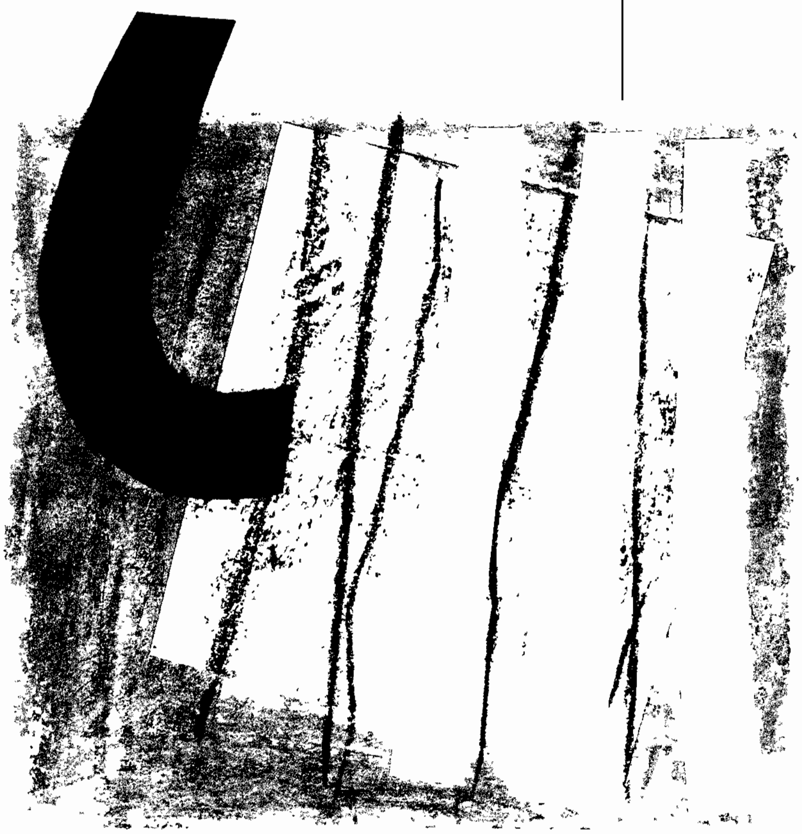
A COMPARISON OF DIFFERENT METHODS OF ESTIMATING AND TESTING THE SPECIFICATION OF RATIONAL EXPECTATION MODELS WITH ONE ENDOGENOUS AND ONE EXOGENOUS VARIABLE

John Denis Sargan

92 - 24



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 92-24
May 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax: (341) 624-9849

A COMPARISON OF DIFFERENT METHODS OF ESTIMATING AND TESTING THE SPECIFICATION OF RATIONAL EXPECTATION MODELS WITH ONE ENDOGENOUS AND ONE EXOGENOUS VARIABLE

John Denis Sargan*

Abstract

This article considers the theory of the estimation and testing of a model with one endogenous variable and one exogenous variable, where the structure of the model assumes a simple rational expectations hypothesis for the determination of the endogenous variable.

Two methods of estimation are considered, the first the method of Maximum Likelihood, and the second the method of Instrumental Variables. The first is asymptotically efficient, the second may be relatively less asymptotically efficient. The first also has the advantage of suggesting suitable tests for the general form of the rational expectations model.

Key words:
Instrumental variables, maximum likelihood estimation.

*Emeritus Professor, London School of Economics. This paper was written while I was at the Statistics and Econometrics Department of the Universidad Carlos III de Madrid as a visiting professor. The visit was sponsored by the Fundación Universidad Carlos III.

1. Introduction.

This article considers the theory of the estimation and testing of a model with one endogenous variable and one exogenous variable, where the structure of the model assumes a simple rational expectations hypothesis for the determination of the endogenous variable. The model used here assumes that a set of entrepreneurs are determining their actions by minimising expected costs where for simp[licity costs are approximated by a quadratic function of the variables. Such models have been considered by ,for example, Muellbauer and Winter(1980)

The theory of such models is slightly simplified by considering the special case where there is only one exogenous variable since it is then not necessary to consider the theory of matrix polynomials. Two methods of estimation are considered , the first the method of Maximum Likelihood, and the second the method of Instrumental Variables. The first is asymptotically efficient , the second may be relatively less asymptotically efficient. The first also has the advantage of suggesting suitable tests for the general form of the rational expectations model.

2. The Model Formulation.

z_t is used to denote the exogenous variable and y_t is used to denote the endogenous variable. The exogenous variable is assumed to be generated by an autoregressive equation of the form

$$z_t = \sum_{i=1}^{p} \varphi_i \; z_{(t-i)} + v_t \qquad (2.1)$$

The equation determining y_t is then of the form

$$y_t = b_1 \; y_{(t-1)} + b*_1 \; E[y_{(t+1)} | t] + c_0 \; z_t + c*_1 E[z_{(t+1)} | t] + u_t \quad (2.2)$$

.

This is an equation of the type derived in Appendix A from a minimising model. Using the arguments of my paper [(1984)], it can be shown that the y_t satisfy an equation of the form

$$y_t = \lambda_1 y_{(t-1)} + \sum_{i=1}^{p} g_i \; z_{(t+1-i)} + u*_t \qquad (2.3).$$

where the λ_i and g_i are determined by the following equations, and $u*_t = (1 + \lambda_1 \lambda_2) u_t$.

λ_1 and $1/\lambda_2$ are the two roots of the quadratic equation

$$b*_1 x^2 - x + b_1 = 0 \qquad (2.4)$$

Both λ_1 and λ_2 should be real and of modulus less than one. This requires that quadratic equation (2.4) should have two real roots ,one with modulus less than one and the other with modulus greater than one, and this in turn ensures that

λ_1 and λ_2 are both unique continuous functions of the parameters of the original model. Conversely b^*_1 and b_1 are defined as functions of λ_1 and λ_2 by the equations

$$b_1 = \lambda_1 / (1 + \lambda_1\lambda_2), \quad b^*_1 = \lambda_2 / (1 + \lambda_1\lambda_2), \tag{2.5}$$

Then it is convenient to define $h = 1 + \lambda_1\lambda_2$ and

$$\varphi(\lambda_2) = \sum_{j=1}^{p} \varphi_j \lambda_2^{j-1}$$

so that

$$g_1 = (c_0 + \varphi(\lambda_2))/(1 - \lambda_2\varphi(\lambda_2)) \tag{2.6}$$

and defining

$$d = c^*_1 + b^*_1 g_1 \tag{2.7}$$

$$g_k = dh \sum_{s=k}^{p} \lambda_2^{s-k}\varphi_s \tag{2.8}$$

In estimating this model by maximum likelihood it is convenient to write x_t for the vector of variables whose elements are y_{t-1} and z_{t-i}, $i = 0$ to $p-1$, in that order. Define the vector ψ to have elements λ_1 and g_i, $i = 1$ to p, so that the equation (2.3) can be written

$$y_t = x_t'\psi + u_t \tag{2.9}$$

and the equations (2.4),(2.6),(2.7) and (2.8) can be summarised as equivalent to the statement that the elements of the vector ψ are functions of the vector θ, whose elements are the parameters b_1, b_1^*, c_0, and c_1^* respectively. Note that ψ depends also on the parameters φ, so that we can write

$\psi = \psi(\theta,\varphi)$.

