A SIMULATION STUDY OF STRUCTURAL ESTIMATORS OF DYNAMIC SIMULTANEOUS EQUATIONS MODELS WITH AUTOREGRESSIVE ERRORS

by

Tomson Ogwang

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
July, 1990

ISBN 0-315-64375-7

TABLE OF CONTENTS

LIST OF TABLES

Abstract

In this study we investigated the small sample properties of six two-step estimators of dynamic simultaneous equation models with autoregressive errors using the Monte Carlo approach. The six estimators were proposed by Hatanaka.

The study focused on the relative performances of the estimators in structural estimation and prediction.

All the six estimators exhibited significant biases. The rankings of the estimators depended on the magnitude of autocorrelation, the coefficient of the lagged endogenous variables and the sample size. Furthermore, the problem of choice among estimators was relatively more important for prediction than for structural estimation. The kernel estimates of the sampling distributions of the estimators were quite similar and were almost symmetric. Significant differences among the estimators emerged only if there were large differences between the autocorrelation coefficients of the equations. The full information estimators generally performed worse than their limited information counterparts when autocorrelation was high and the reverse was true at low levels of autocorrelation. Whereas the asymptotic covariance matrices of the structural parameters were unreliable for purposes of inference in small samples, very large samples were required for the asymptotic covariance matrices of dynamic simulation forecasts to make valid inferences concerning forecasts.

ACKNOWLEDGEMENTS

I would like to thank my research supervisor, Professor U.L.G Rao for his guidance and encouragement in the preparation of this thesis.

My thanks also go to Professors S. Amirkhalkhali, E. Klein and A. Buse who read the thesis and made very valuable comments.

Ms. Heather Lennox enthusiastically handled the logistics of getting the work on paper. To her I owe a permanent debt of gratitude.

Finally, I would like to thank my wife, Molly, and my daughters, Fiona and Lydia, for their patience and support during all stages of the work.

CHAPTER 1: INTRODUCTION

1.1 The choice among several alternative estimators

Econometric modelling involves a series of decisions during various stages of the exercise. Virtually all these decisions require imaginative insights into the problem(s) under investigation taking into consideration any prior knowledge about the underlying economic structure as well as the small sample properties of the available estimators.

Specification of the model involves decisions regarding the variables to be included in the model. Also decisions are made with respect to the functional form of the equation(s) in the model and the assumptions about the distribution of the error terms. The econometrician may choose a single equation model or a simultaneous equations model(SEM). In either case the model might be static or dynamic. The model is dynamic if lagged endogenous variables appear among the explanatory variables in the equation(s). Such models are used to explain the dynamic behaviour of the economy. The specification of dynamic simultaneous equations models has become increasingly common in applied econometric research in recent years. In the estimation stage decisions are made as to which estimators would be appropriate for the specified models. Usually, for a given model specification, there are several alternative estimators to choose from. A choice among these estimators requires some basis for comparing them. One possible approach is to base decisions on the (desirable) properties of these estimators, which may include the large sample properties of asymptotic unbiasedness, consistency, asymptotic efficiency and asymptotic normality, or their small sample properties. For practical applications, however, the asymptotic properties are no more than cold comfort because cost and time constraints often limit the applied researcher to small samples. Perhaps a stronger case for using small samples in applied research arises from the fact that with the passage of time, structural evolution renders obsolete the data that belongs to the remote past. As Brown(1960,p174) explicitly put it:

> "... small samples are the rule with economic time series appropriate for structure estimation of complete models.The passage of time may not correct for this,for structure evolution may force the gradual rejection of data from periods too far in the past".

In view of this problem, evaluating the estimators on the basis of their asymptotic properties is not a useful endeavour. Furthermore, the fact that the asymptotic properties of these estimators may not necessarily reflect their performance in small samples provides a justification for establishing the small sample properties as well. Should the estimators possess the same optimal small sample properties, then factors other than these, e.g.cost, may influence choice among such estimators.

In order to establish the small sample properties of any estimator there are two possible and often complementary procedures that may be followed: analytical derivations and Monte Carlo experiments.

In the former case the exact expressions or approximations for the sampling distributions and/or moments of the sampling distributions of the estimators are derived for a given model specification. This task may be easily accomplished for simple estimators, such as the ordinary least squares (OLS) for the standard linear regression model. For more complex models, analytical derivations of sampling distributions of estimators may be a formidable task. Furthermore, even if the task is successfully accomplished for such estimators, the resulting mathematical expressions are often too complicated, rendering them incomprehensible and thereby undermining their usefulness for purposes of comparison.

Monte Carlo experiments, on the other hand, involve the simulation of the sampling distributions of the estimators by controlled experimentation. This approach, which is also called the distribution sampling method, has the advantage of being feasible over a wider range of error distributions whereas analytical investigations of small sample distributions become even more involved if the structural errors are assumed to follow non-normal distributions.

Assuming that the problems of specification and estimation have been resolved and the estimates of the parameters of the model have been obtained, the researcher would then proceed to the evaluation stage in which decisions are made as to whether the computed estimates are theoretically meaningful on the basis of sign, significance of the coefficients and values of the coefficient of determination, among others. Should the resultant estimates be declared acceptable, they may then be used for either verifying or rejecting the theory being tested as well as for post sample prediction and policy evaluation. Also there is an increasing body of empirical evidence (for example, Raj(1980)) which suggests that the ranking of alternative estimators of the structural parameters according to their small sample properties may differ from their ranking from a post sample prediction point of view. Thus, if the main objective of structural estimation is for post sample prediction, then it might be useful to compare the estimators according to their predictive capabilities and this provides a prima facie case for comparing their predictive capabilities.

It is needless to say that standard errors have direct relevance for drawing inferences about the significance of the estimates of structural parameters. This poses the question of the reliability of the estimated standard errors. If the small sample distributions cannot be derived, the moments cannot be derived either. Since asymptotic distributions of estimators are relatively easier to derive, it is also easy to derive the asymptotic covariance matrix from which the asymptotic standard errors can be computed. In practice estimation and tests of hypotheses are almost invariably based on the asymptotic standard errors. Accordingly, an assessment of the reliability of the asymptotic standard errors of the estimators in small sample situations would be a useful endeavour. Unfortunately, this question was not routinely dealt with in most Monte Carlo studies.

Recent developments in the area of nonparametric density estimation have made it possible to estimate the distributions of the estimators, standard errors and t-ratios using the point estimates obtained in Monte Carlo experiments. For example, Power and Ullah (1984) used the kernel method of nonparametric density estimation to estimate the distribution of limited information rational expectations estimators and their t-ratios. This approach can also be extended to estimate the distributions of prediction errors as well, thus providing a new perspective for comparing alternative estimators in Monte Carlo studies.

As mentioned above each method of estimation has specific problems but in this study we concern ourselves with the comparison of some estimators of dynamic simultaneous equation models (SEM's) in which the errors are assumed to be normally distributed and autocorrelated. The case of non-normal errors is not dealt with in this study.

Before the objectives of this study are articulated it is important to review the problems associated with estimating dynamic SEM's and, in particular, how these problems are compounded when the standard assumption of serial independence of errors is dropped.

1.2 Estimation of SEM and the problem of autocorrelation

Interdependent relationships among economic variables has dominated thinking in economic theory for many years. A survey of Keynesian economics, for example, reveals numerous examples of such interdependent relationships. For over four decades the interest of econometricians has centered on the modelling of such systems and, in most cases, obtaining asymptotically efficient and consistent estimates of the parameters of such models, under given assumptions. Dynamic simultaneous equation models, in particular, are often specified to capture the dynamic behaviour of an economy while highlighting the structural interdependence in the economy.

Recently there has been considerable interest in the specification, estimation and testing of hypotheses in dynamic SEM's with errors generated by vector autoregressive processes. Numerous tests for the presence of autocorrelation in SEM's have been proposed, of which the most relevant for dynamic simultaneous equation models are Guilkey (1975), Maritz (1978) and Godfrey (1976). Whereas these tests apply only in the case of first order autoregressive processes, Godfrey (1978a,1978b) proposed tests that apply for higher order error processes as well. The question which arises is: how do we estimate the structural parameters given that the null hypothesis of no autocorrelation is rejected?

The introduction of serially correlated errors into SEM's poses new estimation problems because the commonly applied estimators, such as two stage least squares(2SLS) and three stage least squares(3SLS) become unsatisfactory, as they yield asymptotically inefficient estimates. Furthermore, the presence of lagged endogenous variables among the predetermined variables causes further estimation problems because these techniques yield both inconsistent and asymptotically inefficient estimates of structural parameters. Inconsistency arises because of the correlation between the lagged endogenous variables and the serially correlated disturbances. Thus, in the presence of autocorrelation, new estimators have to be developed. The derivation of such estimators and their sampling distributions becomes even more difficult in the case of dynamic SEM's with autocorrelated errors.

To fix ideas, we introduce a typical dynamic SEM with an autocorrelated error structure and specify precisely the estimation problem with respect to such models.

1.3 Dynamic SEM with autocorrelation

A typical dynamic SEM consisting of g endogenous variables, g one-period lagged endogenous variables, k_2 exogenous variables and errors generated by a vector autoregressive process may be specified as

$$Y = YB + Y_{-1}C_0 + XC_1 + U \quad (1.1)$$

$$U = U_{-1}R + E$$

where Y is the T x g matrix of observations on g endogenous variables; Y_{-1} is the T x g matrix of observations of one-period lagged values of Y; X is the T x k_2 matrix of observations on k_2 exogenous variables (The t-th row of X is denoted by $x_{t.}$ t = 1,. . .,T); U is the T x g matrix of disturbances generated by the first order autoregressive process defined above. The t-th row of U is denoted by $u_{t.}$. E is the T x g matrix of white noise errors (The t-th row of E is denoted by $e_{t.}$ t = 1,. . .,T). Note that U_{-1} is the T x g matrix of one-period lagged values of U. B, C_0 and C_1 are respectively of orders g x g, g x g and k_2 x g. C_0 has k_1 ($\leq$ g) non-zero rows corresponding to the k_1 lagged endogenous variables entering into the model. The model is dynamic in the sense that C_0 is non-null.

Since one endogenous variable in each equation is considered as the dependent variable, its coefficient may be set arbitrarily equal to unity. To this end all the diagonal elements of B are prespecified to be zero. R is a square matrix of order g x g. Some elements of R may be specified a priori to be zero.

The following assumptions are made regarding model (1.1):

Assumption 1. The matrix (I - B) is non singular. This guarantees that the model is mathematically complete in the sense that the reduced form of the model exists and is unique. Note that I is the g x g identity matrix.

Assumption 2. $e_{t.}$'s are n.i.d. $(0,\Sigma)$, Σ is positive definite.

Assumption 3. The characteristic roots of R and $C_0(I-B)^{-1}$ lie inside the unit circle. This guarantees stability in the model.

Assumption 4. $\lim_{T \to \infty} X'X/T$ is positive definite.

Assumption 5. $\text{plim } 1/T\ [Y_{-1} : X : X_{-1}]'E = 0$

Assumption 6. All equations to be estimated in the model are identified.

The i-th equation embedded in model (1.1) can be written as

$$y_{.i} = Yb_{.i} + Y_{-1}\, c_{0.i} + X\, c_{1.i} + u_{.i} \qquad (1.2)$$

Similarly, at time t, the model (1.1) may be written as

$$y_{t.} = y_{t.}B + y_{-1t.}C_0 + x_{t.}C_1 + u_{t.}$$

$$\text{with } u_{t.} = u_{t-1.}R + e_{t.} \qquad (1.3)$$

The reduced form of the model (1.1), ignoring autocorrelation is

$$Y = \phi^* \Pi^* + V^* \qquad (1.4)$$

where $\phi^* = [Y_{-1}\; X]$; $\Pi^* = [\Pi_1{*}'\; \Pi_2{*}']'$;

$\Pi_1{*} = C_0(I - B)^{-1}$; $\Pi_2{*} = C_1(I - B)^{-1}$ and $V^* = U(I - B)^{-1}$

Similarly the reduced form, taking into account autocorrelation is

$$Y = \phi\Pi + V \qquad (1.5)$$

where $\phi = [Y_{-1}\quad Y_{-2}\quad X\quad X_{-1}]$ and $\Pi = [\Pi_1'\Pi_2'\; \Pi_3'\; \Pi_4']'$;

$\Pi_1 = \{C_0(I-B)^{-1} + (I-B)R(I - B)^{-1}\}$; $\Pi_2 = \{-C_0R(I - B)^{-1}\}$;

$\Pi_3 = C_1(I - B)^{-1}$;$\Pi_4 = -C_1R(I-B)^{-1}$;

and $V = E(I - B)^{-1}$ is the matrix of reduced form errors.

Y_{-2} denotes the matrix of two-period lagged values of Y.

It follows from the above assumptions that the T rows of V are independent random vectors which are normally distributed with mean 0 and covariance matrix $(I-B')^{-1}\Sigma(I- B)^{-1}$.

Application of OLS to (1.5) yields consistent estimates of Π_1, Π_2, Π_3 and Π_4 (and hence consistent unrestricted reduced form predictions of Y). Alternatively, consistent predictions of Y can be obtained using consistent estimates of C_0, B, R and C_1 . However, the application of OLS to (1.4) where the autocorrelation is ignored yields inconsistent estimates of the reduced form parameters, which in turn yields inconsistent predictions of Y.

The identification restrictions may be incorporated into the model (1.1) using special selection matrices. The selection matrices S_{i1}, S_{i2} and S_{i3} are defined below.

For the i-th equation S_{i1} is a selection matrix such that $YS_{i1} = Y_i$ the matrix of endogenous variables included on the RHS of the i-th equation; $Y_{-1}S_{i2} = Y_{-1i}$ is the matrix of lagged endogenous variables included in the i-th equation and $XS_{i3} = X_i$ is the matrix of pure exogenous variables appearing in the i-th equation. This selection matrix has very special properties in that $S_{i1}'b_{.i} = \beta_{.i}$ the elements of $b_{.i}$ not specified a priori to be zero. Similarly $S_{i2}'c_{0.i} = \gamma_{0.i}$ and $S_{i3}'c_{1.i} = \gamma_{1.i}$

If some elements of R are specified to be zero then the non-zero elements can be picked by using appropriate selection matrices. For this we define S_{ri} to be the selection matrix such that $S_{ri}'r_{.i} = r_{*.i}$ the vector of elements of the i-th column not specified a priori to be zero. In this case we estimate $S_{ri}'r_{.i}$ and not $r_{.i}$. It follows that if R is specified to be diagonal then we have $S_{ri}' = [0\ 0.\ .\ .1.\ .\ .0]$ with the number 1 occupying the i-th position in the vector and only the diagonal elements of R need to be estimated. Clearly $U_{-1}S_{ri}$ picks up the columns of U_{-1} which correspond to the elements of $r_{.i}$ not specified a priori to be zero.

With the a priori restrictions incorporated, the i-th structural equation may be written as

$$y_{.i} = ZS_i\delta_{.i} \text{ where } Z = [Y \; Y_{-1} \; X] \tag{1.6}$$

and $S_i = \mathrm{diag}(S_{i1} \; S_{i2} \; S_{i3})$ and $\delta_{.i} = [\beta_{.i}' \; \gamma_{0.i}' \; \gamma_{1.i}']'$ is the vector of coefficients in the i-th equation.
Furthermore, we define $\gamma_{.i} = [\gamma_{0.i}' \; \gamma_{1.i}']'$ to be the vector of coefficients of all included predermined variables in the i-th equation.
Accordingly, the entire model can be written as

$$y = (I \otimes Z)S\delta + u \tag{1.7}$$

where $S = \mathrm{diag}(S_1, S_2, \ldots, S_g)$, $\delta = [\delta_{.1}' \; \delta_{.2}' \ldots \delta_{.g}']'$,
$y = \mathrm{vec}(Y)$ and $\otimes$ is the Kronecker product.
From the equation $U = U_{-1}R + E$, and using the property that $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}\, B$ (See Magnus and Neudecker (1986)), we have

$$u = (R' \otimes I)\, u_{-1} + e \tag{1.8}$$

where $E(e) = 0$ and $\mathrm{cov}(e) = (\Sigma \otimes I)$

It is easy to show that $E(u_{t.}'u_{t.}) = \sum R'^{i}\Sigma R^{i}$ (see, for example, Fomby et. al. (1984, p.548)). Also $\Sigma = \Omega - R'\Omega R$ and therefore $\text{vec}\ \Omega = [I-(R' \otimes R)]^{-1}\ \text{vec}\ \Sigma$.

The estimation problem in the context of SEM with autocorrelated errors is concerned with estimating the elements of B, C_0, C_1, and R in (1.1) not specified a priori to be zero. The estimators proposed in the literature may be classified either as a limited information estimator or as a full information estimator. Limited information estimators use only the information that is specific to the equation we are interested in. Full information estimators, on the other hand, take into account the entire information provided in the model including all a priori restrictions .

Sargan (1961) first proposed limited information maximum likelihood estimators in special cases of SEM with autocorrelated errors. Amemiya (1966) proposed a two-stage least squares analogue of one of Sargan's estimators as well as a truncated version of the analogue. Both Sargan's limited information estimator and Amemiya's 2SLS analogue are consistent and asymptotically efficient if the R matrix is diagonal. It may be pointed out here that Amemiya did not allow for lagged endogenous variables to be present in the model. Fair (1970) and the correction in Fair (1984,p 212-214)) proposed iterated limited information estimators based on selected subsets of instruments required during the first stage regression. This procedure, in most cases, is not asymptotically efficient. The seminal article by Brundy and Jorgenson (1971) provided a breakthrough in devising instrumental variable estimators of the standard SEM which, in turn, constituted the basis for limited information instrumental variable estimators proposed by Fair (1972) and Dhrymes, Berner and Cummins (1974). Fair's limited information instrumental variables efficient (LIVER) estimator is consistent and asymptotically efficient within the class of limited information estimators provided the autoregressive coefficients are known with certainty. If R is not known with certainty, then the LIVER estimator based on an iterative procedure is asymptotically efficient. Hatanaka (1976) devised a way of ingeniously using the residuals from the first stage regressions to obtain three consistent and asymptotically efficient limited information two step estimators. It turns out that one of Hatanaka's limited information methods is an extension and simplification of Amemiya's 2SLS estimator which also takes into account lagged dependent variables. Another estimator is a simplification of the method proposed by Fair (1972). Rao (1986) proposed a Cochrane-Orcutt type Two Stage Least Squares estimator (COT2SLS) and provided an algorithm for estimating such models in the case where R is diagonal. The estimator is applicable in both dynamic and static models and consistency is achieved provided that we begin from initial consistent estimates of the autoregressive parameters. Buse (1989) proposed generalized two-stage least squares estimators of such models.

Sargan (1961) was also the first who proposed full information maximum likelihood (FIML) estimators of SEM for specific cases, e.g., models where all predetermined variables are lagged endogenous. However, it was Hendry (1971, 1974) who derived the full information maximum likelihood estimator and proposed an algorithm for purposes of computation. Chow and Fair (1973) also presented computational algorithms for the FIML estimator. Hendry considered the case of unrestricted autoregressive coefficient matrices (i.e. no zero elements) whereas Chow and Fair considered cases of restricted autoregressive coefficient matrices as well. The full information instrumental variable estimator developed by Brundy and Jorgenson (1971) for the standard SEMs was extended by Fair to include autoregressive errors. Again it was Hatanaka (1976) who proposed three full information two-step estimators, which were in fact extensions of the ideas contained in an earlier paper by Hatanaka (1974). It turns out that the estimators proposed by Dhrymes and Erlat (1974) and Fair (1972) are special cases of Hatanaka's estimators. Dhrymes and Taylor (1976) independently derived one of Hatanaka's full information two step estimators which is a natural extension of Dhrymes (1974) for the single equation case. Rao(1986) proposed a Cochrane-Orcutt type three stage least squares estimator (COT3SLS) and provided an algorithm for estimating such models in the case where R is diagonal. This estimator is consistent provided that we begin from a consistent estimate of Σ (e.g. from COT2SLS described above).

Except for Hatanaka's estimators many of the estimators proposed above are not asymptotically efficient. In addition, the computational algorithms proposed are iterative in character. The problem with iterative procedures is that convergence may not always be guaranteed. In addition, there is always the problem of multiple solutions. The six two-step estimators (three limited information and three full information) proposed by Hatanaka are extremely attractive for applied econometric work in view of their computational simplicity. In particular, the asymptotic covariance matrix of the estimates is easy to compute, which in turn makes testing of hypotheses relatively simple.

Hatanaka proved that all these estimators are consistent and asymptotically efficient in the sense that they attain the asymptotic Cramer-Rao lower bound. As suggested by Hatanaka, the choice among these estimators should be based on their performance when the sample size is small. Although the asymptotic distributions of the two-step estimators have proved analytically tractable, yet the derivation of the exact sampling distributions does not seem feasible. In fact, neither the exact sampling distributions of these estimators nor approximations to them are currently available. This Monte Carlo study purports to fill the gap between what is known about their asymptotic properties and what little is known about their small sample properties.

1.4 The kernel method of nonparametric density estimation

The statistical technique of density estimation is concerned with the methods of estimating a probability density function given a random sample from that distribution.

The methods of density estimation may be classified as a parametric method or as a nonparametric method. In the former case estimates of sample statistics are substituted for the corresponding estimates using the appropriate formulae for the density function from which the sample is taken. This requires prior knowledge of the distribution from which the sample is drawn. The nonparametric method is relatively more flexible for estimating a probability density function; it is a purely data based approach. In this section we concern ourselves with the nonparametric methods of density estimation and, in particular, the kernel method.

The idea of nonparametric density estimation can be traced back to an unpublished paper by Fix and Hodges (1951) who introduced a naive estimator using the property that

$$f(x)=\lim_{h \to 0} (F(x+h/2)-F(x-h/2))/h$$
$$=P(x-h/2 \leq X \leq x+h/2)$$

where F(x) is the cumulative probability distribution of X. Fix and Hodges defined a naive estimator of f(x) given by

$$\hat{f}_T(x)= 1/Th(\text{number of } x_1,x_2,...,x_T \text{ in}[x-h/2,x+h/2]) \qquad (1.9)$$

This can be rewritten as

$$\hat{f}_T(x)=(1/Th) \sum W(x_i-x)/h \qquad (1.10)$$

where $W(\ ) = 1$ if $abs(x_i-x)/h <1/2$
$= 0$ otherwise

Clearly the weight function W is such that $\int(W(z)dz = 1$ where $z_i =(x_i-x)/h$. A major weakness of the naive estimator is that the relevant weight function is discontinuous and this may result in a discontinuous estimate of the density function. To remedy this problem, Rosenblatt (1956) formally introduced the kernel estimator. In particular, Rosenblatt generalized the above estimator by choosing any weight function K, called the kernel function, and obtaining an estimate given by

$$\hat{f}_T(x)=(1/Th)\sum K(x_i-x)/h \qquad (1.11)$$

where K(.) is a non-negative integrable weight function which satisfies, among other conditions, the following: $K(z) \geq 0$; $\int K(z)dz = 1$; $\int zK(z)dz= 0$ where z is as defined before. The quantity h is called the window width and it is chosen such that it approaches 0 as T tends to infinity. The kernel function determines the shape of the curve and the window width determines the smoothness of the curve. If K(.) is continuous, the resultant estimate will also be continuous.

Parzen(1962) extended the analysis to cases where the kernel function need not be non-negative. Any Borel measurable function K(z) satisfying $\int K(z)dz=1$; $\sup|K(z)| < \infty$ and $\int|zK(z)| = 0$ will result in a kernel estimator which is mean square error consistent provided that h_T ---> 0 and Th_T ---> 0 as T tends to infinity.

Silverman (1986) reviewed alternative methods of non-parametric density estimation such as orthogonal series, histograms, maximum penalized likelihood estimators and nearest neighbourhood estimators. However, the kernel method is the most widely used in applied work because it is simple and its properties are relatively well known. It is for these reasons that the kernel method is used to estimate the sampling distributions of the estimators included in this study. Multivariate generalizations of the kernel estimator of the density function were discussed by Cacoullos (1966), among others.

Two main issues arise in practical applications of the kernel method, namely:

1.the choice of the kernel function; and

2.the choice of the window width

Such choices are often based on the asymptotic and/or small sample properties. The most widely used criterion for choosing the kernel function and window width is the mean integrated square error (MISE) defined by

MISE $= \int[\hat{f}_T(x)-f(x)]^2dx$. The MISE measures the global accuracy of the density estimates.

The asymptotic properties are usually based on regularity assumptions about the kernel function and the density function. In addition, we require that the window width h_T depends on T in some way.

In a review article Ullah(1988) outlined some results regarding the choice of the kernel function and the window width in density and regression function estimation from which two important conclusions emerged:

1.The results are not very sensitive to the choice of the kernel function. Any symmetric probability density function (e.g. the standard normal distribution) will suffice.

2.The choice of the window width is relatively more important because of the possibility of a trade off between the bias and the variance. However, Ullah pointed out that a good practical choice of the window width for density estimation in the univariate case is $sT^{-1/5}$, where s is the standard deviation of the sample observations.

In a Monte Carlo study the kernel density estimation method can be used to estimate the distribution of the structural parameters or prediction errors by treating the estimates of each structural parameter (or post-sample prediction errors for each endogenous variable) in all the replications as being i.i.d. observations from their corresponding small sample distributions. In the kernel method the distribution is estimated at selected points. By plotting this distribution one gets an idea of the moments of the distribution. Also, by overlaying the distributions corresponding to different sample sizes, one can visually identify the effect of changes in the sample size on the distribution. Furthermore, it can be shown that the $(1-\alpha)100$ percent confidence interval for $f(x)$ at each point x is given by

$$\hat{f}_T(x) + Z_{\alpha/2}[(1/Th)\hat{f}_T(x)\int K^2(t)\,dt]^{1/2} \qquad (1.12)$$

where $Z_{\alpha/2}$ is the value of Z which leaves an area of $\alpha/2$ on its right under the standard normal curve.

In this study the kernel method is used to estimate and compare the sampling distributions of Hatanaka's two-step estimators and their corresponding distributions of post-sample prediction errors.

1.5 Objectives and plan of this study

In this study the Monte Carlo experiments are conducted to investigate the small sample properties of the three limited information estimators and the three full information estimators proposed by Hatanaka (1976) for dynamic SEMs with autoregressive errors.

As pointed out in Dhrymes and Taylor(1976) the limited information procedures cease to be single equation procedures if the autoregressive matrix is assumed to be non-diagonal. Consequently the design of the Monte Carlo experiments in this study is based on the assumption of a diagonal autoregressive matrix.

The study is divided into two broad parts, namely, structural estimation and prediction.

For structural estimation, the following questions are addressed:

1. How do the small sample properties of these estimators compare in estimating the structural parameters?
2. How reliable are the asymptotic standard errors for purposes of inference in small sample situations?
3. How do the kernel estimates of the sampling distributions of the structural parameters compare?

With respect to prediction the following questions are addressed:

1. How do the small sample properties of these estimators compare in dynamic simulation forecasts?

2. How reliable are the formulae for the asymptotic covariance matrix of forecasts for testing hypotheses about predictions in small sample situations?

3. How do the kernel estimates of the sampling distributions of the dynamic simulation forecasts compare?

Chapter 2 reviews the two step estimators investigated in this study. Also the full information maximum likelihood estimators and the limited information maximum likelihood estimators and their asymptotic distributions are briefly derived. In chapter 3 the asymptotic distribution of dynamic simulation forecasts for dynamic SEM's with autocorrelated errors is derived. In chapter 4 the design of experiments conducted are described. The various nonparametric statistics used to compare the relative small sample properties are spelled out. The chapter concludes with a review of the results of previous Monte Carlo studies which addressed the same issues. In chapter 5 the results of individual experiments performed are summarized. Finally, chapter 6 reports on the conclusions drawn from this study.

CHAPTER 2: STRUCTURAL ESTIMATION

2.1 Introduction

Efficient estimates of model (1.1) specified in chapter 1 may be obtained using the maximum likelihood procedure. Specific algorithms for computing the maximum likelihood estimates of the parameters of the model have been proposed by Hendry (1972) and Chow and Fair (1973). The entire model may be estimated using full-information maximum likelihood (FIML) procedure. Alternatively, each equation in the system may be estimated using limited information maximum likelihood (LIML) procedure. The FIML and the LIML estimators and their asymptotic distributions are derived in Hatanaka (1976). However, Hatanaka does not provide detailed derivations of these estimators and their asymptotic distributions. Brief derivations of the FIML estimator of model (1.1) and its asymptotic distribution and the LIML estimator and its asymptotic distribution are presented in sections 2.2 and 2.3, respectively. The derivations presented in sections 2.2 and 2.3 are essentially Hatanaka's. In section 2.4 the two-step estimators proposed by Hatanaka whose properties are investigated in this study are described.

2.2 Derivation of the full-information maximum likelihood estimator and its asymptotic distribution

Given model 1.1, and assuming normality of errors, the likelihood of observing e is

$$f(e) = (\sqrt{2\Pi})^{-Tg}|\Sigma \otimes I|^{-1/2} \exp -1/2\ e'(\Sigma^{-1} \otimes I)\ e \qquad (2.1)$$

where $|.\ |$ denotes the determinant.

To transform the likelihood from the e-space to the y-space, we need the Jacobian of transformation and it is easy to see that for this particular case the Jacobian matrix is

$\partial e/\partial y = \ |I_T \otimes (I - B)| = |I-B|^T$

Also $|\Sigma \otimes I|^{-1/2} = |\Sigma|^{-T/2}$

$e = u - (R' \otimes I)u_{-1}$

$\quad = y - (R' \otimes I)y_{-1} - [(I \otimes Z) - (R' \otimes Z_{-1})]S\delta$

The log-likelihood function, in terms of y, is, therefore, given by

$$\ln L = (-Tg/2)\ln 2\Pi - (T/2)\ln|\Sigma\ | + T\ln|I-B|$$
$$- (1/2)[y-(I\otimes Z)S\delta(R'\otimes I)(y_{-1}-(I\otimes Z_{-1})S\delta](\Sigma^{-1}\otimes I)\ x$$
$$[y-(I\otimes Z)S\delta-(R'\otimes I)\ (y_{-1}-(I \otimes Z_{-1})S\delta] \qquad (2.2)$$

and

$$\ln L/T = (-g/2)\ln 2\Pi + (1/2)\ln|\Sigma^{-1}| + \ln|I-B|$$
$$-(1/2T)[y - (I \otimes Z)S\delta - (R' \otimes I)(y_{-1} -(I \otimes Z_{-1})S\delta](\Sigma^{-1} \otimes I)[y - (I \otimes Z)S\delta - (R' \otimes I)(y_{-1} - (I \otimes Z_{-1})S\delta] \qquad (2.3)$$

Using the property that tr ABC = (vecC')'(B' $\otimes$ I)vecA (see

Neudecker(1969), p.954), it is easy to show that

$e'(\Sigma^{-1} \otimes I)e = tr\Sigma^{-1}E'E = \Sigma\sigma^{ij}e_{.i}'e_{.j}$

Differentiating (2.3) with respect to Σ^{-1} and equating their derivatives to zero yields $\hat{\Sigma} = E'E/T$

Thus the concentrated log-likelihood function is given by

$T^{-1}lnL* = -(g/2)ln2\Pi + ln|I-B| - (1/2)ln|\hat{\Sigma}| - (1/2)Te'(\hat{\Sigma}^{-1} \otimes I)e$

But $(1/2T)e'(\Sigma^{-1} \otimes I)e = tr\hat{\Sigma}^{-1}E'E/2T = trT\hat{\Sigma}^{-1}\hat{\Sigma}/2 = g/2$

$$\text{Thus } T^{-1}lnL* = -g/2[ln2\Pi + 1] + ln|I-B| - 1/2ln|\hat{\Sigma}| \qquad (2.4)$$

$$= c \quad + \quad p(\theta) \quad + q(\theta)$$

where $\theta = [\delta' \ r']'$, $c = -g/2[In2\Pi + 1]$ and $r = vec\ R$

(2.4) should be maximized with respect to the unknown elements of B, C_0 , C_1 and R.

Differentiating (2.4) w.r.t δ, we obtain

$$T^{-1}\partial lnL*/\partial\delta = T^{-1}[S'(I\otimes Z')(\hat{\Sigma}^{-1}\otimes I)e - S'(R'\otimes Z_{-1})(\hat{\Sigma}^{-1} \otimes I)\ e] + c$$

$$= T^{-1}S'[(I\otimes Z') - (R'\otimes Z_{-1})'](\hat{\Sigma}^{-1} \otimes I)e + c \qquad (2.5)$$

where

$c = [c_1' \ c_2' \ldots c_g']'$and $c_i = \partial ln|I-B/\partial\beta_{.u} = S_{u1}\partial ln|I-B|/\partial b_{.u}$

Also differentiating (2.4) w.r.t. r, we obtain

$$T^{-1}\partial lnL*/\partial r = T^{-1}[\hat{\Sigma}^{-1} \otimes U_{-1}']e \qquad (2.6)$$

To derive the asymptotic covariance matrix, we need to obtain

$\text{plim } -T^{-1}\partial^2 \ln L^*/\partial\theta\partial\theta'$

It turns out that

$$\text{plim } T^{-1}\partial^2 \ln L/\partial\delta\partial\delta'$$

$$= -\text{plim}T^{-1}S'[(I\otimes\overline{Z})-(R\otimes Z_{-1})]'(\Sigma^{-1}\otimes I)[(I\otimes\overline{Z})-(R'\otimes Z_{-1})S] \quad (2.7)$$

where $\overline{Z}=[\overline{Y}\ Y_{-1}\ X]$ and $\overline{Y} = A\Pi$ from formula 1.5 in chapter 1.

$$\text{Also plim } T^{-1}\ \partial^2 \ln L^*/\ \partial r\partial r' = -[\Sigma^{-1}\otimes\Omega] \quad (2.8)$$

Similarly, $\text{plim } T^{-1}\partial^2 \ln L^*/\partial r\partial\delta'$

$$= -(I\otimes U_{-1}')(\Sigma^{-1}\otimes I)(I\otimes Z)S + (I\otimes U_{-1}')(\Sigma^{-1}R'\otimes I)(I\otimes Z^{-1})S$$

$$= -(I\otimes U_{-1}')(\Sigma^{-1}\otimes I)(I\otimes\overline{Z})S + (I\otimes U_{-1}')(\Sigma^{-1}\otimes I)(R'\otimes I)(I\otimes Z_{-1})S$$

$$= -(I\otimes U_{-1}')(\Sigma^{-1}\otimes I)[(I\otimes\overline{Z}) - (R'\otimes Z_{-1})]S \quad (2.9)$$

and

$$\text{plim } T^{-1}\partial^2 \ln L/\partial\delta\partial r' = -S'[(I\otimes\overline{Z})-(R'\otimes Z_{-1})](\Sigma^{-1}\otimes I)(I\otimes U_{-1}) \quad (2.10)$$

Finally combining (2.7), (2.8), (2.9) and (2.10) we have established that

$$-\text{plim}T^{-1}\partial^2 \ln L^*/\partial\theta\partial\theta' = \begin{bmatrix} M & P1 \\ P1' & P2 \end{bmatrix} \quad (2.11)$$

where

$$M = \text{plim } T^{-1}S'[(I\otimes\overline{Z})-(R'\otimes Z_{-1})]'(\Sigma^{-1}\otimes I)[(I\otimes\overline{Z})-(R'\otimes Z_{-1})]S$$

$$P1 = \text{plim } T^{-1}S'[(I \otimes \overline{Z}) - (R'\otimes Z_{-1})]'(\Sigma^{-1}\otimes I)(I \otimes U_{-1})$$

$$P2 = \Sigma^{-1} \otimes \Omega$$

The inverse of the matrix in (2.11) is the asymptotic covariance matrix of the FIML. If some elements of R are specified a priori to be zero then we delete the rows and columns of the information matrix (2.11) which correspond to the elements of R which are prespecified to be zero before inverting. It follows that if R is diagonal then replace the sub-matrix $(I \otimes U_{-1})$ in (2.11) by $\text{diag}[u_{.1,-1} \ . \ . \ . \ u_{.g,-1}]$ and the asymptotic covariance of the unknown elements of B, C_0, C_1 and R is given by the inverse of the information matrix after making the appropriate substitutions.

