

SENSITIVITY OF ECONOMETRIC
ESTIMATES OF SYSTEM PARAMETRES
TO CHANGES IN SAMPLING
INTERVALS, MEASUREMENT ERROR,
AND PROCESS ERROR

By

Ole Fugleberg

Norwegian School of
Management

David Andersen

State University
of New York at
Albany

ABSTRACT

Synthetic data methods are used to test the robustness of estimators of the parameters within a simple linear oscillator. Econometric methods are used to estimate the known parameters in the model.

The major result is that the deviations from the estimates to the true values of the parameters increase with sample interval. The influence from stochastic inputs is marginal.

This is due to, that for great sample interval, the lag in the causal dependency is relatively small compared to the interval between the observations involved.

I. DIFFERENCES IN PARADIGMS

This paper is influenced by the fact, that modelers building system dynamic and econometric models have relied on widely differing methods to model social, economic, and administrative systems. These differences have led to conflicts and misunderstandings concerning what methods are appropriate for estimating parameters under what circumstances, Meadows (1980).

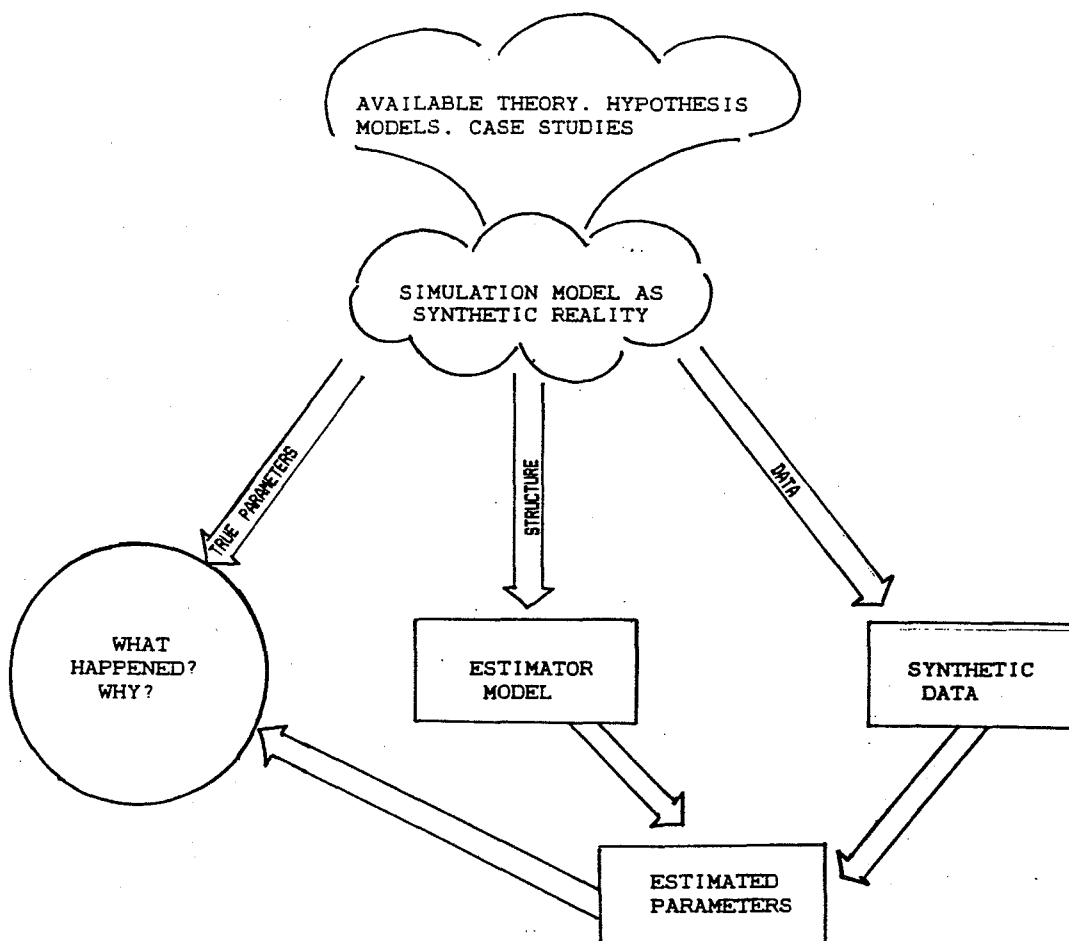
The work presented here is designed to present an empirical investigation of this "cross paradigm" discussion. Specifically, a system dynamic model with known structures and parameters is run many times with different sample intervals, and driven with various types of statistical disturbances corresponding to process error and measurement error. Output from this system is then sampled and observed under differing conditions. The model output thus generated is taken as "synthetic data" to be used in an econometric estimation designed to recover as closely as possible the parameters known in advance to exist within the system dynamic model.