If φ were known then asymptotically efficient estimates of θ would be obtained by estimating equation (2.9) by non-linear least squares. With φ unknown it would be necessary to first estimate φ by least squares and then to estimate ψ from equation (2.9) by non-linear least squares. These are not asymptotically efficient estimates and the standard errors of the estimates of θ must allow for the extra error caused by having to estimate φ. Alternatively we can obtain efficient estimates of both θ and φ by maximising a suitable likelihood function with respect to both sets of parameters simultaneously. This was discussed in an earlier paper (Sargan,1984), and this method will not be discussed in this paper.It is convenient to have mnemonics for all the different methods of estimation of this paper and the method of non-linear least squares when φ is assumed to be known will be denoted by NLS and if φ is assumed to be estimated will be denoted by FNLS.

These estimation methods can be compared with various methods of Instrumental Variable estimation.The equation to be estimated must first be converted to the form

$$y_t = b_1 y_{t-1} + b_1^* y_{t+1} + c_0 z_t + c_1 z_{t+1} + u_t - b_1^* u_{t+1}^* - (b_1^* g_1 + c_1^*) v_{t+1} \quad (2.10)$$

since $y_{t+1} = E(y_{t+1}|t) + b_1^*(u^*_{t+1} + g_1 v_{t+1})$

and $z_{t+1} = E(z_{t+1}|t) + v_{t+1}$.

Equation 2.10 can be estimated by IV in several ways . First consider the case where φ_i are known a priori. In this case v_{t+1} is an observable variable and can be used as an IV. Since optimal predictors of all the variables in 2.10 can be expressed as linear combinations of y_{t-1} and of z_{t-j}, $j = 0$ to $p-1$, (except for v_{t+1}, which is discussed below) the set of IV listed above is the set of Instrumental variables which will be discussed first and will be denoted in the subsequent theory by the $p+1 \times T$ matrix Z. The corresponding IV estimators will be referred to as simple IV estimators. Note that all these instrumental variables are uncorrelated with v_{t+1}, so that its coefficient cannot be estimated consistently by this set of IV but the term in v_{t+1} is included in the overall error in the equation. The error on equation (2.10) is of moving average form, but the errors on the prediction equations (2.3) and (2.4) are serially independent, so that there is no need to introduce serially transformed instrumental variable estimators. A direct proof of the efficiency of IV estimators of the Sargan type (Sargan 1988b), where the variables in the equation are transformed but not the Instrumental Variables will be given below.

These simple IV estimators can be modified in several ways. If the coefficient of v_{t+1} is denoted by d in equation 2.10, then it improves the efficiency of the IV estimates to include v_{t+1} in the equation while retaining the constraint

$$d = -(b_1^* g_1 + c_1^*) \qquad (2.11).$$

This leads to non-linear IV estimators, which will be denoted by constrained or CIV estimators. These will be shown to be fully efficient asymptotically, provided that the equation (2.11) is suitably transformed by an inverse MA transformation so that the error on the transformed equation is u_t and so is a white noise error(this type of estimator will be denoted by SCIV). Simpler computations are obtained by ignoring the constraint 2.11, but adding z_{t+1} to the set of IV. These will be denoted extended or SEIV estimates. These are as efficient as SCIV and also allow an asymptotically powerful test for the constraint 2.11 which provides one good test of the rational expectations model.

Unfortunately these results are of only theoretical interest since the φ_i are not known and must be estimated by OLS. If for v_{t+1} is substituted its OLS estimator the efficiency of estimation is reduced . Denoting these feasible estimators by FCIV and FEIV it will be shown that both are equally efficient, with an efficiency equal to that of the FNLS estimators.

From these estimators tests for the restrictions implied by the rational expectations model will be derived.

3.A Comparison of NLS and CIV estimators.

Writing y for the vector wirh elements y_t, t=1 to T,and Z for the T×(p+1) matrix defined in the previous section, whose row vectors are the x_t defined by equation 2.9,and u* for the vector of errors on that equation,and considering u* as a function of the parameters θ the first order conditions for the NLS estimators obtained by minimising u*'u* as a function of θ is

$$\partial\psi'/\partial\theta(Z'u^*) = 0. \qquad (3.0)$$

which when divided by $(1 + \lambda_1\lambda_2)$ gives

$$\partial\psi'/\partial\theta(Z'u) = 0)$$

Now if **w** is the vector whose elements are $u_t - \lambda_2 u_{t+1}$, this is the vector of errors on equation (2.10), and if Λ is the TxT matrix

$$\begin{pmatrix} 1 & -\lambda_2 & 0 & 0 & 0 \ldots\ldots\ldots\ldots\ldots & 0 \\ 0 & 1 & -\lambda_2 & 0 & 0 \ldots\ldots\ldots\ldots\ldots & 0 \\ 0 & 0 & 1 & -\lambda_2 & 0 \ldots\ldots\ldots\ldots\ldots & 0 \\ 0 & 0 & 0 & 1 & -\lambda_2 \ldots\ldots\ldots\ldots\ldots & 0 \\ \ldots & & & & & \\ \ldots & & & & & \\ 0 & 0 & 0 & 0 & 0 \ldots\ldots\ldots\ldots 1 & -\lambda_2 \end{pmatrix},$$

then $u = \Lambda^{-1}\mathbf{w}$.

Thus 3.0 is equivalent to

$$\partial\psi'/\partial\theta(Z'\Lambda^{-1}\mathbf{w}) = 0 \qquad (3.1)$$

or writing X for the Tx5 data matrix with elements $(y_{t-1}, y_{t+1}, z_t, z_{t+1}, v_{t+1})$, and $\mathbf{y^*} = \Lambda^{-1}\mathbf{y}$, $\mathbf{X^*} = \Lambda^{-1}\mathbf{X}$, $\theta^* = (\ b_1, b_1^*, c_0, c_1^*, d^*)$, where $d^* = -(b_1^* g_1 + c_1^*)$.