Computing the maximum likelihood estimates involves solving non-linear equations. To this end iterative procedures are required. Furthermore, the convergence of these procedures are usually not guaranteed or rather the speed of convergence may not be fast enough. These computational difficulties associated with maximum likelihood estimators are partly resolved using linearized versions of the maximum likelihood estimator similar to the one proposed in Rothenberg and Leenders (1964), which are easily computable if we start from initial consistent estimates and which require only one iteration if the Hessian matrix is known. The linearized maximum likelihood estimator simplifies, from a computational point of view, the method of Newton which may require more than one iteration. Alternatively, if determining the information matrix is not difficult, we may use the method of scoring in which the Hessian in Newton's method is replaced by the information matrix. Clearly if one starts from an initial consistent estimate, only one iteration is required for the scoring estimator.

It is easy to prove that the linearized maximum likelihood and the scoring estimator have the same asymptotic distribution as the FIML estimator.

2.3 Derivation of the limited information maximum likelihood estimator and its asymptotic distribution

The limited information estimators proposed by Hatanaka are based on the assumption that the R matrix is diagonal. As mentioned above these procedures would not be termed single equation procedures if the R matrix is not diagonal. This is because limited information methods use only the information pertaining to the equation of interest and ignore a priori restrictions on the parameters of the remaining equations. Ignoring the restrictions related to other equations does not only simplify the calculations but also constitutes the essential difference between LIML and FIML. In particular, we assume that the i-th diagonal element is non-zero and the rest of the elements in the i-th column are specified a priori to be zero. For i = 1 this is identical to Amemiya's (1966) Λ_0 specification.

To derive the limited information maximum likelihood estimator of, say, the first equation it is necessary to eliminate the parameters of all the other equations by concentrating them out.

The reduced form of model (1.1) can be rewritten in a more convenient form as

$$Z^*A = E \tag{2.12}$$

where $A = [A^{0\prime}\ A^{1\prime}\ A^{2\prime}\ A^{3\prime}\ A^{4\prime}]'$ and $Z^* = [Y\ Y_{-1}\ Y_{-2}\ X\ X_{-1}]$; $A^0 = (I-B)$, $A^1 = \{C_0+(I-B)R\}$, $A^2 = C_0R$, $A^3 = -C_1$ and $A^4 = C_1R$ Furthermore, we partition $A^i = [a_{.1}{}^i\ A_1{}^i]$, $i = 0, 1, \ldots, 4$ and $E = [e_{.1}\ E_1]$

Therefore, using the new partition, the first equation may be specified as

$$Ya_{.1}{}^0 + Y_{-1}a_{.1}{}^1 + Y_{-2}a_{.1}{}^2 + Xa_{.1}{}^3 + X_{-1}a_{.1}{}^4 = e_{.1} \tag{2.13}$$

The remaining g-1 equations may be specified collectively as

$$YA_1{}^0 + Y_{-1}A_1{}^1 + Y_{-2}A_1{}^2 + XA_1{}^3 + X_{-1}A_1{}^4 = E_1 \tag{2.14}$$

In order to disregard the a priori specification of the second through the g-th columns of (I-B), C_0, C_1 and R, we need to transform the system such that

1) the sub-systems (2.13) and (2.14) are mutually independent; and

2) we do not disturb the parameters of the first equation

This is achieved through a transformation H such that

$AH = [a_{.1}\ A_1]H = [a_{.1}\ A_1^*]$

$$H'\Sigma H = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & I_{g-1} \end{bmatrix}$$

where $a_{.1} = [a_{.1}^{0\prime}\ a_{.1}^{1\prime}\ a_{.1}^{2\prime}\ a_{.1}^{3\prime}\ a_{.1}^{4\prime}]'$

and $A_1^* = [A_1^{0*\prime}\ A_1^{1*\prime}\ A_{12}^{*\prime}\ A_1^{3*\prime}\ A_1^{4*\prime}]'$

This transformation leaves the parameters of the first equation unchanged and renders the first equation independent of the g-1 transformed equations. The transformation is possible [See for example Schmidt (1976, Lemma 17, p.187) or Dhrymes 1970, pp. 330-332).

Given the independence of the T rows of E it is easy to see that

$$E(e_{t.}H)'(e_{s.}H) = \delta_t^s H'\Sigma H \qquad t,s = 1,2, \ldots, T \qquad (2.15)$$

where $\delta_t s$ denotes the Kronecker delta.

From (2.15) we conclude that the vectors $e_{t.}H$ are mutually independent.

Since transformation by the non-singular matrix H does not affect the value of the likelihood function, we can write the log-likelihood function in terms of the transformed system as

$$\ln L = -(gT/2)\ln 2\Pi \quad -(T/2)\ln |H'\Sigma H| \quad + T\ln|(I - B)H|$$
$$- (1/2)\mathrm{tr}(H'\Sigma^{-1}H)(Z*AH)'(Z*AH) \qquad (2.16)$$

Note that the Jacobian of transformation from $e_{t.}H$ to $y_{t.}$ is given by $J = |I\otimes A_0 H| = |I\otimes (I-B)H|$. Using the above properties of H, the log-likelihood (2.16) is

$$\ln L = -(gT/2)\ln 2\Pi \ -(T/2)\ln\sigma_{11} \ +T\ln|a_{.1}{}^0 \ A_1{}^0{*}|$$
$$-(1/2\sigma_{11})a_{.1}{}'Z*'Z*a_{.1} - 1/2\mathrm{tr}A_1{*}'Z*'Z*A_1{*} \qquad (2.17)$$

We eliminate A_1* from (2.17) by maximizing with respect to its elements, ignoring all restrictions on A_1*

$\partial \ln L/\partial A_1*$

$$= T\partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|\partial A_1{*} \ -(1/2)\partial\mathrm{tr}\ A_1{*}'Z*'Z*A_1{*}/\partial A_1{*} \qquad (2.18)$$

But $\partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial A_1{*} = \begin{bmatrix} \partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial A_1{}^0{*} \\ \cdots\cdots\cdots \\ \partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial A_1{}^4{*} \end{bmatrix}$

and $\partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial A_1{}^i{*} = 0$ for $i = 1,2,3,4$

The only derivative to evaluate is $\partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial A_1{}^0{*}$

But $\partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial[a_{.1}{}^0 \ A_1{}^0{*}| = [a_{.1}{}^0 \ A_1{}^0{*}|'^{-1}$

$$= \partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial a_{.1}{}^0 \ \partial\ln|a_{.1}{}^0 \ A_1{}^0{*}|/\partial A_1{}^0{*}$$

Using the fact that $\partial|A|/\partial A = A'^{-1}$, it follows that

$$\partial \ln|a_{.1}^{0}\ A_1^{0}*|/\partial A_1^{0}*= [a_{.1}^{0}\ A_1^{0}*|'^{-1}[0\ \ I_{g-1}]'=J_2 \tag{2.19}$$

The matrix $[0\ \ I_{g-1}]'$ eliminates the first column of $[a_{.1}^{0}\ A_1^{0}*]'^{-1}$ and therefore the matrix J_2 essentially consists of the last g-1 columns of $[a_{.1}^{0}\ \ A_1^{0}*]'^{-1}$.

Also, using the fact that $\partial \text{tr}\ X'AX/\partial X = 2AX$, it follows that

$\partial \text{tr}\ A_1*'Z*'Z*A_1*/\partial A_1* = 2Z*'Z*A_1*$

Substituting these in (2.18) yields

$$\partial \ln L/\partial A_1* = [(TJ_2\ )'\ 0'\ 0'\ 0'\ 0']'$$
$$-Z*'Z*\ [\ A_1^{0}*'A_1^{1}*\ '\ A_1^{2}*\ '\ A_1^{3}*\quad A_1^{4}*'\ ]' = 0 \tag{2.20}$$

From (2.20) it follows that $TJ_2 = Z*'Z*A_1^{0}*$. Furthermore, it can be shown that

$A_1*'Z*'Z*A_1* = T \times I_{g-1}$

Thus $\text{tr}\ A_1*'Z*'Z*A_1* = T\text{tr}\ I_{g-1} = T(g-1)$

and $-(1/2)\text{tr}\ A_1*'Z*'Z*A_1* = -(T/2)(g-1)$ which is a constant.

Also disentangling the $Z*'Z*$ matrix in (2.20) and noting that $Z* = [Y\ Y_{-1}\ Y_{-2}\ X\ X_{-1}] = [Y\ \phi]$ we have

$$\begin{bmatrix} TJ_2 \\ 0 \end{bmatrix} = \begin{bmatrix} Y'Y & Y'\Phi \\ \Phi'Y & \Phi'\Phi \end{bmatrix} \begin{bmatrix} A_1^{0}* \\ A_1** \end{bmatrix} \tag{2.21}$$

where $\Phi = [Y_{-1}\ Y_{-2}\ X\ X_{-1}]$ and $A_1** = [A_1^{1}*\ A_1^{2}*\ A_1^{3}*\ A_1^{4}*]'$

(2.21) can be written in expanded form as

$$\begin{aligned} TJ_2 &= Y'YA_1^{0}* + Y'\Phi A_1** \\ 0 &= \Phi'YA_1^{0}* + \Phi'\Phi A_1** \end{aligned} \tag{2.22}$$

Eliminating A_1** from the second equation in (2.22) yields

$$A_1^{**} = -(\Phi'\Phi)^{-1}\Phi'YA_1^{0*}$$

Substituting this in the first equation in (2.22) yields

$$TJ_2 = Y'YA_1^{0*} - Y'\Phi(\Phi'\Phi)^{-1}\Phi'YA_1^{0*}$$
$$= [Y'Y - Y'\Phi(\Phi'\Phi)^{-1}\Phi'Y]A_1^{0*}$$
$$= WA_1 0^* \quad \text{where } W = [Y'Y - Y'\Phi(\Phi'\Phi)^{-1}\Phi'Y]$$

Therefore $J_2 = W'A_1^{0*}/T$.

We can derive the identity

$$T\ln|a_{.1}^0\ A_1^{0*}| = \{(T/2)\ln|a_{.1}^0\ A_1^{0*}|\}^2 + (T/2)\ln|W| - (T/2)\ln|W|$$
$$= T/2\ln\{|a_{.1}^0\ A_1^{0*}|^2.|W|\} - T/2\ln|W| \qquad (2.23)$$

But $|W|$ is a constant in the sense that it does not depend on any parameters we are interested in estimating.

Also $|a_{.1}^0\ A_1^{0*}|^2.|W| = |[a_{.1}^0\ A_1^{0*}]'|\ |W|\ |a_{.1}^0\ A_1^{0*}|$

$$= \begin{bmatrix} a_{.1}^{0\prime}Wa_{.1}^0 & a_{.1}^{0\prime}WA_1^{0*} \\ A_1^{0*\prime}Wa_{.1}^0 & A_1^{0*\prime}WA_1^{0*} \end{bmatrix}$$

Now consider the individual components of this matrix.

It can be shown that $a_{.1}^{0\prime}WA_1^{0*} = 0$

Also $A_1^{0*\prime}WA_1^{0*} = I_{g-1}$

Substituting the results in (2.23) we conclude that

$$T\ln|a_{.1}^0\ A_1^{0*}| = (T/2)\ln|a_{.1}^0 Wa_{.1}^0| - (T/2)\ln|W| \qquad (2.24)$$

Inserting the maximizing values of (2.24) and into (2.17) we obtain the concentrated log-likelihood function

$$\ln L^* = c - (T/2)\ln\sigma_{11} + (T/2)\ln|a_{.1}^{0\prime}Wa_{.1}^{0}|$$
$$- (1/2\sigma_{11})a_{.1}'Z^{*\prime}Z^*a_{.1} \qquad (2.25)$$

where $c = -(Tg/2)(\ln 2\Pi + 1) + (T/2)(1 - \ln|W|)$

The concentrated log-likelihood function (2.25) is now expressed in terms of the parameters of the first equation and we have to maximize it, imposing all a priori restrictions related to the first equation.

Differentiating (2.25) with respect to σ_{11} we have

$\partial \ln L^*/\partial\sigma_{11} = -(T/2\sigma_{11}) + (1/2\sigma_{11}^2)a_{.1}'Z^{*\prime}Z^*a_{.1} = 0$

Thus $\sigma_{11} = a_{.1}'Z^{*\prime}Z^*a_{.1}/T$

Concentrating out σ_{11} from (2.25), we obtain

$$\ln L^{**} = c-(T/2)\ln a_{.1}'Z^{*\prime}Z^*a_{.1}-(T/2)\ln T+(T/2)\ln|a_{.1}^{0}Wa_{.1}^{0}|-T/2$$
$$= c_1-(T/2)\ln a_{.1}'Z^{*\prime}Z^*a_{.1}/T + (T/2)\ln|a_{.1}^{0}Wa_{.1}^{0}| \qquad (2.26)$$

where $c_1 = (Tg/2)(\ln 2\Pi + 1) - (T/2)(\ln|W| + \ln T)$

Now $a_{.1}^{0} = [1\ 0\ 0\ .\ .\ .\ 0]' - b_{.1} = e_1 - S_{11}\beta_{.1}$

Noting that $a_{.1}^{1} = -S_{12}\gamma_{0.1} - r_{11}e_1 + r_{11}S_{11}\beta_{.1}$; $a_{.1}^{2} = r_{11}S_{12}\gamma_{0.1}$; $a_{.1}^{3} = -S_{13}\gamma_{1.1}$ and $a_{.1}^{4} = r_{11}S_{13}\gamma_{1.1}$, we may write

$Z^*a_{.1} = [Y_{.1} - ZS_1\delta_{.1}] - r_{11}[Y_{.1,-1} - r_{11}Z_{-1}S_1\delta_{.1}]$

This implies that

$$a_{.1}'Z^{*\prime}Z^*a_{.1} = \{[Y_{.1} - ZS_1\delta_{.1}] - r_{11}[Y_{.1,-1} - r_{11}Z_{-1}S_1\delta_{.1}]\}'$$
$$\{[Y_{.1} - ZS_1\delta_{.1}] - r_{11}[Y_{.1,-1} - r_{11}Z_{-1}S_1\delta_{.1}]\} \qquad (2.27)$$

Substituting (2.27) into the concentrated log-likelihood function (2.26), we obtain

$$L^{**} = c_1-(T/2)\ln \{ [y_{.1} - ZS_1\delta_{.1}] - r_{11}[y_{.1,-1} - r_{11}Z_{-1}S1\delta_{.1}]\}' \{[y_{.1} - ZS_1\delta_{.1}]-r_{11}[y_{.1,-1} - r_{11}Z_{-1}S_1\delta_{.1}]\} +(T/2)\ln(e_1-S_{11}\beta_{.1})'W(e_1-S_{11}\beta_{.1}) \qquad (2.28)$$

A slight rearrangement of (2.28) yields

$$\ln L^{**} = c + (T/2)\ln (e_1 - S_{11}\beta_{.1})'W(e_1 - S_{11}\beta_{.1}) -(T/2)\ln[(y_{.1} - r_{11}y_{.1,-1})+Y_1{*}\beta_{.1} -X_1{*}\gamma_{0.1} -Y_{-11}{*}\gamma_{1.1}]' [(y_{.1} -r_{11}y_{.1,-1}) +Y_1{*}\beta_{.1} -X_1{*}\gamma_{0.1} -Y_{-11}{*}\gamma_{1.1}] \qquad (2.29)$$

where $Y_1{*} =Y_1-r_{11}Y_{-1\ 1}$, $X_1{*} =X_1-r_{11}X_{-11}$, $Y_{-11}{*} =Y_{-11}-r_{11}Y_{-21}$.

$$\ln L^{**} = c_1+(T/2)\ln(e_1-S_{11}\beta_{.1})W(e_{.1}-S_{11}\beta_{.1})-(T/2)\ln[Y{*}(e_1-S_{11}\beta_{.1}) -Y_{-1}{*}S_{12}\gamma_{0.1}-X{*}S_{13}\gamma_{1.1}]'[Y{*}(e_1-S_{11}\beta_{.1})-Y_{-1}{*}S_{12}\gamma_{0.1}-X{*}S_{13}\gamma_{1.1}] \qquad (2.30)$$

Using the partitioned matrix $X_1{*}=[(Y_{-1}-r_{11}Y_{-2})S_{12}:(X-r_{11}X_{-1})S_{13}]$ and letting $\gamma_{.1} = [\gamma_{0.1}' \ \ \gamma_{1.1}']'$, the vector of coefficients of the predetermined variables included in the first equation, and $Y{*} = (Y - r_{11}Y_{-1})$, the concentrated log-likelihood function (2.30) can be written as

$$\ln L^{**} = c_1 + (T/2)\ln(e_1 - S_{11}\beta_{.1})'W(e_1 - S_{11}\beta_{.1})- (T/2)\ln[Y{*}(e_1-S_{11}\beta_{.1})-\gamma_{0.1}-X_1{*}\beta_{.1}]'[Y{*}(e_1-S_{11}\beta_{.1})-X_1{*}\gamma_{.1}] \qquad (2.31)$$

Differentiating (2.31) w.r.t $\gamma_{.1}$, we obtain

$\gamma_{.1} = -(X_1*'X_1*)^{-1}X_1*'Y*(e_1 - S_{11}\beta_{.1})$ and substituting it in (2.31) we get $\ln L** = c + (T/2)\ln(e_1-S_{11}\beta_{.1})'W(e_1-S_{11}\beta_{.1})$
$-(T/2)\ln(e_1-S_{11}\beta_{.1})'Y*'[I - X_1*(X_1*'X_1*)^{-1}X_1*]Y*$

$= c + (T/2)\ln(e_1-S_{11}\beta_{.1})'W(e_1-S_{11}\beta_{.1})$

$\quad - (T/2)\ln(e_1-S_{11}\beta_{.1})'W_1(e_1 - S_{11}\beta_{.1})$

where $W_1 = Y*'(I - X_1*(X_1*'X_1*)^{-1}X_1*')Y*$

$$=c+(T/2)\ln[(e_1-S_{11}\beta_{.1})'W_1(e_1-S_{11}\beta_{.1})/(e_1-S_{11}\beta_{.1})'W(e_1-S_{11}\beta_{.1})] \qquad (2.32)$$

It should be noted that (2.32) takes the form in Amemiya (1966, equation 14, p. 288) adjusted for the presence of lagged dependent variables.

The LIML estimator of $\delta_{.1}$ and r_{11} are obtained by maximizing the concentrated log-likelihood function under an appropriate normalization rule using a similar procedure to that outlined in Amemiya (1966), namely:

1.Find the smallest root $\hat{\lambda}*$ of the determinantal equation $|W - \lambda W_1| = 0$ where W* and W_1 are as defined above.

2.Find $\hat{\beta}_{.1}$ for which $(W - \hat{\lambda}W_1)\hat{\beta}_{.1} = 0$ subject to an arbitrary normalization rule. i.e. $\hat{\beta}_{.1}$ is the eigenvector corresponding to the smallest characteristic root in step(1).

3.Substitute $\hat{\beta}_{.1}$ in the expression $(X_1*'X_1*)^{-1}X_1*Y*\hat{\beta}_{.1}$ and this is the estimator of $\hat{\gamma}_{.1}$.

4.Substitute the values of $\hat{\beta}_{.1}$ and $\gamma_{.1}$ to get an estimate of r_{11} given by $\hat{r}_{11} = [Y_{-1}(e_1 - S_{11}\hat{\beta}_{.1})-Y_{-2}S_{12}\hat{\gamma}_{0.1} - X_{-1}S_{13}\hat{\gamma}_{1.1}]'$ $[Y(e_1 -$

$S_{11}\hat{\beta}_{.1}) - Y_{-1}S_{12}\hat{\gamma}_{0.1} - XS_{13}\hat{\gamma}_{1.1}] / [Y_{-1}(e_1 - S_{11}\hat{\beta}_{.1}) - Y_{-2}S_{12}\hat{\gamma}_{0.1} - X_{-1}S_{13}\hat{\gamma}_{1.1}]'$
$[Y_{-1}(e_1 - S_{11}\hat{\beta}_{.1}) - Y_{-2}S_{12}\hat{\gamma}_{0.1} - X_{-1}S_{13}\hat{\gamma}_{1.1}]$

It can be shown that the LIML estimator is consistent and asymptotically efficient within the limited information class.

To establish the asymptotic covariance matrix of LIML we take the second derivatives of (2.31) and obtain $-\text{plim } T^{-1}\partial^2 L^{**}/\partial\theta_1\partial\theta_1'$, where $\theta_1 = [\delta_{.1}' \; r_{11}]'$.

It turns out that $-\text{plim } T^{-1}\partial^2 L^{**}/\partial\delta_{.1}\partial\delta_{.1}'$

$$=1/\sigma_{11}\text{plim}T^{-1}[S_1'(\bar{Z}-r_{11}Z_{-1})'][(\bar{Z}-r_{11}Z_{-1})S_1] \qquad (2.33)$$

Also $-\text{plim } T^{-1}\partial^2 \ln L^{**}/\partial\delta_{.1}\partial r_{11}$

$$= [(ZS_1)' - r_{11}(Z_{-1}S_1)']u_{.1,-1}/e_{.1}'e_{.1}$$

$$= 1/\sigma_{11}\text{plim } T^{-1}[(\bar{Z}S_1)' - r_{11}(Z_{-1}S_1)']u_{.1,-1} \qquad (2.34)$$

Also $-\text{plim}T^{-1}\partial^2 \ln L^{**}/\partial r_{11}^2 = \omega_{11}/\sigma_{11}$ (2.35)

Combining (2.33), (2.34) and (2.35) we obtain

$$-\text{plim}T^{-1}\partial^2 L^{**}/\partial\theta_1\partial\theta_1' = \begin{bmatrix} m_1 & p_1 \\ p_1' & p_2 \end{bmatrix} \qquad (2.36)$$

where $m_1 = (1/\sigma_{11})\text{plim}T^{-1}[S_1'(\bar{Z}-r_{11}Z_{-1})'][(\bar{Z}-r_{11}Z_{-1})S_1]$

$p_1 = (1/\sigma_{11})\text{plim}T^{-1}[(\bar{Z}S_1)' - r_{11}(Z_{-1}S_1)']u_{.1,-1}$

and $p_2 = \omega_{11}/\sigma_{11}$

The inverse of the matrix given in (2.36) is the asymptotic covariance matrix of the LIML estimator.

It is apparent from the above discussion that obtaining LIML is a tedious process. The three two-step limited information estimators proposed by Hatanaka have the same asymptotic distribution as the LIML and yet are much easier to compute.

In the next section we discuss the six two-step estimators proposed by Hatanaka whose small-sample properties are investigated in this study. Three of these are full-information estimators which are denoted briefly as HF1, HF2 and HF3. The remaining three are limited information estimators denoted briefly as HL1, HL2 and HL3. All these estimators are asymptotically efficient in the sense that they attain the Cramer-Rao lower bound.

2.4 Hatanaka's two-step estimators investigated in this study

The first stage which is common to all the estimators proposed involves instrumental variable estimation of model (1.1) using the appropriate instruments for the current and lagged endogenous variables. This yields consistent estimates of the structural coefficients in each of the equations in the system. For the i-th equation we obtain

$$\tilde{\delta}_{.i} = \{(WS_i)'ZS_i\}^{-1}(WS_i)'y_{.i} \tag{2.37}$$

where $W = [M\ M_{-1}\ X]$, the M denotes the $T \times g$ matrix of instruments for Y. These estimates are used to obtain the residuals, $\tilde{u}_{.i} = y_{.i} - ZS_i\tilde{\delta}_{.i} \quad i=1,2,\ldots,g$.

If some elements of R are specified a priori to be zero then the elements of $r_{.i}$ not specified a priori to be zero are estimated by $S_{ri}'\tilde{r}_{.i} = (S_{ri}'\tilde{U}_{-1}'\tilde{U}_{-1}S_{ri})^{-1}S_{ri}'\tilde{U}_{-1}'\tilde{u}_{.i}$.

It turns out that if R is diagonal then $\tilde{r}_{ii} = (\tilde{u}_{.i-1}'\tilde{u}_{.i})/\tilde{u}_{.i-1}\tilde{u}_{.i-1} \quad i = 1,2,\ldots,g.$ and $\tilde{R} = \text{diag}\ (\tilde{r}_{11},\ldots,\tilde{r}_{gg})$; $E = \tilde{U} - \tilde{U}_{-1}\tilde{R}$ and $\tilde{\Sigma} = \tilde{E}'\tilde{E}/T$. These are used to obtain restricted reduced form predictions, $\overline{Y}^R$, of the current endogenous variables which are used in the second stage for some estimators. For other estimators we need to obtain the unrestricted reduced form predictions, $\overline{Y}^u$, by applying ordinary least squares to (1.5). The first stage estimates are used to form the following matrices and vectors:

$\overline{Z}_i^{u+} = [\overline{Y}_i^u - \tilde{r}_{ii}Y_{-1i} : Y_{-1i} - \tilde{r}_{ii}Y_{-2i} : X_i - \tilde{r}_{ii}X_{-1i}]$

$\overline{X}^{u+} = [\text{diag}(\overline{Z}_1^{u+},\ldots,\overline{Z}_g^{u+}):\text{diag}(\tilde{u}_{.1-1},\ldots,\tilde{u}_{.g-1})]$

$\overline{Z}_i^{R+} = [\overline{Y}_i^R - \tilde{r}_{ii}Y_{-1i} : Y_{-1i} - \tilde{r}_{ii}Y_{-2i} : X_i - \tilde{r}_{ii}X_{-1i}]$

$\overline{X}^{R+} = [\text{diag}(\overline{Z}_1^{R+},\ldots,\overline{Z}_g^{R+}):\text{diag}(\tilde{u}_{.1-1},\ldots,\tilde{u}_{.g-1})]$

$\overline{Z}_i^{+} = [\overline{Y}_i - \tilde{r}_{ii}Y_{-1i} : Y_{-1i} - \tilde{r}_{ii}Y_{-2i} : X_i - \tilde{r}_{ii}X_{-1i}]$

$\overline{X}^{+} = [\text{diag}(Z_1^{+},\ldots,Z_g^{+}):\text{diag}(\tilde{u}_{.1-1},\ldots,\tilde{u}_{.g-1})]$

$\overline{X}_i^{u+} = [\overline{Z}_i^{u+}:\tilde{u}_{.i-1}]$, $\overline{X}_i^{R+} = [\overline{Z}_i^{R+}:\tilde{u}_{.i-1}]$, $\overline{X}_i^{+} = [\overline{Z}_i^{+}:\tilde{u}_{.i-1}]$ and $\tilde{r} = [\tilde{r}_{11}\ldots\tilde{r}_{gg}]'$

The six estimators differ in the second stage regressions as described below:

Full information estimators

HF1: a) Form the joint generalized least squares(GLS) estimator

$$[\bar{\delta}^{1\prime}\ \bar{r}^{1\prime}]' = [\bar{X}^{u+}(\tilde{\Sigma}^{-1} \otimes I)\bar{X}^{u+}]^{-1}\bar{X}^{u+}(\tilde{\Sigma}^{-1} \otimes I)(y - (\tilde{R}' \otimes I)y_{-1})$$

b) Construct the residual adjusted estimator

$$[\hat{\delta}^{1\prime}\ \hat{r}^{1\prime}]' = [\ \bar{\delta}^{1\prime}\ (\bar{r}^{1} + \tilde{r})']'$$

HF2: a) Form the joint GLS estimator

$$[\bar{\delta}^{2\prime}\ \bar{r}^{2\prime}]' = [\bar{X}^{R+}(\tilde{\Sigma}^{-1} \otimes I)\bar{X}^{+}]^{-1}\bar{X}^{R+\prime}(\tilde{\Sigma}^{-1} \otimes I)(y - (\tilde{R}' \otimes I)y_{-1})$$

b) Construct the residual adjusted estimator

$$[\hat{\delta}^{2\prime}\ \hat{r}^{2\prime}]' = [\bar{\delta}^{2\prime}\ (\bar{r}^{2} + \tilde{r})']'$$

HF3: a) Form the joint GLS estimator

$$[\bar{\delta}^{3\prime}\ \bar{r}^{3\prime}]' = [\bar{X}^{R+}(\tilde{\Sigma}^{-1} \otimes I)\bar{X}^{R+}]^{-1}\bar{X}^{R+\prime}(\tilde{\Sigma}^{-1} \otimes I)(\tilde{u}-(\tilde{R}'\otimes I)\tilde{u}_{-1})$$

b) Construct the residual adjusted estimator

$$[\hat{\delta}^{3\prime}\ \hat{r}^{3\prime}]' = [(\bar{\delta}^{3} + \tilde{\delta})'\ (\bar{r}^{3} + \tilde{r})']'$$

Limited information estimators of the i-th equation:

HL1: a) Form the OLS estimator

$$[\bar{\delta}_{.i}^{1\prime}\ \bar{r}_{ii}^{1\prime}]' = [\bar{X}_i^{u+\prime}\bar{X}_i^{u+}]^{-1}\bar{X}i^{u+\prime}(y_{.i}-\tilde{r}_{ii}y_{.i-1})$$

b) Construct the residual adjusted estimator

$$[\hat{\delta}_{.i}^{1\prime}\ \hat{r}_{ii}^{1\prime}]' = [\bar{\delta}_{.i}^{1\prime}\ (\bar{r}_{ii}^{1} + \tilde{r}_{ii})']'$$

HL2: a) Form the OLS estimator

$$[\bar{\delta}_{.i}^{2\prime}\ \bar{r}_{ii}^{2\prime}]' = [\bar{X}_i^{R+\prime}\bar{X}_i^{+}]^{-1}\bar{X}_i^{R+\prime}(y_{.i}- \tilde{r}_{ii}y_{.i-1})$$

b) Construct the residual-adjusted estimator

$$[\hat{\delta}.i^{2\prime}\ \hat{r}_{ii}^{2\prime}]' = [\bar{\delta}_{.i}^{2\prime}\ (\bar{r}_{ii}^{2} + \tilde{r}_{ii})']'$$

HL3: a) Form the OLS estimator

$$[\bar{\delta}_{.i}3'\ \bar{r}_{ii}^{3\prime}]' = [\bar{X}_i^{R+\prime}\bar{X}_i^{R+}]^{-1}\bar{X}_i^{R+\prime}(\tilde{u}_{.i} - \tilde{r}_{ii}\tilde{u}_{.i-1})$$

b) Construct the residual adjusted estimator

$$[\hat{\delta}_{.i}^{3\prime}\ \hat{r}_{ii}^{3\prime}]' = [(\bar{\delta}_{ii}^{3} + \tilde{\delta}_{.i})'\ (\bar{r}_{ii}^{3} + \tilde{r}_{ii})']'$$

In fact, HL1, HL2 and HL3 are limited information analogues of HF1, HF2 and HF3, respectively. The asymptotic covariance matrix for HF1, HF2 and HF3 are given by the inverse of (2.11) and for HL1, HL2 and HL3 by the inverse of the matrix in (2.36).

If asymptotic properties of these estimators are a useful guide to their small-sample properties, we would expect the full information estimators HF1, HF2 and HF3 to be more efficient than their limited information counterparts for both structural estimation and dynamic simulation forecasting. The purpose of the Monte Carlo study is to compare the performances of these estimators when small samples are used. In Chapter 3 we discuss the problem of dynamic simulation forecasting in dynamic SEM's with errors generated by a vector-autoregressive process.

CHAPTER 3: PREDICTION

3.1 Introduction

One of the objectives of specifying a dynamic SEM is for dynamic simulation forecasting, provided that the model which generated the observations in the sample period remains valid in the period of prediction. An important characteristic of the dynamic simulation forecasts is that the values of the lagged endogenous values are replaced by their corresponding forecasted values for purposes of post-sample predictions. It is needless to say that in the presence of autocorrelated errors, the standard formulae for the asymptotic distribution of the dynamic simulation forecasts for the uncorrelated errors (e.g. Schmidt (1974)) become invalid. Therefore, in order to make valid inferences about the reliability of the post-sample predictions, the standard asymptotic formulae must be adjusted to take into account the feature of autocorrelated errors in the model.

Baillie (1982) has derived the asymptotic distribution of the dynamic simulation forecasts from dynamic simultaneous equations models with autocorrelated errors in the general case using Taylor's expansion and recommends exploration of some of the recursions in the matrix elements of the asymptotically negligible terms for practical applications. In the next section we derive the asymptotic distribution of the dynamic simulation forecasts in the dynamic simultaneous models with errors generated by a first order vector autoregressive process. The approach adopted in deriving the asymptotic distribution of dynamic simulation forecasts is similar to that in Schmidt (1974) and makes use of the distribution of the reduced form parameters. If the distribution of the reduced form parameters can be determined, then the distribution of the dynamic simulation forecasts can easily be obtained. It turns out that the distribution of the restricted reduced form can easily be derived in the autoregressive case (see, for example Knight (1982)), where the reduced form is estimated by FIML. By setting R = 0, the formula specializes to the uncorrelated case considered by Schmidt (1974).