Since the structure and parameter values for the parent system are known with absolute precision, this experiment can serve as a sort of controlled test of the ability of an econometric model to recover parameters from a dynamic feedback driven system.

II. THE OVERALL LOGIC OF SYNTHETIC DATA EXPERIMENTS

Figure II.1 below presents a schematic that overviews the logic of synthetic data experiments. Beginning at the top of that diagram, available theory, hypotheses, models, and case studies are used to create a computer simulation model of the social, economic, or administrative system under investigation.

FIGUR II.1: THE LOGIC OF SYNTHETIC DATA EXPERIMENTS

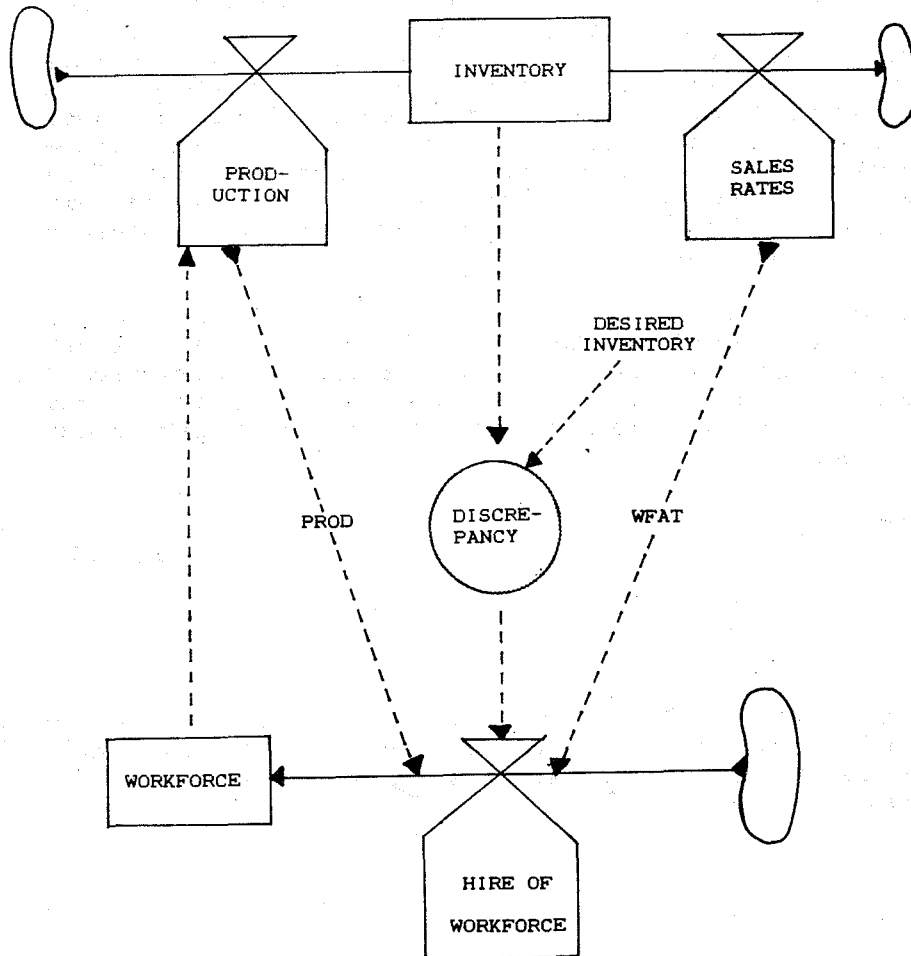


The simulation model is observed longitudinally with various sample intervals and disturbed by stochastic variables designed to simulate process and measurement error. These processes reflect biases, and other complications designed to mimic data and measurement problems that would certainly exist in any real world estimation problem.