Equation 3.1 is then

$$\partial\psi'/\partial\theta(Z'X^*\theta^* - Z'y^*) = 0.$$

These are similar in form to the nonlinear IV estimators of equation 2.10 which would be obtained by minimising

$$\mathbf{w}'\Lambda'^{-1}[Z(Z'Z)^{-1}Z']\Lambda^{-1}\mathbf{w},$$

or

$$(\theta^{*\prime}X^{*\prime}Z - y^{*\prime}Z)(Z'Z)^{-1}(Z'X^*\theta^* - Z'y)$$

with first order conditions

$$\partial(\theta^{*\prime}X^{*\prime}Z - y^{*\prime}Z)/\partial\theta\ (Z'Z)^{-1}(Z'u) = 0 \quad (3.2).$$

Now considering for all values of θ the identity

$$Z'\Lambda^{-1}(X\psi^* - y) = Z'u = Z'u^*/(1 + \lambda_1\lambda_2).$$

Considered as functions of θ with the observed variables y,X and Z treated as constants these identities can be differentiated to give

$$\partial(Z'\Lambda^{-1}(X\psi^* - y))/\partial\theta' = \partial(Z'u^*/(1 + \lambda_1\lambda_2))/\partial\theta'$$

$$= (Z'u^*)\partial(1/(1 + \lambda_1\lambda_2))/\partial\theta' + (1/(1+ \lambda_1\lambda_2))\partial(Z'u^*)/\partial\theta'$$

$$= \quad \text{ditto} \quad + (1/(1 + \lambda_1\lambda_2))Z'Z(\partial\psi/\partial\theta'). (3.3)$$

Dividing this equation by T and taking plims on each side, since $\text{Plim}(Z'u^*/T) = 0$. Then

$$\text{Plim}\{\partial(\theta^{*\prime}X^{*\prime}Z - y^{*\prime}Z)/\partial\theta\ (Z'Z)^{-1}\partial(Z'X^*\theta^* - Z'y^*)/\partial\theta'/T\} =$$

$$(1+\lambda_1\lambda_2)^{-2}\text{Plim}\ (\partial\psi'/\partial\theta\ (Z'Z)\ \partial\psi/\partial\theta'/T).$$

The white noise error on equation 2.10 is u_t with standard deviation σ, say, and the error on equation (2.3) is u^*_t with standard deviation $\sigma^* = \sigma(1 + \lambda_1\lambda_2)$. Then

$$\sigma^2\text{Plim}\{\partial(\theta^{*\prime}X^{*\prime}Z - y^{*\prime}Z)/\partial\theta\ (Z'Z)^{-1}\partial(Z'X^*\theta^* - Z'y^*)/\partial\theta'/T\}^{-1}$$

which is the AVM of the SCIV estimator and is equal to

$$\sigma^{*2}\text{Plim}\{\partial\psi'/\partial\theta\ (Z'Z)\ \partial\psi/\partial\theta'\}^{-1},$$

which is the AVM of the NLS estimator.

An alternative proof , using the methods of Sargan(1988b) p.102, can show from equation 3.3 that the difference between the two estimators is O(1/T), again meaning that both estimators are asymptotically equivalent.

4. A Comparison with the SEIV Estimator.

If the constraint 2.11 is no longer used, and d is estimated by IV the number of unknown parameters is now 5, and the number of instrumental variables must then clearly be at least 5. If $p>3$ or z_{t-i} and y_{t-j} with $i>p$ or $j>1$ are used as instrumental variables, these have the property of being asymptotically independent of v_{t+1}, so that the coefficient of v_{t+1} cannot be identified by using this set of IV. One simple way of dealing with this problem is then to treat the term in v_{t+1} as an addition to the random error on the equation. Since all the IV are independent of the new combined error it is possible to estimate the equation 2.10 omitting the variable v_{t+1} from the equation. This leads to a simple linear IV estimator which will be labelled the SLIV estimator. On the other hand when φ_i are assumed known then v_{t+1} is observable and can be used as an IV (giving the SEIV estimator). This then leads to an estimate of d that allows the constraint 2.11 to be tested.

The SLIV estimator can be defined by writing X_1 for the T×4 matrix obtained by omitting the last column of X and $X_1{*}$ for the corresponding serially transformed variables, and writing $\hat{\theta}\dagger_1$ for the SLIV estimator

$$X_1{*}'Z(Z'Z)^{-1}Z'X_1{*}\hat{\theta}\dagger_1 = X_1{*}'Z(Z'Z)^{-1}Z'y{*} \qquad (4.1)$$

This can alternatively be written

$$[(X_1{*}'Z(Z'Z)^{-1}Z'X_1{*})/T]T^{½}(\hat{\theta}\dagger_1 - \theta)=X_1{*}'Z(Z'Z)^{-1}Z'(u+d\ v_1{*})T^{-½} \qquad (4.2)$$

where $v_1{*}$ represents the vector of serially transformed elements equal to v_{t+1}. Note that all the factors have been written so that they are of order one. Then $T^{-½}(Z'u)$ and

$T^{-½}(Z'v_1*)$ are asymptotically independent since the series u_t and the series v_{t+1} are completely independent stochastically. It follows that the AVM of

$$T^{-½}(Z'u) + d\,[T^{-½}(Z'v_1*)]$$

can be written

$$Plim\,(Z'Z/T)\sigma^2 + d^2\,Plim\,(Z'\Lambda'^{-1}\Lambda^{-1}Z/T)\sigma_v{}^2.$$

The second term can be simplified since $\Lambda\Lambda' = \Omega$, where Ω is the variance matrix of the first order MA stochastic process with moving average coefficient $-\lambda_2$. Then using Cramer's general linear transformation theorem and defining the following symbols ; $V = [Plim\ X_1*'Z(Z'Z)^{-1}Z'X_1*/T]$, $Q = Plim(X_1*'Z(Z'Z)^{-1})$, $B = Plim(Z'\Omega^{-1}Z/T)$; the variance of the SLIV estimators can be written

$$\sigma^2 V^{-1} + d^2\sigma_v{}^2\ V^{-1}QBQ'V^{-1} \quad . . . (4.3)$$

It is easily seen that the first term here is the AVM of the SCIV estimator when account is taken that $Plim(Z'v_1*)/T) = 0$, and the second term represents the loss of efficiency from treating the effect of v_{t+1} as a addition to the error term on the equation 2.10 rather then including it as a variable in the equation. However this comparison is not very interesting since normally the φ_i are not known so that it is necessary to consider feasible estimators such as FEIV or FNLS estimators.