3.2 Asymptotic distribution of dynamic simulation forecasts

To derive the asymptotic distribution of dynamic simulation forecasts we rewrite the reduced form given in (1.5) in chapter 1 as follows:

$$y_{t.}^* = y_{t-1.}^*\Pi_1^* + x_{t.}^*\Pi_2^* + v_{t.}^* \tag{3.1}$$

where $y_{t.}^* = [y_{t.} \; y_{t-1.}]$, $y_{t-1.}^* = [\, y_{t-1.} \; y_{t-2.}]$,

$x_{t.}^* = [x_{t.} \; x_{t-1.}]$ and $v_{t.}^* = [v_{t.} \quad 0]$.

$$\Pi_1^* = \begin{bmatrix} \Pi_1 & I_g \\ \Pi_2 & 0 \end{bmatrix} \qquad \Pi_2^* = \begin{bmatrix} \Pi_3 & 0 \\ \Pi_4 & 0 \end{bmatrix}$$

I_g is the identity matrix of order g. Note that the first g columns realize the reduced form and the last g columns of (3.1) represent an identity.

On repeated substitution this yields

$$y_{t+h.}^* = y_{t.}^*\Pi_1^{*h} + \sum x_{t+h-j.}^*\Pi_2^*\Pi_1^{*j} + \sum v_{h-j.}^* \quad \Pi_1^{*j} \qquad (3.2)$$

Unless otherwise indicated all the summations in this chapter run from j=0 to j=h-1.

Accordingly, the vector of forecasts for the h-th post sample period is given by

$$\hat{y}_{h.}^* = y_{0.}^*\hat{\Pi}_1^{*h} + \sum x_{h-j.}^*\hat{\Pi}_2^*\hat{\Pi}_1^{*j} \qquad (3.3)$$

But $y_{h.}^* = [y_{h.} \quad y_{h-1.}]$ and this implies that $y_{h.} = y_{h.}^*D$ where D $= [I_g \quad 0]'$ is a matrix which selects the first g columns of the matrices in question.

Thus $\hat{y}_{h.} = y_{0.}^*\hat{\Pi}_1^{*h}D + \sum x_{h-j.}^*(\hat{\Pi}_2^*\hat{\Pi}_1^{*j})D$ (3.4)

(3.4) can be written in matrix notation as

$$\hat{y}_{h.} = W_h\hat{A}_h \qquad (3.5)$$

where $W_h = [y_{0.}^* \quad x_{1.}^* \quad . \quad . \quad . \quad x_{h.}^*]$

and $\hat{A}_h = [(\hat{\Pi}_1^{*h}D)':(\hat{\Pi}_2^*\hat{\Pi}_1^{*h-1}D)':(\hat{\Pi}_2^*\hat{\Pi}_1^{*h-2}D)'...(\hat{\Pi}_2^*D)']'$

$$\hat{\Pi}_1^* = \begin{bmatrix} \hat{\Pi}_1 & I_g \\ \hat{\Pi}_2 & 0 \end{bmatrix} \qquad \hat{\Pi}_2^* = \begin{bmatrix} \hat{\Pi}_3 & 0 \\ \hat{\Pi}_4 & 0 \end{bmatrix}$$

Since $v_{h-j.}^* = v_{h-j.}D'$ we can write the following relation corresponding to (3.5)

$$y_{h.} = y_{0.}^*\Pi_1^{*h}D + \sum x_{h-j.}^*\Pi_2^*\Pi_1^{*j} D + \sum v_{h-j.}D'\Pi_1^{*j} D \qquad (3.6)$$

(3.6) can also be rewritten as $y_{h.} = W_hA_h + \sum v_{h-j.}D'\Pi_1^{*j}D$ where

$A_h = [\ (\Pi_1^{*h}D)':(\ \Pi_2^*\Pi_1^{*h-1}D)':(\Pi_2^*\Pi_1^{*h-2}D)'\ldots(\Pi_2^*D)']'$

and $W_h = [y_{0.}^*\ x_{1.}^*\ \ldots\ x_{h.}^*]$

Applying the column stacking operator, vec, to (3.6) we get

$$y_{h.}' = \text{vec}\ y_{h.} = (I \otimes W_h)\text{vec}A_h + \sum (D'\Pi_1^{*j*}D)'v_{h-j.}' \qquad (3.7)$$

From (3.5) and (3.7) it follows that

$$y_{h.}' - \hat{y}_{h.}' = (I\otimes W_h)\ \text{vec}(A_h-\hat{A}_h) + \sum\ (D'\Pi_1^{*j}D)'v_{h-j.}' \qquad (3.8)$$

With known coefficients ($A_h = \hat{A}_h$), the asymptotic covariance matrix of the h-th post-sample forecast error is given by

$$\begin{aligned} &E(y_{h.} - \hat{y}_{h.})'(y_{h.} - \hat{y}_{h.}) \\ &= E\sum[(D'\Pi_1^{*j}D)v_{h-j.}'v_{h-j.}(D'\Pi_1^{*j}D) \\ &= \sum\ [(D'\Pi_1^{*j}D)'(I - B)^{-1\prime}\Sigma(I - B)^{-1}(D'\Pi_1^{*j}D) \qquad (3.9) \end{aligned}$$

In most cases the parameters of $\hat{\Pi}_1^*$ and $\hat{\Pi}_2^*$ are unknown and therefore have to be estimated through an appropriate procedure, such as maximum likelihood or least squares. More specifically, we need to obtain estimates of Π_1 and Π_2 of Π_1* and Π_3 and Π_4 of Π_2*.

We now consider the distribution of individual elements of $(A_h - \hat{A}_h)$. Note that $\hat{A}_h$ has the same structure as A_h except for the fact that Π_1 and Π_2 in Π_1* and Π_3 and Π_4 in Π_2* are replaced by their estimated values.

Using the results of Schmidt (1973) we know that

$$(\hat{\Pi}_1*^h - \Pi_1*^h) = \sum \hat{\Pi}_1*^j(\hat{\Pi}_1* - \Pi_1*)\Pi_1*^{h-1-j}$$

Therefore $(\hat{\Pi}_1*^h - \Pi_1*^h)D = \sum \hat{\Pi}_1*^j(\hat{\Pi}_1* - \Pi_1*)\Pi_1*^{h-1-j}D$

$$= \sum \hat{\Pi}_1*^j (\hat{\Pi}_1* - \Pi_1*)D(D'\Pi_1*^{h-1-j}D) \qquad (3.10)$$

The expression (3.10) follows from the fact that the upper and lower right hand blocks of $(\hat{\Pi}_1* - \Pi_1*)$ are zero which makes $(\hat{\Pi}_1* - \Pi_1*) = (\hat{\Pi}_1* - \Pi_1*)DD'$.

Therefore, $\mathrm{vec}(\hat{\Pi}_1*^h - \Pi_1*^h)D$

$$= \sum (D'\Pi_1*^{h-1-j}D) \otimes \hat{\Pi}_1*^j \mathrm{vec}(\hat{\Pi}_1* - \Pi_1*)D$$

$$= B_h \ \mathrm{vec} \ (\hat{\Pi}_1* - \Pi_1*)D \qquad (3.11)$$

where $B_h = \sum (D'\Pi_1*^{h-1-j}D)' \otimes \hat{\Pi}_1*^j$

Similarly for $i = 1,2, \ldots, h-1$ we have

$$\mathrm{vec}(\hat{\Pi}_2*\hat{\Pi}_1*^i - \Pi_2*\Pi_1*^i)D = [C_i \ \ D_i] \begin{bmatrix} \mathrm{vec}(\hat{\Pi}_1* - \Pi_1*)D \\ \mathrm{vec}(\hat{\Pi}_2* - \Pi_2*)D \end{bmatrix}$$

where $C_i = \sum(D'\Pi_1{*}^jD{*})' \otimes \hat{\Pi}_2{*}\hat{\Pi}_1{*}^{i-1-j}$ and $D_i{*} = (D'\Pi_1{*}^iD)' \otimes I$

Defining $A_h{*} = \begin{bmatrix} \text{vec}(\hat{\Pi}_1{*}^h - \Pi_1{*}^h)D \\ \text{vec}(\hat{\Pi}_2{*}\hat{\Pi}_1{*}^{h-1} - \Pi_2{*}\Pi_1{*}^{h-1})D \\ \cdots\cdots\cdots\cdots \\ \text{vec}(\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix}$

we may write

$$A_h{*} = Q_h \begin{bmatrix} \text{vec}(\hat{\Pi}_1{*} - \Pi_1{*})D \\ \text{vec}(\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix}$$

where $Q_h = \begin{bmatrix} B_h' & : & C_1' & \cdots & :C_{h-1}' & :\cdots & : 0' \\ 0 & : & D_1' & \cdots & : D_{h-1}' & :\cdots & : I \end{bmatrix}'$

We know that $\begin{bmatrix} \text{vec}(\hat{\Pi}_1{*} - \Pi_1{*})D \\ \text{vec}(\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix} = P \text{ vec} \begin{bmatrix} (\hat{\Pi}_1{*} - \Pi_1{*})D \\ (\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix}$ (3.12)

where $P = [(I_1 \otimes F_1)' : (I_2 \otimes F_2)']'$, and $F_1 = [I_1 \;\; 0]$, $F_2 = [0 \;\; I_2]$, and I_1 and I_2 are , respectively, $2g \times 2g$ and $2k_2 \times 2k_2$ identity matrices.

Therefore

$$A_h{*} = Q_h \begin{bmatrix} \text{vec} (\hat{\Pi}_1{*} - \Pi_1{*})D \\ \text{vec} (\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix} = Q_h \; P \;\; \text{vec} \begin{bmatrix} (\hat{\Pi}_1{*} - \Pi_1{*})D \\ (\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix}$$

This can be written as $A_h{*} = Q_h P \text{ vec } (\hat{\Pi} - \Pi)$

$$\text{Furthermore, } A_h{*} = G_h \text{ vec} \begin{bmatrix} (\hat{\Pi}_1{*}^h - \Pi_1{*}^h)D \\ (\hat{\Pi}_2{*}\hat{\Pi}_1{*}^{h-1} - \Pi_2{*}\Pi_1{*}^{h-1})D \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \\ (\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix}$$

$$\text{or} \quad A_h{*} = G_h \text{ vec}(\hat{A}_h - A_h) \tag{3.13}$$

where $G_h = [(I \otimes V_1)', (I \otimes V_2)', \ldots, (I \otimes V_{h+1})]'$ and $V_1 = [I\ 0\ \ldots\ 0]$, $V_2 = [0\ I\ 0\ \ldots\ 0], \ldots, V_g = [0\ 0\ \ldots\ I]$; V_1 is of order $2g \times 2(g + hk_2)$ and V_i $(i = 2, \ldots, h+1)$ are of order $2k_2 \times 2(g + hk_2)$.

An inspection of the matrix G_h indicates that $G_h'G_h$ is the identity matrix of order $2(g + hk_2)$ and premultiplying (3.13) by $G_h' = [(I\otimes V_1) : (I\otimes V_2) : \ldots, (I\otimes V_{h+1})]'$ we obtain $G_h'A_h{*} = \text{vec}(\hat{A}_h - A_h)$

$$\text{Therefore vec}(\hat{A}_h - A_h) = G_h'Q_h \text{ vec} \begin{bmatrix} (\hat{\Pi}_1{*} - \Pi_1{*})D \\ (\hat{\Pi}_2{*} - \Pi_2{*})D \end{bmatrix}$$

$$= G_h'Q_h P \text{ vec } (\hat{\Pi} - \Pi)$$

Knight (1982) proved the asymptotic distribution of $\sqrt{T}\text{vec}(\hat{\Pi} - \Pi)$, where $\hat{\Pi}$ is the restricted reduced form estimator of Π based on FIML, is $N(0, J\Phi J')$ where J and Φ are defined in

Knight(1982,pp. 588-589). In order to take into account the diagonality of R in ($J\Phi J'$) we replace $I \otimes U_{.1}$ by diag($u_{.1,-1}$, . . ., $u_{.g-1}$). We also note that the full information two-step estimators have the same asymptotic distribution as the FIML. Thus the asymptotic covariance matrix of the forecast error ($y_{h.} - \hat{y}_{h.}$) based on HF1, HF2 or HF3 is given by

$$E(\hat{y}_{h.} - y_{h.})'(\hat{y}_{h.} - y_{h.})$$
$$= \theta_h/T + \sum[(D'\Pi_1{*}^jD)'(I - B)^{-1\prime}\Sigma(I - B)^{-1}(D'\Pi_1{*}^jD) \qquad (3.14)$$

where $\theta_h = (I \otimes W_h)G_h'Q_hP(J\Phi J')P'Q_hG_h(I \otimes W_h')$

We also note that if δ and r are based on LIML then the expression (24) in Knight's article becomes

$$\sqrt{T}\ [(\hat{\delta} - \delta)'(\hat{r} - r)']' \sim N(0,\Phi*)$$

where $\Phi* = \text{plim}(1/TX^{+\prime}X^{+})^{-1}X^{+\prime}(\Sigma \otimes I)X^{+}/T(X^{+\prime}X^{+}/T)^{-1}$ and $X_{+} = [\text{diag}(Z_1^{+},\ldots,Z_g^{+}) : \text{diag}(u_{.1-1},\ldots,u_{.g-1})$. It follows that the asymptotic covariance matrix of the forecast error based on HL1, HL2 and HL3 is given by

$$E(\hat{y}_{h.} - y_{h.})'(\hat{y}_{h.} - y_{h.})$$
$$= \theta_h{*}/T + \sum[(D'\Pi_1{*}^jD)'(I - B)^{-1\prime}\Sigma(I - B)^{-1}(D'\Pi_1{*}^jD) \qquad (3.15)$$

where $\theta_h{*}$ is the same expression as θ_h but with Φ replaced by

Φ^*. Although the first terms in (3.14) and (3.15) are asymptotically negligible, it is not exactly clear what the impact of omitting this term would be in small sample situations. Accordingly, in our study, we have included these asymptotically negligible terms in estimating the asymptotic covariance matrices of the dynamic simulation forecasts based on each of the six estimators.

CHAPTER 4: THE DESIGN OF EXPERIMENTS

4.1 Introduction

A typical Monte Carlo experiment designed to compare the small sample properties of estimators of economic models consists of the following steps:

1. Specify an appropriate model.
2. Assign numerical values to all the structural parameters including the variance-covariance matrix of the errors, taking into account any a priori restrictions that the parameters are expected to obey. In other words, we specify a structure within the model.
3. Specify T numerical values for each exogenous variable in the model and, in addition, specify the initial values of the lagged endogenous variables if the model is dynamic. Note that T represents the sample size.
4. Generate a random sample of size T of errors from the desired probability distribution and combine it with the specified model-structure to generate a sample of size T of current values of the endogenous variables. In the case of dynamic SEM's current values of the endogenous variables are generated from the structural reduced form of the model.
5. For purposes of estimation a sample of observations consists of the specified T values of each exogenous variable, the T current values of the endogenous variables generated as

in step 4 and, if the model is dynamic, the T lagged values of the endogenous variables, which include the initial values. These data constitute a sample which is used to estimate the parameters of the model by alternative estimators included in the study.

6. The process described in steps 4 and 5 is repeated several times, e.g. 1000 times, thereby generating 1000 point estimates of each parameter by each of the estimators. The point estimates of any given parameter and estimator are used to construct an empirical distribution of that estimator. These empirical distributions of estimators are compared to shed light on their relative small sample properties. The descriptive statistics and the ranking statistics upon which the comparison of the estimators is based are described later in this chapter.

4.2 Description of the Experiments

The experiments conducted in this study were based on a two-equation model used by Wang and Fuller (1982). The particular model used in the Monte Carlo experiments is a special case of model (1.1) with the B, C_0 and C_1 specified as follows:

$$B = \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix} \quad C_0 = \begin{bmatrix} c_{11}^0 & 0 \\ 0 & c_{22}^0 \end{bmatrix} \quad C_1 = \begin{bmatrix} c_{11} & c_{21} & c_{31} & 0 & 0 \\ c_{12} & 0 & 0 & c_{42} & c_{52} \end{bmatrix}'$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \qquad R = \begin{bmatrix} r_{11} & 0 \\ 0 & r_{22} \end{bmatrix}$$

The model consists of two endogenous variables, two lagged endogenous variables, five exogenous variables which include the dummy variable that accounts for the intercept parameter. Hereafter, for convenience, we refer to the endogenous variable in the first column of Y as y1 and the second second column of Y as y2. The model satisfies all the assumptions listed in chapter 1. The equations are identified in the sense that the number of purely exogenous variables that are not redundant when lagged exceeds, 2, the number of equations. See Hendry (1976) for a discussion of the issue of identification. Since it has only two equations, the total number of parameters to be estimated is kept reasonable, which in turn facilitates comparison of the estimators, and, in addition, allows us to conveniently extend the analysis to compare the prediction performances of these estimators.

The model consists of 10 structural coefficients(5 in each equation), 2 autocorrelation coefficients (r_{11} and r_{22}) and 3 distinct elements of Σ (σ_{11}, σ_{12} and σ_{22}). Thus, for each structure, there is a total of 15 parameters to be estimated. For convenience we identify the parameters by numbers as follows:

Equation 1:	Parameter :	β_{21}	c_{11}^0	c_{11}	c_{21}	c_{31}
	Coefficient No.:	1	2	3	4	5

Equation 2: Parameter	:	β_{12}	c_{22}^{0}	c_{12}	c_{42}	c_{52}
Coefficient No.:		6	7	8	9	10
R and Σ : Parameter	:	r_{11}	r_{22}	σ_{11}	σ_{12}	σ_{22}
Coefficient No.:		11	12	13	14	15

Thus, we refer to β_{21} as coefficient number 1, c_{11}^{0} as coefficient number 2, etc. The coefficients of the two lagged endogenous variables are identified as coefficients 2 and 7, respectively.

Four different structures, differing in the elements of C_0 and R were used for purposes of experimentation. These model-structures henceforth referred to as models 1 through 4, are specified below.

Structure	1	2	3	4
C_0	$\begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$
R	$\begin{bmatrix} 0.9 & 0.0 \\ 0.0 & 0.9 \end{bmatrix}$	$\begin{bmatrix} 0.9 & 0.0 \\ 0.0 & 0.3 \end{bmatrix}$	$\begin{bmatrix} 0.9 & 0.0 \\ 0.0 & -0.6 \end{bmatrix}$	$\begin{bmatrix} 0.9 & 0.0 \\ 0.0 & 0.0 \end{bmatrix}$

For the 4 structures, the B, C_1 and Σ matrices which remain the same are specified below.

$$C_1 = \begin{bmatrix} 1.0 & 2.0 & 1.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.9 & 4.0 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0 & -1.0 \\ 1.0 & 0.0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1.00 & 1.21 \\ 1.21 & 2.21 \end{bmatrix}$$

Two sample sizes, T=30 and T=60, were used to study the effect of varying T on the sampling distributions of the structural parameters and forecasts. Thus, we had four major experiments each consisting of two sub-experiments. Experiment 1 was based on model 1 and the sub-experiments 1A and 1B were based on samples of sizes 30 and 60, respectively. A similar intepretation holds true of experiments 2, 3 and 4.

Generation of sample data

The following steps were followed in generating samples of dat on the endogenous variables

1) Generation of uniform random numbers

Two hundred and ten uniform random numbers on the open interval (0,1) were generated by the multiplicative congruential method. The pseudo random numbers were generated using the RAN subroutine that is available in VAX Fortran version 4.0. The 210 random numbers were rearranged into a 105 x 2 matrix.

2) Generation of standard normal numbers

The random numbers generated in step (1) were transformed into independent standard normal numbers using the Box-Mueller (1958) method and were arranged into a 105 x 2 matrix, NS. The Box-Mueller method is used to transform any pair of independent uniform random numbers, say U_1 and U_2, into a pair of independent standard normal numbers, say, N_1 and N_2. The following formulae were used to generate N_1 and N_2.

$$N_1 = (-2\ln U_1)^{1/2} \cos (2\Pi U_2)$$
$$N_2 = (-2\ln U_1)^{1/2} \sin (2\Pi U_2)$$

3)Generation of independent normal vectors with given covariance matrix Σ

The matrix, NS, of standard normal deviates obtained in step (2) were transformed into a matrix E whose row vectors are bivariate independent normal vectors with the specified covariance matrix Σ using an appropriate lower triangular matrix Q such that $Q'Q = \Sigma$. This is achieved using the relation $E = NS \times Q$. The 105 x 2 error matrix U was obtained from E as follows:

$$u_{1i} = \sqrt{(1.0 - r_{ii}^2)}\, e_{1i} \quad \text{for } i=1,2$$
$$u_{ki} = r_{ii}u_{k-1i} + e_{ki} \quad \text{for } k=2,\ldots,105,\ i=1,2$$

4)Generation of exogenous variables

The 5 exogenous variables used in the simulation study were the same as those used in Wang and Fuller (1982) which are reproduced here.

x_1: 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0

x_2:	1.0	3.0	0.0	9.0	1.0	6.0	3.0	8.0	6.0	7.0
x_3:	0.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0
x_4:	0.0	10.0	8.0	0.0	16.0	0.0	0.0	4.0	10.0	14.0
x_5:	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0

The first exogenous variable was taken to represent the constant term. The data were repeated to obtain a total of 105 observations on each exogenous variable. The exogenous variables were kept fixed in all replications and for all experiments.

5) Generation of endogenous variables

105 observations on each endogenous variable were generated using the specified data on the exogenous variables and the error matrix U via the reduced form. The first row of Y_{-1} matrix was calculated from the reduced form given in (1.4) using the formula $y_{-11.} = \bar{x}C_0(I-B)^{-1}$ where $\bar{x}$ is a row vector of the means of the the 10 repeated observations on each exogenous variable that are given above (i.e. $\bar{x}$ = [1.0 4.4 0.5 6.2 0.6])

Generation of samples for simulation

The first 40 observations of the 105 observations on Y, Y_{-1} and X were discarded to eliminate the effect of the initial values of the lagged endogenous variables. Observations 71 to 100 constituted the sample size 30 and observations 41 to 100 constituted the corresponding samples of size 60. In all cases observations 101-105 were used to investigate the properties of the estimators for dynamic simulation. The above steps were repeated each time we wished to generate a sample. In our study the number N of samples or replications was fixed at 1000. The use of nested/overlapping samples was not only to improve the reliability of the comparisons of the estimators between the two sample sizes but also to economize on the number of random numbers generated. In addition, it ensured that the five post-sample prediction periods 101 to 105 considered in the study remained the same in both the sample periods used for structural estimation.

Variance reduction

We decided to use the direct simulation approach discussed by Smith (1972) rather than implement the antithetic variate or control variate approach discussed, for example, in Hendry (1984). The antithetic variate technique was used in the study of limited information estimators of dynamic SEM's with autocorrelated errors by Moazzami and Buse (1986). However, as pointed out by Hendry and Harrison (1974), in a single equation context, the antithetic method offers little efficiency gain for autoregressive schemes such as the one considered here. The control variate approach, on the other hand, requires that we find an auxiliary estimator which is positively correlated with the estimators whose properties are being investigated but whose first few moments can be obtained analytically. However, it is not immediately clear what the best control variate for estimators of dynamic SEM's with autocorrelated errors would be.

Since simulation efficiency can also be increased in a relatively straightforward manner by increasing the number of replications in the experiment, we opted for this approach by increasing the number of replications from the typical number of 500 to the much larger number of 1000.

Programming and execution

For purposes of estimation a sample consisted of data on the current and lagged endogenous variables as well as the exogenous variables. The variables used as instruments in the first stage regression were chosen from current and /or one-period lagged exogenous variables. All experiments were programmed in VAX Fortran, using the IMSL matrix inversion subroutines LINDS and LINRG and executed on Dalhousie University's VAX 8800 VMS. For each sample and for each estimator,we obtained point estimates of the following:

i) structural coefficients of the two equations;

ii) autoregressive parameters which include the non-zero elements of R (r_{11} and r_{22}) and the three distinct elements σ_{11}, σ_{12} and σ_{22} of Σ;

iii) the asymptotic covariance matrix of the structural coefficients and the autoregressive parameters;

iv) dynamic simulation forecasts for up to 5 post-sample periods; and

v) estimates of the asymptotic covariance matrix of the dynamic simulation forecasts

As the number of replications was 1000, we obtained, for each estimator 1000 point estimates of each structural parameter and 1000 forecasts of each endogenous variable and for each of the five periods beyond the sample run. Treating the 1000 point estimates of the structural parameters and dynamic simulation of forecast errors as a random sample from the small sample distributions of the respective estimators, we computed the kernel density estimates of their distributions at 50 equally spaced points within 4 standard deviations from the respective means (which is 0 for post-sample predictions). The kernel function used for this purpose is the standard normal distribution.

4.3 Descriptive statistics used in the design of experiments

As pointed out in the previous section, each experiment, given T, resulted in 1000 point estimates of each structural parameter, 1000 dynamic simulation forecasts of each endogenous variable for each of the 5 post-sample periods, and 1000 estimates of their asymptotic covariance matrices. These estimates, which vary from replication to replication, constitute a random sample from their respective finite sampling distributions. In order to compare the sampling distributions of the estimators we need to answer two basic questions:

1. How close are the estimates of the parameters to their corresponding true values?
2. How large/small is the spread of these estimates around their corresponding true values?

It turns out that these questions can be answered with the aid of the statistical measures of central tendency and dispersion. In this section we introduce the basic statistics that are used to compare the relative performances of the six estimators investigated in this study. These are described separately for structural estimation and prediction.

a) Summary statistics for structural estimation

1.Mean bias(MB)

Let a_1, a_2,. . .,a_N be the N (=1000) estimates from the sampling distribution of a certain estimator for a particular parameter whose true value is, say, α. Furthermore, let us denote the mean of these estimates by $\bar{a}$. The mean bias of these estimates is measured by:

$$MB = \bar{a} - \alpha$$

The mean bias describes, on average, the extent by which the estimates differ from the true values. The smaller the value of the mean bias the better the estimator. The mean bias is used to test the hypothesis that the bias = 0. A major deficiency with the mean bias is that the mean of the sampling distribution might not exist. However, in the context of a Monte Carlo study, the existence of moments of the sampling distributions of estimators can be verified following the procedure suggested by Hendry (1984). Suppose we wish to verify the existence of the mean, the first moment about the origin. To this end sample means based on increasing sample sizes say 200, 400, 600 and 1000 are computed and compared. If the sample means converge, within limits of sampling variability, it is most probable that the moment of interest exists. If, on the other hand, the sample means are not well-behaved in the sense that they fluctuate widely, it is most probable that the moment of interest does not exist. This procedure was used in this study to verify the existence of the first four moments of the sampling distributions of estimators and those of post-sample prediction errors. We verified the existence of moments by varying the number of replications. It turned out that the first four sample moments were not sensitive to the variations in the number of replications, thus confirming, with some relief, the existence of moments. Also the bias is an inadequate measure in the sense that an estimator with a smaller bias does not necessarily mean that it has a smaller variance. Since the mean-squared error measures the spread of the estimates around the true value, we have used it to compare the dispersions of the estimators.

Testing the significance of the bias

In a Monte Carlo study the true value of the structural parameter, which is known, can be used to perform a large sample two tailed test of the null hypothesis H_0: Bias = 0 against the alternative H_a: Bias $\neq$ 0. This is done using the test statistic

$Z = MB/SE(\bar{a})$

where $SE(\bar{a}) = (s^2/T)^{1/2}$, where the sample variance $s^2 = 1/N\sum(a_i - \bar{a})^2$. N is the number of estimates a_i which is the same as the number of replications. The summation runs from i=1 to i=N. Note Z is AN(0,1).

The Mean Square Error(MSE)

The MSE is given by

$$MSE = 1/N\sum(a_i - \alpha)^2$$

where the summation runs from i=1 to N.

The MSE, which measures dispersion about the true value of the parameter, is a better measure in the sense that it incorporates both bias and dispersion. It is easy to show that

$$MSE = (MB)^2 + \text{Variance}$$

If one estimator has a larger mean bias but a smaller variance than the other, it is clear that a trade off between these two characteristics is taken care of by the MSE. The smaller the values of the MSE the better the estimator.

Frequency statistics for structural estimation

The relative performances of the estimators in testing hypotheses about the structural coefficients are also assessed. This is achieved by computing the studentized t-ratios based on the standard errors computed from the asymptotic covariance matrix of the estimator of structural coefficients. The studentized t-ratio is $t = (a_i-\alpha)/SE(a_i)$ ---> AN(0,1). We calculated the number of type I errors based on the z-statistics defined above in order to assess the reliability of estimators in testing hypotheses concerning structural parameters. These statistics are refered to as frequency statistics. At the 5 percent level of significance, the expected number of type I errors in 1000 replications is 50. Using the binomial test the number of type I errors falling between 36 and 64 in 1000 replications was not considered to be significantly different from 50. Thus, if for a particular coefficient, and an estimator, the number of type I errors was less than 36, the estimator was deemed to have performed better than expected and if the number was greater than 64, it was deemed to have performed worse than expected.

b) Summary statistics for Prediction

In a simulation study of predictions the true values of the endogenous variables vary from sample to sample. Let y_{hi} be the true value of an endogenous variable for the h-th post sample period in the i-th replication (i=1,. . .,N) and denote an estimate of the dynamic simulation forecast by $\hat{y}_{hi}$. Then the Mean Prediction Bias (MPB) and the Mean Squared Error of Prediction (MSEP) are defined as follows:

$$\text{MPB} = \sum(y_{hi} - \hat{y}_{hi})/N \ , \ i = 1,2\ldots,N \ (=1000)$$

The MPB measures on average the closeness of the forecasts to the true values of the endogenous variable.

$$2.\ \text{MSEP} = \sum(y_{hi} - \hat{y}_{hi})^2/N \ , i = 1,2,\ldots,N \ (=1000)$$

The MSEP measures the dispersion of the forecasts around the true values of the endogenous variable.

Frequency statistics for predictions

We also counted the number of type I errors that occured in the 1000 replications for each post sample period, and for each endogenous variable by estimators for each individual experiment. In a typical test of hypothesis concerning predictions the null hypothesis for the test is given by H_0: $y_{hi} - \hat{y}_{hi} = 0$ and the alternative hypothesis is given by H_a: y_{hi}

- $\hat{y}_{hi} \neq 0$. Under H_0 $(y_{hi}-\hat{y}_{hi})/SE(y_{hi}-\hat{y}_{hi})$ is asymptotically N(0,1). By dividing each forecast error by the corresponding estimate of its asymptotic standard error (including the asymptotically negligible term), we were able to determine the number of type I errors in the 1000 replications and compare with the expected number of type I errors at the 5 percent level of significance, which is 50.

The procedure discussed above in the context of structural parameters is equally valid here. Accordingly, the number of type I errors lying between 36 and 64 were not considered to be significantly different from 50 at the 5 percent level of significance.

Skewness and kurtosis statistics

The asymmetry and peakedness of the sampling distributions of the estimators /dynamic simulation forecasts may be described with the help of the sample measures of skewness and kurtosis. These measures were computed from the 1000 point estimates of the structural parameters and the 1000 dynamic simulation forecasts for each of the six estimators. These estimators were compared by these statistics to shed some light on the shapes of the sampling distributions.

Skewness

The coefficient of skewness (CS) is given by

$$CS = [1/N\sum(a_i - \bar{a})^3]/s^3$$

For symmetric distributions the coefficient of skewness is zero; for positively skewed distributions the coefficient of skewness is greater than zero and for negatively skewed distributions the coefficient of skewness is less than zero. Although there is no upper or lower limit for the coefficient of skewness, marked skewness is indicated when the absolute value of the coefficient is greater than 2. The skewness statistics were used in conjunction with the kernel density estimates in analyzing the empirical distributions of the estimators for both the structural parameters and forecast errors.

Kurtosis

The kurtosis measure describes the peakedness of the distributions, i.e., the degree to which the curve tends to be pointed or peaked. For this purpose we computed the coefficient of Kurtosis (CK) which is given by

$$CK = [1/N\sum(a_i-\bar{a})^4]/s^4$$

The standard for comparing the kurtosis measures among distributions is the standard normal distribution. For this distribution the kurtosis measure is 3. For flatter curves the value is less than 3 and for more peaked curves the value is greater than 3. The higher the value, the more peaked is the distribution. The kurtosis statistics were also used in conjunction with the kernel density estimates in interpreting the empirical distributions of the estimators of structural parameters and forecast errors.

c)Ranking statistics

Since the ranking of the estimators may depend on the structural parameter or forecast period chosen, we need an indication of the strength of the overall ranking. This is achieved by using Kendall's coefficient of Concordance, W. The W is used to test the extent to which the rankings of the estimators agree or disagree. The null hypothesis of the test is that there is no association among the rankings of the estimators. This is tested against the alternative that there is an association.

The ranking procedure followed is described here. To fix ideas this discussion focuses on ranking the estimators of a structural coefficient for a chosen statistic, say, the MSE though the same procedure is usually valid for ranking the forecasting performances of estimators.

Let $\grave{r}_{jk}$ denote the ranking of the jth structural coefficient for the k-th estimator (k=1,2,...,6). For each structural coefficient, the estimator which yields the smallest mean square error is assigned rank number 1. The next largest is assigned rank number 2, etc. The implication is that an estimator which performs consistently well on the basis of each structural coefficient or forecast is ranked consistently higher and that which performs worse is ranked consistently lower. This means that the sums of the ranks serve as indexes to compare the performances of the estimators. Accordingly, if there are n coefficients on which to base the ranking of the estimators or forecasts, and if HF1 performed consistently the best, then the total sum of the ranks is n. If HL3 performed consistently the worst, the sum of the ranks corresponding to HL3 is 6n. If there is absolutely no association we would expect the rank sums to be approximately equal. In this case the ranking of the estimators is not consistent over all the structural parameters considered, i.e, estimators that are ranked first according to a given structural parameter have lower ranks according to other structural parameters.

Kendall's coefficient of concordance, W, is given by

$$W = S/ \ [\{k^2(L^3-L)\}/12]$$

where

k is the number of parameters considered

L is the number of estimators to be ranked (6 in this case)

$S = \sum(R_j-\bar{R})^2$ where

R_j is the rank sums for the j'th estimator, where the summation used to obtain R_j is over all the structural parameters considered. $\bar{R} = \sum R_j/L$ is the mean of the rank sums.

Accordingly, S is the sums of squares of deviations of the rank sums from the mean of the rank sums. It turns out that for structural estimation k = 8 (2 intercept parameters excluded) and for prediction k = 5.

If there is perfect agreement or perfect association W=1. If there is perfect disagreement (i.e. complete absence of association) W=0. However, if there are more than 2 ranks to be considered, a situation of perfect disagreement does not arise. For example, if there is perfect disagreement between HF1 and HF2 and HF3 then HF2 and HF3 must agree. i.e., complete disagreement is impossible.