III. THE DATA GENERATING MODEL

Figure III.1 presents a DYNAMO flow chart for the model used in the synthetic data experiment. A simple workforce-inventory model, which is a version of an elementary pedagogical model first used by Forrester (1955) in Principles of Systems and subsequently often used in introductory courses as an example of a simple linear, harmonic oscillator. The model shown in Figure III.1 has a totally linear structure with one exogenous variable, Sales Rate. The endogenous response of the system is sustained oscillations involving the two state variables, Inventory and Workforce.

FIGURE III.1: DYNAMO FLOW CHART FOR A SIMPLE WORKFORCE INVENTORY MODEL



The Inventory is assumed to be added to by a Production Rate and depleted by a Sales Rate. Sales Rate is taken to be exogenous, in the case reported a simple ramp with an stochastic disturbance. Production is assumed to be a simple linear function of the workforce. That is, the model assumes that labor is the only factor of production and that all units of this factor have a uniform productivity given by the constant, Productivity (PROD). The parameter, PROD, is one of the several parameters to be estimated within the econometric model.

The workforce is adjusted over time by a Hiring Rate which is a linear function of the Inventory. Actual Inventory is compared to Desired Inventory to arrive at a Discrepancy. This discrepancy is divided by PROD to scale for the relative productivity per worker of the workforce. Finally, the discrepancy is divided by an assumed Workforce Adjustment Time (WFAT) to convert a desired number of workers into a rate of hiring, measured in workers per month. The model assumes that WFAT is four months.

Actual workforce is corrupted by random term to arrive at Observed Workforce. This auxiliary variable adds an amount of measurement between the true workforce operating in the model and the workforce variable observed and used within the regression model. In the real world, this would correspond to a random rather than purely determined measure of the amount of productivity that can be obtained from a single worker.

Since this system is a purely linear one, it can easily be rewritten as a set of simultaneous first order linear differential equations. Using matrix notation, the two basic equations for the system are seen to be:

$$\begin{array}{c} \text{INV}' \\ \text{WF}' \end{array} = \begin{array}{cc} \begin{array}{c} 0 \\ -1 \\ \hline \text{PROD*WFAT} \end{array} & \begin{array}{c} \text{PROD} \\ 0 \end{array} \end{array} \begin{array}{c} \text{INV} \\ \text{WF} \end{array} + \begin{array}{c} -\text{SR} \\ \hline \text{DESINV} \\ \text{PROD*WFAT} \end{array} + \begin{array}{c} \epsilon_{\text{INV}} \\ \epsilon_{\text{WF}} \end{array} \quad (\text{III.1})$$

where INV, PROD, WF and SR denote respectively inventory, production, workforce and sales rate. ϵ_{INV} and ϵ_{WF} denote error terms.

The left hand side of this matrix equation represents the first derivative of the vector of two state variables. INV and WF.

IV THE ECONOMETRIC CORRESPONDENCE TO THE SYSTEM DYNAMIC MODEL

In order to arrive at equations that can be used in the estimation model, the continuous derivative on the left hand side of equation (III.1) must be approximated discretely. Concentrating on the single equation for inventory, we see that this approximation can be formulated in two different ways. In the first formulation, a lagged variable is entered on the right side of the equation. We arrive at the following formulation, where LAGINV denote lagged inventory:

$$INV = (PROD * WF - SR + LAGINV)h + \mu_{LINV} \quad (IV.1)$$

where the term h is a discrete time interval equal to the sampling period used to observe the data. μ_{LINV} is the error

Similarly, the difference formulation is given by:

$$DINV = (PROD * WF - SR)h + \mu_{DINV} \quad (IV.2)$$

Using similar logic, equation (IV.3) below defines the lagged formulation for workforce and equation (IV.4) defines the difference formulation for workforce:

$$WF = \left[\frac{DESINV}{PROD * WFAT} - \frac{I}{PROD * WFAT} * INV + LWF \right]h + \mu_{LWF} \quad (IV.3)$$

and

$$DWF = \left[\frac{DESINV}{PROD * WFAT} - \frac{I}{PROD * WFAT} * INV \right]h + \mu_{DWF} \quad (IV.4)$$

For sample interval one, h is equal to one.