Consider for example the FNLS estimators equation 3.1 defines the estimators but now in the next transformation in defining the set of variables in X, v_{t+1} is replaced by v^+_{t+1}, where this denotes the OLS estimator of v_{t+1}, and using v_1 and v^+

to denote the corresponding T×1 vectors and Z_1 for the T×p matrix with elements z_{t-i}, t = 1,...,T, i= 0,...,p-1,

$$v^{\dagger} = v_1 - Z_1(Z_1'Z_1)^{-1}Z_1'v_1$$

Then the equivalent of the equation following (3.1) is

$$\partial\psi'/\partial\theta[(Z'X^*\theta^* - Z'y^*) - d^*\{(Z'\Lambda^{-1}Z_1)(Z_1'Z_1)^{-1}(Z_1'v_1)\}]=0$$

and combining the arguments following equation 3.1 with the arguments of the last section it follows that the AVM of the FNLS estimator can be written

$$\sigma^2V^{-1} + d^2\sigma_v^2V^{-1}QCQ'V^{-1} \quad ... \quad (4.4)$$

where $C = \text{Plim}[(Z'\Lambda^{-1}Z_1)(Z_1'Z_1)^{-1}(Z_1'\Lambda'^{-1}Z)/T]$.

The difference in the two AVMs is

$$d^2\sigma_v^2 \; V^{-1}Q(B-C)Q'V^{-1}$$

where $B - C = \text{Plim}\ \{(Z'\Lambda^{-1})[I - Z_1(Z_1'Z_1)^{-1}Z_1'](\Lambda'^{-1}Z)/T\}$.

Since the matrix in square brackets above is an idempotent matrix of rank one B-C is always non-negative definite. This shows that FNLS estimators, in general, are more efficient than the SLIV estimators.

Now consider the simpler IV estimators where the equations are not linearly transformed to obtain a serially independent error.The estimators where no attempt is made to restrict d will be written $\theta^{\dagger}_2$ and satisfies the equation

$$(X_1'Z)(Z'Z)^{-1}(Z'X_1)\theta^{\dagger}_2 = (X_1'Z)(Z'Z)^{-1}(Z'y) \quad . (4.5).$$

This simple linear estimator will be denoted by LIV. Finally a more efficient untransformed IV estimator may be obtained by taking estimators where $v^{\dagger}_{t+1}$ is included in the set of variables in the equation and z_{t+1} is included in the set of

IV, but no serial transformation is carried out. This will be denoted the EIV estimator.

The well known inequalities for IV estimators with serially correlated errors shows that the LIV estimator is worse than the SLIV, and that the EIV estimator is worse than the SEIV.

To summarise this section the order of asymptotic efficiency for these various estimators is as follows:- NLS would be fully efficient if the φ_i were known. Among the feasible estimators FNLS,FCIV,FEIV are all equally efficient, SLIV and EIV are less efficient, and LIV is least efficient.

2. Testing the Model.

One method of testing this model depends upon the comparison of the estimation of the NLS estimates of equation 2.3 with the corresponding unconstrained equation of this form. In fact a simple test of these constraints, depending on the difference btween the constrained and unconstrained estimates of the sum of squares of the errors u_t is not valid since the constrained estimate depends on the estimated φ_i. If the constraints on the g_i were capable of explicit formulation then it would not be difficult to compute an appropriate Wald test provided that the influence of the estimated φ_i was taken into account in computing the AVM of the constraints. Explicit constraints are only available when p=3, and if p>3 some approximation technique for the constraints would yield an approximation to the Wald test.

A simpler test for mispecification would test that z_{t-i} with i>p-1 have zero coefficients in equation 2.3 by using an appropriate F-ratio test. A more specific test would test the validity of the constraint 2.11 when equation 2.10 has been estimated by an estimation procedure such as SEIV which allows unconstrained estimates of d to be made. After some algebraic manipulation it can be shown that this test is equivalent to testing that u^*_t is uncorrelated with v_{t+1}, and that with this model it is permissible to replace these errors with their estimated values from the NLS estimates of equation 2.11 and the OLS estimates of equation 2.1. Denoting these estimates in vector form by u† and v†₁ then the criterion

$$t_d = T^{-\frac{1}{2}}(u^{\dagger\prime}v^{\dagger}_1)/s_u s_v$$

is asymptotically distributed as a t-ratio on the null hypothesis,where s_u and s_v are the usual estimates of the standard deviations of u^*_t and v_t.

Finally it is possible to test the restriction that the MA coefficient in the equation 2.11 is equal to $-\lambda_2$. (This follows since $b_1^*u_{t+1}^* = \lambda_2 u_{t+1}$, and it is possible to replace the resulting forward moving average representation by an alternative backward moving average representation with the same coefficient. The most asymptotically powerful test against the alternative of a different MA coefficient is obtained by defining a vector of errors

$$u_\Lambda = \Lambda^{-1}u^\dagger - u^\dagger,$$

whose t th element is equal to

$$\Sigma \lambda_2^k u_{t+k+1}$$

The criterion

$$T^{½} (u_t' u_\wedge) / \sqrt{(u_t' u_t)(u_\wedge' u_\wedge)}.$$

is asymptotically distributed as a t-ratio on the null hypothesis, and is asymptoiclly powerful. An alternative criterion which tests for the same alternative hypothesis but is not so powerful is the first order autocorrelation of u_t. A suitable criterion is the Sargan modification of the Durbin test statistic, defined by taking u_{t1} as the vector of elements $u_{t_{t-1}}$.The criterion is then defined by

$$T^{½} (u_{t1}' u_t) / \sqrt{(u_t' u_t)(u_{t1}' u_{t1} - u_{t1}' Z(Z'Z)^{-1} Z' u_{t1})}.$$

This version of the Durbin criterion has the advantage that the expression under the square root sign is always non-negative so that the criterion is always well-defined, and asymptotically is distributed as a t-ratio.