A test of hypothesis concerning W consists of checking whether the observed W exceeds the critical value. The stronger the ranking the closer the value of W to +1 and the weaker the ranking the closer the value of W to 0. The critical values of W are given in Siegel (1956, Table R, p.286). When the observed sets of rankings are in close agreement, the computed value of W tends to be large. If there is perfect disagreement W tends to be small. Sufficiently large values of W lead us to reject the null hypotheses of no association. If L is greater or equal to 7, the test may be based on the X^2 distribution. For a detailed description of this test see Siegel (1956,chapter 8).

To examine the strength of the rankings we computed the Kendall's W and tested it for significance in rankings of the estimators by various criteria like the mean bias, MSE, etc. in both structural estimation and prediction.

4.4 Computational Problems

As noted above, the two-step estimators considered in this study are free from the convergence problems which are characteristic of the maximum likelihood and other iterative procedures. Also we did not encounter any estimates of the autoregressive coefficients which are greater than 1 in absolute value. However, one of the problems encountered was the occurence of negative elements on the diagonal of the asymptotic covariance matrices of the dynamic simulation forecasts. This problem arose as a result of the near-singularity of the matrices to be inverted in order to obtain the covariance matrices. The problem tended to occur whenever the reciprocal of the condition number of one of the matrices to be inverted was less than 10^{-10}. Consequently, such a sample was dropped and replaced with a new sample. However this problem resulted in the rejection of no more than 2 of the 1000 samples in any given experiment.

Table 4-1 shows the number of samples rejected by experiments.

Table 4-1: Number of rejected samples by experiment

	T = 30	T = 60
Experiment 1	1	2
Experiment 2	2	1
Experiment 3	2	0
Experiment 4	2	2

4.5 Review of previous Monte Carlo studies

A brief review of earlier Monte Carlo studies which focussed on the small sample properties of SEM's with autocorrelated errors is presented below.

Hurd (1972) considered the small sample properties of OLS and 2SLS in a two equation static simultaneous equation autocorrelated model. He concluded that at low degrees of autocorrelation in the error structure, OLS is generally more efficient than 2SLS. The study also found that the single equation Prais-Winsten estimator which corrects for autocorrelation is still more efficient than OLS at fairly large degrees of autocorrelation by yielding smaller MSE's. However, this study is not relevant in the context of the present study because it did not deal with dynamic SEM's. Goldfeldt and Quandt (1972) also considered static SEM's.

Table 4-2 provides a summary of the Monte Carlo studies that deal specifically with dynamic SEM's with vector autoregressive errors.

Hendry and Harrison (1974) used the control variate approach and found that the two inconsistent estimators, OLS and 2SLS, could lead to very biased results when applied to dynamic SEM's with autocorrelated errors. This justifies the use of consistent estimators like the two-step estimators considered here.

Hendry and Srba(1977) utilized the control variate approach to compare the small sample properties of the following limited information estimators: autoregressive instrumental variable estimator (AIV), 2SLS, OLS and autoregressive least squares (ALS). The ranking of the estimators according to the mean squared error criterion was that AIV outperformed the other estimators at very large degrees of (positive or negative) autocorrelation and large sample sizes (T>55). 2SLS ranked first at modest degrees of autocorrelation $|r|<0.5$ and large sample sizes. OLS performed best at modest degrees of autocorrelation and small sample sizes whereas ALS performed best at large degrees of autocorrelation and small sample sizes.

Wang and Fuller (1982) performed a Monte Carlo study to compare the small sample properties of two full information estimators and one limited information estimator. The two full information estimators considered were the autoregressive three stage least squares 1 (A3SLS1) and the autoregressive three stage least squares 2 (A3SLS2). The limited information estimator considered was the autoregressive two stage least squares (FA2SLS). Like Hatanaka's estimators, the three estimators are also residual adjusted but are not necessarily fully efficient.Using a two equation model (without identities) Wang and Fuller found no significant difference between A3SLS1 and A3SLS2 and that FA2SLS performed quite well.

Moazzami and Buse (1986) performed a Monte Carlo study of a number of estimators of dynamic simultaneous equation models with autocorrelated errors. In addition to the three limited information two-step estimators considered in this study, they also included the S2SLS, Fair's (1970) modification of S2SLS, the instrumental variable and iterative 2SLS estimators proposed by Dhrymes, Berner and Cummins (1974) and a modified version of Theil's (1958) estimator which has been modified to take into account of autocorrelation and a practically unrealistic estimator which uses known autocorrelation coefficients, which they dubbed "True Dhrymes". Using a three equation model and the antithetic variate technique, they found that, on the criterion of aggregate bias, Hatanaka's two-step estimators were inferior to the 2SLS, True Dhrymes and Theil's estimators with HL2 (which behaved abnormally in certain cases) coming dead last. The other two, HL1 and HL3 were, however, superior to Fair's estimator. There were no substantial changes in the rankings when other criteria were used although when the median bias was used HF2 dominated all the other estimators.

An important characteristic of all these studies was that they dealt exclusively with structural estimation. In addition, no results were reported on the relative post-sample prediction performances of estimators of dynamic SEM's with autocorrelated errors. Furthermore, although some results of Monte Carlo studies were reported regarding the reliability of the asymptotic standard errors in dynamic SEM's satisfying the standard assumptions, (e.g. Basmann et. al. (1974), Maddala(1974)), this issue has not been explored in the case of dynamic SEM's with autocorrelated errors. In this study we address these questions, viz, structural parameter estimation, reliability of asymptotic standard errors in hypothesis testing and forecasting performance.

TABLE 4-2:SUMMARY OF MONTE CARLO STUDIES OF ESTIMATORS OF DYNAMIC SEM's WITH AUTOCORRELATED ERRORS

	Hendry and Harrison(1974)	Hendry and Srba(1977)	Wang and Fuller(1982)	Moazzami and Buse(1986)
1.No. of equations	2	2	2	3
(lagged endogenous)	1	1	2	2
2.Sample size	20 to 80	20,40,60,80	30 ,60	30,60
3.Estimation techniques compared	OLS,2SLS	AIV,ALS,OLS,TSLS	FA2SLS,FA3SLS,A3SLS	Includes HL1,HL2,HL3
4.Prediction	No	No	No	No
5.Main conclusions	Significant bias of OLS and 2SLS	Autocor-relation / Sample size / Best estimator High / large / AIV Low / large / 2SLS Low / small / OLS Low / large / ALS	A3SLS1 and A3SLS2 not significantly different; FA2SLS performed well Results consistent with asymptotic theory	No significant difference between HL1 and HL3 with HL2 performing consistently worse.However some other estimators considered performed better than HL1, HL2 and HL3.

OLS - Ordinary Least Squares
2SLS - 2 Stage Least Squares
AIV - Autoregressive Instrumental Variable Estimator
ALS - Autoregressive Least Squares Estimator
A3SLS1 - Autoregressive three stage least squares 1(For a description see Wang and Fuller(1982)
A3SLS2 - Autoregressive three stage least squares 2(For a description see Wang and Fuller(1982)
FA2SLS - Autoregressive two stage least squares(For a description see Wang and Fuller(1982)

CHAPTER 5: SUBSTANTIVE RESULTS

In this chapter we report the results of the major experiments performed. As stated earlier, four major experiments, differing in the coefficients of the lagged endogenous variables and/or autoregressive coefficients, were performed. Each major experiment consisted of two sub-experiments, denoted by A (sample size 30) and B (sample size 60). Accordingly, we conducted 8 experiments. These experiments allow us to investigate the effect of different magnitudes of the coefficients of the lagged endogenous variables, autoregressive coefficients and the sample size. Model 1 is characterized by high autocorrelation coefficients in that the autoregressive coefficient in both equations is 0.9. In model 2 the coefficient of the lagged endogenous variable in the first equation is reduced from 0.8 to 0.2 and the autoregressive coefficient of the second equation is reduced from 0.9 to 0.3. This is designed to capture the combined effect of changing the coefficient of the lagged endogenous variable and the autoregressive coefficient. Model 3 is characterized by significant differences in the autocorrelation coefficients of the two equations (0.9 for the first equation and -0.6 for the second equation). This allows us to investigate the effect of differences in the autocorrelation coefficients on the small sample properties of estimators. In model 4, the errors in the second equation are specified to be serially independent.

It must be mentioned that, in reporting the mean biases, the MSE and other statistics, the intercept parameter, i.e., coefficients numbered 3 and 8, were excluded. This is because the estimate of the intercept parameter is not directly relevant in empirical work. In the next sections we report the results of the experiments conducted under the broad headings of structural estimation, prediction, ranking statistics and density estimates. Hereafter we use the "model" and "experiment" interchangeably. For example, we may refer to the results of experiment 1 as the results of model 1, etc. Furthermore, we refer to the coefficients of the variables in the two equations as the structural coefficients. Also the non-zero elements of R, i.e, r_{11} and r_{22} and the distinct elements σ_{11}, σ_{12} and σ_{22} of the error covariance matrix Σ are referred to as the autoregressive coefficients. In this case the structural parameters include the structural coefficients and the non-zero coefficients of R and the distinct elements of Σ.

5.1 Structural estimation

Bias

Tables 5.1-1 and 5.1-2 show the mean biases of the structural coefficients and the autoregressive parameters, respectively for all the experiments.

TABLE 5.1-1:MEAN BIASES OF COEFFICIENT PARAMETERS*

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	-0.0053	0.0105	-0.0186	0.0056	0.0241	-0.0252	-0.0054	0.0015	-0.0009	0.0001	-0.002	-0.0043
	2	0.0064	0.0131	-0.0494	0.0207	0.0342	-0.0463	0.0009	0.0055	-0.0022	0.0072	0.0061	-0.0026
	4	0.0094	0.0107	-0.0453	0.0075	0.0315	-0.0424	0.0001	0.0022	-0.0024	0.0005	-0.0016	-0.0056
	5	-0.0051	0.0616	-0.1798	0.0772	0.1064	-0.1857	-0.0006	0.029	-0.0132	0.0319	0.0283	-0.0149
	6	0.0237	-0.0033	0.0001	0.0471	-0.0007	0.0078	0.0114	-0.0011	0.0024	0.0223	0.0002	0.0051
	7	-0.0085	0.0097	-0.0079	-0.0308	0.0043	-0.0175	-0.0042	0.0062	-0.0009	-0.0151	0.002	-0.0056
	9	-0.0101	0.0056	0.0002	-0.0201	0.0036	-0.0057	-0.0058	0.0037	0.0013	-0.0109	0.0003	-0.0021
	10	-0.0209	0.0696	-0.0743	-0.1259	0.0167	-0.1917	-0.021	0.0305	-0.0084	-0.0717	-0.012	-0.0631
2	1	-0.0226	0.0189	-0.032	-0.0096	0.0164	-0.0334	-0.0036	0.0003	-0.0018	0.0003	-0.0008	-0.003
	2	-0.0456	-0.0413	-0.0433	-0.0366	-0.0414	-0.0338	0.0038	0.0071	-0.0058	0.0091	0.0088	-0.0045
	4	-0.0157	0.0126	-0.0394	-0.0124	0.0122	-0.0377	0.002	0.0001	-0.0063	0.003	0.0019	-0.0048
	5	-0.291	-0.2137	-0.1243	-0.2316	-0.2337	-0.1021	0.02[illegible]7	0.0546	-0.0304	0.0474	0.0455	-0.0409
	6	0.0121	-0.0241	-0.0267	0.0321	-0.0163	-0.0186	0.0062	-0.011	-0.011	0.0119	-0.0122	-0.0114
	7	0.0106	0.0363	0.0362	-0.0059	0.0227	0.0256	0.0034	0.0146	0.0141	0.0005	0.0161	0.0145
	9	-0.0041	0.0225	0.0188	-0.0075	0.0165	0.0164	-0.0029	0.0084	0.0078	-0.003	0.0089	0.0078
	10	0.0419	0.21	0.3184	-0.0665	0.1301	0.2034	-0.037	0.0369	0.0296	-0.0346	0.0508	0.0457
3	1	-0.0061	0.0103	0.0061	-0.0018	0.0084	0.0056	-0.0042	0.0033	0.0006	-0.0021	0.0021	0.0006
	2	0.0178	0.0267	0.0213	0.0167	0.0229	0.0186	0.0079	0.0122	0.0091	0.0066	0.0095	0.007
	4	0.0001	0.0109	0.0062	-0.0005	0.0107	0.0069	-0.0007	0.0041	0.0017	-0.001	0.0037	0.0018
	5	0.0859	0.1342	0.107	0.0772	0.0963	0.0777	0.0494	0.0702	0.0582	0.0333	0.0421	0.0345
	6	0.0116	-0.0577	-0.0386	0.0248	-0.0337	-0.0424	-0.0049	-0.0334	-0.0259	-0.0012	-0.0397	-0.0293
	7	-0.0041	0.0171	-0.0395	-0.0118	0.0146	-0.0282	0.0092	0.0133	-0.0035	0.0081	0.0135	-0.0038
	9	-0.0081	0.031	0.0063	-0.016	0.0286	0.0037	0.0027	0.0185	0.0124	0.0004	0.0181	0.0106
	10	-0.037	0.3185	0.1294	-0.1019	0.3535	0.1601	0.0474	0.1881	0.1357	0.0228	0.208	0.1408
4	1	-0.0048	0.0036	0.0003	0.0018	0.0051	0.0019	-0.0043	-0.0013	-0.0024	-0.0007	0.0005	-0.0008
	2	0.0171	0.0215	0.0173	0.0158	0.0173	0.0118	0.0069	0.0087	0.0059	0.0061	0.0066	0.0031
	4	0.0051	0.0062	0.0005	0.0041	0.0075	0.0027	0.0001	0.0009	-0.0016	0.0008	0.0021	0.0001
	5	0.0687	0.1044	0.076	0.0661	0.0712	0.0462	0.0373	0.0487	0.0324	0.0277	0.0294	0.0142
	6	0.0054	-0.0304	-0.0246	0.021	-0.038	-0.0301	-0.0022	-0.0127	-0.0118	0.0024	-0.0201	-0.0181
	7	0.0186	0.0373	0.0172	0.0086	0.0376	0.015	0.0152	0.0202	0.[illegible]12	0.0149	0.024	0.0139
	9	0.0036	0.0252	0.0168	-0.0028	0.027	0.0171	0.0049	0.0107	0.0[illegible]38	0.0038	0.0144	0.0166
	10	0.0439	0.2307	0.1642	-0.0483	0.2567	0.1779	0.046	0.1024	0.0851	0.0156	0.1285	0.1088

TABLE 5.1-2:MEAN BIASES OF THE AUTOREGRESSIVE COEFFICIENTS

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	11	-0.1315	-0.1328	-0.1438	-0.1243	-0.122	-0.1292	-0.0531	-0.0534	-0.0581	-0.0543	-0.0539	-0.0568
	12	-0.1335	-0.1222	-0.1396	-0.1325	-0.1142	-0.1319	-0.0607	-0.057	-0.0617	-0.0632	-0.0574	-0.0614
	13	-0.0306	3.7787	4.0659	-0.2132	24.1411	3.7886	-0.098	-0.083	-0.0034	-0.1133	-0.1076	-0.0219
	14	-0.2589	9.2453	-0.0954	-0.3933	-0.3406	-0.4357	-0.1311	-0.1083	-0.0831	-0.2022	-0.1781	-0.1844
	15	-0.4305	-0.2251	4.9185	-0.4746	-0.3132	15.4902	-0.2559	-0.1956	0.1246	-0.3218	-0.248	0.0867
2	11	-0.0299	-0.0127	-0.0549	-0.0813	-0.062	-0.1103	-0.0356	-0.0391	-0.0421	-0.063	-0.0637	-0.0655
	12	-0.0789	-0.0645	-0.0809	-0.0818	-0.0638	-0.0712	-0.023	-0.0194	-0.0211	-0.0241	-0.0178	-0.0188
	13	0.3131	0.4379	0.1475	0.3843	0.6112	0.2012	-0.0666	-0.085	-0.063	-0.1115	-0.1092	-0.0881
	14	-0.2065	-0.2049	-0.2074	-0.3645	-0.3476	-0.3471	-0.125	-0.1105	-0.1092	-0.2014	-0.1771	-0.1765
	15	-0.4544	-0.3561	-0.3193	-0.5433	-0.4213	0.4046	-0.2389	-0.1861	-0.176	-0.2837	-0.211	-0.2075
3	11	-0.0879	-0.0952	-0.0969	-0.1165	-0.1186	-0.121	-0.0377	-0.0414	-0.0407	-0.0573	-0.0585	-0.0588
	12	0.1724	0.1701	0.191	0.1895	0.1941	0.2157	0.0721	0.0723	0.0866	0.0875	0.0937	0.1077
	13	-0.196	-0.1905	-0.1683	-0.224	-0.2191	-0.2006	-0.0986	-0.0936	-0.0828	-0.1176	-0.1153	-0.1072
	14	-0.2866	-0.2297	-0.2536	-0.4082	-0.3486	-0.3728	-0.1168	-0.0829	-0.0857	-0.1982	-0.1667	-0.1771
	15	-0.2039	0.0294	1.2138	-0.2683	0.0112	0.843	0.1053	0.012	0.0514	-0.1452	0.012	0.0317
4	11	-0.0844	-0.0926	-0.0965	-0.1157	0.1188	-0.1214	-0.034	-0.0373	-0.0394	-0.0571	-0.0582	-0.0589
	12	0.0305	0.0367	0.0451	0.0413	0.0566	0.0648	0.0097	0.0114	0.0183	0.0169	0.0253	0.0324
	13	-0.1865	-0.1885	-0.1561	-0.2226	-0.2205	-0.1972	-0.0957	-0.0953	-0.0859	-0.1178	-0.1165	-0.1107
	14	-0.257	-0.2323	-0.2468	-0.383	-0.3395	-0.3518	-0.1145	-0.105	-0.1053	-0.1932	-0.1733	-0.1754
	15	-0.3947	-0.3028	-0.159	-0.4687	-0.3191	-0.2027	-0.1951	-0.1617	-0.1256	-0.2404	-0.1714	-0.1502

We note that both positive and negative biases were observed. In addition, the null hypothesis that the bias = 0 was rejected at the 5 percent level for all the structural parameters in the four models and for all the six estimators. The biases generally decreased as the sample size was increased from 30 to 60 observations implying that the estimators were asymptotically unbiased.

The salient points regarding the biases encountered in individual experiments are reported below.

Experiment 1:

No estimator completely dominated the others according to the bias criterion though HF2 and HF3 and their limited information counterparts had generally larger biases than HF1 and its limited information counterpart. Furthermore, the limited information estimators had generally larger biases than their full-information analogues for T = 30 but this observation did not consistently hold for T = 60. There are cases when the biases were as much as 40 percent of the true values. All the six estimators tended to perform equally well when T = 60.

Experiment 2:

Generally the observations made on the biases encountered in experiment 1 were also valid for this experiment. No estimator completely dominated the others although, in general, the full information estimators had larger biases than the limited information estimators for the coefficients of the first equation and the reverse was true for the coefficients of the second equation. The biases generally decreased as the sample size was increased from 30 to 60.

Experiment 3:

The ranking of the estimators according to the biases were fairly consistent for this experiment and for both sample sizes. HF2 and HL2 were relatively more biased than the other estimators for a number of coefficients for T = 30 and T = 60. Also the full information estimators had generally larger biases than their limited information counterparts. The biases decreased considerably as the sample size was increased.

Experiment 4:

Again HF2 and HL2 were generally the most biased but the results varied substantially among the coefficients. The biases generally decreased as the sample size was increased from 30 to 60. The full information estimators had generally smaller biases than their limited-information counterparts.

Comparison of the biases of the structural coefficients across experiments

In models 2, 3 and 4 where there are differences in the autoregressive coefficients of the two equations, the full information estimators tended to have relatively larger biases for the equation with high autocorrelation and the reverse was true for the equation with low autocorrelation. However, the differences in the biases became less apparent when T=60. There was, however, no noticeable pattern in the magnitude of the biases across experiments except that the absolute values of the biases were generally greater for experiment 2 than the corresponding values for experiment 1. This suggests that changing the coefficient of the lagged endogenous variable and the autoregressive coefficient might have significant effects on the biases. Furthermore, the ordering of estimators according to the biases was most consistent for model 3.

Mean Biases of the autoregressive coefficients

The biases of all the autoregressive parameters were significant at the 5 percent level. Furthermore, the biases encountered were generally negative in all the four experiments. In every experiment the results were mixed in the sense that no estimator completely dominated the others according to this criterion and this was true of all the autoregressive coefficients. For experiments 3 and 4 the biases of r_{11} and r_{22} were opposite in sign. However, these biases tended to decrease considerably for all models as the sample size was increased. The absolute value of the biases were generally the highest for model 1 for which the autocorrelation coefficient in both equations was 0.9. In some cases the biases were as much as twice the true values of the corresponding elements of R and Σ. The absolute values of the biases of the elements of Σ were higher than those of R for every estimator and in all the experiments.

Mean Square Errors

Tables 5.1-3 and 5.1-4 give the MSE's of the structural coefficients and the autoregressive parameters, respectively. As the entries indicate, there were no substantial differences among the six estimators as the MSE's of parameters were close to each other. Also the MSE's of estimators generally decreased as the sample size was increased; this was true for all models.

The main observation regarding the MSE's encountered in the individual experiments are reported below:

Experiment 1:

Although there were no large differences in the magnitudes of the MSE's of the structural parameters, the limited information estimators had generally larger MSE's than the full-information estimators but this result was not true of all the 8 structural coefficients considered. This

TABLE 5.1-3:MEAN SQUARE ERRORS OF COEFFICIENT PARAMETERS

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	0.0062	0.0085	0.0066	0.0061	0.008	0.0066	0.003	0.0066	0.0035	0.0029	0.0062	0.0034
	2	0.0208	0.0233	0.0262	0.0253	0.0288	0.0299	0.0173	0.0219	0.0237	0.0211	0.0268	0.0268
	4	0.0124	0.0163	0.0132	0.0115	0.0137	0.0132	0.0066	0.0128	0.0081	0.006	0.0096	0.0084
	5	0.4622	0.4951	0.6413	0.5969	0.6611	0.8189	0.3531	0.4198	0.6847	0.4982	0.6042	0.9099
	6	0.0601	0.0307	0.0362	0.0644	0.0338	0.0358	0.0603	0.0313	0.0416	0.0653	0.0333	0.0422
	7	0.0384	0.0303	0.0538	0.047	0.0378	0.0591	0.0449	0.0368	0.0637	0.0558	0.0466	0.0711
	9	0.0191	0.0126	0.0094	0.0219	0.0163	0.0103	0.0169	0.0097	0.0099	0.0195	0.0117	0.0113
	10	2.0807	1.5241	1.1662	2.5147	2.0465	1.4613	1.5741	1.0778	0.8683	1.8691	1.2662	1.0266
2	1	0.0177	0.0143	0.0121	0.0164	0.0142	0.0123	0.0183	0.0126	0.0105	0.017	0.0115	0.0104
	2	0.038	0.0401	0.034	0.0461	0.0498	0.0424	0.0565	0.0617	0.0546	0.0659	0.0731	0.0618
	4	0.0174	0.0172	0.017	0.0172	0.0173	0.0173	0.0147	0.009	0.014	0.0131	0.0085	0.0137
	5	1.0089	1.0566	0.8549	1.2279	1.3395	1.0895	1.5091	1.6436	1.2749	1.811	2.0284	1.5349
	6	0.0102	0.0121	0.0121	0.011	0.0123	0.0123	0.0047	0.0056	0.006	0.0048	0.0053	0.0055
	7	0.0093	0.0113	0.0116	0.0113	0.0129	0.013	0.0051	0.0063	0.0066	0.0058	0.0064	0.0065
	9	0.0053	0.0063	0.0061	0.0059	0.007	0.0069	0.0025	0.0029	0.0028	0.0028	0.003	0.0031
	10	0.5325	0.6397	0.7092	0.666	0.7505	0.7911	0.2619	0.3058	0.3733	0.3196	0.338	0.3614
3	1	0.0157	0.0258	0.016	0.0169	0.0284	0.0161	0.0186	0.0469	0.0213	0.0198	0.0489	0.0198
	2	0.0153	0.0264	0.0175	0.0165	0.0275	0.0202	0.0155	0.027	0.0147	0.0166	0.0289	0.0169
	4	0.0149	0.0224	0.0152	0.019	0.0264	0.0181	0.0152	0.0265	0.0145	0.0184	0.0288	0.0155
	5	0.2496	0.5406	1.2699	0.3429	0.6329	1.4681	0.1889	0.6382	1.7817	0.2542	0.7318	2.0473
	6	0.0123	0.0216	0.0157	0.0144	0.0249	0.0182	0.0058	0.0104	0.0072	0.0062	0.0109	0.0084
	7	0.0129	0.0152	0.0109	0.0111	0.0139	0.0117	0.0076	0.008	0.0034	0.0043	0.0046	0.0039
	9	0.0061	0.0104	0.0071	0.007	0.0106	0.0082	0.0034	0.0068	0.0031	0.003	0.0046	0.0036
	10	0.4215	0.7369	0.5516	0.5266	0.8735	0.673	0.2117	0.365	0.2353	0.2323	0.3659	0.2855
4	1	0.0145	0.0118	0.0118	0.015	0.0137	0.0123	0.0162	0.0138	0.0123	0.0166	0.0144	0.0123
	2	0.0153	0.019	0.0163	0.0175	0.0211	0.0188	0.0135	0.0156	0.017	0.0162	0.0191	0.0199
	4	0.0163	0.0138	0.0168	0.018	0.0157	0.0179	0.0159	0.0085	0.016	0.0166	0.0087	0.016
	5	0.2543	0.3607	0.5564	0.3621	0.4481	0.7086	0.1572	0.3033	0.5699	0.2566	0.3884	0.6682
	6	0.0118	0.0166	0.0149	0.0138	0.0187	0.0174	0.0065	0.0079	0.0072	0.0059	0.0075	0.0075
	7	0.012	0.0152	0.0094	0.0112	0.014	0.0123	0.0102	0.0127	0.0049	0.0053	0.0061	0.0058
	9	0.0061	0.0088	0.0068	0.0069	0.0092	0.0085	0.0034	0.0051	0.0029	0.0029	0.0036	0.0035
	10	0.6168	0.8241	0.7919	0.7559	0.9597	0.9387	0.344	0.4309	0.4115	0.3431	0.4031	0.4112

TABLE 5.1-4:MEAN SQUARE ERRORS OF THE AUTOREGRESSIVE PARAMETERS

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	11	0.085	0.0891	0.0862	0.0928	0.0964	0.0956	0.0467	0.05	0.0466	0.0469	0.0491	0.0482
	12	0.1425	0.0806	0.0779	0.1629	0.112	0.0788	0.0749	0.0401	0.0396	0.0799	0.0516	0.0373
	13	0.1442	0.1619	0.1654	0.1508	0.1656	0.1718	0.0997	0.1096	0.1153	0.1031	0.1168	0.1343
	14	0.3153	0.2962	0.3368	0.4276	0.4041	0.4323	0.1876	0.1775	0.2642	0.2783	0.2467	0.3483
	15	0.9805	4.6112	1.2906	1.6302	14.1923	1.9064	0.3794	1.3348	0.5523	0.5388	3.0588	0.7[illegible]92
2	11	0.0757	0.0778	0.0622	0.1057	0.1106	0.0953	0.0886	0.1011	0.0633	0.1036	0.1161	0.0768
	12	0.0368	0.0367	0.0324	0.0521	0.0523	0.048	0.0137	0.0138	0.0125	0.0219	0.0221	0.0208
	13	2.0356	4.2258	0.8347	5.1463	12.7825	2.0176	4.2507	9.7964	2.165	7.5276	17.894	3.9515
	14	0.2235	0.2311	0.2332	0.2563	0.2654	0.2603	0.1381	0.1437	0.1406	0.1169	0.132	0.1202
	15	0.5946	0.6022	0.6184	0.612	0.5986	0.594	0.2664	0.2706	0.2833	0.2726	0.2686	0.2678
3	11	0.0548	0.0587	0.049	0.0746	0.0815	0.0809	0.0447	0.0756	0.025	0.0531	0.097	0.0395
	12	0.0243	0.0268	0.0285	0.0356	0.0399	0.0338	0.0102	0.0109	0.0163	0.0146	0.0158	0.0147
	13	0.6674	4.0513	0.7508	0.7977	7.2406	0.734	0.7119	13.1233	1.5529	0.9137	18.9006	1.5542
	14	0.238	0.282	0.2441	0.2778	0.3004	0.2953	0.1286	0.2275	0.1186	0.1146	0.1514	0.1222
	15	0.6214	0.794	0.608	0.6423	0.7825	0.6291	0.2749	0.3617	0.2763	0.2778	0.3115	0.2911
4	11	0.0459	0.0432	0.0382	0.0721	0.0709	0.071	0.0348	0.0378	0.0224	0.0465	0.0535	0.0351
	12	0.0335	0.0341	0.0284	0.0497	0.0506	0.0472	0.0147	0.0152	0.0132	0.0226	0.0231	0.0228
	13	0.4575	0.7923	0.4238	0.6332	1.2419	0.4717	0.4577	0.9256	0.6889	0.6244	1.7162	0.7371
	14	0.2202	0.215	0.2125	0.2642	0.2611	0.2566	0.1321	0.1272	0.1115	0.1085	0.1169	0.1046
	15	0.5886	0.6258	0.6147	0.6159	0.6232	0.6184	0.2879	0.3078	0.2852	0.2767	0.2828	0.2828

observation remained true as the sample size was increased from 30 to 60. The results varied from parameter to parameter and no estimator completely dominated the others on the basis of MSE.

Experiment 2:

The limited information estimators generally had larger MSE's than their full information analogues for both sample sizes though there was no consistency over all the structural parameters. The MSE's of the coefficients of the second equation were generally smaller than the MSE's of the coefficients of the first equation. In all cases the MSE's decreased as the sample size was increased.

Experiment 3:

This experiment produced fairly consistent rankings of the estimators by MSE's for both T = 30 and T = 60. HF2 and HL2 performed poorly especially with sample size 30. Also the MSE's decreased as the sample size was increased. The limited information estimators performed worse than their full information counterparts for both sample sizes. The differences in the MSE's were not very apparent when sample size 60 was used.

Experiment 4:

In this experiment the errors in the second equation were specified to be serially independent. The limited information estimators generally performed worse than their full-information counterparts for T = 30 but not for T = 60. The MSE's decreased as the sample size was increased from 30 to 60. There was no consistent ranking of the estimators for all structural coefficients.

Comparison of the MSE's across experiments

The MSE's of the coefficients of the second equation were generally smaller for models 2, 3 and 4 than for model 1. This suggests that smaller absolute values of the autocorrelation coefficient in the equations might reduce the MSE. The MSE's generally decreased as the sample size was increased from 30 to 60.

MSE of the autoregressive parameters

Like the biases of the autoregressive coefficients the results for the MSE were rather mixed. In all the 4 models there was a systematic decrease in MSE as the sample size was increased from 30 to 60. The results of individual experiments are discussed below.

Experiment 1:

The limited information estimators generally performed worse than their full information counterparts for both sample sizes and that no estimator completely dominated the others for all the autoregressive coefficients.

Experiment 2:

The full information estimators generally performed better than the limited information estimators. However, HF2 and HL2 had generally larger MSE's than the other estimators. The differences in MSE were less apparent for $T = 60$.

Experiment 3:

Generally the limited information estimators performed worse than their full information counterparts. HF2 and HL2 performed consistently worse than the other estimators for both sample sizes. Furthermore, for HF2 and HL2, the MSE of some autoregressive coefficients tended to increase as the sample size was increased from 30 to 60.

Experiment 4:

The observations were generally the same as for experiment 3. The limited information estimators performed worse than their full information counterparts and the MSE's decreased as the sample size was increased from 30 to 60. HF2 and HL2 had generally the largest MSE's and in some cases their MSE's increased as the sample size was increased.

Comparison of the MSE's of the autoregressive coefficients

The MSE's were extremely high for the parameters in model 1 where the autoregressive coefficients for both equations were high. The MSE's of r_{11} and r_{22} were generally smaller for model 2 than for model 1 but the MSE's of the distinct elements of Σ were larger for model 2 than for model 1. The MSE's of r_{11} and r_{22} were lower in models 3 and 4 than in model 1. Generally the MSE's of the distinct elements of Σ did not exhibit any noticeable pattern across experiments. Except for HF2 and HL2 in some cases, the MSE's generally decreased as the sample size was increased from 30 to 60. These results suggest that reducing the absolute value of the autocorrelation coefficient also reduces the MSE's of the estimates of these coefficients.

Hypothesis testing

To assess the usefulness of the asymptotic standard errors in hypothesis testing, we computed the number of type I errors for each coefficient parameter and for each estimator in the 1000 replications. Based on the binomial test at the 5 percent level of significance, we expect between 36 and 64 type I errors in 1000 replications. However, in all the experiments and for all the six estimators, the number of type I errors observed for any given coefficient parameter is zero and hence not reported in tabular form. The failure to find any rejections of the null hypothesis is very surprising as it indicates that the standard normal distribution which is used to set up the critical region $|Z_\alpha| > 1.96$ is perhaps thicker in tails than the sampling distributions of estimators. In fact, the quality of a test of hypothesis is measured by its power, and in the present context, the power of the test could be increased by adjusting the critical region. The adjustment that should be made here is to choose a smaller critical value than $|Z_\alpha| = 1.96$, which, in turn, might increase the power of the test.

The experience with the standard linear model suggests that, in principle, adjustments to the asymptotically valid test statistics could be made so that it has the correct size in small samples. Two such adjustments are: 1) to implement a degrees of freedom correction in estimating the error covariance matrix and 2) to compute the second order approximations to the distributions of the test statistics.

In their Monte Carlo study of the impact of alternative degrees of freedom in 2SLS and 3SLS estimators, Binkley and Nelson (1984) had concluded that a degrees of freedom correction should be used in estimating Σ, the covariance matrix of errors. They stated: "While the use of a normal approximation for making inferences does not appear to be highly accurate for any of the procedures examined in this study (in small samples), it was less so when no correction was used." Unfortunately, such a degrees of freedom correction is not useful in the present context as it would generate even smaller values of the z statistic and hence would not result in rejecting the null hypothesis.

Size corrections which make use of the second order approximations to the distributions of the test statistics may be a reasonable way to proceed by analogy with inferential procedures used in the standard linear model (see, for example, Rothenberg (1984), (1987), Evans and Savin(1982), (1987) and Harris (1985). Thus, size-adjusted values based on these approximations could be computed. However, as pointed out by Rothenberg (1984, p.259), in the context of the standard linear model, useful general results on the magnitude of the size corrections do not seem to be feasible as the answer depends very much on the specific model and the null hypothesis. A resolution of the problem of zero type I errors observed in this study is feasible, in principle, though the required size adjustment is not clear.

Another plausible explanation for the poor performance of the test statistic could be the two step nature of the estimator. Dhrymes and Taylor (1976) recommended iterating the two step estimator at least once in practical applications. While iteration makes no difference asymptotically it might as well be crucial in small samples. This issue needs further investigation.