The error term in the model of discrete form, μ_{LINV} , μ_{DINV} , μ_{LWF} and μ_{DWF} will be continuous time records of the error terms ϵ_{INV} and ϵ_{WF} in the continuous form.

V. ECONOMETRIC PROBLEMS OF ESTIMATING PARAMETERS FROM DISCRETE OBSERVATION TO A CONTINUOUS MODEL

Bergstrom (1976) p. 3 and 4 proposes that the following two points give problems in econometrics of time continuous models based on discrete observations:

- 1) The data comprise series of observations made at longer intervals than the intervals between the decisions that they reflect.
- 2) The minimum lag in any causal dependency is equal to the interval between the variables involved.

Based on this we formulate the following hypothesis for doing econometrics of the synthetic data generated from the discrete time model approximated from the continuous system dynamic model.

The deviations from the true estimates will be much greater regarding variations in sample interval than to variations in processing and measurement error. The deviations will increase by sample interval.

VI RESULT FROM THE TIME SERIES EXPERIMENTS

In this section, we present the runs of the synthetic data experiment. The runs were executed for the combinations of sampling intervals, processing error and measurement error as the table below shows:

TABLE VI.1: A SURVEY OF THE RUNS OF THE SYNTHETIC DATA EXPERIMENTS

NAME OF THE DATA SET	SAMPLING INTERVAL	PROCESSING ERROR	MEASUREMENT ERROR
SI25	0.25	50.00	20.00
SI50	0.50	50.00	20.00
BASLON	1.00	50.00	20.00
SI200	2.00	50.00	20.00
SI300	3.00	50.00	20.00
PE10	1.00	10.00	20.00
PE30	1.00	30.00	20.00
BASLON	1.00	50.00	20.00
PE70	1.00	70.00	20.00
PE90	1.00	90.00	20.00
ME00	1.00	50.00	0.00
ME10	1.00	50.00	10.00
BASLON	1.00	50.00	20.00
ME30	1.00	50.00	30.00
ME40	1.00	50.00	40.00
ME50	1.00	50.00	50.00

The runs were executed with OLS for both lagged formulation and difference formulation, and with 2SLS for lagged formulation.

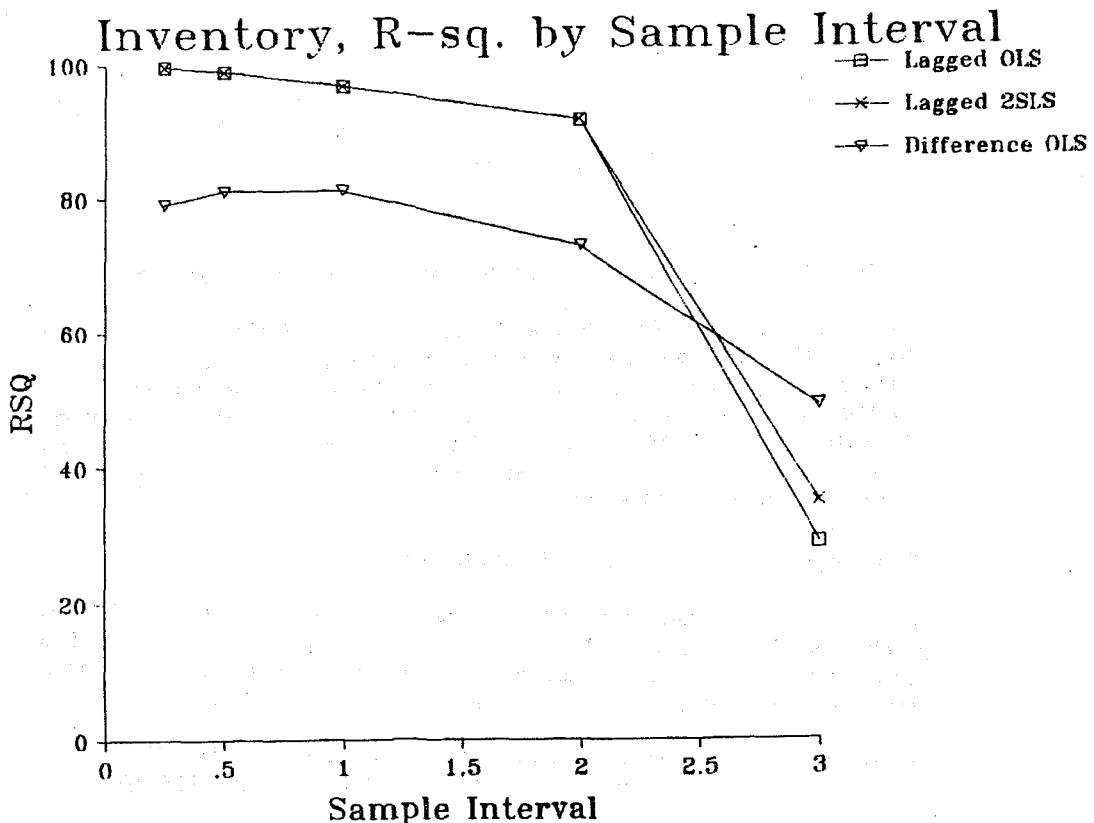
Comments concerning the statistical observations:

R-squared

- By sample interval

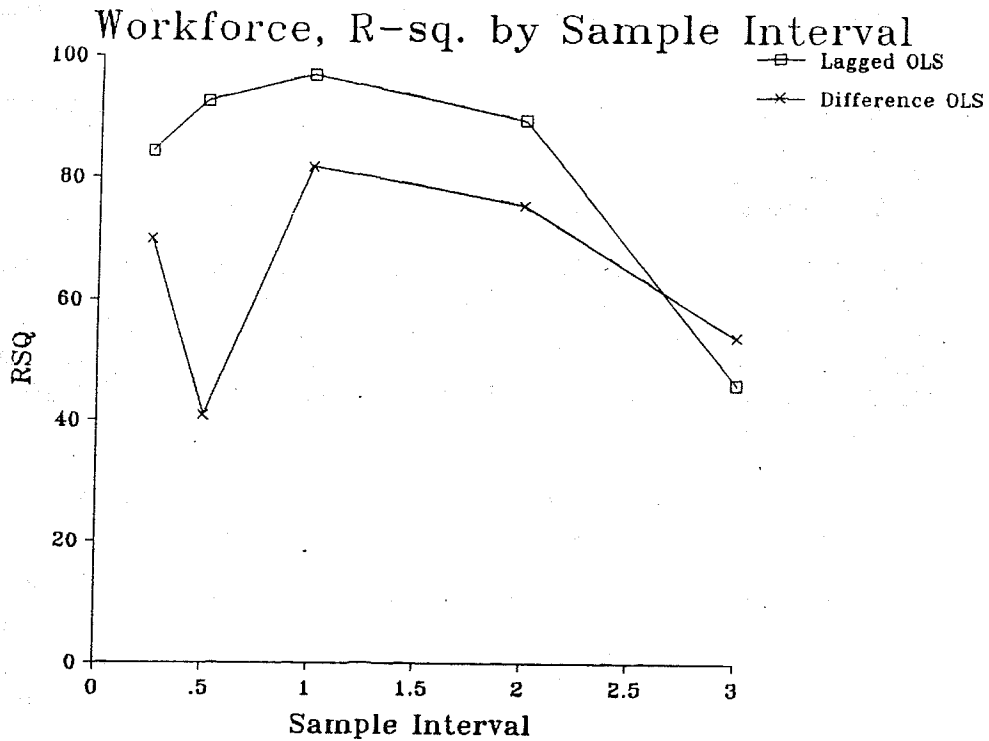
INVENTORY EQUATION: R-squared is quite good with sample interval, with the exception of a sample interval of three months, where it falls dramatically. This applies to both OLS and 2SLS-estimation, but slightly better for 2SLS.

Figure VI.1: Inventory equation, R-squared by sample interval.