5. A Monte-Carlo Simulation.

In order to consider the finite sample properties of these estimators and test statistics some simple models were simulated. To take advantage of the storage capacities of personal computers with hard discs a special program was written which would store the second moments of the data generated from a model consisting of the two equations 2.1 and 2.10. For greater efficiency it was arranged that the program generated a continuous stream of variables x_t and y_t using a standard quasi-normal deviate generating sub-routine for t = 0 to infinity, and this was cut up into appropriate lengthed samples for which second moments were calculated. In practice it was decided to consider sample of length 20, 50 and 100 observations. In addition it was decided to omit at least 30 observations between each sample so as to minimise the autocorrelation between successive samples.

The stream of data was thus cut up into lengths of 1,040 observation, each of these was cut up into both 8 lengths of 130 observation and 13 lengths of 80 observations.From each length of 130 obsevations one sample of length 100 observations was extracted, and from each length of 80 observations a sample of length 50 and a sample of length 20 observations was extracted.The total number of simulations was chosen by taking 3,846 of the lengths of 1,040 observations. This meant that the total number of replications of sample size 20 and 50 was 13×3,846= 49,998, and the total number of replications of samples of size 100 was 30,768. These proved of adequate size to give sufficient accuracy in the estimation of the empirical frequency distribution functions. In order to save space on the hard disc it was decided to store the moments as rescaled integers, thus requiring only 2 bytes or 32 bits to store each moment.(If they had been stored as single-length floating point real numbers 4 bytes would have been required.) In order to carry out a reasonable truncation the moments were multiplied by 3,000 before being set to the value ±31,500 if the scaled value lay outside the limits

±30,000. This was a crude attempt to give a representative value for the moments lying outside the limits ±10.0, whenever the original moments lay outside these limits. In order that such truncation was only very rarely required it was necessary to provide that if some variable had a second moment which had a statistical expectation greater than 3.0 then this variable was scaled down by an adequate factor.In practice it proved unnecessary to scale down the x_t variable but necessary to scale down the y_t variables in the models which were studied here.

The program was written in a general form suitable for a form of equaton 2.10 with general p and the possibilty of more than one lag on the y_t variable. For the models studied in this paper only one lag on the y_t variable is required for generating the data but the IV estimating procedures require the use of moments involving more than one time lag. Thus the second moments stored were the covariance for any two variables from the following sets of variables; y_{t-i}, i=0,1,2, and x_{t-i}, i=0,...,5. This makes 45 covariances for each sample. The moments were stored as covariances since it was regarded as more appropriate to assume that an unconstrained constant term was included on each equation. The total storage space required for each sample size was 4.3M for sample sizes 20 and 50, and 2.6M for sample size 100.

In this study only the case p = 3 is reported. This is because if p>3 then the NLS estimators require numerical optimisation methods for calculation , whereas when p=3 optimum estimates of ψ are obtained by taking unconstrained OLS estimates of equation 2.3 and optimum estimates of φ by OLS estimates of equation 2.1. Then corresponding estimates of θ are obtained by solving $\psi(\theta,\varphi) = \psi$ for θ. Conversely if θ is estimated by some form of IV then the corresponding estimates of θ are obtained directly from the same equation.

A program was written which read the covariances from the hard disc, used sub-routines to calculate the values of various statistics expressed as functions of these covariances, and then calculated simulation means , variances, and standard deviations of these sample statistics and also the standard errors of these simulation statistics. It also produces empirical distribution functions, recording the proportions of the simulation samples which lie between given limits, these corresponding to given probability limits on the corresponding statistic's asymptotic distribution. This makes possible a direct comparison between the statistic's estimated distribution function and its theoretical asymptotic distribution function. For this study where it is desired to compare the efficiency of various estimators two sets of subroutines were written. The first computed the estimates of ψ and θ using first NLS and then using LIV. This gives 16 different statistics, since each vector has 4 components. The second set of subroutines calculates the EIV estimates of θ, then t-ratios for the NLS estimators of ψ and θ, and t-ratios for the LIV estimators of θ, and finally the two specification test statistics t_d and the Sargan/Durbin test for serial correlation. This gives simulation of a further 18 statistics.

6. A Model and some Results.

It proved a little difficult to choose suitable models for simulation. In order to make it possible to estimate the parameters of the model at all accurately and to be able to discriminate between different forms of the model and to test specification powerfully it is necessary to have coefficients sufficiently large compared with their standard errors of estimation. In particular both φ_3 and g_2 should be relatively large (say. greater than .3 in absolute value) since otherwise it will often be found to give large errors for the estimated θ. But in the case of third order autoregressive equations the last coefficient, being the product of the latent roots of the autoregressive latent roots equation, must be small unless at least one of these roots is large. For example if all the latent roots have moduli less than .7 then $\varphi_3 < .343$. But if the latent roots have large moduli then it is to be expected that the variance of z_t will be large, and especially in the likely case where all the periods of oscillation are large compared with the unit time period and only slightly damped, i.e. the case where all the roots are close to one. In such models the various lagged values of z_{t-i} for different i are highly correlated and the standard errors of the estimated g_i are relatively high. All these characteristics were found in the first model which was simulated, resulting in all methods of estimation being poor and having sample variances much greater than that predicted by asymptotic sampling theory. So the model discussed in this paper was chosen so that all latent roots have modulus about 0.75 or more but not near one. The equation determining z_t is

$$z_t = -.4\ z_{t-1} - .5\ z_{t-2} - .5\ z_{t-3} + v_t \quad \ldots \quad (6.1)$$

where $v_t \sim IIN(0,1)$.

Then the structural equation was chosen so that $\lambda_1 = .5$ and $\lambda_2 = .8$. The corresponding coefficients of the structural equation are: $b_1 = .3571$, and $b_1* = .5714$, $c_0=1.$, and $c_1* =1.$ $u_t* \sim IIN(0,1)$.

From these parameters it follows that the vector ψ has elements $\lambda_1 = .5$, and g_i -.0886,-1.196,-.664. In storing the covariances y_t were scaled down by a factor 3.0. Thus the rescaled $y*_t$ was generated by the equation

$$y*_t = .5\ y*_{t-1} -.0295\ z_t - .399\ z_{t-1} -.222\ z_{t-2} + u**_t \quad (6.2)$$

where the standard deviation of $u**_t$ is .333.

Equations 6.1 and 6.2 generated the moments for storage. In analysing the results it is clear that for the IV estimators and for the NLS estimates of θ no moments exist since the IV estimators are of the just identified type where the number of instrumental variables is equal to the number of estimated coefficients, and the θ estimators are functional tranformations of the direct estimators(see Sargan 1988). Thus when the mean and variances of these statistics were calculated they were found to be very large, and to be increasing proportionally with the size of the simulation sample. So the means and variances are only reecorded here for the NLS estimates of ψ, and the means are in fact recorded in the following table as biases, by subtracting the true values of the coefficients.The figures in brackets are the corresponding standard errors.