It is concluded that the standard errors computed from the asymptotic covariance matrices of the six two step estimators are of doubtful validity for purposes of inference regarding the true values of the coefficient parameters. The failure to observe any type I errors calls for some size adjustments to the distributions of the test statistics, which in turn, might increase the power of the test.

Skewness

Tables 5.1-5 and 5.1-6 report the skewness statistics of the structural coefficients and the autoregressive parameters, respectively.In all experiments we notice that there were no substantial differences in skewness. Most values of the coefficient of skewness were less than 1 in absolute value which suggest that the sampling distributions of the structural parameters were almost symmetric. Both positive and negative values for the coefficient of skewness were observed in all experiments.

We note the following in regard to the results of individual experiments.

Experiment 1:

Although the skewness tended to vary from coefficient to coefficient, the full-information estimators were generally more skewed than their limited information counterparts for

TABLE 5.1-5:SKEWNESS STATISTICS FOR COEFFICIENT PARAMETERS

Model	Coeff.	T = 30 HF1	HF2	HF3	HL1	HL2	HL3	T = 60 HF1	HF2	HF3	HL1	HL2	HL3
1	1	-0.405	-0.0501	-0.1101	-0.3033	0.0244	-0.0396	-0.3484	0.3961	0.2739	-0.2947	0.3748	0.284
	2	0.3317	0.5198	0.2911	0.1166	0.2438	0.1078	0.7218	0.8916	0.5309	0.4671	0.6112	0.3951
	4	-0.1902	-0.1033	-0.1549	-0.205	-0.1415	-0.1589	0.0434	0.418	0.1839	0.0187	0.4009	0.1501
	5	0.8197	0.966	0.9019	0.5228	0.6241	0.6261	1.0443	1.1985	0.9694	0.6819	0.8036	0.7271
	6	0.7007	0.4973	0.5395	0.5803	0.245	0.4673	0.6143	0.8076	0.7574	0.5154	0.6443	0.6944
	7	-0.3192	-0.5144	-0.367	-0.1438	-0.3134	-0.264	-0.2494	-0.5413	-0.3263	-0.0941	-0.3502	-0.2239
	9	-0.7543	-0.2667	-0.723	-0.5553	0.0782	-0.573	-0.7342	-0.852	-1.0751	-0.554	-0.5025	-0.8532
	10	-0.6288	0.3584	0.6947	-0.3465	-0.0099	-0.4275	-0.7745	-0.8389	-1.0728	-0.5298	-0.4752	-0.8115
2	1	-2.0928	-1.3641	-1.7115	-2.0205	-1.2502	-1.5615	-2.6416	-2.1284	-2.5568	-2.6629	-2.1033	-2.4542
	2	1.0924	1.036	1.5677	0.6857	0.5789	1.0881	1.198	1.1254	1.5268	0.9395	0.8823	1.2695
	4	-0.8869	-0.276	-0.3733	-0.7689	-0.3179	-0.3753	-1.382	-0.4325	-0.5074	-1.3069	-0.4545	-0.4557
	5	0.8381	0.912	1.573	0.4768	0.4794	1.0617	0.9621	1.1054	1.6857	0.7338	0.8669	1.3522
	6	-0.5338	-0.4991	-0.5046	-0.6403	-0.6254	-0.6224	-0.2948	-0.3115	-0.3193	-0.4214	-0.4366	-0.4367
	7	0.1943	0.1944	0.2275	0.2626	0.2673	0.2725	0.0694	0.0652	0.1047	0.1754	0.201	0.2037
	9	0.2953	0.3681	0.3856	0.3594	0.3981	0.4109	0.252	0.2362	0.246	0.3058	0.333	0.3429
	10	0.1501	0.2065	0.2396	0.1248	0.1501	0.1682	-0.0143	0.0005	0.0151	-0.038	-0.023	0.0017
3	1	-2.3313	1.3846	-0.2058	-2.117	1.1514	-0.1912	-2.6207	1.4884	-0.3221	-2.4236	1.3347	-0.322
	2	-0.3257	1.2588	0.7919	-0.2758	1.1388	0.7473	-0.4348	2.0336	1.8079	-0.2073	2.1817	1.817
	4	-1.2791	1.0366	-0.0058	-1.099	0.7091	-0.0968	-1.7508	1.5827	0.1488	-1.5196	1.364	-0.0609
	5	0.2639	2.3924	1.7479	0.4552	2.0953	1.4278	-0.117	2.8028	2.0973	0.4302	3.0578	1.9297
	6	-0.378	[illegible]	-0.5159	-0.4939	-0.5221	-0.6274	-0.2067	-0.1635	-0.2258	-0.3412	0.2597	-0.2885
	7	-0.1268	0.097	-0.4416	0.0102	0.3027	-0.3126	-0.1138	0.0141	0.3279	0.2177	0.3017	0.0181
	9	0.2534	0.4044	0.4368	0.4027	0.5633	0.4913	0.0903	0.1577	0.3081	0.3576	0.4737	0.4097
	10	0.0863	0.283	0.2345	0.3345	0.5042	0.4901	0.0771	0.1218	0.0562	0.2763	0.3441	0.2502
4	1	-2.0997	-0.439	-0.8921	-1.9045	-0.3702	-0.7569	-2.4483	-1.0423	-1.6511	-2.2742	-0.9121	-1.4428
	2	-0.3662	0.1088	-0.0465	0.3347	0.0553	-0.0243	-0.316	0.6221	0.3962	-0.0969	0.868	0.5381
	4	-1.1095	-0.098	-0.3024	-0.9274	-0.1483	-0.2483	-1.4497	-0.1248	-0.3487	-1.3123	-0.2643	-0.3649
	5	0.173	0.7362	0.996	0.3331	0.6975	0.8365	0.1696	1.3704	1.7301	0.5861	1.7313	1.7152
	6	-0.3856	-0.3465	-0.3658	-0.5486	-0.5074	-0.5016	-0.2663	-0.1808	-0.1813	-0.397	-0.3433	-0.3212
	7	0.1147	0.183	0.1658	0.3749	0.4916	0.4057	-0.0479	-0.0425	-0.0164	0.2795	0.3254	0.265
	9	0.3186	0.4449	0.4456	0.416	0.4887	0.4736	0.1654	0.1902	0.2467	0.4177	0.4725	0.457
	10	0.1155	0.1833	0.2047	0.3275	0.3617	0.3387	0.0501	0.0318	0.0578	0.2353	0.248	0.2508

TABLE 5.1-6:SKEWNESS STATISTICS FOR THE AUTOREGRESSIVE PARAMETERS

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	11	-0.6235	-0.6069	-0.6452	-0.5263	-0.5022	-0.5062	-0.9214	-0.919	-0.8389	-0.6812	-0.6833	-0.6505
	12	-0.7536	-0.0532	-0.702	-0.3534	0.2474	-0.4658	-0.9682	-0.6366	-1.2934	-0.692	-0.2733	-0.3974
	13	2.2646	2.6838	3.5801	3.5398	3.1542	4.0473	3.0152	2.8265	4.3839	4.3166	3.5931	5.8745
	14	16.5265	16.0122	15.2281	12.983	11.0169	11.9695	37.2101	34.5388	24.2193	24.176	22.0639	16.6758
	15	3.2724	3.5725	2.2293	3.5589	3.9378	2.0485	1.2007	2.5272	1.2693	1.8979	3.8879	1.6769
2	11	-1.165	-0.8445	-1.4786	-0.6045	-0.2997	-0.7854	-1.7345	-1.2049	-1.9747	-1.1312	-0.7119	-1.3059
	12	-0.0101	0.0633	-0.0561	-0.0815	-0.0148	-0.144	-0.0658	-0.0447	-0.1357	-0.2016	-0.1741	-0.2858
	13	9.3673	9.0607	7.9047	11.909	11.7375	10.1926	6.4093	5.8577	4.5849	5.9522	6.4381	7.3306
	14	21.3096	20.2311	20.4748	22.077	19.3418	20.1267	40.6565	37.5973	39.7264	53.337	44.4222	50.3959
	15	0.8751	0.943	1.0526	0.7458	0.8761	0.8754	0.5511	0.543	0.626	0.507	0.5356	0.5473
3	11	-0.8638	-0.0341	-1.2743	-0.6229	0.0042	-0.8308	-1.567	0.0316	-1.6874	-1.3027	-0.0177	-1.1688
	12	0.3256	0.145	0.4197	0.3574	0.183	0.4444	0.4081	0.327	0.3618	0.4434	0.404	0.4293
	13	8.5191	6.7144	3.4703	8.764	9.5704	4.0764	3.751	7.1618	4.3121	4.5204	7.2232	4.8386
	14	19.9535	15.8645	18.9698	21.348	17.603	17.2368	44.3655	20.2975	51.2044	63.065	46.2151	49.4872
	15	1.3373	1.6978	1.0784	1.453	1.8482	1.0976	0.5217	0.7167	0.5044	0.5324	0.7094	0.5954
4	11	-0.7049	-0.3294	-0.8506	-0.5199	-0.305	-0.5648	-1.4721	-0.8571	-1.4552	-1.1425	-0.6643	-0.9568
	12	0.2284	0.2823	0.0457	0.1759	0.2292	0.1223	0.1741	0.1986	-0.0423	0.0214	0.0527	-0.0226
	13	6.1265	8.181	2.5422	7.5179	7.9299	3.9055	3.1981	4.3629	3.3997	3.7829	5.1231	4.1191
	14	21.7891	22.3905	22.6688	23.179	21.4027	20.9051	43.3855	45.8579	56.5718	68.087	64.3404	62.5194
	15	0.9222	0.9854	1.012	0.8166	0.9618	0.9484	0.5474	0.526	0.5392	0.5046	0.5521	0.5533

both sample sizes. Furthermore, there were no substantial changes in skewness as the sample size was increased from 30 to 60. Also there was not much variability in skewness across the coefficients.

Experiment 2:

Generally the coefficients of the first equation were more skewed than those of the second equation as judged by the relative magnitudes of the coefficient of skewness. For each coefficient the values of the skewness statistics were quite similar for all the estimators suggesting no marked differences among the estimators. For the first equation, the full information estimators were more skewed than their limited information counterparts and for the second equation the limited information estimators were more skewed for both sample sizes.

Experiment 3:

For all estimators and for both sample sizes the values of the coefficient of skewness were greater for the coefficients of the first equation than those of the second equation. No estimator was consistently the most skewed for all the structural coefficients, though, for the parameters of the second equation, the limited information estimators were relatively more skewed.

Experiment 4:

The skewness was generally greater for the coefficients of the first equation than for the coefficients of the second equation for all the six estimators.
The coefficients of skewness for the coefficients of the lagged endogenous variables were much closer to zero indicating near perfect symmetry in these cases. However, the coefficient of the endogenous variable of the first equation (i.e. coefficient 1) tended to give numerically larger values of the coefficient of skewness, especially for HF1 and HF2.

Comparison of the skewness statistics across experiments

In general the values of the coefficient of skewness encountered were much higher for the parameters in model 2 than those in model 1. For experiments 1 and 2, the sampling distributions tended to become more skewed as the sample size was increased from 30 to 60. However, as indicated earlier, the low values of the coefficients of skewness in Table 5.1-5 suggest near symmetry of the sampling distributions.

Skewness statistics for autoregressive parameters

The distributions of estimators of the parameters r_{11} and r_{22} were almost symmetric in all experiments. However, the distributions of the estimators of σ_{11}, σ_{12} and σ_{22} were highly skewed and tended to yield very large values especially when the sample size was increased and this was true in all experiments. The most serious case of skewness arose for the σ_{12} parameter. The limited information estimators were generally more skewed than the full information estimators for models 3 and 4 and less skewed than the full information estimators for models 1 and 2.

As indicated earlier, for cases where the kernel density estimates were not provided, we relied on the skewness statistics in interpreting the shapes of the empirical distributions as the skewness statistics provide the same information as the kernel density estimates in regard to the symmetry characteristic of the sampling distributions.

Kurtosis

Tables 5.1-7 and 5.1-8 present the kurtosis statistics for the structural coefficients and the autoregressive parameters, respectively.

The kurtosis statistics of the coefficient parameters presented in Table 5.1-7 were markedly close to each other most of the time. In most cases, the kurtosis statistics were greater than 3, suggesting that the distributions of the structural parameters were more peaked than the standard normal distribution.

The following observations are noteworthy regarding the results of individual experiments.

TABLE 5.1-7:KURTOSIS STATISTICS FOR COEFFICIENT PARAMETERS

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	4.6057	3.0903	3.1884	4.1201	2.9317	3.0611	4.0387	3.3506	3.4661	3.8276	3.3756	3.4986
	2	3.4947	3.7308	3.5562	3.2019	3.2784	3.3673	3.7703	4.0657	3.5935	3.2939	3.4543	3.3693
	4	3.2354	2.8944	2.9985	3.1917	2.862	2.9381	3.2315	3.4831	3.2848	3.0711	3.5887	3.2931
	5	4.2476	4.6401	4.1445	3.4186	3.5765	3.4877	4.8894	5.3804	4.495	3.9243	4.2208	3.9793
	6	3.3126	3.5882	3.5218	3.1895	3.5269	3.4517	2.7698	3.5519	3.4238	2.6926	3.3827	3.2998
	7	2.6523	3.0733	2.8431	2.6092	2.9389	2.8027	2.2655	2.7974	2.5166	2.2207	2.6011	2.4268
	9	3.9545	3.8956	4.2151	3.7236	3.6728	3.7286	3.2213	4.0967	4.5237	3.0246	3.6521	3.9375
	10	3.346	3.2947	3.7296	3.1077	3.0767	3.2801	3.4554	4.2565	4.5914	3.2076	3.77	3.9592
2	1	10.0588	8.[illegible]	[illegible]192	10.0556	8.1929	8.5897	11.321	11.9439	13.3879	11.5497	11.6124	12.6362
	2	8.6667	9.0859	10.[illegible]505	7.7755	8.3044	8.5863	8.5641	9.0842	10.1305	7.8082	8.4258	9.3351
	4	4.9724	3.891	3.[illegible]362	4.6024	3.8855	3.4821	6.3063	4.5258	3.8528	6.2433	4.3258	3.683
	5	8.2528	8.8173	10.0548	7.1091	7.7716	7.8106	8.0749	9.2363	11.0553	7.5357	8.7643	10.1763
	6	3.9503	3.6609	3.7451	4.2865	4.17	4.2061	3.3214	3.2393	3.2495	3.495	3.4728	3.4601
	7	3.0562	2.9609	2.9181	3.1464	3.1172	3.0673	3.236	3.2893	3.255	3.1762	3.2368	3.2367
	9	3.1606	3.2883	3.3431	3.1872	3.2947	3.323	3.1144	3.0415	3.086	3.0954	3.1015	3.1082
	10	3.1227	3.1523	3.1426	3.1497	3.1419	3.0626	2.9605	2.9103	2.8956	3.1556	3.2232	3.1489
3	1	11.974	7.4659	3.5421	10.861	6.4563	3.4496	11.479	6.85	4.0376	10.2463	6.2086	3.8519
	2	3.836	9.1559	6.1904	3.2809	8.3775	5.9329	3.7655	11.2112	9.1872	9.7325	12.3518	9.6062
	4	6.6589	5.9092	2.8733	5.7172	4.8661	2.8598	7.6975	7.5437	3.569	6.6456	6.9703	3.2383
	5	5.9302	17.2371	8.7281	5.5555	13.8376	7.1184	3.9187	15.1203	9.4697	4.4924	17.076	9.0211
	6	3.5992	3.2233	3.957	3.9714	3.5576	4.3895	3.0271	3.2926	3.2401	3.2475	3.303	3.3455
	7	3.4229	3.5246	3.8085	3.5434	3.7565	3.7195	3.0729	3.0975	3.6295	3.2743	3.4331	3.4834
	9	3.0562	3.1706	3.3535	3.2406	3.3059	3.5142	2.9943	3.33	3.5359	3.0406	3.4579	3.5842
	10	3.2322	3.1178	3.4654	3.5389	3.5822	3.8858	3.1102	3.3763	3.4974	3.3818	3.4756	3.6365
4	1	10.5305	6.2664	5.894	9.5901	5.314	5.3411	10.7746	9.4837	10.1749	9.6306	7.6161	8.2449
	2	3.7363	4.1847	3.6431	3.983	4.4106	4.0754	3.6492	5.3188	5.1461	3.9839	6.5487	5.8313
	4	5.8274	4.0768	3.4295	5.0078	3.3627	3.2056	6.559	4.8265	3.7931	5.8427	4.1854	3.5678
	5	4.2083	5.3536	4.9553	4.5282	5.2733	4.6794	3.5913	6.9434	7.9141	4.6632	9.5207	9.0683
	6	3.4364	3.2647	3.4801	4.0521	3.7636	3.8804	3.2231	3.0679	3.1088	3.4461	3.2931	3.2763
	7	3.1094	3.066	3.08	3.3627	3.4344	3.338	2.9436	3.0445	2.9717	3.0873	3.1196	3.0266
	9	2.9662	3.1326	3.1531	3.1071	3.1059	3.1167	2.9943	2.926	2.9651	3.1524	3.1827	3.1621
	10	3.0892	2.9966	2.9896	3.4383	3.221	3.1968	3.0389	3.0305	3.035	3.4639	3.4486	3.4367

TABLE 5.1-8:KURTOSIS STATISTICS FOR THE AUTOREGRESSIVE PARAMETERS

		T = 30						T = 60					
Model	Coeff.	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	11	3.5536	3.6479	3.4103	3.1927	3.1779	3.1124	4.7248	4.9884	4.3958	3.859	3.9807	3.8766
	12	4.1942	3.3792	4.609	3.4362	3.0263	3.5986	3.9637	3.8747	5.5366	3.4417	3.4303	4.5142
	13	8.0285	16.2721	22.0179	12.026	15.106	17.5867	12.9546	13.0614	17.6238	16.4606	15.1338	25.8134
	14	53.4939	51.7675	47.8229	41.753	37.7523	36.6744	143.526	130.107	81.9958	85.5354	76.4622	51.8725
	15	29.4176	21.3251	13.225	24.96	23.3013	9.5268	6.6368	15.0272	6.6344	11.2345	26.7652	10.1907
2	11	7.0148	6.9274	8.4678	5.6161	5.7493	6.0669	9.1187	8.1722	11.0892	6.7164	6.2692	8.014
	12	3.1395	3.067	2.9797	2.8658	2.8238	2.7179	3.5731	3.5326	3.6436	3.2539	3.2567	3.2774
	13	130.809	109.581	111.739	183.88	167.926	155.438	58.421	40.3965	27.6563	42.9896	49.0448	80.5463
	14	80.2329	75.1268	81.4664	82.199	70.5123	73.8599	172.401	154.127	169.627	232.704	185.252	216.313
	15	4.1438	4.3005	4.8752	3.7855	4.0627	4.0876	3.2861	3.1961	3.5055	3.1702	3.1696	3.1991
3	11	5.7915	4.6384	7.4968	4.7463	3.8924	5.3297	7.3224	4.4397	8.2809	6.3007	3.8291	6.2668
	12	4.0584	3.9007	3.5177	3.4271	3.3091	3.3186	3.6502	3.5027	3.4885	3.4453	3.4747	3.5698
	13	137.174	56.3272	14.105	130.94	128.577	16.9299	17.1074	63.1921	17.2193	24.8648	66.7066	20.4289
	14	75.3542	58.8164	69.2632	77.09	61.2303	59.4215	192.343	78.898	219.068	285.169	192.743	207.669
	15	6.6695	8.7303	4.8698	7.9517	10.2286	4.8848	3.2333	3.854	3.1543	3.2459	3.8121	3.3939
4	11	4.88	4.5547	4.9645	4.3117	4.0877	4.3158	7.1687	6.067	7.9181	5.6852	4.8623	5.6787
	12	3.2403	3.2559	3.1026	2.9352	2.9965	2.8626	3.2744	3.2798	3.1722	3.0314	3.0264	2.9631
	13	76.7268	119.511	10.0882	91.237	91.9764	20.1264	13.824	22.1863	13.791	16.0118	30.8038	17.8329
	14	82.874	84.7496	86.3503	84.846	77.1554	74.853	187.748	200.9	253.5	313.124	291.68	280.45
	15	4.197	4.3041	4.5359	3.939	4.2819	4.2397	3.2997	3.1693	3.2333	3.1965	3.2504	3.2452

Experiment 1:

The kurtosis statistics were close to each other for all the structural parameters and for all the six estimators. However, the coefficient of kurtosis tended to increase as the sample size was increased from 30 to 60, indicating greater than normal peakedness.

Experiment 2:

The kurtosis statistics for the coefficients of the first equation were generally higher than those of the corresponding coefficients of the second equation. The values of kurtosis statistic for the coefficients of the lagged endogenous variables had generally increased as the sample size was increased. Again there was a general increase in the kurtosis statistics as the sample size was increased from 30 to 60.

Experiment 3:

For the coefficients of equation 1, the full information estimators had generally more peaked distributions than the limited information analogues. For the second equation the limited information estimators had more peaked distributions. These observations were true for both sample sizes.

Experiment 4:

The observations were similar to experiment 3 though the values of the coefficient of kurtosis were more uniform for the coefficients of the two equations. For equation 1, the full information estimators had generally more peaked distributions than their limited information analogues.

Comparison of the kurtosis statistics across experiments

Generally it seemed that decreasing the autocorrelation coefficient decreased the kurtosis of the coefficients in that equation whose autocorrelation coefficient was decreased (compare the results of model 1 with those of models 2, 3 and 4). In every experiment, the full information estimators generally yielded greater absolute values for the coefficient of kurtosis compared to the limited information estimators. The numerical values were extremely high for the full information estimators, suggesting that they are more peaked than the limited information estimators. Comparing the results of experiments 1 and 2 we note that the effect of decreasing the coefficient of the lagged endogenous variable and the autoregressive coefficient was to increase the peakedness of the distribution. Also peakedness tended to increase as the sample size increased, and this was true of every estimator.

Kurtosis statistics for autoregressive parameters

The kurtosis statistics indicate no major differences among the estimators for all the autoregressive coefficients. For all models the peakedness was the greatest for the elements of Σ. Also for the elements of R and Σ, the peakedness was low for model 1 and high for models 2,3 and

4 suggesting that a decrease in the size of the autocorrelation coefficients tends to increase the peakedness of the distributions of the autoregressive coefficients. The kurtosis generally increased as the sample size was increased from 30 to 60 in the case of all models and for all estimators.

5.2:Prediction

Bias

Tables 5.2-1 and 5.2-2 present the mean prediction biases (MPB) for y1 and y2, respectively.

A test of the significance of prediction biases indicated that the biases of all estimators were significant at the 5 percent level. This observation remained valid for both endogenous variables, all the 5 forecast periods in all the four models and for all the six estimators. For all the experiments and estimators both positive and negative biases were observed. The biases tended to decrease as the sample size was increased. However, there was no noticeable pattern in the prediction biases of the estimators as the forecast period was extended into the future, although, in some cases it tended to increase. In particular, the biases tended to be very large for the fifth post sample prediction when T = 30.

The results of individual experiments are discussed below.

TABLE 5.2-1:MEAN PREDICTION BIAS FOR Y1

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	0.0492	0.0408	0.2228	0.0561	-0.0354	0.1569	0.0012	0.0104	0.1249	0.0137	-0.032	0.0864
	2	0.0955	0.0882	0.1877	0.1294	0.0091	-0.162	0.0379	0.022	0.2623	0.032	-0.0406	0.1581
	3	0.225	0.1338	-1.8241	0.2319	0.0438	-3.0079	0.0502	-0.0089	0.3519	0.0556	-0.0739	0.1159
	4	-0.0001	0.1655	-26.107	0.1297	0.1613	-30.937	-0.0283	-0.0216	0.3455	-0.0131	0.1005	-0.3385
	5	0.2551	0.2833	-278.32	0.1129	0.5565	-287.95	0.0587	0.0415	0.2375	0.0214	0.0393	-1.8369
2	1	-0.1815	0.295	-0.3242	-0.1435	-0.3154	-0.3432	0.0001	-0.0383	-0.0269	0.009	-0.0472	-0.0365
	2	-0.0516	-0.1369	-0.2383	-0.1472	-0.2737	-0.3233	-0.0227	-0.0499	-0.0576	-0.0031	-0.0411	-0.0475
	3	0.4477	0.1454	0.0648	0.3601	0.1295	0.0847	-0.0525	-0.1103	-0.0924	-0.0364	-0.0996	-0.0792
	4	-0.7372	-0.4859	-0.5654	-0.7197	-0.6266	-0.6479	-0.1085	-0.0537	-0.0425	-0.0792	-0.0548	-0.044
	5	-0.1903	-0.1685	-0.3017	-0.317	-0.2581	-0.3556	0.0434	0.0424	0.0367	0.0383	0.0501	0.042
3	1	-0.094	-0.091	-0.0549	-0.0952	-0.1056	-0.0823	-0.0535	-0.0464	-0.0377	-0.054	-0.0524	-0.0467
	2	-0.0163	-0.0454	-0.0421	-0.0267	-0.051	-0.0355	-0.0255	-0.0331	-0.0316	-0.0261	-0.034	-0.0373
	3	0.0343	-0.0989	0.1792	0.0386	-0.0916	0.0696	-0.0429	-0.0905	-0.0575	-0.0435	-0.0909	-0.0627
	4	-0.0229	0.1083	-0.4132	-0.0337	0.1025	0.0485	-0.0401	0.0341	0.0588	-0.0456	0.0267	0.0428
	5	0.0707	0.0758	3.1625	0.0465	0.0775	0.6403	0.068	0.068	0.0589	0.0528	0.062	0.0452
4	1	-0.038	-0.0557	-0.024	-0.0228	-0.0632	-0.0377	-0.0293	-0.0323	-0.0146	-0.0195	-0.0354	-0.0261
	2	0.0293	-0.0017	0.0553	0.0073	-0.0234	0.0132	-0.0042	-0.0108	-0.0133	-0.011	-0.022	-0.016
	3	0.0401	-0.0618	0.0466	0.0236	-0.0893	0.0014	-0.0289	-0.0562	-0.0158	-0.0409	-0.0799	-0.0538
	4	-0.0289	0.0202	0.1427	-0.0464	0.0029	0.0998	-0.0551	-0.0337	0.0196	-0.0643	-0.0395	-0.0023
	5	0.1158	0.1184	0.194	0.0568	0.0867	0.1496	0.0761	0.077	0.1014	0.0493	0.0626	0.0735

TABLE 5.2-2:MEAN PREDICTION BIAS FOR Y2

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	0.059	0.0139	0.2667	0.1483	0.0719	0.3239	0.0367	0.0349	0.1224	0.0703	0.0371	0.1428
	2	0.0759	0.0209	0.6517	0.1189	0.0352	0.8606	0.0384	0.0324	0.174	0.0623	0.0324	0.2381
	3	0.0138	-0.0344	3.1552	0.1768	-0.0429	4.1467	0.0335	0.0153	0.2458	0.0963	0.0288	0.5049
	4	-0.0303	-0.1185	29.1469	-0.0732	-0.2501	34.0331	-0.0151	-0.0004	0.396	-0.0179	0.0076	1.2329
	5	-0.0856	-0.1245	298.715	-0.0161	-0.6572	315.562	0.0311	0.0234	0.7926	0.0279	0.0387	3.5259
2	1	-0.1148	-0.24	-0.2829	-0.039	-0.2411	-0.2903	0.0223	0.0057	-0.0088	0.0502	-0.0011	-0.0136
	2	-0.2009	-0.3371	-0.331	-0.0214	-0.0518	-0.1735	0.0007	-0.0066	-0.0104	0.0119	-0.0157	-0.0157
	3	-0.121	-0.311	-0.3217	0.1177	-0.1252	-0.1585	0.0077	-0.0087	-0.0326	0.0351	-0.0234	-0.0423
	4	-0.5398	-0.5147	-0.5434	-0.3545	-0.2768	-0.3443	0.0008	0.0291	-0.004	-0.0066	0.0298	-0.0016
	5	0.1201	0.0413	0.084	0.3233	0.3191	0.3042	0.0619	0.0571	0.041	0.0608	0.0626	0.0467
3	1	0.0292	0.0168	0.0356	0.0217	-0.0393	0.0013	0.013	-0.0028	0.011	0.008	-0.0111	0.0009
	2	0.0867	0.048	0.0449	0.098	0.0501	0.0766	0.0524	0.0302	0.0359	0.0557	0.03	0.0389
	3	0.0753	-0.0214	0.2784	0.1245	-0.0198	0.1606	0.0317	-0.0095	0.0303	0.041	-0.0177	0.0307
	4	-0.025	0.1305	-0.5806	-0.0466	0.15	0.0041	0.048	0.1134	0.1351	0.0329	0.1217	0.1403
	5	0.021	-0.076	3.4325	0.0494	-0.0458	0.6044	0.0645	0.0013	-0.0191	0.0764	0.0152	0.0039
4	1	-0.038	-0.0577	-0.024	-0.0228	-0.0632	-0.0377	0.0236	0.0178	0.0364	0.0256	0.0031	0.0081
	2	0.0293	-0.0017	0.0553	0.0073	-0.0234	0.0132	0.0392	0.0376	0.0488	0.0351	0.0321	0.0349
	3	0.0401	-0.0618	0.0466	0.0236	-0.0893	0.0014	0.0069	-0.0051	0.0164	0.0194	-0.0165	0.0056
	4	-0.0289	0.0202	0.1427	-0.0464	0.0029	0.0998	0.0145	0.0374	0.0534	0.0059	0.0539	0.0726
	5	0.1158	0.1184	0.194	0.0568	0.0867	0.1496	0.0761	0.0646	0.0489	0.0847	0.0747	0.0556

Experiment 1:

The biases for y1 and y2 were generally positive for both sample sizes though there was some evidence of large negative biases for HF3 and HL3. For every prediction period, HF3 and HL3 were generally the most biased and the other estimators were relatively, albeit slightly, less biased. These biases generally decreased when the sample size was increased from 30 to 60. Generally the limited information estimators had larger biases than their full information analogues, but the differences were only marginal. In some cases the limited information estimators outperformed the full information estimators. These results suggest that the limited information estimators be preferred to the full information estimators according to the bias criterion.

Experiment 2:

The biases observed were generally negative for both y1 and y2. Although the results were somewhat mixed, HF3 and HL3 were generally the most biased for both y1 and y2. However, the full information estimators did not uniformly dominate the corresponding limited information estimators for all forecast periods and for both sample sizes. As the sample size was increased to 60 the biases decreased and the differences between the estimators became less apparent.

Experiment 3:

The ranking of the biases of the estimators over all the five forecast periods was fairly consistent for this experiment. The biases were generally negative for both y1 and y2. HF3 and HL3 generally exhibited the most significant biases for both y1 and y2 and for all the five forecast periods, especially when T = 30. HF2 and HL2 did not perform well either. Generally the full information estimators performed better than their limited information analogues for both y1 and y2 and for most of the forecast periods. However, no specific estimator completely dominated the other estimators for all forecast periods and this observation was true irrespective of the sample size. In fact, the difference among the estimators became less apparent as the sample size was increased from 30 to 60.

Experiment 4:

The biases for y1 and y2 were generally negative for all the five forecast periods. The limited information estimators generally yielded the largest biases for y1 and y2 for all forecast periods compared to the full information estimators though this observation was not true for all the forecast periods. The differences became less apparent as the sample size was increased from 30 to 60.

Comparison of the prediction biases across experiments

For each forecast period the magnitude of the bias generally decreased considerably for experiment 2 compared to those of experiment 1 for both y1 and y2 as the sample size was increased. This observation was particularly true for T = 60, not for T = 30. Experiment 4 produced generally the smallest absolute biases for all forecast periods and for all models compared to experiments 1, 2 and 3. These results seem to suggest that prediction biases are likely to be smaller when the autocorrelation coefficients are smaller.

Mean Square Error of Predictions(MSEP)

Table 5.2-3 and 5.2-4 show the MSEP for y1 and y2, respectively. For each estimator and for each model the MSEP increased as the forecast period extended. This result agrees with the conjecture that the MSEP increases with the distance of the forecast period. In addition, the MSEP's were generally larger for sample size 30 than for sample size 60 for all estimators. Again the MSEP was highest for model 1 compared to other models for all the forecast periods.

From the results of individual experiments we note the following.