WORKFORCE EQUATION: For lagged formulation, R-squared by sample interval follows the same pattern as for inventory equation, but in fact it is better for OLS. For difference formulation R-squared is relatively poor also for small sample intervals.

Figure VI.2: Workforce equation, R-squared by Sample Interval.



- With processing and measurement error

For lagged formulation, both for inventory and difference formulation, R-squared is quite good by processing and measurement error. It is no considerable difference between OLS and 2SLS. For difference formulation it is relatively poor for great measurement error.

The estimated coefficients

The true coefficients vary with the sample interval, but of course not with the processing error and the measurement error. The true coefficients are as follows:

TABLE VI.2: THE TRUE COEFFICIENTS WITH SAMPLE INTERVAL, INVENTORY

DATA SET	SAMPLE INTERVAL	WF	SR	LAGGED INVENTORY
S125	0.25	0.25	-0.25	1.00
S150	0.50	0.50	-0.50	1.00
BASLON	1.00	1.00	-1.00	1.00
SI200	2.00	2.00	-2.00	1.00
SI300	3.00	3.00	-3.00	1.00

TABLE V.3: THE TRUE COEFFICIENTS WITH SAMPLE INTERVAL, WORKFORCE

DATA SET	SAMPLE INTERVAL	INVEN- TORY	LAGGED WORKFORCE
S125	0.25	-0.06	1.00
S150	0.50	-0.13	1.00
BASLON	1.00	-0.25	1.00
SI200	2.00	-0.50	1.00
SI300	3.00	-1.00	1.00

Comparison of the estimated coefficients with the true coefficients

INVENTORY EQUATION: Both the estimated workforce and sales rate coefficients are for either formulations nearly the true value by sample interval, with the exception of two and three months sample intervals, especially of 3 months. The corresponding 2SLS estimates give the same pattern, but the distance to the true values are slightly greater.

By processing and measurement error for lagged formulation, the estimated workforce coefficients are nearly the true values, especially 2SLS estimates. For difference formulation, the estimated workforce coefficients have slightly lower values than the true values, and the estimates decrease slightly by increased processing and measurement error.

The estimated sales rate coefficients decrease from the true value of minus one by increasing processing error, especially for the difference formulation. For measurement error the estimates are slightly lower than the true values. There are marginal differences between OLS and 2SLS estimates.

The estimated lag inventory coefficients show great deviations from the true values for two and three sample intervals. By processing and measurement error, there are in all slightly higher values of the estimates than the true values.

Figure VI.3: Inventory equation, estimated workforce coefficient by sample interval.