Table 6.1.NLS Estimator of ψ. Biases and S.Deviations.

T	B	SD	B	SD	B	SD	B	SD
20	-.052	.170	.019	.246	.028	.246	-.029	.240
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
50	-.019	.096	.009	.139	.011	.138	-.010	.139
	(.000)	(.000)	(.001)	(.000)	(.001)	(.000)	(.001)	(.000)
100	-.010	.066	.004	.095	.006	.094	-.006	.096
	(.000)	(.000)	(.001)	(.000)	(.001)	(.000)	(.001)	(.000)

These biases are not large and although the standard deviations are somewhat above the asymptotic standard errors of the estimators the discrepancy is not large.

Turning now to the other estimators study of their empirical distribution functions shows that the spread of the distributions is larger than might be expected from the asymptotic standard errors. To summarise this compactly is difficult so that the following tables merely records the probabilities of being below a certain limit, denoted by L, and of being above a certain limit,denoted by U. These lower and upper limits are chosen to be the lower and upper limits corresponding to lower and upper tail probabilities of 2½% in the asymptotic distributions of the appropriate NLS estimators.

Table 6.2 Tail Probabilities for Estimators of θ.

T		FNLS		LIV		FEIV	
		L	U	L	U	L	U
20	θ_1	.136	.096	.280	.188	.202	.095
	θ_2	.164	.075	.203	.143	.125	.154
	θ_3	.112	.106	.112	.133	.088	.815
	θ_4	.060	.173	.139	.188	.130	.626
50	θ_1	.037	.090	.202	.103	.148	.058
	θ_2	.134	.019	.180	.060	.083	.130
	θ_3	.093	.042	.101	.063	.039	.920
	θ_4	.010	.155	.054	.183	.035	.824
100	θ_1	.019	.068	.154	.089	.177	.029
	θ_2	.102	.012	.145	.033	.042	.182
	θ_3	.072	.024	.079	.029	.010	.977
	θ_4	.006	.122	.028	.152	.006	.942

No standard errors are quoted but they are all less than .003.

On this criterion it is clear that LIV is rather worse than FNLS and that FEIV has strong biases for θ_3 and θ_4. Comparing the NLS and the FEIV estimators of ψ the NLS estimators have finite moments summarised in table 6.1 whereas the FEIV estimators have infinite moments, and large probailities on the tails, for example for T=100 ψ_1 had lower and upper tail probabilities equal to .380 and .278 respectively. There is no doubt for this model that NLS give better estimates of ψ.

Considering the t-ratio statistics, for all of these the second moments of the statistics exist, and so a summary in terms of the means and standard deviations is given in table 6.3. Note that the t-ratios are all given in the form the estimator divided by its estimated standard error, so that if the true value of the coefficient is non- zero then the asymptotic distributionOf the t-ratio has a non-zero mean. This type of t-ratio was studied to get some indication of the relative powers of the different estimator's t-ratios to reject a non-valid null hypothesis. The t-ratios were calculated for the NLS estimators of ψ, and the FNLS estimators and the LIV estimators of θ.

Table 6.3.t-Ratios for Different Estimators.

T		M	SD	M	SD	M	SD	M	SD
20	NLSψ	3.31	1.63	-.35	1.23	-6.14	2.10	-3.63	1.62
	FNLSθ	2.48	1.68	1.48	1.70	2.38	1.80	1.28	.91
	LIVθ	.51	.57	.50	.70	2.12	2.37	.57	.64
50	NLSψ	5.43	1.52	-.62	1.08	-9.46	1.87	-5.27	1.47
	FNLSθ	4.86	1.85	2.23	1.82	3.84	2.04	2.10	.81
	LIVθ	1.21	.83	.94	.93	4.42	3.54	1.15	.74
100	NLSψ	7.78	1.49	-.93	1.04	-13.34	1.83	-7.31	1.37
	FNLSθ	7.51	1.91	3.19	1.91	5.60	2.18	3.04	.76
	LIVθ	2.04	.98	1.48	1.07	6.78	4.32	1.82	.76

The standard errors of these means and standard deviations are not quoted but are all less than .01.

Note that since these are t-ratios their asymptotic standard deviations should be one , although for non-central t-ratios the standard deviations may be somewhat greater than one. The biases upwards are largest for small samples and high non-centraiities. If the symmetric 95% asymptotic confidence interval is used to accept the null hypothesis that the coefficient is zero , then the probability of accepting an incorrect null hypothesis can be compared for the FNLS and LIV estimators of θ.

Table 6.4.Probability of Accepting the Zero Coefficient Hypothesis.

	θ_1		θ_2		θ_3		θ_4	
T	FNLS	LIV	FNLS	LIV	FNLS	LIV	FNLS	LIV
20	.410	.975	.681	.953	.444	.604	.789	.963
50	.065	.818	.489	.861	.196	.300	.439	.864
100	.005	.506	.282	.705	.059	.119	.073	.603

Clearly the probability of accepting the invalid null hypothesis is greater for the LIV t-ratio than for the FNLS t-ratio for all coefficients and sample sizes, so that the latter is a more powerful test for all cases simulated here.

Finally the distributions of the two specification test statistics seem to be well approximated by their asymptotic distributions. Their means and standard deviations are summarised in table 6.5.

Table 6.5. Means and Standard Deviations for t_d and Serial correlation Test Criterion.

	T = 20		T = 50		T = 100	
	M	SD	M	SD	M	SD
t_d	-.084	1.126	-.032	1.043	-.021	1.016
SD	-.002	1.125	.001	1.039	-.003	1.011

Sargan / Durbin test.

and the probability of being outside the asymptotic 95% confidence interval is given below.

Table 6.6.Tail Probability for the Test Statistics.

	T=20	T=50	T=100
t_d	.082	.060	.054
SD	.081	.059	.052

Clearly both test statistics give tests of the expected size rather accurately even in samples of size 20 for this model.

7.General Conclusions.

Although this paper only reports results for one model these results support the general statement that the greater asymptotic efficiency of NLS estimators of this type of model compared with the efficiency of IV estimators is realised in these models even for sample sizes down to 20 observations. Of course to validate this for a wider range of models requires the study of models with p>3 and possibly with more than one exogenous variables. But programs have now been written which are relatively efficient for studying simultaneously a large number of statistics generated from the same model, which could be used advantageously for further studies.