Experiment 1:

HF2 and HL2 possessed the largest MSEP's relative to the other estimators. In all cases the full information estimators

TABLE 5.2-3:MSEP FOR Y1

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	2.4695	2.5664	2.4413	2.3582	2.4481	2.3248	2.0816	2.1501	1.9845	1.9707	2.0001	1.9016
	2	6.7482	7.193	6.6701	6.7111	7.6066	6.5269	5.7158	5.8747	5.5378	5.4982	5.8302	5.3442
	3	12.1231	13.3968	11.9732	12.1241	15.7686	11.5806	10.4458	10.5773	10.2088	9.8792	10.8807	9.6273
	4	14.7421	20.407	14.7149	16.5956	30.8863	14.9527	11.8644	13.6958	12.4219	12.024	16.1606	12.1877
	5	15.9523	26.7609	14.601	20.4493	58.4508	14.9995	12.8476	15.0726	12.7405	13.2276	21.246	12.56
2	1	2.4172	2.4373	2.2575	2.4945	2.5478	2.2806	2.1471	2.2024	1.9594	2.2076	2.2704	1.9677
	2	2.7853	2.8388	2.8483	2.9965	3.1151	3.0037	2.7312	2.8461	2.9639	2.8945	3.0538	3.1077
	3	3.4163	3.8435	3.1238	4.2388	5.4867	3.4121	3.0752	3.5935	2.9219	3.5089	4.2637	3.1243
	4	3.85	5.1979	3.2267	6.1729	11.2701	3.8342	4.0729	5.5397	3.7598	5.0788	7.35	4.3015
	5	6.3472	11.0575	3.9374	15.9691	41.0851	5.6294	6.1226	11.1705	5.0201	8.89	17.1715	6.2247
3	1	1.979	2.0117	1.9456	2.09	2.1385	1.9901	1.8723	2.0407	1.8749	1.9519	2.2203	1.8815
	2	2.3618	2.5518	2.4432	2.406	2.699	2.4483	2.1776	2.496	2.3793	2.2215	2.8554	2.4012
	3	2.8373	3.5112	2.7543	2.8948	4.0592	2.7788	2.6557	3.628	2.6153	2.6622	4.8151	2.6696
	4	3.1704	5.567	2.8795	3.4562	7.3087	2.8743	2.7884	6.5651	2.875	3.0579	10.6445	2.8267
	5	4.1803	10.7756	3.2315	4.7584	15.9208	3.2216	3.2489	13.881	2.8733	2.7957	27.0662	2.8756
4	1	2.0015	1.9492	1.8777	2.1251	2.0552	1.9606	1.8894	1.8232	1.7232	1.9508	1.897	1.7574
	2	3.0549	3.0875	3.187	3.1571	3.1783	3.2571	2.8202	2.7924	3.0882	2.8703	2.9169	3.1562
	3	3.4632	3.5619	3.6029	3.5886	3.7841	3.6651	3.2742	3.3035	3.5639	3.2616	3.4939	3.6582
	4	3.6186	3.8336	3.6192	4.0952	4.6261	3.7996	3.3213	3.4326	3.6755	3.5393	4.2283	3.917
	5	4.238	4.814	3.976	5.3331	6.7139	4.2995	3.6546	4.2513	3.7478	4.2143	5.993	4.0756

TABLE 5.2-4:MSEP FOR Y2

Model	h	HF1 (T = 30)	HF2 (T = 30)	HF3 (T = 30)	HL1 (T = 30)	HL2 (T = 30)	HL3 (T = 30)	HF1 (T = 60)	HF2 (T = 60)	HF3 (T = 60)	HL1 (T = 60)	HL2 (T = 60)	HL3 (T = 60)
1	1	0.7662	0.8192	0.8115	0.9312	1.0418	0.9609	0.687	0.5955	0.7618	0.8151	0.6952	0.8761
	2	0.8815	1.1677	0.9139	1.072	1.6201	1.0882	0.7293	0.8307	0.9151	0.8214	0.9643	1.0217
	3	1.8746	2.3304	2.0559	2.5025	3.6773	2.5092	1.7033	1.7456	2.3337	2.1072	2.1743	2.6982
	4	2.7035	4.1494	2.1847	3.3696	7.6508	2.5656	2.321	2.6469	2.1096	2.4121	3.1646	2.2535
	5	2.8658	5.603	2.8737	4.2172	14.5689	3.3487	2.4892	3.3852	2.6594	2.658	4.5783	2.9022
2	1	0.5491	0.5201	0.4483	0.6963	0.7872	0.5573	0.6272	0.7324	0.4355	0.7521	0.8325	0.5067
	2	0.8687	1.0122	0.8354	1.0483	1.2984	0.9784	1.0625	1.2448	1.1852	1.2193	1.4197	1.3256
	3	1.7843	2.5635	1.4348	2.6414	4.5024	1.8676	2.2165	3.2829	1.9981	2.7271	4.0196	2.3317
	4	3.2933	5.8718	2.2653	6.209	14.3768	3.2161	4.2314	7.4335	3.8664	5.711	9.983	4.6921
	5	6.5557	15.5975	3.2488	18.4361	57.6427	5.9895	8.2659	18.0094	6.8633	12.4772	26.9313	8.8628
3	1	0.5085	0.6145	0.4755	0.6248	0.7273	0.5839	0.5224	0.7778	0.4775	0.6424	0.8813	0.5638
	2	2.6729	2.959	2.9272	2.8624	3.2415	3.0208	2.5294	3.0279	3.0178	2.7001	3.4624	3.0548
	3	3.6371	5.1978	3.7927	3.8992	5.769	3.8578	3.4436	6.1743	3.7965	3.7507	7.5125	3.8667
	4	5.5078	11.066	5.1054	5.8868	13.314	5.1601	4.9324	13.3097	5.2123	5.3811	18.7857	5.4135
	5	7.1456	23.4551	4.663	7.9826	29.673	4.6819	6.3051	30.3093	4.506	7.2205	47.8986	4.5087
4	1	0.4708	0.4452	0.3727	0.6043	0.585	0.4968	0.4728	0.453	0.3238	0.601	0.5689	0.41
	2	1.741	1.7106	1.7985	1.9008	1.9181	1.9296	1.6855	1.5593	1.9898	1.7635	1.7521	2.0197
	3	3.1499	3.3581	3.3383	3.5374	3.8313	3.6001	2.9616	3.1679	3.6703	3.297	3.7533	3.8954
	4	3.5253	5.2694	4.6982	5.1999	6.5268	4.9029	4.1423	4.8491	5.5268	4.5335	6.1173	5.775
	5	5.6649	7.4161	4.6736	7.6091	11.0525	5.2199	5.2522	6.9863	5.6573	6.0877	10.2073	6.3414

were just marginally better than their limited information counterparts for each sample size, although for some forecast periods the limited information estimators outperformed their full information counterparts. For all forecast periods, and both sample sizes, HF2 and HL2 performed consistently the worst though the differences between these two and the others were not substantial.

Experiment 2:

Generally the limited information estimators tended to perform worse than their full information counterparts for T = 30. In virtually all cases HF2 and HL2 were outperformed by the other four estimators, and tended to behave erratically(in that the MSEP increased as the sample size increased) for some forecast periods. For the other estimators, the MSEP decreased as the sample size was increased. This observation remained valid even when the sample size was increased. Overall the full information estimators outperformed their limited information counterparts.

Experiment 3:

Again HF2 and HL2 were outperformed by the other estimators. The full information estimators performed better than their limited information counterparts for all forecast periods and for both sample sizes. The MSEP's of HF2 and HL2 tended to behave erratically for some forecast periods as the sample size was increased. The differences amongst the estimators for this experiment where one of the autocorrelation coefficients is negative were substantial.

Experiment 4:

The dominance of the full information estimators over the corresponding limited information estimators was very clear for both y1 and y2. HF2 and HL2 still performed the worst. Overall no large differences among the MSEP's of the six estimators were observed.

Comparison of the MSEP's across experiments

The MSEP's in experiment 2 were smaller compared to those of experiment 1. This observation was true especially for forecast periods 3, 4 and 5. Generally models 1 and 3 yielded larger MSEP's than models 2 and 4, especially for sample size 30. The MSE's were comparable across experiments for sample size 60.

Hypothesis testing

Tables 5.2-5 and 5.2-6 provide the number of type I errors for y1 and y2, respectively, by estimators in the 1000 replications. The number of type I errors between 36 and 64 were not significantly different from 50 at the 5 percent level. Clearly, the number of type I errors were generally

TABLE 5.2-5:NUMBER OF TYPE 1 ERRORS FOR Y1*

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	93	86	88	89	79	89	64	80	79	74	71	71
	2	105	95	113	107	90	104	82	70	83	71	63	75
	3	105	89	115	96	79	95	96	71	97	76	65	77
	4	113	90	118	103	91	103	81	62	94	70	51	74
	5	73	60	73	54	54	58	69	50	71	39	38	44
2	1	102	102	97	99	102	96	77	79	75	80	78	71
	2	86	85	102	101	101	113	83	85	97	75	77	103
	3	61	65	67	57	54	57	41	43	48	38	38	42
	4	63	64	68	56	57	70	43	44	48	46	43	53
	5	33	31	35	24	27	29	24	23	29	24	23	34
3	1	82	83	82	98	87	91	85	79	68	89	83	76
	2	98	77	71	96	80	82	74	54	62	75	59	70
	3	44	32	37	28	29	36	47	27	28	38	26	32
	4	75	76	66	79	72	78	61	47	48	60	55	61
	5	6	5	3	4	3	7	11	5	13	5	4	7
4	1	80	74	82	97	93	91	82	72	68	87	80	73
	2	88	81	86	103	104	106	80	70	80	83	76	93
	3	53	52	52	54	55	54	50	43	45	48	39	41
	4	70	60	56	68	64	65	58	50	48	55	48	59
	5	23	19	15	19	19	21	21	21	18	16	17	24

* The number of type 1 errors between 36 and 64
are not significantly different from 50 at the 5 percent level

TABLE 5.2-6:NUMBER OF TYPE I ERRORS FOR Y2*

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	81	71	71	68	51	62	81	50	65	79	45	58
	2	43	41	50	33	29	45	42	42	61	33	29	57
	3	40	31	35	25	20	22	67	65	85	52	48	62
	4	62	70	52	34	37	32	86	100	65	58	60	45
	5	20	23	21	13	14	12	24	29	21	15	20	18
2	1	87	88	72	76	75	62	73	75	54	67	70	46
	2	74	76	70	60	57	63	44	42	60	55	47	62
	3	21	22	20	17	16	20	26	30	24	21	28	24
	4	54	55	55	46	50	50	34	35	33	36	40	32
	5	19	18	19	16	16	17	8	7	10	13	15	14
3	1	81	68	51	58	58	36	82	69	32	67	63	33
	2	84	71	88	98	78	91	76	44	61	74	45	55
	3	26	17	20	19	15	24	37	23	26	27	21	23
	4	74	62	73	83	78	85	50	31	42	56	41	49
	5	2	2	2	0	0	2	8	1	2	4	1	4
4	1	90	82	53	64	63	42	83	60	34	76	66	36
	2	83	76	84	86	82	78	73	52	67	78	58	68
	3	25	24	26	33	30	36	36	29	30	34	27	27
	4	77	74	71	85	82	94	46	32	43	62	42	57
	5	8	6	9	15	12	17	11	6	5	12	9	9

* The number of type 1 errors between 36 and 64 are not significantly different from 50 at the 5 percent level

greater for sample size 30 than for sample size 60, suggesting that very large size samples might be required for accurate testing of hypotheses concerning forecasts. The model 1, in which the autoregressive parameters in both equations assumed the high value of 0.9, resulted in the greatest number of rejections compared to the other three models for both sample sizes. To give an overall idea of the effect of changes in the coefficient of the lagged endogenous variables, autocorrelation and sample size on the reliability of the asymptotic standard errors, the number of significant predictions defined as those yielding more than 64 type I errors in 1000 replications for all the estimators and for all the five forecast periods and for both y1 and y2 are given below.

The number of significant predictions for both y1 and y2

	T = 30	T = 60
Model 1	31	32
Model 2	24	17
Model 3	35	16
Model 4	29	19

The results of individual experiments are discussed below.

Experiment 1:

The number of type I errors was significant in the sense that the estimators performed worse than expected, for many of the forecast periods, especially for y1. The number of significant forecasts of both y1 and y2 for the five forecast periods combined was 31 (out of 60) for sample size 30, and 32 for sample size 60. However, the number of type I errors tended to decrease as the sample size was increased. Clearly, when T = 30, the asymptotic standard errors led to too many rejections compared to the the case when T = 60. In terms of the relative performances of the estimators, for each sample size the limited information estimators tended to perform marginally better than the full information estimators for virtually all the forecast periods.

Experiment 2:

The usefulness of the asymptotic standard errors were quite good especially for 3 period ahead and 5 period ahead forecasts. The number of type I errors decreased considerably. The difference between the limited information estimators and the full information estimators became less pronounced though the limited information estimators were still marginally better. The number of significant predictions of both y1 and y2 for all the five forecast periods combined was 24 for T=30, and 17 for T=60. The number of type I errors decreased as the sample size was increased from 30 to 60.

Experiment 3:

The full information estimators dominated the limited information estimators. The number of significant predictions of y1 and y2 for all the five forecast periods combined was 35 for sample size 30 and 16 for sample size 60.

Experiment 4:

The full information estimators dominated the limited information estimators when one of the autocorrelation coefficients is zero, especially for sample size 30. However, the differences between the full information estimators and the limited information estimators became even smaller as the sample size was increased. The number of significant predictions of y1 and y2 for all the five forecast periods combined was 29 for sample size 30, and 19 for sample size 60.

Comparison of the number of type I errors across experiments

The number of type I errors were generally lower in experiments 2, 3 and 4 than those in experiment 1. Furthermore, the number of type I errors generally decreased as the sample size was increased from 30 to 60.

These results suggest that at low degrees of autocorrelation the full information estimators are more reliable in tests of hypotheses concerning forecasts, and at high degrees of autocorrelation, the asymptotic standard errors of the limited information estimators performed better. This might be due to the fact that computation of the asymptotic covariance matrix of the full information estimators involves more matrix inversions than in the limited information case and this increases the possibility of rounding errors.

The main conclusion that has emerged is that inferences concerning forecasts are not reliable in cases where autocorrelation is high or when the sample size is small. If the autocorrelation coefficients are high, then very large samples are required for valid and useful inferences about the forecasts. Even a sample of size 60 used in this Monte Carlo study yielded too many rejections for model 1 where autocorrelation was extremely severe.

Skewness

Tables 5.2-7 and 5.2-8 provide skewness statistics for y1 and y2, respectively. An inspection of these statistics revealed both positive and negative skewness. These values were fairly close to zero for all estimators (except, in some cases for HL1 and HL2 in experiment 2). However, as the sample size was increased the distribution became almost symmetric as was evidenced by the smaller values of the skewness statistic. In all experiments HL2 had generally the most skewed distributions thus confirming the speculation of Moazzami and Buse (1986) that this estimator might be significantly skewed. There was no noticeable pattern in skewness of the empirical distribution of forecasts of estimators as the forecast period was extended into the future.

TABLE 5.2-7:SKEWNESS STATISTICS FOR Y1

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	-0.0083	-0.0067	-0.0391	0.0405	0.0763	0.0279	0.0306	0.0676	0.0631	0.0423	0.0762	0.0818
	2	-0.0202	0.0289	-0.0537	0.0022	0.1492	-0.0412	0.0108	0.0288	0.0259	0.0098	0.0601	0.029
	3	0.0459	0.1822	0.0024	0.0798	0.3488	-0.0119	0.0718	0.0508	0.0717	0.0493	0.0662	0.0597
	4	0.0572	0.676	-0.0323	0.3191	1.0095	-0.0488	-0.0172	-0.0037	0.011	0.0198	0.191	-0.0008
	5	0.1432	1.5508	-0.0641	0.7908	1.4276	-0.0795	-0.0394	-0 1331	-0.021	-0.0145	-0.0472	-0.0522
2	1	-0.0742	-0.034	-0.1411	0.0791	0.1908	-0.0617	-0.1324	-0.031	-0.0813	-0.0576	0.0671	-0.0354
	2	0.052	0.0696	0.0529	0.0448	0.1372	0.0182	-0.0577	-0.0899	-0.0614	-0.1291	-0.1482	-0.1346
	3	0.467	1.024	0.101	1.3557	3.0143	0.1907	0.0946	0.3648	-0.0214	0.1988	0.5521	-0.077
	4	0.8253	2.653	-0.116	3.539	7.4567	-0.0453	-0.1893	0.3236	-0.5737	-0.2117	0.3617	-0.8358
	5	4.2231	7.6015	0.1481	10.1444	13.0066	0.8127	-0.2378	1.4238	-0.9566	-0.2287	1.2874	-1.4496
3	1	-0.1507	0.1279	-0.0374	-0.106	0.2367	-0.0215	-0.1023	0.1295	0.0198	-0.0521	0.2791	0.0209
	2	-0.0084	0.0507	0.023	-0.0405	0.1093	0.0204	-0.0147	-0.0121	0.0284	-0.0432	0.1187	0.0182
	3	0.0773	0.1092	0.0063	0.0614	0.6013	0.009	0.0613	0.1694	0.0487	0.0429	0.7973	0.0641
	4	0.4752	0.866	-0.0581	0.5037	2.085	-0.0557	-0.0769	1.6577	-0.0459	0.0121	2.8872	-0.0524
	5	1.5031	0.8688	0.1376	1.6569	3.6849	0.1363	0.1074	3.251	0.2033	0.4366	5.4902	0.2036
4	1	-0.1686	-0.0968	-0.0664	-0.1183	-0.0212	-0.0543	-0.1288	-0.0445	-0.0063	-0.0562	0.048	0.0253
	2	0.0006	0.0139	0.0033	-0.0161	0.0031	-0.0227	-0.0073	-0.027	-0.0169	-0.018	-0.036	-0.0248
	3	0.0883	0.1216	0.0414	0.0843	0.1702	0.0243	0.1009	0.0457	0.0777	0.0772	0.0347	0.0628
	4	0.1254	0.2378	-0.1035	0.4109	0.6937	-0.1235	-0.0721	-0.0098	-0.093	-0.0215	0.1739	-0.181
	5	0.5971	1.1212	-0.0223	1.3284	2.0677	-0.2908	0.1007	0.188	-0.0154	0.2481	0.5541	-0.2623

TABLE 5.2-8:SKEWNESS STATISTICS FOR Y2

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	0.3094	0.4505	0.1942	0.2241	0.5052	0.1369	0.3474	0.2226	0.1663	0.2453	0.2158	0.0957
	2	0.1465	0.6747	-0.0267	0.2314	0.9101	-0.0171	0.1092	0.2077	-0.0311	0.1252	0.3636	0.0053
	3	0.3205	0.8416	0.1004	0.3231	0.8777	0.1925	0.177	0.2597	0.1234	0.1503	0.2774	0.1643
	4	0.4287	1.8092	0.0754	0.8772	2.1464	0.2941	-0.0836	0.2551	-0.1174	-0.0131	0.5485	0.0494
	5	0.3326	2.0332	0.08	1.0933	2.0356	0.2242	-0.1166	0.0096	0.0251	-0.0525	-0.0798	0.3084
2	1	0.895	0.3565	0.8895	0.8816	0.1473	1.021	1.0065	0.5457	0.9184	1.1942	0.6514	1.2155
	2	0.1367	-0.2521	0.2296	-0.0016	-0.9132	0.5173	0.5026	0.3922	0.802	0.7245	0.6674	1.0473
	3	-1.1538	-2.7013	0.0405	-2.566	-5.1646	0.2941	0.0441	-0.6188	0.9281	0.1898	-0.3992	1.3837
	4	-2.6866	-5.3547	-0.0825	-5.7316	-9.0384	0.0575	0.2286	-0.8153	1.5239	0.237	-0.6374	2.0492
	5	-6.2957	-9.2119	-0.6597	-10.4724	-12.8631	-1.3997	0.2586	-1.3485	2.0379	0.0844	-1.2606	2.6276
3	1	0.7677	-0.3942	0.2352	0.8881	-0.2627	0.2438	0.9256	-0.63	0.4942	1.3042	-0.4853	0.6186
	2	0.1421	-0.1884	0.0731	0.1607	-0.3073	0.0639	0.1025	-0.2869	0.0468	0.1204	-0.3806	0.0688
	3	-0.0514	-0.1373	0.0182	0.0164	-0.6648	0.0496	0.0593	-0.5049	-0.0052	0.0356	-0.8207	0.0232
	4	-0.4722	-0.1002	-0.015	-0.4981	-1.869	-0.0194	-0.0076	-1.3755	-0.0646	-0.099	-2.0874	-0.0565
	5	-2.2928	0.3349	-0.0233	-2.0746	-3.3956	-0.0335	-0.0572	-2.868	-0.0828	-0.3485	-3.975	-0.0316
4	1	0.8192	0.122	0.4192	0.9557	0.2981	0.4812	0.9655	0.2963	0.4635	1.3002	0.5957	0.6512
	2	0.3082	0.2049	0.2023	0.2261	0.0039	0.2116	0.3262	0.1638	0.3352	0.3248	0.1199	0.3715
	3	-0.0618	-0.0742	0.0927	-0.0794	-0.2945	0.2449	0.0921	0.0816	0.2884	0.1116	0.0074	0.3984
	4	-0.4769	-0.5128	0.0612	-0.9493	-1.34	0.2091	0.0311	0.0216	0.3748	-0.0435	-0.2607	0.5516
	5	-1.9473	-2.0001	0.1978	-3.3596	-3.6182	0.3032	0.0179	-0.2552	0.7729	-0.2337	-0.8592	1.1867

The distributions of forecast were more skewed in model 2 than in model 1 which suggests that decreasing the coefficient of the lagged endogenous variable and autocorrelation seems to increase skewness. In fact, the absolute values of the skewness statistic were much higher for models 2, 3 and 4 than for model 1. Furthermore, the skewness was most uniform across estimators for model 4. Again it must be borne in mind that, in virtually all cases, the values of the skewness statistic were fairly close to zero.

Kurtosis

Tables 5.2-9 and 5.2-10 show the kurtosis statistics for y1 and y2, respectively.

In all experiments the values of the kurtosis statistic were greater than 3 implying that they have more peaked distributions compared to the standard normal distribution. However, as the sample size was increased from 30 to 60, the values of the kurtosis statistic generally decreased though still greater than 3. Although no estimator consistently had the highest values of the kurtosis statistic for all the forecast periods, HF2 and HL2 tended to have the most peaked distributions in all experiments and for both sample sizes. The kurtosis generally decreased as the sample size was increased and this was true in all experiments.

TABLE 5.2-9:KURTOSIS STATISTICS FOR Y1

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	3.0113	2.997	3.0662	3.022	3.0895	3.1169	2.9964	2.9572	3.0434	2.987	2.9954	3.0404
	2	3.0887	3.0617	3.0796	3.0948	3.3564	3.0981	3.1219	3.0836	3.144	3.096	3.3057	3.1168
	3	2.8926	3.2008	2.8604	2.9946	4.6441	2.8053	2.9721	2.9456	3.0116	2.8813	3.5195	2.9339
	4	3.0673	5.4959	2.8665	4.1466	10.0805	2.7575	2.88	3.1643	3.0157	2.8894	5.7239	2.9624
	5	3.8407	12.7521	3.0551	9.3557	21.2604	3.026	3.0967	3.9863	3.2184	3.2554	10.5731	3.2618
2	1	3.8141	3.8061	3.5512	4.117	4.259	3.5421	3.3722	3.395	3.1628	3.2944	3.3147	3.0902
	2	3.0905	3.0885	3.0735	3.0945	3.2768	3.0592	3.047	3.1812	3.0445	3.2478	3.3986	3.2146
	3	4.8031	8.2313	3.2325	11.7677	29.0179	3.4969	3.2686	5.027	3.079	4.3668	6.5229	3.416
	4	8.9592	26.6985	3.187	43.7007	102.977	5.1054	5.1292	11.8249	4.5153	8.1897	15.789	6.002
	5	56.8952	112.015	6.4103	180.247	225.063	20.7864	8.7054	31.5842	7.2113	13.7831	31.8886	9.6128
3	1	3.3395	3.2473	2.818	3.3779	3.4596	2.8511	3.3406	3.1478	2.8584	3.3853	3.3725	2.8493
	2	2.9063	3.0209	2.8741	2.8333	3.1154	2.8541	2.9379	3.3942	2.9073	2.8956	4.0692	2.8909
	3	3.227	6.0623	3.147	2.9887	6.3313	3.0163	3.0197	6.2305	2.9357	3.076	10.3457	2.9495
	4	6.4869	17.1052	2.952	6.0684	17.1867	2.9863	3.1336	21.7959	2.8874	3.5224	35.5323	2.9239
	5	17.2746	43.9959	3.2348	14.893	37.6831	3.0761	3.5479	57.2016	3.2249	4.9271	90.3921	3.237
4	1	3.3763	3.1956	3.0928	3.3504	3.2646	3.1267	3.2065	3.0117	2.9423	3.1866	3.0704	2.9453
	2	2.9791	2.9753	3.0117	2.9273	2.8945	3.0012	3.0064	2.9738	2.9292	2.9578	2.9346	2.889
	3	3.1386	3.1953	3.0867	2.9024	3.0654	2.953	2.9133	2.9211	2.9481	2.8869	2.9635	2.9403
	4	3.6076	4.0102	2.8541	4.9586	5.5394	2.8669	2.7497	2.9427	2.913	2.9716	3.9236	3.2892
	5	6.3259	11.4611	3.2904	12.2537	17.9096	4.5386	3.0993	4.3726	3.5451	3.771	7.3072	4.4697

TABLE 5.2-10:KURTOSIS STATISTICS FOR Y2

		T = 30						T = 60					
Model	h	HF1	HF2	HF3	HL1	HL2	HL3	HF1	HF2	HF3	HL1	HL2	HL3
1	1	3.1452	3.8305	3.1496	3.0101	3.9722	3.1006	3.3339	2.9959	3.015	3.1604	3.1268	2.9911
	2	3.4015	5.5691	2.8256	3.9636	7.2913	2.9345	3.3272	3.9807	3.031	3.3653	4.8627	3.149
	3	3.4205	7.3798	2.7366	4.0736	10.6341	2.9771	2.9644	3.6373	2.8899	3.1307	4.9125	3.1409
	4	5.3111	13.119	2.9884	7.8304	18.643	3.4003	3.3996	4.2937	3.3031	3.4539	6.5457	3.6775
	5	5.3088	18.1823	3.5377	17.3707	31.2828	4.1633	3.1041	3.8742	3.6081	3.2895	8.8862	4.8877
2	1	7.4778	8.4138	6.8649	8.1601	9.9117	7.452	7.1341	6.983	6.5355	7.9137	7.885	7.8414
	2	5.0257	7.7031	4.8345	5.7265	12.5594	5.4051	5.377	7.1647	5.4858	6.1817	7.4409	6.3583
	3	11.9492	27.6837	4.625	27.6533	60.7561	8.224	7.0993	13.6771	6.8692	8.571	12.5241	8.8282
	4	30.1302	65.4265	7.0086	77.2435	131.0016	16.9755	9.8504	21.3538	9.4576	13.2638	20.419	12.8716
	5	89.8429	138.55	15.3411	175.724	215.9977	39.797	15.1811	35.5701	13.3982	19.8964	33.2927	17.3987
3	1	7.0122	6.792	3.7249	7.0003	6.6423	3.6776	6.9226	5.8989	4.0693	8.458	6.7305	4.4777
	2	3.429	4.274	2.9653	3.5633	4.3957	3.0197	3.2927	3.8966	2.9372	3.434	4.4256	2.9995
	3	3.4887	10.0972	2.9176	3.4424	8.1486	2.9627	3.0875	9.1204	2.9215	3.2548	10.6198	2.9228
	4	5.4758	26.264	3.2532	5.3083	18.9671	3.1874	3.6168	19.377	2.9931	4.067	24.5925	3.0876
	5	26.0913	67.6158	2.8842	20.8174	39.6803	2.9213	4.362	45.065	2.8006	5.8002	56.6479	2.9181
4	1	7.1681	6.2039	4.5745	7.5404	6.6138	5.0194	6.9834	6.5082	4.8055	8.1779	7.8089	5.5148
	2	4.4374	4.4578	3.6121	4.4133	4.4539	3.5754	3.8375	3.5904	4.3701	4.0029	3.986	4.4176
	3	3.7084	4.4706	3.0264	4.4851	4.878	3.3008	3.1096	3.668	3.8467	3.3281	4.2316	4.0954
	4	5.7885	7.9302	3.6517	9.9197	10.7725	4.1209	3.1942	4.173	5.0078	3.6967	5.7246	5.8927
	5	21.9121	29.7474	3.4842	40.2248	33.4365	4.058	3.3584	6.6108	8.7033	5.1462	10.9218	11.8241

5.3 Ranking statistics

The ranking statistics presented here show how changes in the coefficients of the lagged endogenous variables and autocorrelation affect the MSE's of the structural parameters and forecasts. We discuss the ranking statistics for structural estimation and prediction separately.

Ranking statistics for structural estimation

The ranking statistics for the structural parameters based on the MSE are given in Table 5.3-1. The sums of the ranks given in Table 5.3-1 are based on all the structural coefficients (excluding the two intercept parameters). The estimators are first ranked in individual experiments and then an overall ranking is provided for all the experiments combined. The values and the significance of Kendall's coefficient of Concordance (W) are also given in the last column of the table. These values are used to compute the ranks of the estimators displayed in Table 5.3-1 (a).

Although HF1 emerged as the winner in experiment 1, the value of Kendall's W is low and insignificant at the 5 percent level indicating a lack of evidence of systematic ranking of the estimators. This observation is true for both sample sizes 30 and 60. HF1 also dominated in all the other models (i.e., models 2, 3 and 4). For all models, the ranking of the estimators became weaker when the sample size was increased from 30 to 60. Furthermore, the ranking of the estimators based on all models did not produce any clear winner due to the low value of Kendall's W (0.327 for sample size 30 and 0.129 for sample size 60). Based on this criterion HF1 came out as the overall winner both for sample size 30 and sample size 60. HF3, HL1 and HF2 also occupied the top four positions based on this criterion. However, the low value of Kendall's W for both sample sizes were not significant enough to support the hypothesis of a systematic ranking of the six estimators. The ranking of the estimators was the strongest for model 3 wherein the two autoregressive parameters assumed opposite signs. In this case HF1 emerged as the overall winner for both sample sizes, with its limited-information counterpart HL1 followed as a close second and this, in tr[illegible] was followed by HF3 and HL3. HF2 and HL2 occupied the last two positions. This ranking was the strongest for T=30 (W=0.721) and for T=60 (W = 0.584). The rankings remained basically the same for the other models. Thus, despite the rather weak rankings of estimator performance for models 1, 2 and 4, there was no doubt that based on the MSE criterion, HF2 and HL2 (occasionally HL3) performed consistently worse. However, HF2 performed surprisingly well in model 2, sample size 60, where the coefficient of the lagged dependent variable assumed a lower value. Kendall's W (0.089) for this particular case led to a clear rejection of the hypothesis of a systematic ranking among the estimators. Thus the strongest ranking was obtained in cases where the difference in the autocorrelation coefficients was larger. Incidentally, HF2 and HL2 ranked last according to this criterion. Similar observations are valid of the ranking statistics for the autoregressive coefficients indicated in Table 5.3-2. Except for model 3, the ranking was very weak for T = 60. The fact that significant differences among the estimators occurred in small samples can be seen from the results for the autoregressive coefficients in model 4. Again HF1 and HF3 and their limited information counterparts occupied the top four positions with HF2 and its limited information counterpart HL2 occupied the last two positions. HF1 performed relatively the best in estimating the structural parameters. Again the ranking was the strongest for model 3. Note that significant differences among the estimators emerged only when the autocorrelation coefficients were larger and when the sample size was smaller. This suggests that the choice among these estimators becomes important when either the sample size is small or when the magnitudes of the autocorrelation coefficients are relatively large.

Ranking statistics for predictions

Table 5.3-3 gives the ranking statistics for the predictions based on MSEP. The last column of the table gives the values of Kendall's Coefficient of Concordance, W. The sums of the ranks relate to both y1 and y2 combined over the five forecast periods.

TABLE 5.3-1: RANKING STATISTICS FOR STRUCTURAL PARAMETERS BASED ON MSE

a)Sum of the ranks of the structural coefficients(excluding the intercept)

		HF1	HF2	HF3	HL1	HL2	HL3	W
	Model 1	24	24	26	30	32	32	0.067
	Model 2	14	27.5	21.5	23	44	38	0.561*
T = 30	Model 3	11	39	22	20	45	31	0.721*
	Model 4	16	29.5	21.5	26	29	36	0.339*
	All Models	65	120	91	99	160	137	0.327
	Model 1	23	23	29	29	31.5	32.5	0.076
	Model 2	20	27	29.5	27.5	32	32	0.089
T = 60	Model 3	16	40	20	19.5	43	29.5	0.584*
	Model 4	22	33	24.5	22.5	31.5	34.5	0.143
	All Models	81	123	103	98.5	138	128.5	0.129

W denotes Kendall's Coefficient of Concordance
* Significant at the 5 percent level of significance

b) Ranking of estimators based on the sums in 5.3-1a(Best to worst)

	Model 1	HF1 / HF2 HF3 / HL1 HL2 HL3
	Model 2	HF1 HF3 HL1 HF2 HL3 HL2
T = 30	Model 3	HF1 HL1 HF3 HL3 HF2 HL2
	Model 4	HF1 HF3 HL1 HF2 HL3 HL2
	All Models	HF1 HF3 HL1 HF2 HL3 HL2
	Model 1	HF1 / HF2 HF3 / HL1 HL2 HL3
	Model 2	HF1 HF2 HL1 HF3 HL2 / HL3
T = 60	Model 3	HF1 HL1 HF3 HL3 HF2 HL2
	Model 4	HF1 HL1 HF3 HL2 HF2 HL3
	All Models	HF1 HL1 HF3 HF2 HL3 HL2

/ between two estimators denotes a tie between these estimators

TABLE 5.3-2: RANKING STATISTICS FOR THE AUTOREGRESSIVE PARAMETERS BASED ON MSE

a)Sum of ranks of non-zero elements of R and Σ

		HF1	HF2	HF3	HL1	HL2	HL3	W
	Model 1	10	15	13	21	24	22	0.36
	Model 2	11	15	12	24	27	16	0.488*
T = 30	Model 3	7	20	10	20	29	19	0.717*
	Model 4	11	16	6	25	26	21	0.726*
	All Models	39	66	41	90	106	78	0.511*
	Model 1	11	18	14	18	23	21	0.223
	Model 2	13	22	14	20	24	12	0.301
T = 60	Model 3	10	24	14	13	27	17	0.506*
	Model 4	16	23	12	14	24.5	15.5	0.293
	All Models	50	87	54	65	98.5	65.5	0.258

W denotes Kendall's Coefficient of Concordance

* Significant at the 5 percent level of significance

b) Ranking of estimators based on the sums in 5.3-2a(Best to worst)

	Model 1	HF1	HF3	HF2	HL1	HL3	HL2
	Model 2	HF1	HF3	HF2	HL3	HL1	HL2
T = 30	Model 3	HF1	HF3	HL3	HF2 / HL1		HL2
	Model 4	HF3	HF1	HF2	HL3	HL1	HL2
	All Models	HF1	HF3	HL3	HL1	HF2	HL2
	Model 1	HF1	HF3	HF2 / HL1		HL3	HL2
	Model 2	HL3	HF1	HF3	HL1	HF2	HL2
T = 60	Model 3	HF1	HL1	HF3	HL3	HF2	HL2
	Model 4	HF3	HL1	HL3	HF1	HF2	HL2
	All Models	HF1	HL1	HF3	HL3	HF2	HL2

/ between two estimators indicates a tie between these estimators

TABLE 5.3-3: RANKING STATISTICS BASED ON MSEP

a)Sum of the ranks for all the 5 forecast periods

		HF1	HF2	HF3	HL1	HL2	HL3	W *
	Model 1	24	47	18	36	58	27	0.656*
	Model 2	25	38	12	49	60	26	0.881*
T=30	Model 3	21	47	18	38	60	26	0.768*
	Model 4	19	29	22	45	55	40	0.563*
	All Models	89	161	70	168	233	119	0.636*
	Model 1	25	42	32	28	51	32	0.304*
	Model 2	22	43	14	39	59	33	0.726*
T=60	Model 3	17	49	21	29	60	34	0.787*
	Model 4	19	28	33	35	50	45	0.311*
	All Models	83	162	100	131	220	144	0.423*

W denotes Kendall's Coefficient of Concordance
* Significant at the 5 percent level of significance

b)Ranking of estimators based on the rank sums in 5.3-3a (Best to worst)

	Model 1	HF3	HF1	HL3	HL1	HF2	HL2
	Model 2	HF3	HF1	HL3	HF2	HL2	HL2
	Model 3	HF3	HF1	HL3	HL1	HF2	HL2
	Model 4	HF1	HF3	HF2	HL3	HL1	HL2
T = 30	All Models	HF3	HF1	HL3	HF2	HL1	HL2
	Model 1	HF1	HL1	HF3	HL3	HF2	HL2
	Model 2	HF3	HF1	HL3	HL1	HF2	HL2
	Model 3	HF1	HF3	HL1	HL3	HF2	HL2
	Model 4	HF1	HF3	HF2	HL3	HL1	HL2
T = 60	All Models	HF1	HF3	HL1	HL3	HF2	HL2

A comparison of the entries in table 5.3-3 with those in tables 5.3-1 and 5.3-2 suggests that the rankings of the estimators according to the MSEP were much stronger than the rankings of the estimators based on the structural parameters for both sample sizes. However, the rankings were stronger for T = 30 than for T = 60. Again, for all models, HL2 came dead last with HF2 not very far behind. The full information estimators HF1 and HF3 occupied spots 1 and 2 interchangeably with their limited information counterparts following immediately behind. The ranking was extremely strong for model 3 where there is a substantial difference in the autoregressive coefficients. The rankings were fairly consistent for all the four models. Thus using the MSEP criterion HL2 should possibly not be considered seriously for purposes of prediction. The rankings of the estimators on the basis of all the four models produced weak rankings (W=0.327 for sample size 30 and W = 0.129 for sample size 60). As for the reliability of asymptotic covariance formulae, the full information estimators completely dominated in model 4 where one of the autocorrelation coefficients is zero. Accordingly the full information estimators tended to be relatively better for prediction if the autocorrelation coefficient was low. This ranking is what we would expect from asymptotic theory. However, the relatively low value of W for model 4 (T=60) precludes the advantage which the full information estimators might have over the limited information estimators for testing hypotheses about predictions. Also the weak overall ranking of the estimators suggest that we should be concerned with the choice of the estimators either when the sample size is very small or when the differences in the autocorrelation coefficients are large as in model 3.