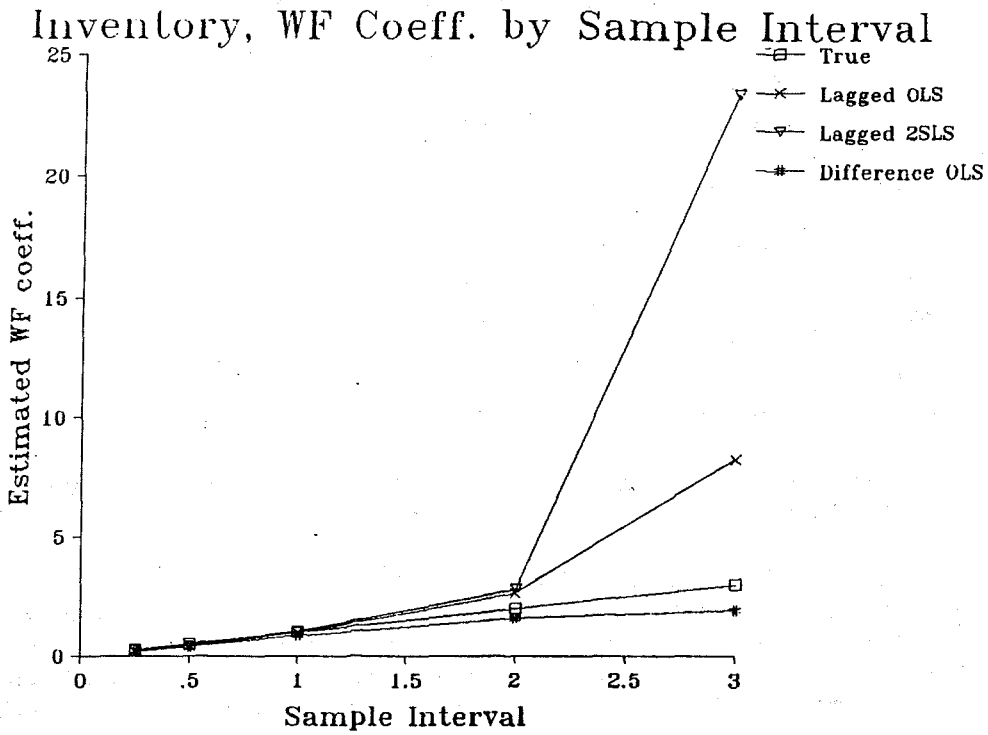
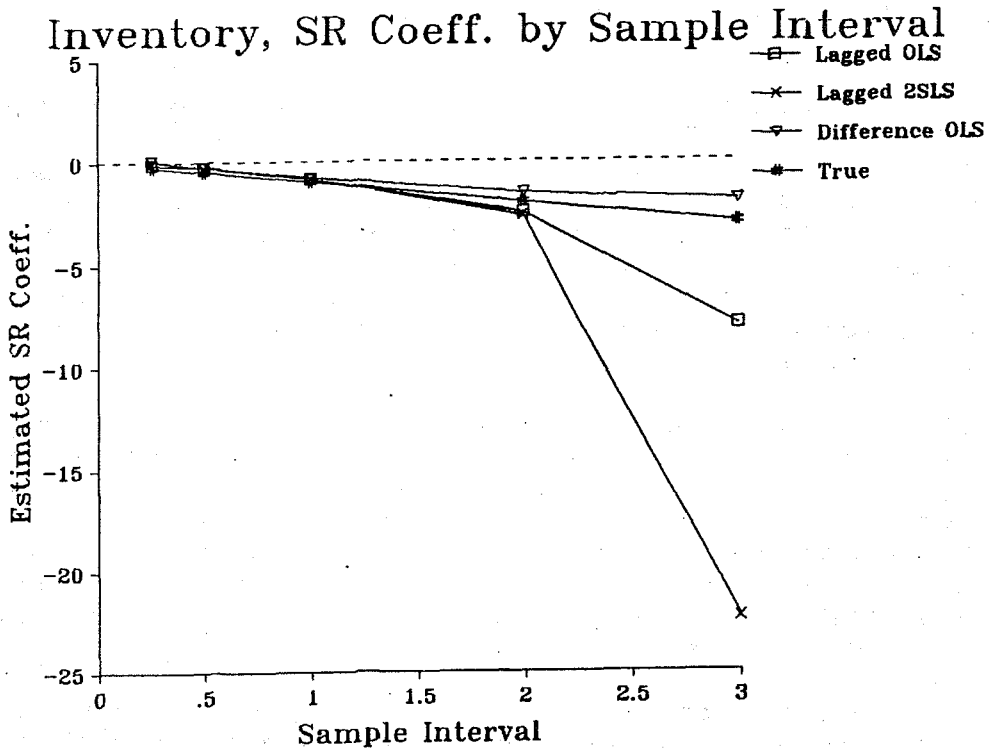