References.

Lucas, Robert.E. and Thomas J. Sargent (eds)(1981)"Rational Expectations and Econometric Practice". George Allen and Unwin, London.

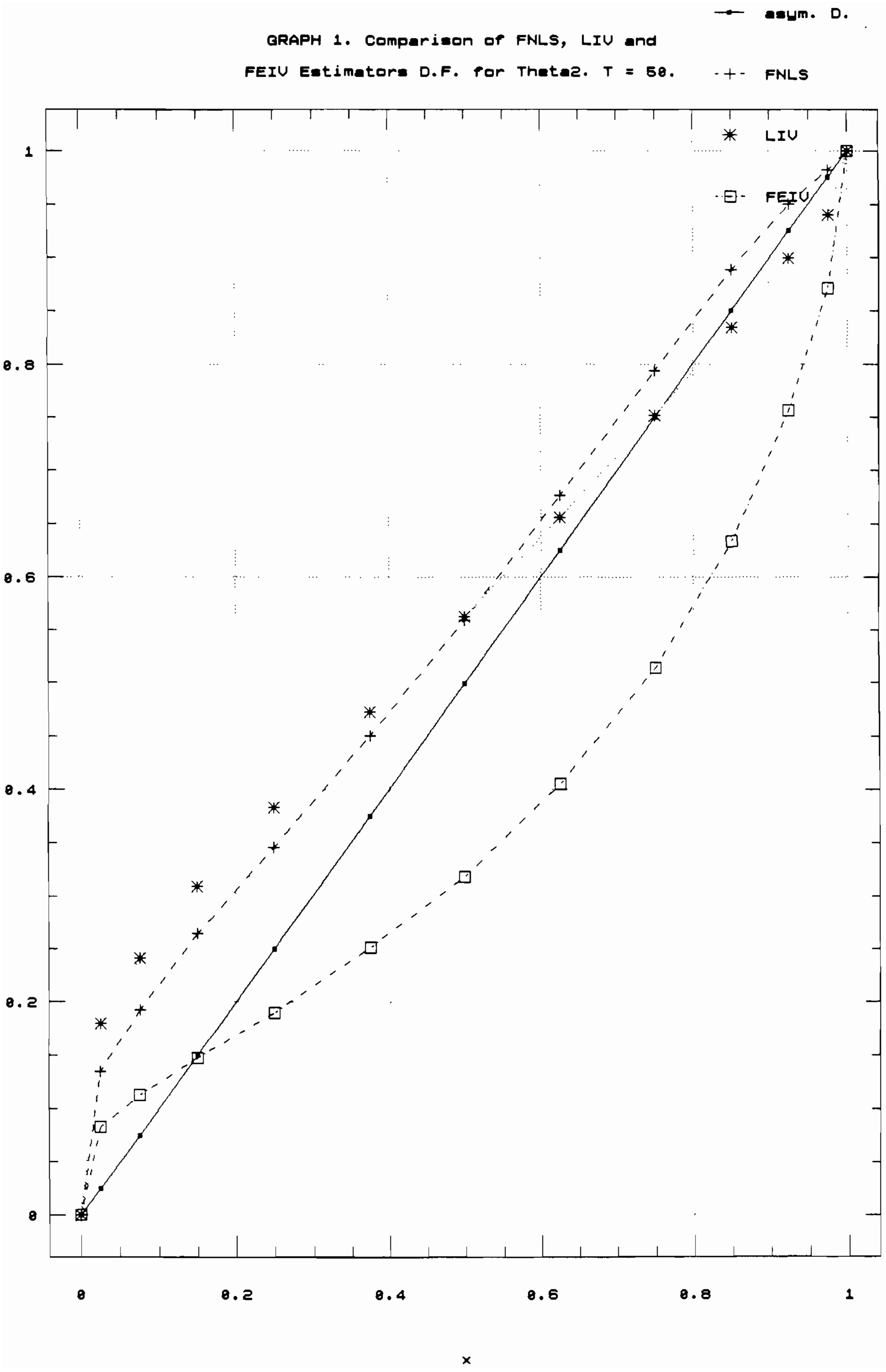
Muellbauer, John and D.Winter,(1980)"Unemployment,Employment and Exports in British Manufacturing", European Economic Review,13,383-415

Pesaran. M.Hashem,(1987)"The Limits of Rational Expectations",Basil Blackwell ,Oxford.