5.4 Density estimates

The kernel density estimates of the sampling distributions of the structural parameters and the forecasts are reported below in separate sections.

To facilitate comparison of the estimators all the density estimates of the structural estimators/forecasts are drawn to the same scale. Also, in reporting the density estimates, the horizontal scale consists of the difference between the estimated value and the true value of the corresponding structural coefficient. However, it turns out that the direction of the biases are not very apparent from the density estimates. To avoid the complication which would arise in summarizing all the information regarding each experiment in a single diagram, the diagrams are presented in sets. Each set consists of 8 diagrams, each depicting the results of one of the eight sub-experiments corresponding to a particular structural coefficient or forecast, namely: 1A, 1B, 2A, 2B, 3A, 3B, 4A and 4B. For example, for a particular structural coefficient or forecast, experiment 1A, depicts the result, for that structural parameter or forecast, for model 1, sample size 30. The A refers to sample size 30 and B refers to sample size 60. Similar interpretations are made for experiments 2, 3 and 4. Because of the similarities in the densities of the six estimators studied here, we chose to make general observations concerning the behaviour of the density estimates rather than venture into ranking the estimators on the basis of such density estimates.

Density estimates of structural parameters

Since the density estimates are quite similar and due to the space limitations we only report the density estimates of the coefficient of one lagged endogenous variable (coefficient number 2) and 2 exogenous variables (coefficients number 4 and 10). A major limitation of the density estimates is that the degree of comparability depends on the scale used. For this reason we chose to interpret the skewness and kurtosis statistics together with the density estimates. Although the results for other structural parameters are not reported here, it must be mentioned that they are as suggested by the skewness and the kurtosis statistics presented in tables 5.1-5 to 5.1-8.

Density estimates of coefficient number 2

The set of 8 diagrams, referred to as set 1, provides the density estimates of coefficient number 2, the coefficient of the lagged endogenous variable in the first equation.

A careful inspection of the density estimates indicates a great deal of similarity among the six estimators. As noted from the summary statistics above, the biases of coefficient number 2 were generally negative and significant. Although this information was not very apparent from the density estimates (because of the scale used) there was nevertheless a slight evidence of negative bias, especially in experiments 2, 3 and 4. Also the dispersions of the estimators were generally greater for model 3 than for other models. Furthermore, the distributions appeared to be almost symmetric and exhibited sharper peaks. However, as the sample size was increased from 30 to 60, the peakedness of the distributions of the structural parameters increased considerably thus confirming the information provided by the skewness and kurtosis statistics. The value of the kurtosis statistic increased as we increased the sample size from 30 to 60 and that all values were greater than 3 indicating that the distributions were more peaked than the standard normal distributions.

Density estimates of coefficient number 4

Set 2 summarizes the density estimates of the coefficient of the exogenous variable (coefficient no. 4). An inspection of the density estimates reveals marked differences in the distributions of this coefficient, especially for model 3 where the autocorrelation coefficients differed widely between the two equations. Again the distributions of the six estimators were almost symmetric. The density estimates also reveal that the peakedness of this distribution was not as high as for the corresponding values of coefficient No. 2. For model 3 we note that HF1 and HF2 have the most peaked distributions. Once again the density estimates have confirmed the information provided by the skewness, the kurtosis and the ranking statistics.

Density estimates of coefficient number 10

Set 3 presents the density estimates for the coefficient of the exogenous variable (coefficient number 10). The density estimates revealed near symmetry of the distributions of estimators in all models. The pronounced peakedness of the distributions relative to the standard normal distribution was also confirmed. Again the biases were not very apparent from the density estimates (due to the scale) but the most significant differences among the estimators occurred in model 3. For both sample sizes, the estimators had the largest dispersions for model 1 and this was picked up by the density estimates. Also the peakedness of the distributions of the estimators increased and the MSE decreased as the sample size was increased. This information, as provided by the density estimates, was also contained in the descriptive statistics.

Prediction

For prediction we only present the results for the three-step ahead and the five-step ahead forecasts for each of the two endogenous variables.

Density estimates of three-step ahead forecasts

Set 4 presents the density estimates of the three-step ahead forecast errors for the first endogenous variable, y1. Again the direction of the biases is not easy to see from the density estimates. An inspection of the density estimates, however, reveals that the distributions were almost symmetric. However, there was little change in the shape of the distributions of the estimators as we increased the sample size from 30 to 60.

Referring to the skewness statistics in tables 5.2-7 and 5.2-8, the absolute value of the skewnwess was fairly close to zero indicating near symmetry. Increasing the sample size from 30 to 60 resulted in almost near symmetry of the distributions as the skewness statistics decreased. The kurtosis, however, was fairly constant as the sample size was increased. This information was also revealed by the density estimates.

Set 5 presents the density estimates of the three-step ahead forecast errors for y2. The density estimates of the three step ahead forerecast for y2 were almost identical to the three-step ahead forecast for y1 and hence not discussed in detail.

b) Density estimates of five-step ahead forecasts

Set 6 presents the density estimates of the five-step ahead forecast errors for y1. Again as noted above the biases were quite significant although it was not apparent from the density estimates. The density estimates became more symmetric as the sample size was increased from 30 to 60. However, the value of the kurtosis statistic decreased as the sample size was increased. This information was also contained in the density estimates.

Set 7 presents the density estimates of the five-step ahead forecast errors for y2. Again the observations are quite similar to that of the five period ahead forecast for y1.

5.5 Summary of major observations:

a) Structural estimation

1. The null hypothesis that the bias is equal to zero was rejected at the 5 percent level of significance for all estimators irrespective of the parameter, sample size or the model. The bias and the MSE generally decreased as the sample size was increased from 30 to 60. These observations were also true of the elements of R and Σ. The biases of the autoregressive parameters were generally negative. Furthermore, r_{11} and r_{22} had relatively smaller biases than the elements of Σ. The absolute values of the biases tended to decrease with the absolute magnitudes of the autocorrelation coefficients.

2. For structural estimation the full information estimators HF1 and HF3 and their limited information counterparts performed relatively better than HF2 and HL2 according to the MSE criterion. However, the performance of HF2 and HL2 improved in cases where the differences between the autocorrelation coefficients were not very large. Considering the limited information estimators in isolation we note that HL2 performed consistently worse than HL1 and HL3. These results were similar to the findings of Moazzami and Buse(1986).

3. The rankings of the estimators were very strong for model 3 in which there was a large difference between the autocorrelation coefficients of the first and the second equation. Thus it seems that differences in the relative performances of estimators emerge in cases where the autocorrelation coefficients of the different equations in a model vary considerably. The rankings became weaker as we increased the sample size suggesting that the choice among estimators is relatively more important when the sample size is small.

4. The number of type I errors encountered in each experiment was zero for all the six estimators. This anomalous result questions the reliability of the asymptotic standard errors computed from the asymptotic covariance matrix of estimators for testing hypotheses in small sample situations. More specifically, there is need to make significant small sample

adjustments in tests of hypotheses concerning structural parameters if the tests are to be of the right power.

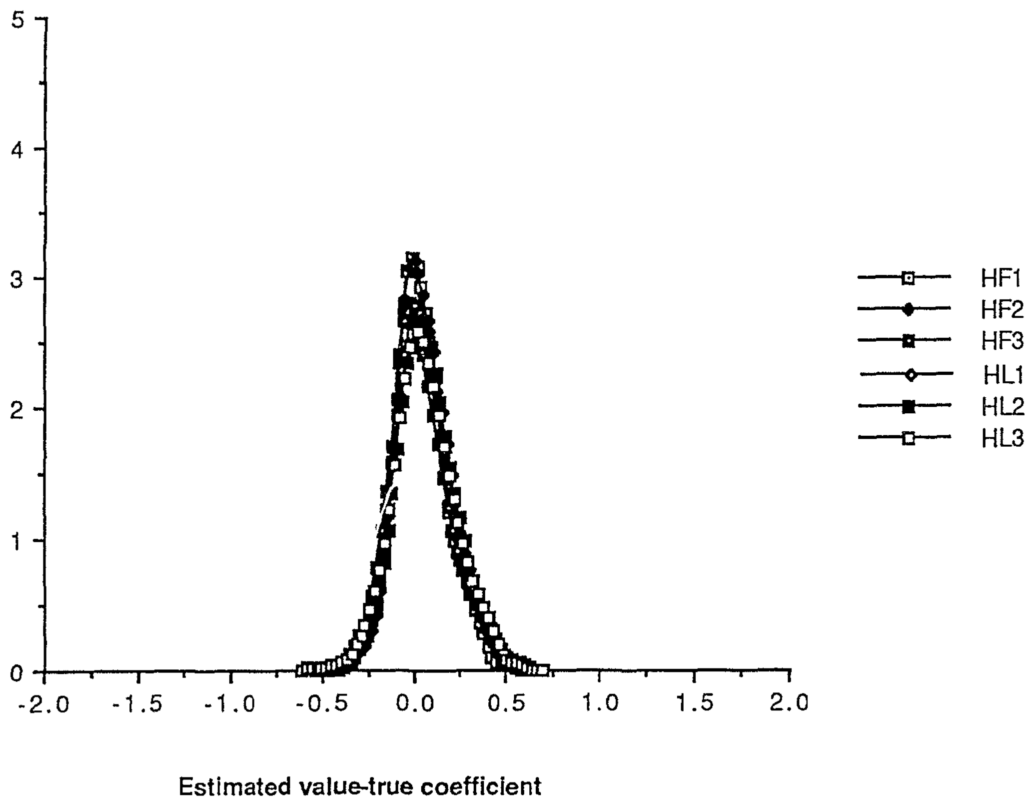
Prediction

1. Both positive and negative biases were observed. The biases and the MSEP's generally decreased as the sample size was increased. In addition, the MSEP increased as the forecast period was extended.

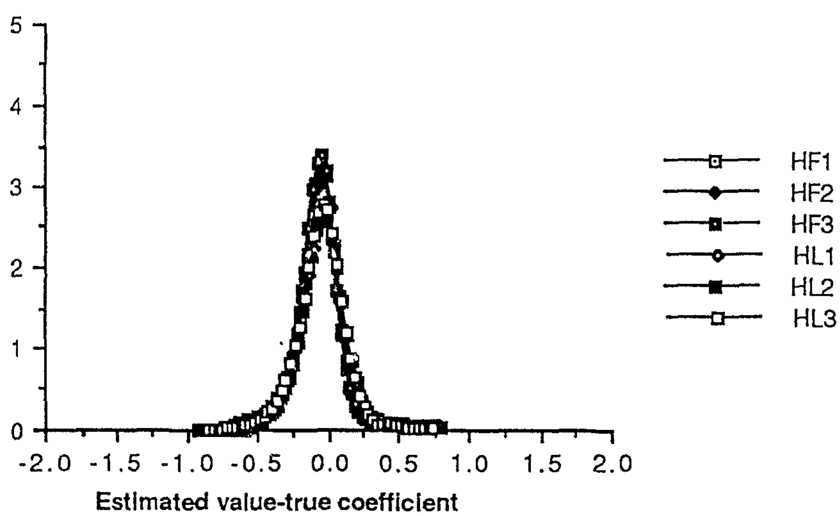
2. Using the MSEP criterion we found that the full information estimators HF1 and HF3 performed as well as their limited information analogues. The ranking remained basically the same even if we considered each endogenous variable separately. However, the ranking became weaker as the sample size was increased as evidenced by a low value of Kendall's W. In particular the ranking became extremely weak for model 4 in which the autocorrelation coefficient in the second equation was zero. These were the only cases in which the relative performance of HF2 showed dramatic improvement. Considering the limited information estimators in isolation (i.e. HL1, HL2 and HL3), our results corroborated the findings of Moazzami and Buse (1986) who found that HL2 performed worse than HL1 and HL3. It must be mentioned, however, that Moazzami and Buse ranked the estimators on the basis of structural estimation. Thus there is some reason to believe that the ranking of the two step estimators is the same in both structural estimation and prediction. However, the

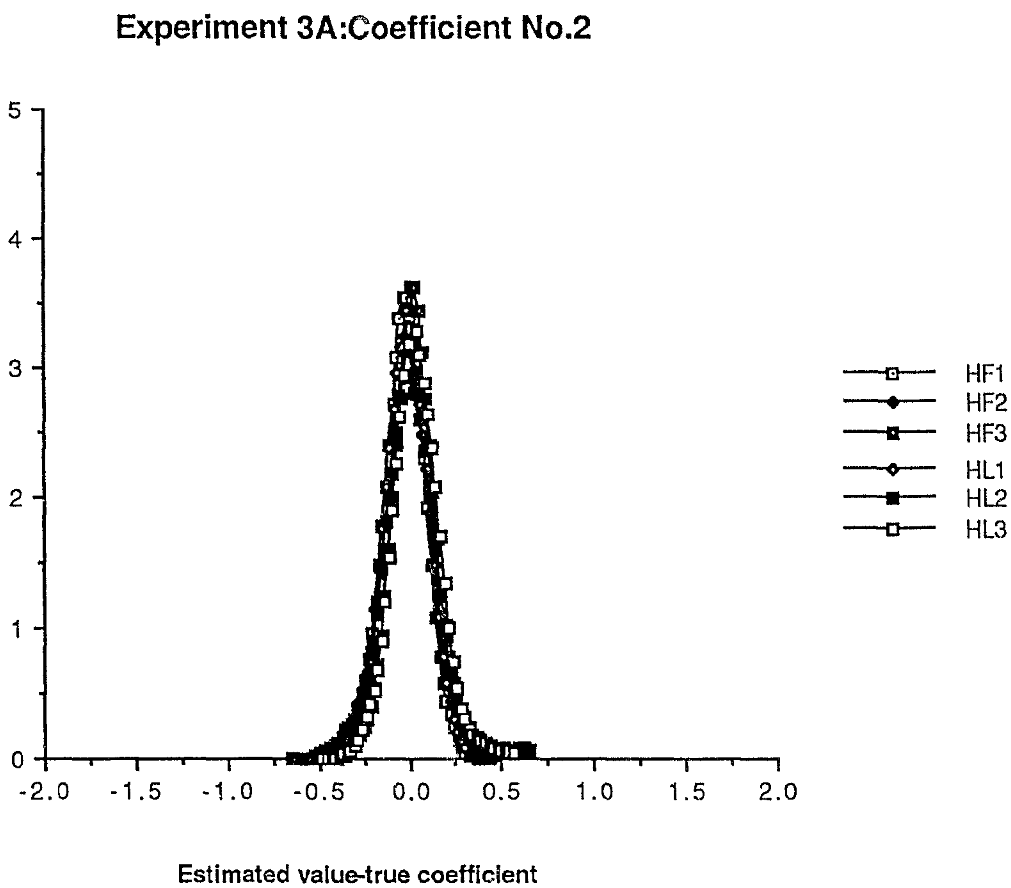
ranking of the estimators according to MPB alone was mixed. Since the MSEP incorporates both the bias and the variance, it seems quite reasonable to rely on the MSEP results. Noting that HF1, HF3 and their limited information counterparts HL1 and HL3 performed superior to HF2 and HL2 and that HF1, HF3, HL1 and HL3 used the unrestricted reduced form predictions in the second stage (whereas HF2 and HL2 used restricted reduced form predictions in the same stage) provide some support for the use of the unrestricted reduced form predictions rather than the restricted reduced form predictions in the second stage. 3. Regarding the reliability of the asymptotic covariance matrix of the dynamic simulation forecasts, the limited information predictors tended to perform, in general, better than the full information predictors as evidenced by the number of type I errors for y1 and y2. This result is especially true in cases where the autocorrelation coefficients in both equations were large. This result (for model 1) is not surprising as the computation of the asymptotic covariance matrices for the full information predictors involve a greater number of matrix inversions and this increases the cumulative effect of rounding errors, and lead to an increased possibility of near singularity of the relevant matrices. Although there were too many rejections for the 1 step ahead forecasts, the results tended to stabilize as the forecasts were made further ahead into the future. Even HF2 and HL2 performed well according to this criterion. Thus, if the main purpose of structural estimation is for testing hypothesis and constructing confidence intervals for forecasts, the limited information estimators should be used as they performed relatively better.

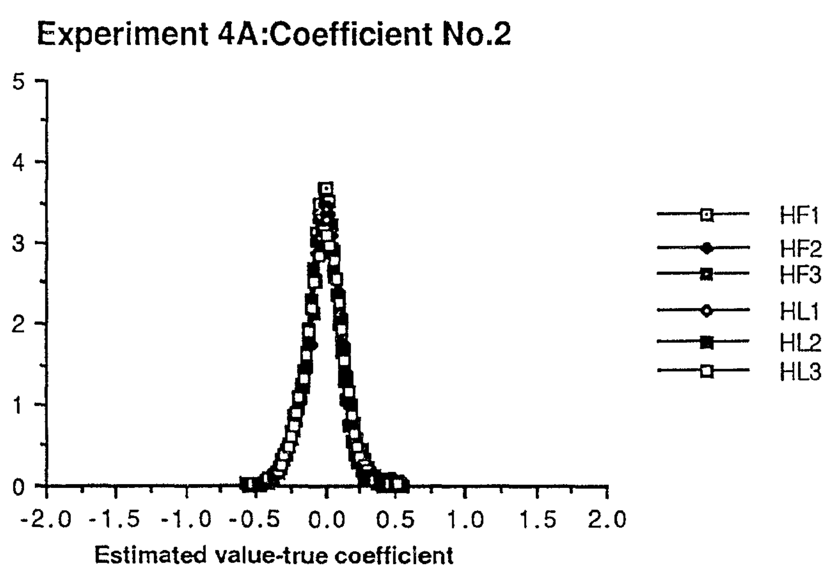
SET 1: **Experiment 1A:Coefficient No.2**

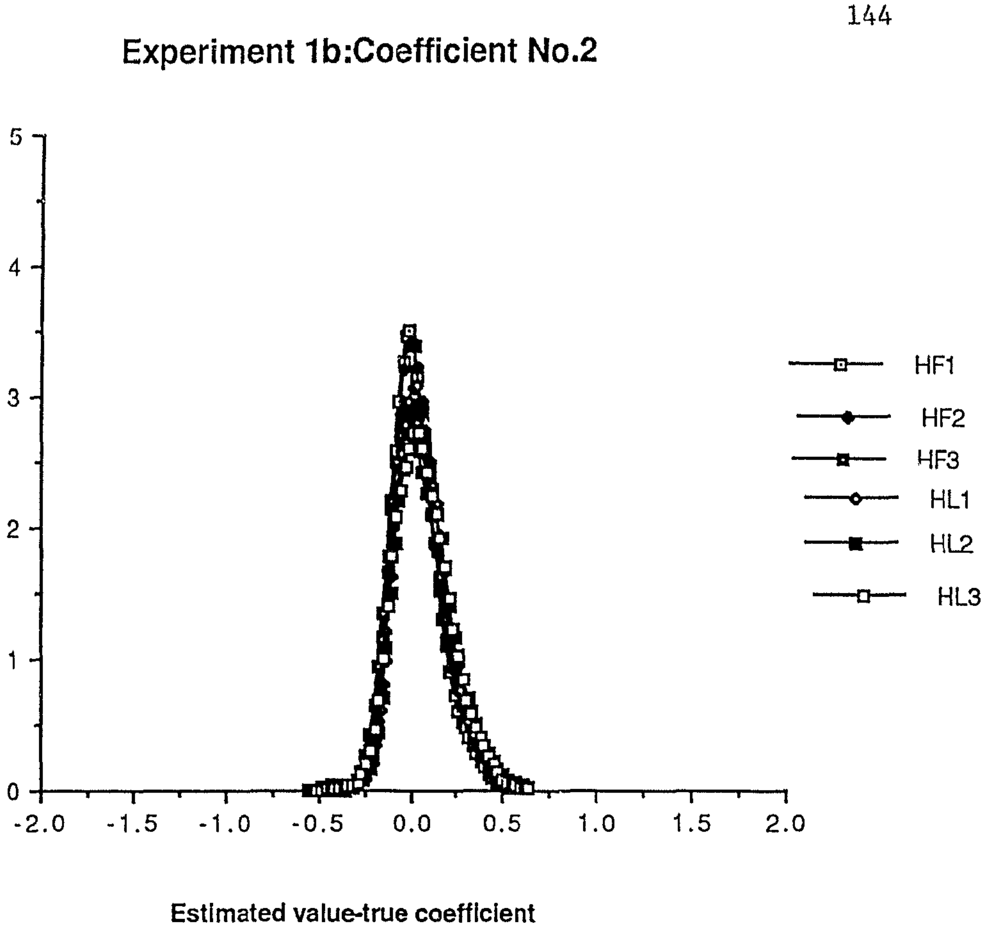


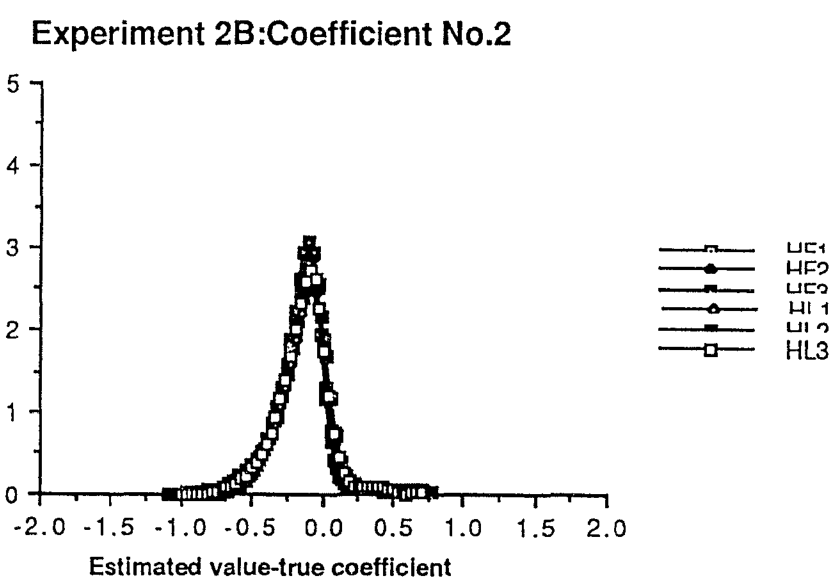
Experiment 2A:Coefficient No.2



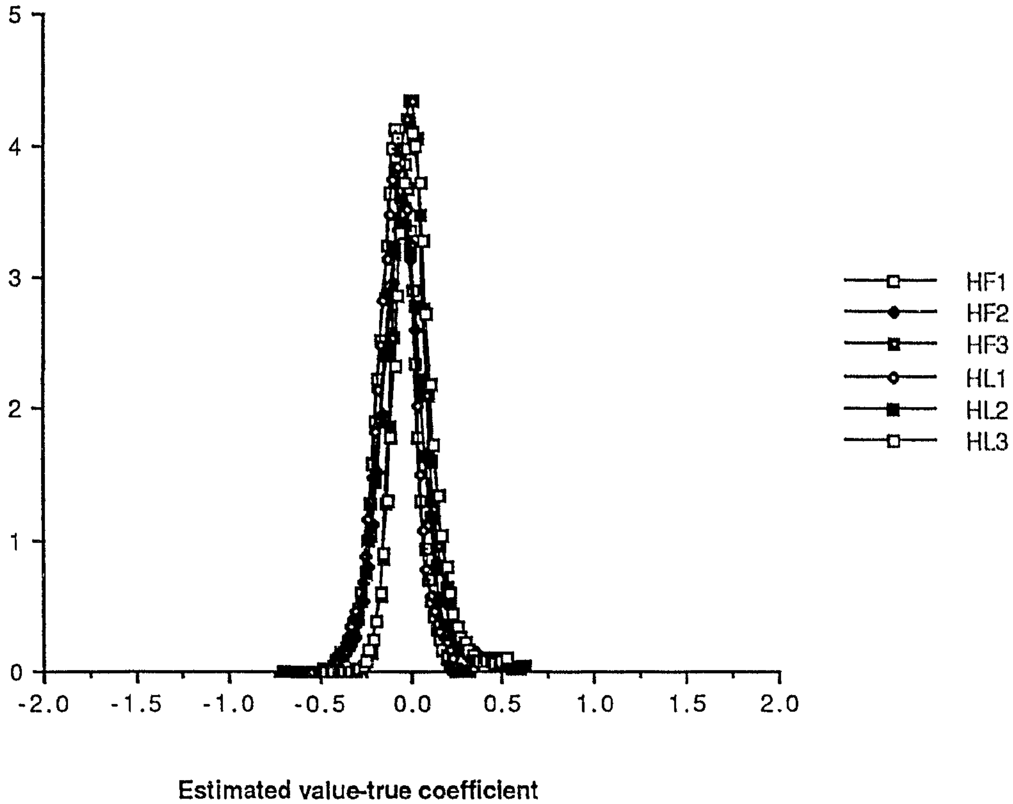

Experiment 3A:Coefficient No.2
5
4
3
2
1
0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated value-true coefficient
HF1
HF2
HF3
HL1
HL2
HL3


Experiment 4A:Coefficient No.2
5
4
3
2
1
0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated value-true coefficient
HF1
HF2
HF3
HL1
HL2
HL3

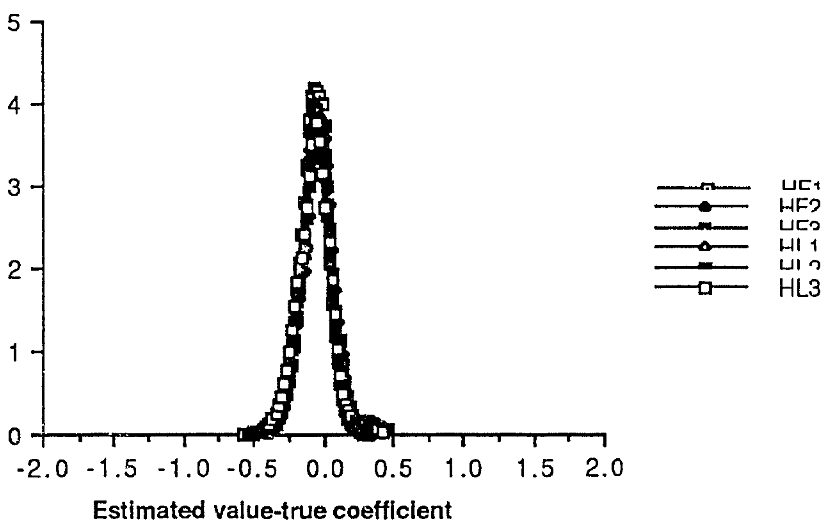

Experiment 1b:Coefficient No.2
5
4
3
2
1
0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated value-true coefficient
HF1
HF2
HF3
HL1
HL2
HL3


Experiment 2B:Coefficient No.2
5
4
3
2
1
0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated value-true coefficient
HF2
HL3

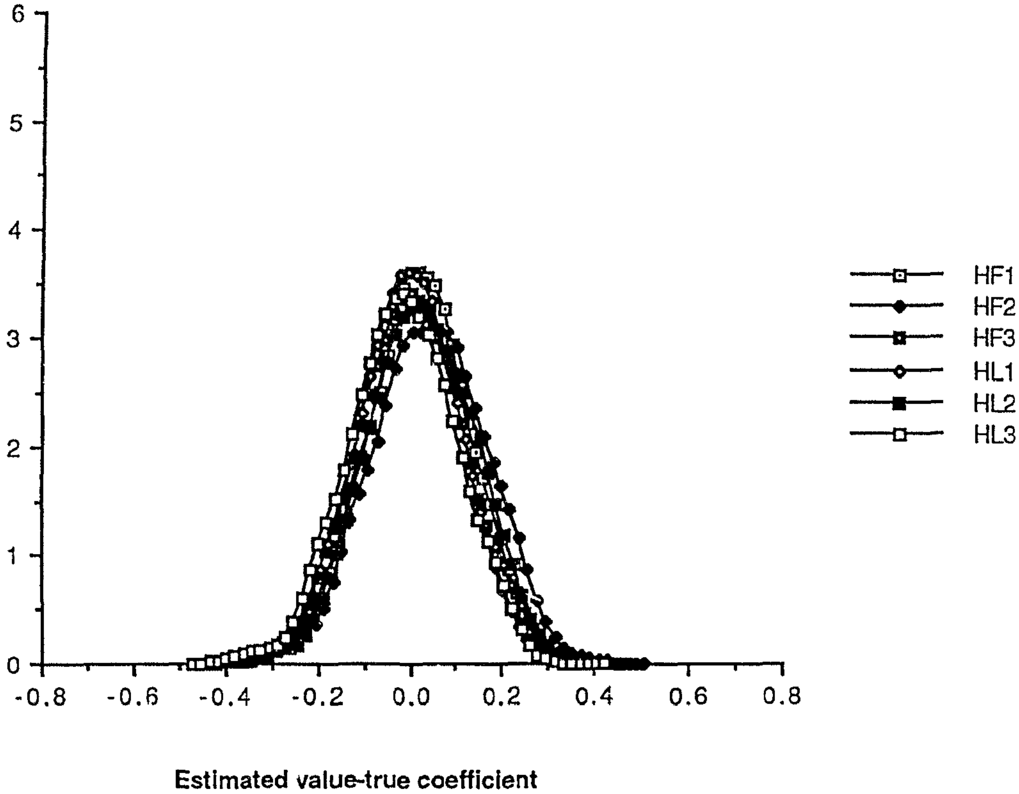
Experiment 3B:Coefficient No.2



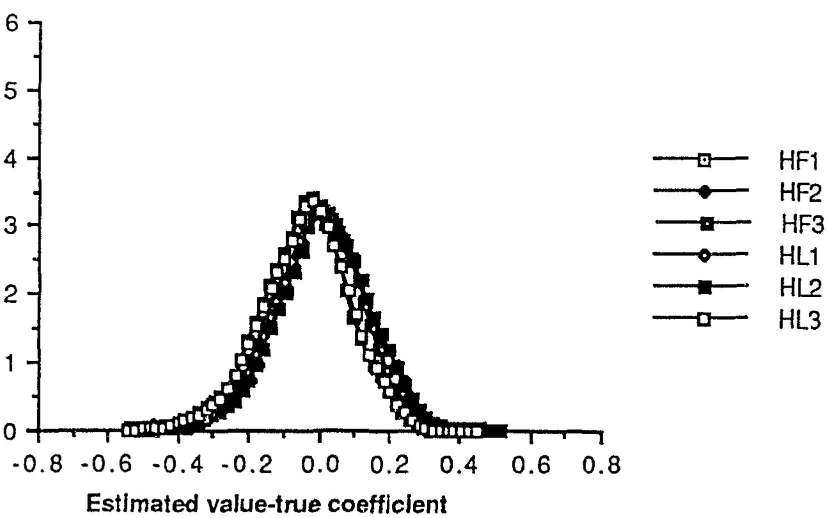
Experiment 4B:Coefficient No.2



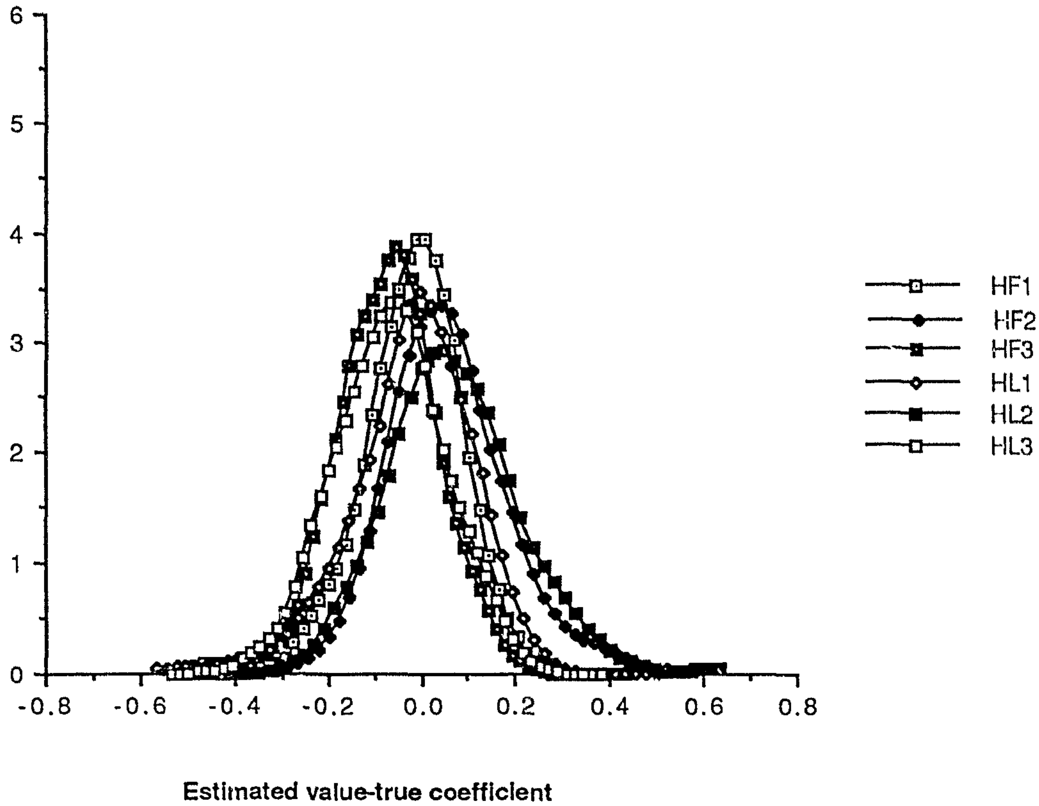
SET 2: **Experiment 1A:Coefficient No.4**