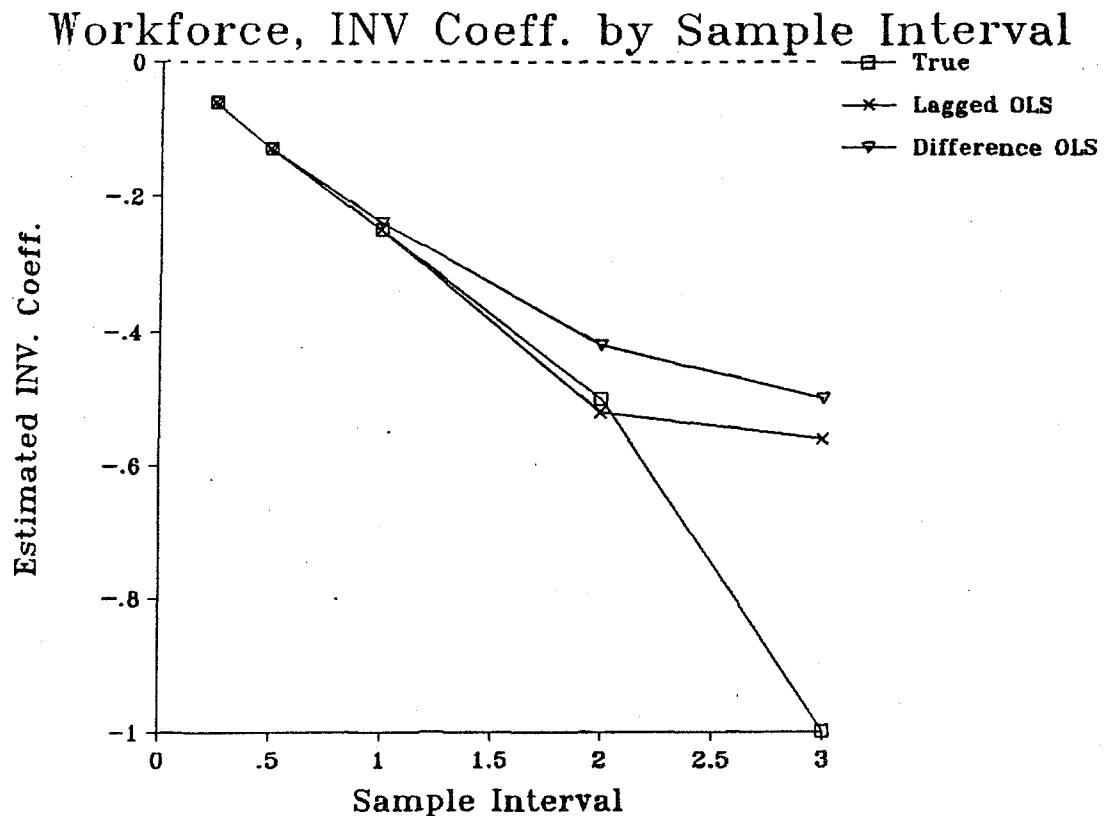
Figure VI.4: Inventory equation, estimated sales rate coefficient by sample interval



WORKFORCE EQUATION: The estimated inventory coefficients by sample interval are nearly the true values with the exception of 3 months sample interval. This applies to both formulations and to OLS and 2SLS estimation. There are small differences between OLS and 2SLS estimation.

By processing and measurement error, these estimates are in all cases nearly the true values, both by OLS and 2SLS estimation. The estimated lag workforce coefficients are in all cases nearly the true values.

Figure VI.5 Workforce equation, estimated inventory coefficient by sample interval.



Standard deviation:

There are only marginal differences among the standard deviations of the estimates by both formulations, sample interval, processing and measurement error. There are marginal differences between OLS and 2SLS estimates.

VII. FINAL CONCLUSION CONCERNING THE ESTIMATED COEFFICIENTS

The OLS-estimated coefficients based on data generated from the linear harmonic oscillator model are nearly the true values for relatively small sample intervals. The deviations to true values increase by sample interval, especially for three months sample interval.

For three months sample interval the deviations to the true values are less for the difference formulation than the lagged formulation. This is due to the fact that the interval between the observations involved is less compared to the lag in the causal dependency in the difference formulation.

By processing and measurement error, the deviations between the estimated and true values are much less compared with sample interval. And between processing error and measurement error, the deviations to the true values are slightly greater for processing error.

There are marginal deviations among OLS and 2SLS estimates.

The observator R-squared follows the same pattern as the estimated coefficients with regard to variation in sample interval, processing and measurement error and by estimation techniques.

These results are in accordance with the referenced results from theoretical works of estimation in continuous time economic models by discrete data.

REFERENCES

Forrester, J.W (1955): "Principles of Systems", MIT Press, Cambridge, Massachusetts.

Meadows, D.H. (1980): "The Unavoidable a Priori", In Elements of the System Dynamics Method (Ed. Randers J.). MIT Press. Cambridge, Massachusetts.

Bergstrom, A.R (1976): "Statistical Inference in Continuous time Economic Models", North Holland Publishing Company-Amsterdam Oxford.