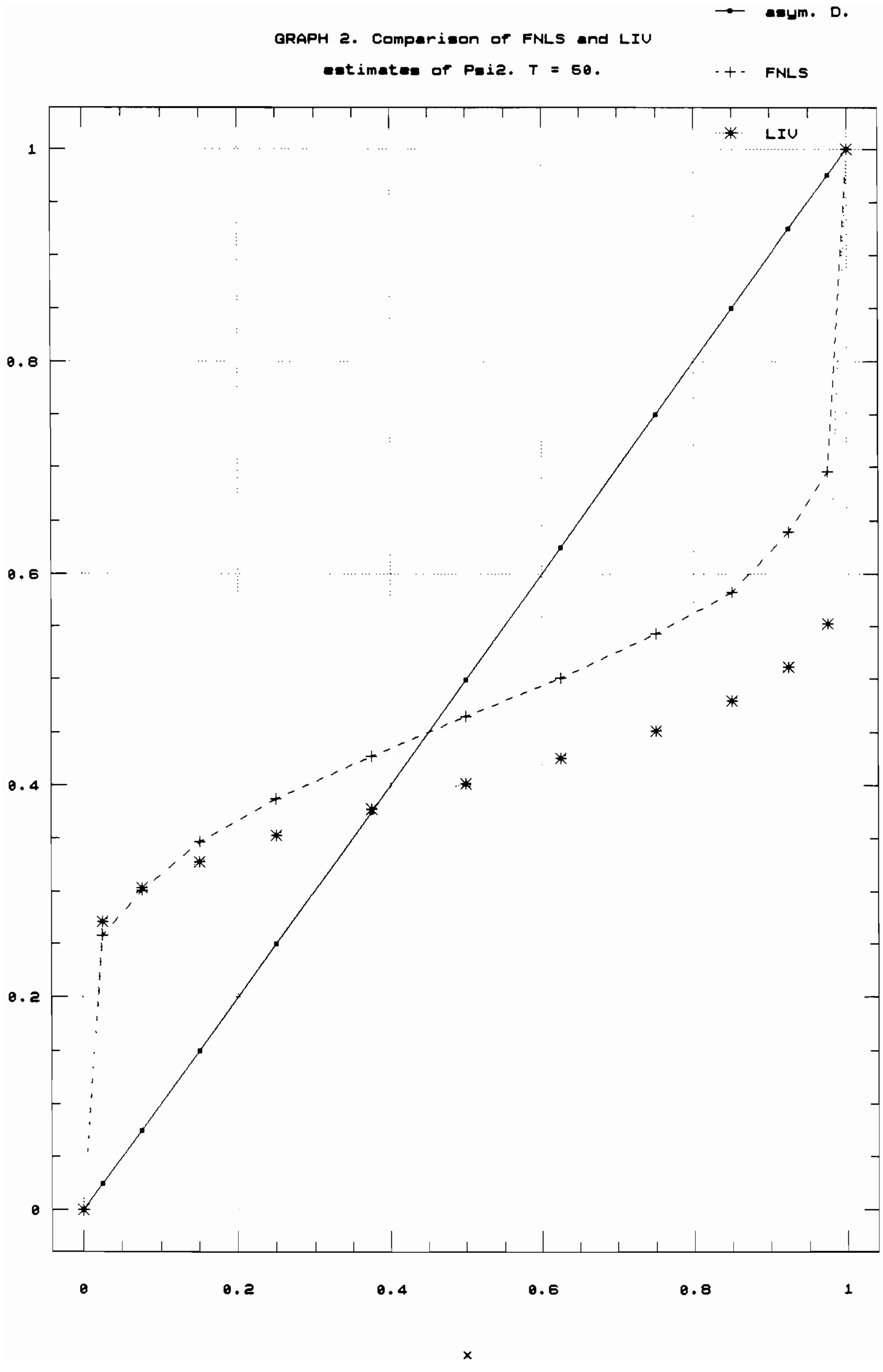
Sargan, J.Denis, (1988)"The Existence of the Moments of Estimated Reduced Form Coefficients" in "Contributions to Econometrics vol.2" ,C.U.P, Cambridge.

Sargan, J.Denis and Satwant S. Marwaha(1984) "Some Single Endogenous Variable Rational Expectation Models. Asymptotic Theory and Finite Sample Simulations Compared."LSE DEMEIC Econometrics Programme, discussion paper A.60.

Sargan J. Denis, (1988b) "Advanced Econometric Theory", Blackwell ,Oxford



GRAPH 1. Comparison of FNLS, LIV and
FEIV Estimators D.F. for Theta2. T = 50.
asym. D.
FNLS
LIV
FEIV
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
x



GRAPH 2. Comparison of FNLS and LIV
estimates of Psi2. T = 50.
asym. D.
FNLS
LIV
0
0.2
0.4
0.6
0.8
1
x

Appendix A.- Some Optimal Control Models.

There are many alternative models which can be used as simple optimal control models for business management or economic behaviour. In order to achieve a general form suppose that we have an exogenous variable z_t generated by a general autoregressive equation

$$z_t=\sum_{i=1}^{k}\phi_i z_{t-i}+v_t \tag{A1}$$

and y_t is a variable which it is costly to change, and which is used to control some third variable x_t. There is also a lag in the determination of x_t, which is also partly determined by z_{t-1}, and also by a further variable ω_{t-1} (which will be discussed later), so that

$$x_t= ay_t+ by_{t-1}+ cz_{t-1}+ d + \omega_{t-1}. \tag{A2}$$

It is desired to equate x_t to a target x^*_t which in turn is determined by

$$x^*_t= ez_t.$$

Then a loss function is set up as

$$\sum_{t=1}^{T} K^t[(x_t-x^*_t)^2+A(y_t-y_{t-1})^2].$$

ω_t is also regarded as an exogenous variable.

Then the FOC give the equations

$$\begin{aligned}&[a(ay_t+ by_{t-1}+ cz_{t-1}- ez_t+ d+ \omega_{t-1})\\ &+A(y_t-y_{t-1})+Kb(ay_{t+1}+ by_t+ cz_t\\ &- ez_{t+1}+ d + \omega_t)- KA(y_{t+1}- y_t)]\\ &=[a^2+Kb^2+A(1+K)]y_t+K(ab-A)y_{t+1}\\ &+(ab-A)y_{t-1}+(Kbc-ae)z_t-(Kbe)z_{t+1}\\ &+acz_{t-1}+(a+Kb)d+Kb\omega_t+a\omega_{t-1}= 0\end{aligned} \tag{A3}$$

It is assumed that the ω_t variables are exogenous and known to the decision taker both in period t and t-1, but that y_{t+1} and z_{t+1} are replaced by their expectations in period t, and the working equation is

$$y_t=b_1y_{t-1}+b_1^*E\,(y_{t+1}|t)+c_0z_t+c_1z_{t-1}+c_1^*E\,(z_{t+1}|t)+u_t$$

and

$$b_1=-\frac{K(ab-A)}{(a^2+Kb^2+A(1+K))},$$

$$b_1^*=-\frac{(ab-A)}{(a^2+Kb^2+A(1+K))},$$

$$c_0=\frac{(ae-Kbc)}{(a^2+Kb^2+A(1+K))},$$

$$c_1=\frac{-ac}{(a^2+Kb^2+A(1+K))},$$

$$c_1^*=\frac{Kbe}{(a^2+Kb^2+A(1+K))}.$$

and

$$u_t=-\frac{(Kb\omega_t+a\omega_{t-1})}{(a^2+Kb^2+A(1+K))}.$$

If we treat ω_t as white noise then u_t is a moving average error, but it is probably simplest to assume that u_t is a white noise error.

This is of the form of equation estimated if ac = 0, and, since a = 0 does not make much sense, it is appropriate to put c = 0, which gives a possible form of the model.

An alternative generalized model has a loss function of the form

$$\sum_{t=1}^{T} K^t\left[(x_t-x_t^*)^2+A(y_t-y_{t-1})^2+B(x_t-x_t^*)\,(y_t-y_{t-1})\right]$$

which leads to FOC equations of much the same form.

As an application a firm is deciding on the employment of labour which is denoted by y_t, z_t is the demand for product, and x_t is an output variable partly determined by y_t and y_{t-1}.

The model allows x_t to be partly determined by lagged z_t (this may be thought of as some external variable which is affected by the general level of demand in the economy. The same form of structural equation is obtained by allowing x_t^* to depend on z_{t-1}.