_t)$

(A9) $E(Y_t\ y_{t-2}) = \frac{1}{1-ab}(ac'\Lambda_1^2 E(Z_t\ y_t) + adE(y_t\ y_{t-1}) + f'\Lambda^2 E(y_t\ w_t) + ar^2 E(y_t\ u_t))$

(A10) $E(y_t\ y_{t-2}) = bE(Y_t\ y_{t-2}) + c'\Lambda_1^2 E(Z_t\ y_t) + dE(y_t\ y_{t-1}) + r^2 E(y_t\ u_t)$

REFERENCES

Amemiya, T. (1966) "Specification Analysis in the Estimation of a Simultaneous Equations Model with Autoregressive Residuals", Econometrica, 34, 283-306.

Box, G.E.P. and Jenkins G.M. (1970) Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.

Cochran, W.G. and Cox, G.M. (1957) Experimental Designs, John Wiley & Sons, New York.

Cooper, R.L. (1972) "The Predictive Performance of Quarterly Econometric Models of the United States". p.813-947 in Hickman (1972).

Dixon, L.C.W. (1972) Non-Linear Optimization, English Universities Press, London.

Goldfeld, S.M. and Quandt, R.E. (1972) Nonlinear Methods in Econometrics, North-Holland.

Hammersley, M.J. and Handscomb, D.C. (1964) Monte Carlo Methods, Methuen.

Harrison, R.W. and Hendry, D.F. (1974) "NAIVE: Numerical Analysis of Instrumental Variable Estimators - Users' Manual", mimeographed London School of Economics.

Hendry, D.F. (1974a) "Stochastic Specification in an Aggregate Demand Model of the United Kingdom", Econometrica, 42, 559-578.

Hendry, D.F. (1974b) "The Structure of Simultaneous Equations Estimators", Invited paper Grenoble Meeting of the Econometric Society (forthcoming in Journal of Econometrics).

Hendry, D.F. (1975) "The Consequences of Mis-Specification of Dynamic Structure Autocorrelation and Simultaneity in a Simple Model with an Application to the Demand for Imports", Chapter 11 in Modelling the Economy, Ed. A. Renton, Heinemann, London.

Hendry, D.F. and Trivedi, P.K. (1971) "Maximum Likelihood Estimation of Difference Equations with Moving Average Errors: A Simulation Study", Review of Economic Studies, 39, 117-136.

Hendry, D.F. and Harrison, R.W. (1974) "Monte Carlo Methodology and the Finite Sample Behaviour of Ordinary and Two-Stage Least Squares", Journal of Econometrics, 2, 151-174.

Hickman, B.G. (1972) Econometric Models of Cyclical Behaviour, National Bureau of Economic Research, Studies in Income and Wealth No. 36, Columbia University Press.

Kakwani, N.C. (1967) "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimator", Journal of the American Statistical Association, 62, 141-142.

Malinvaud, E. (1966) Statistical Methods of Econometrics, North-Holland, Amsterdam.

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

Nicholls, D.F., Pagan, A.R. and Terrell, R.D., (1975) "The Estimation and Use of Models with Moving Average Disturbance Terms: A Survey", International Economic Review, 16, 113-134.

Naylor, T.H. (1971) Computer Simulation Experiments with Models of Economic Systems, Wiley.

Sargan, J.D. (1959) "The Estimation of Relationships with Auto-Correlated Residuals by the Use of Instrumental Variables", Journal of the Royal Statistical Society, B, 21, 91-105.

Sargan, J.D. (1964) "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", Colston Papers, 16, Butterworths Scientific Publications.

Sargan, J.D. (1972) "The Identification and Estimation of Sets of Simultaneous Stochastic Equations", mimeographed, London School of Economics.

Sargan, J.D. (1974) "The Validity of Nagar's Expansion for the Moments of Econometric Estimators", Econometrica, 42, 169-176.

Sargan, J.D. (1975) "Monte Carlo Studies when Moments are Infinite", mimeographed, London School of Economics.

Shenton, L.R. and Johnson, W.L. (1965) "Moments of a Serial Correlation Coefficient", Journal of the Royal Statistical Society, B, 308-320.