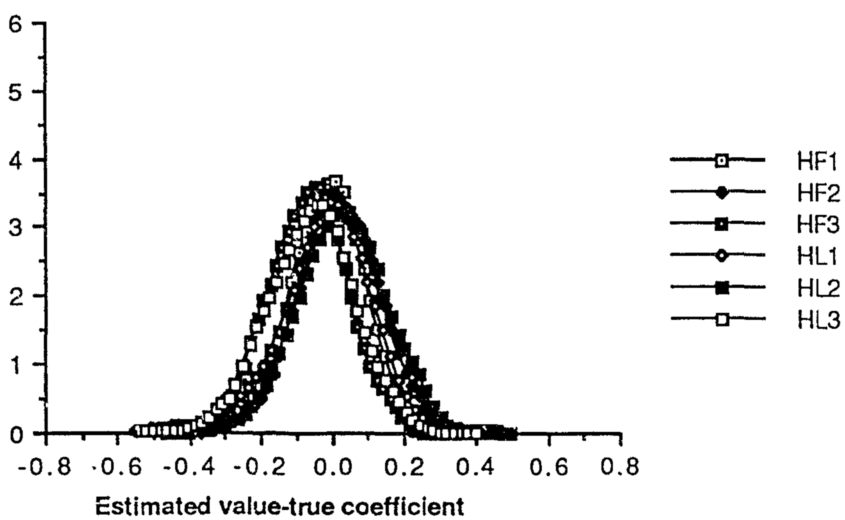
Experiment 2A:Coefficient No.4



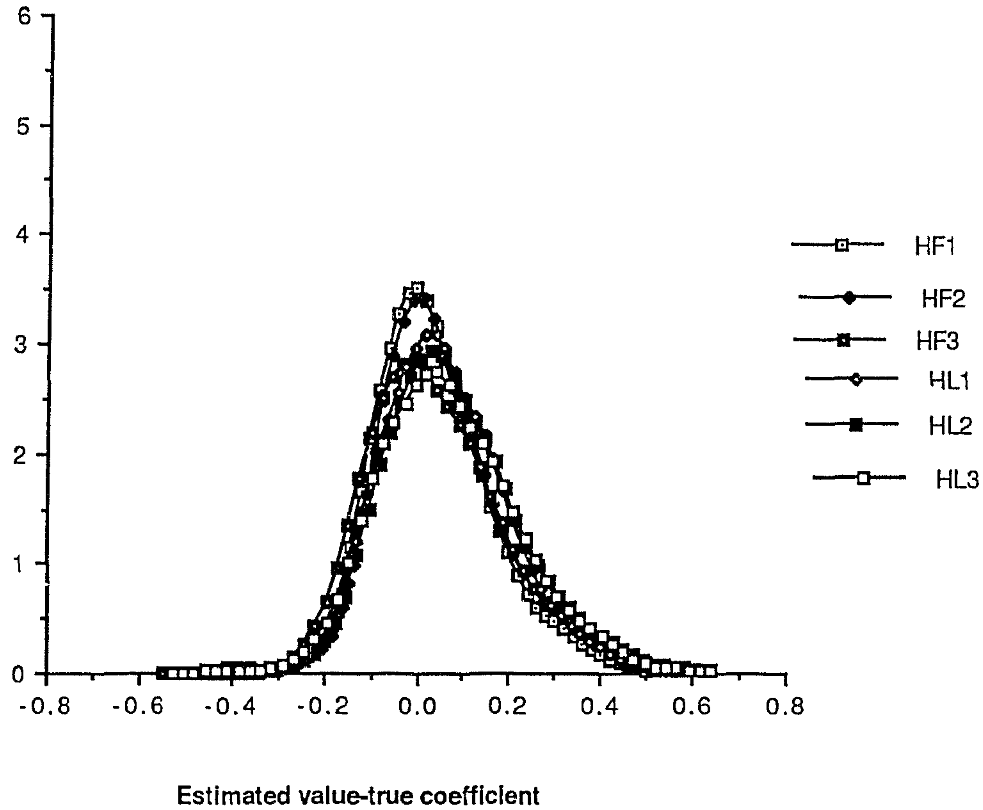
Experiment 3A:Coefficient No.4



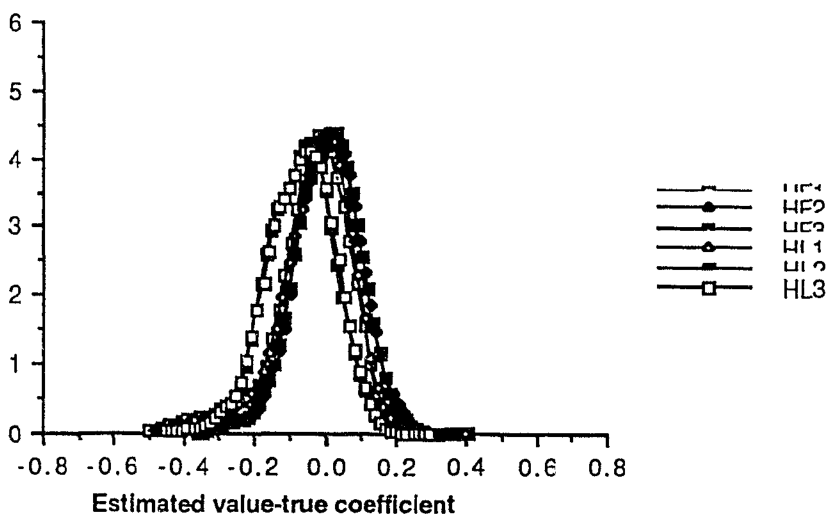
Experiment 4A:Coefficent No.4



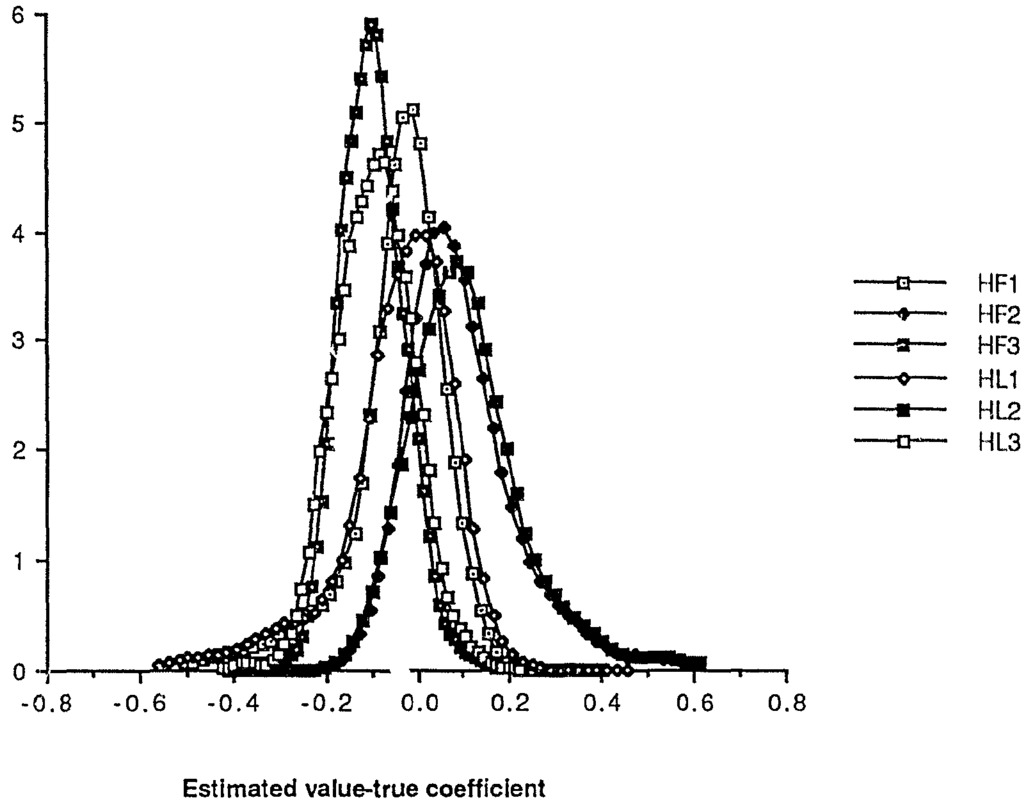
Experiment 1b:Coefficient No. 4



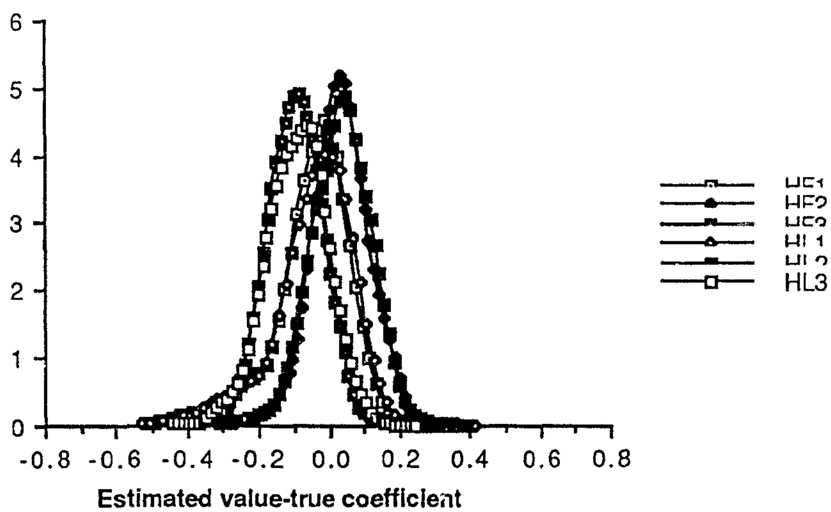
Experiment 2B:Coefficient No.4



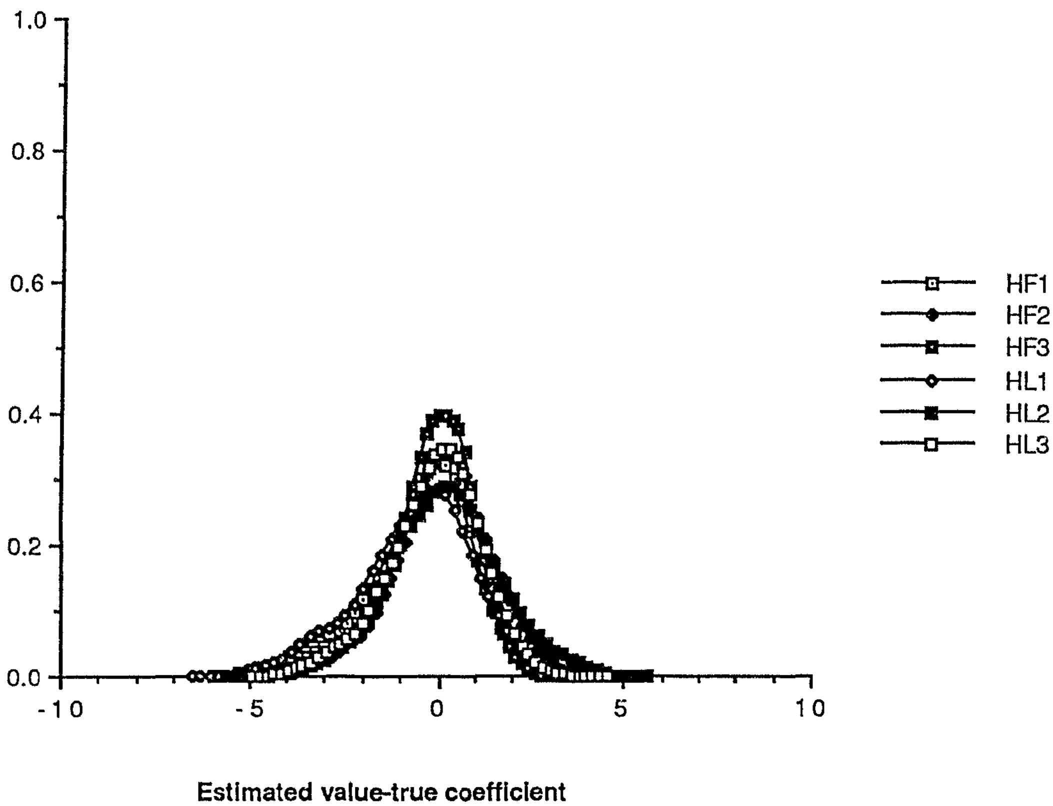
Experiment 3B:Coefficient No.4



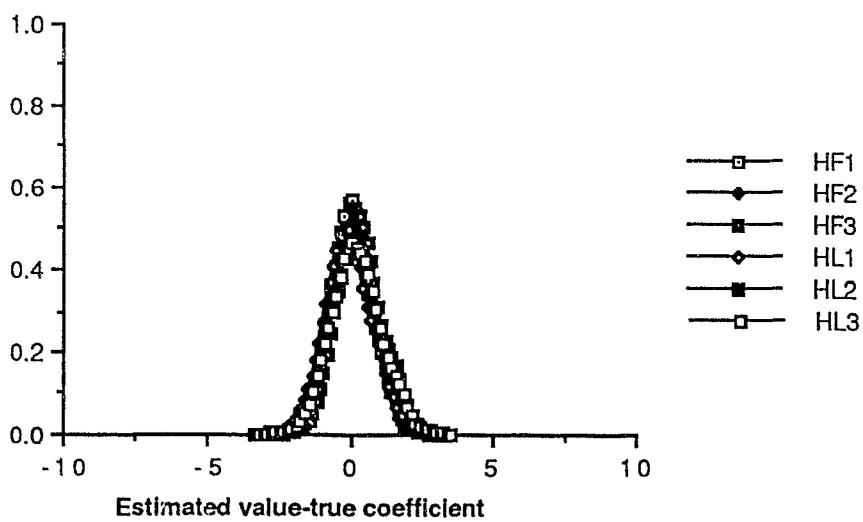
Experiment 4B:Coefficient No.4



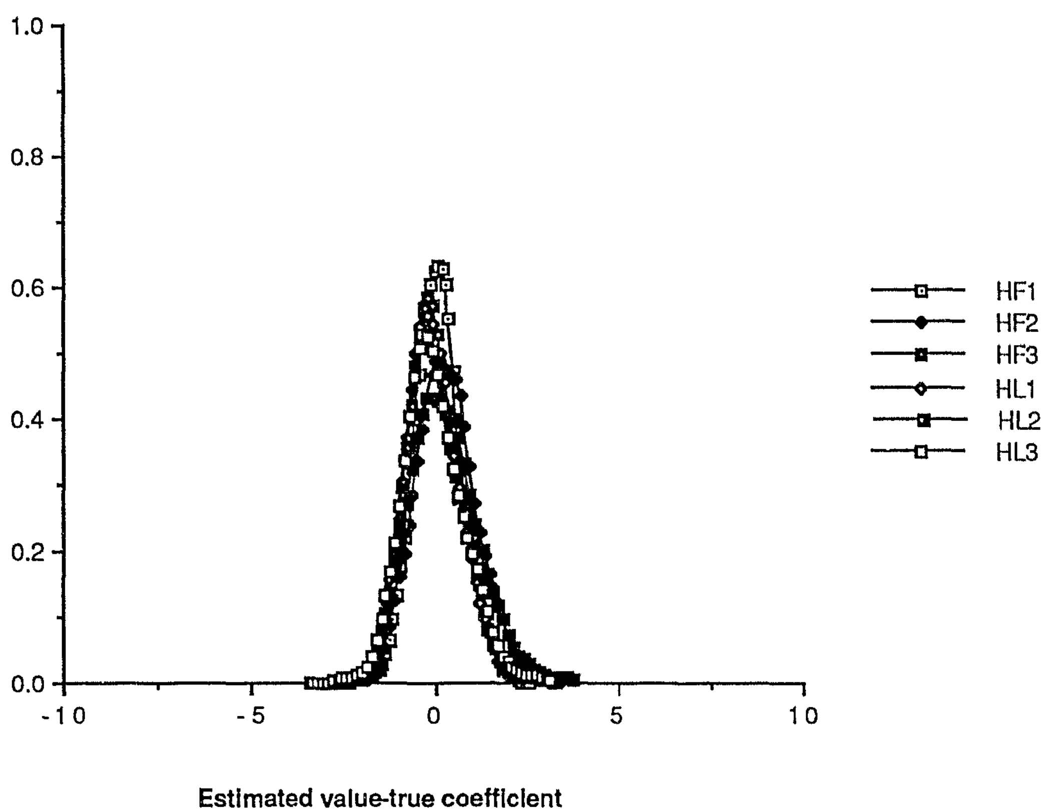
SET 3: **Experiment 1A:Coefficient No.10**



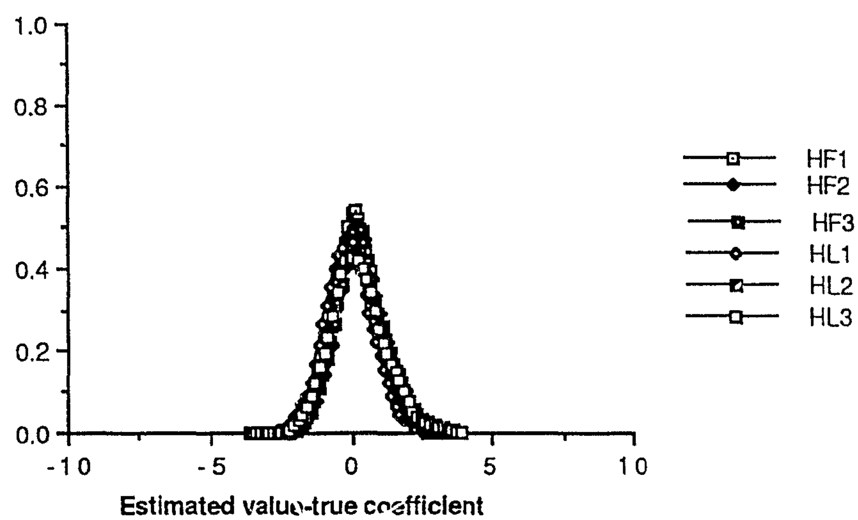
Experiment 2A:Coefficient No.10



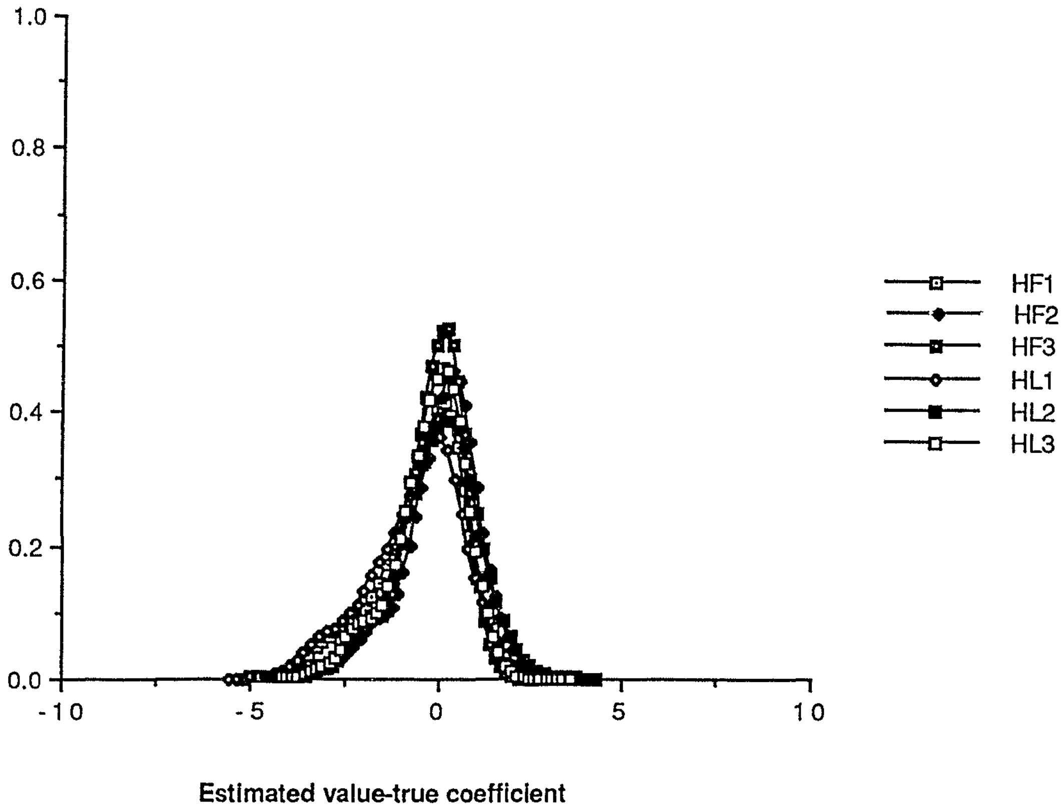
Experiment 3A:Coefficient No.10



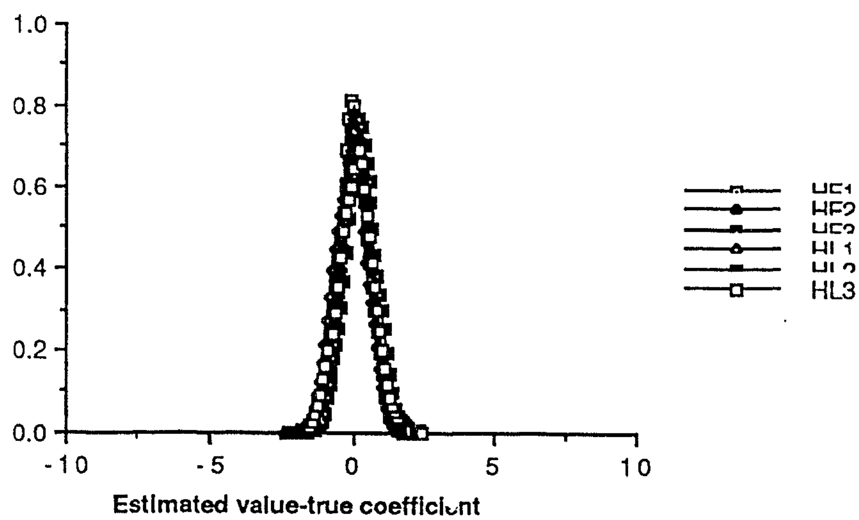
Experiment 4A:Coefficient No.4



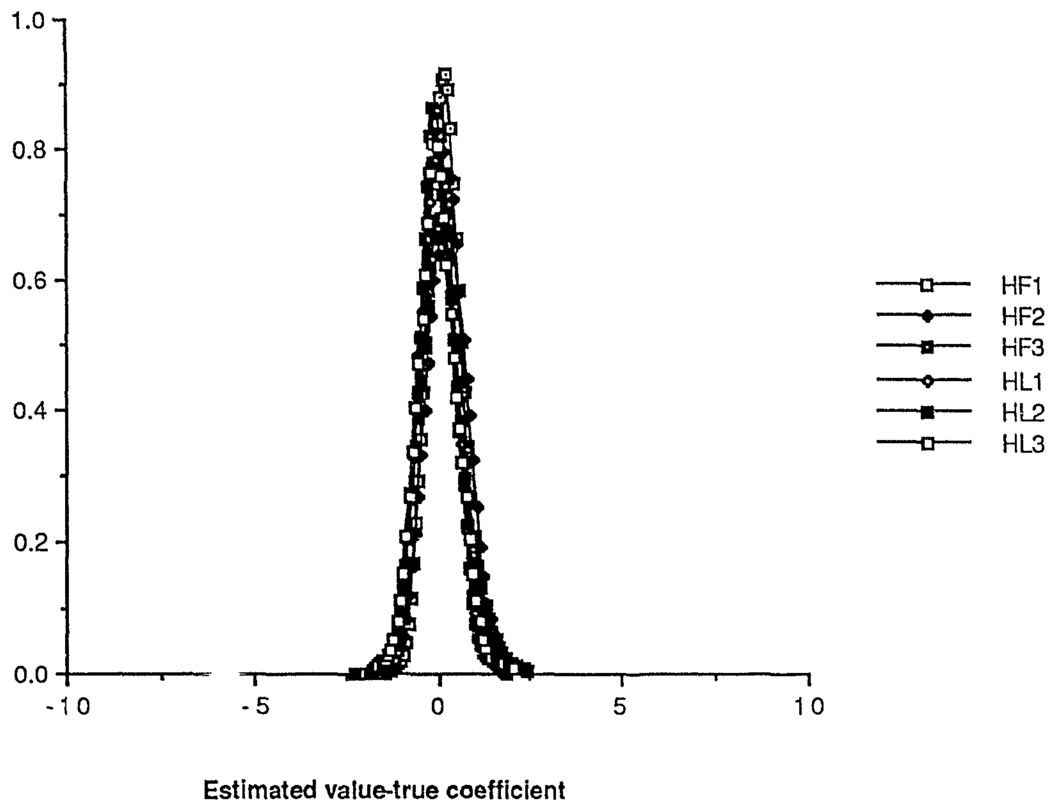
Experiment 1B:Coefficient No.10



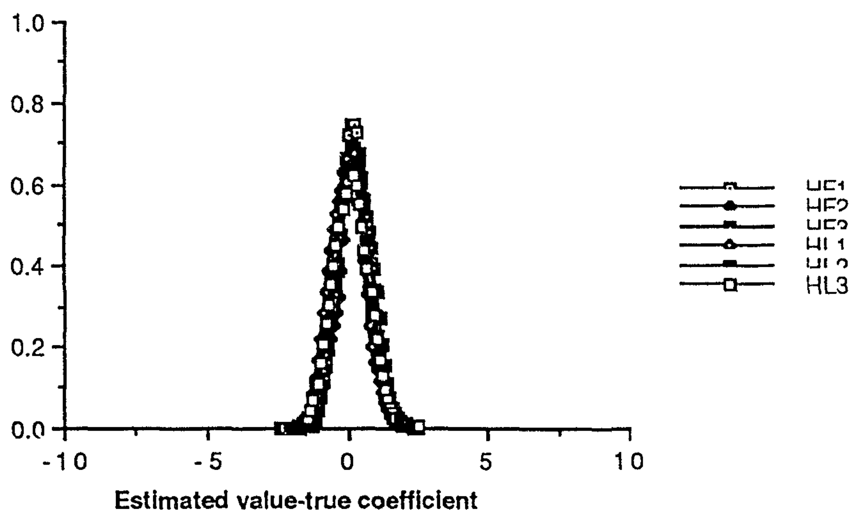
Experiment 2B:Coefficient No.10



Experiment 3B:Coefficient No.10

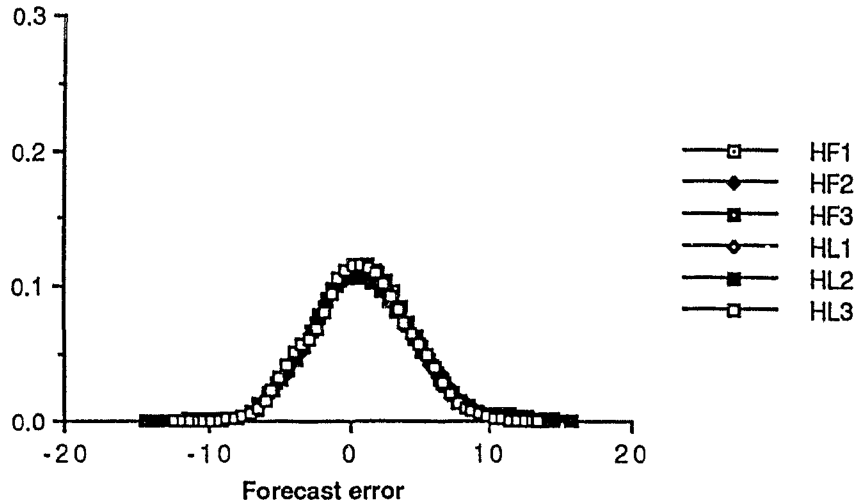


Experiment 4B:Coeff ient No.10

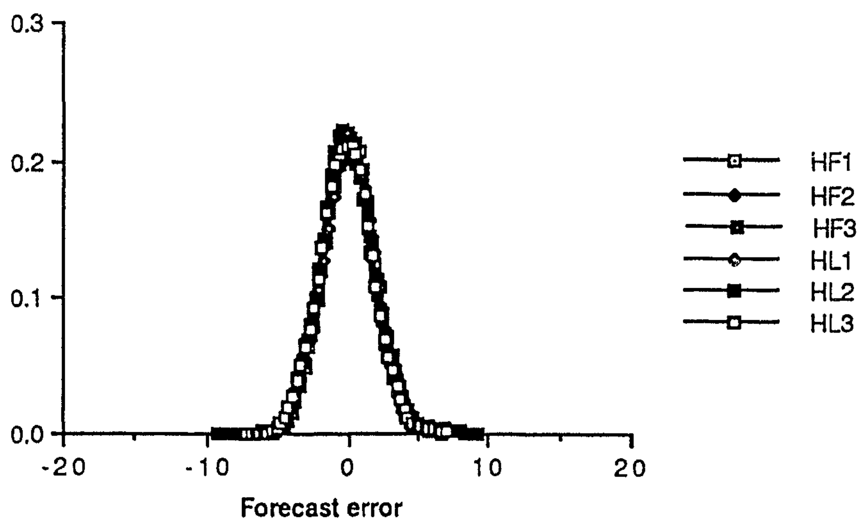


SET 4:

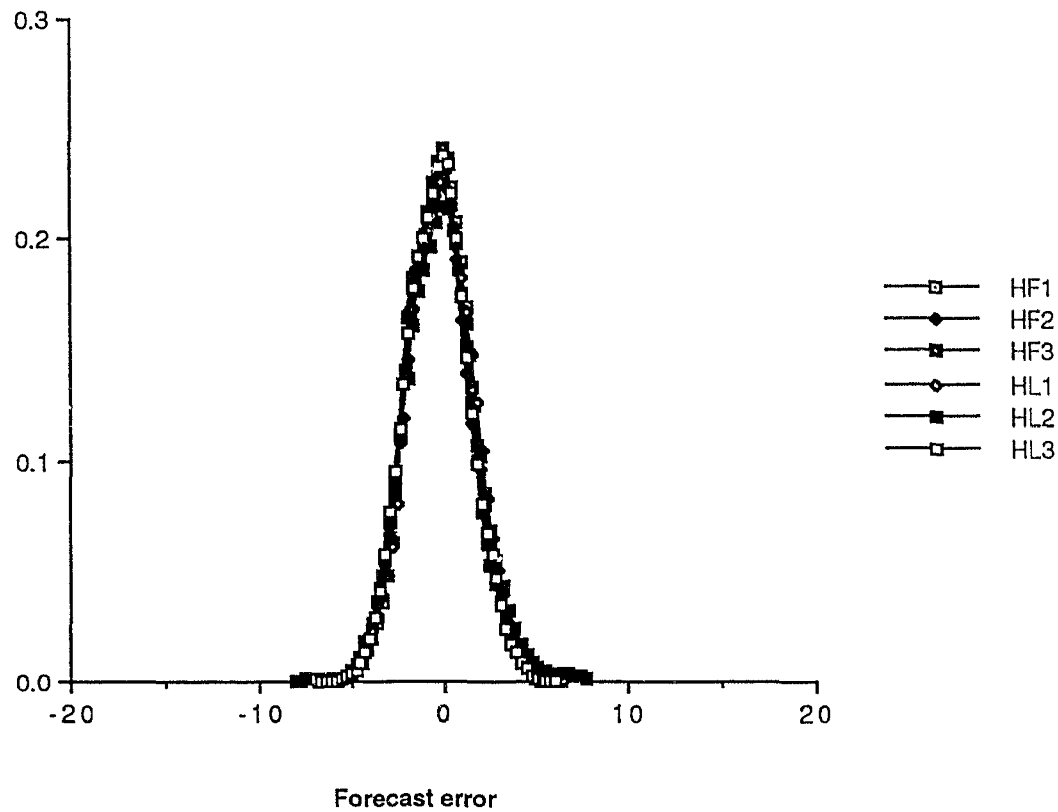
Experiment 1A:3 step ahead forecast for y1



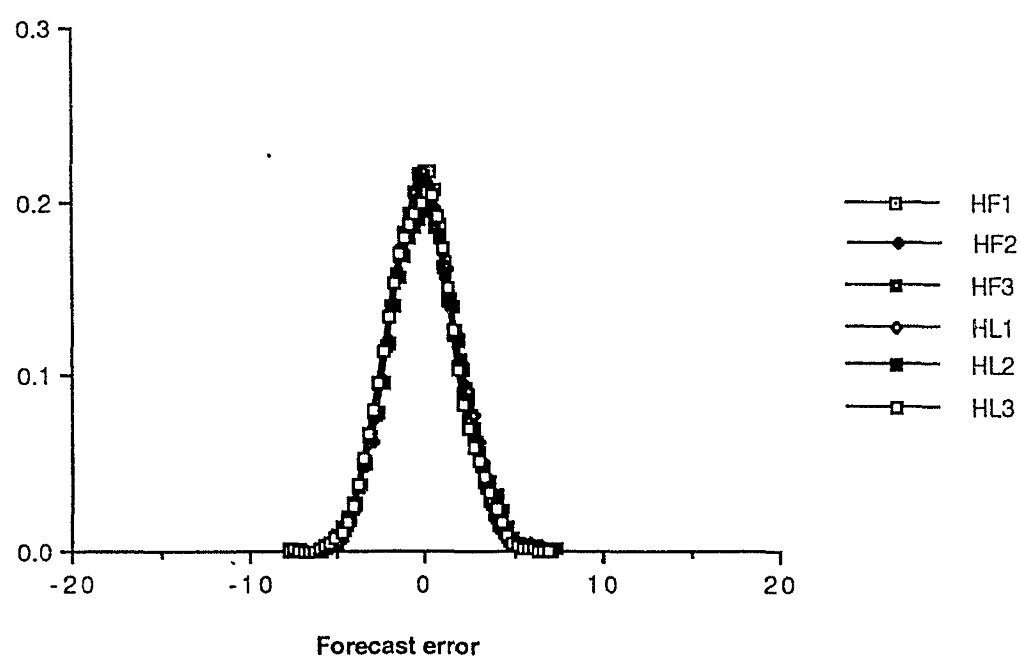
Experiment 2A:3 step ahead forecast for y1



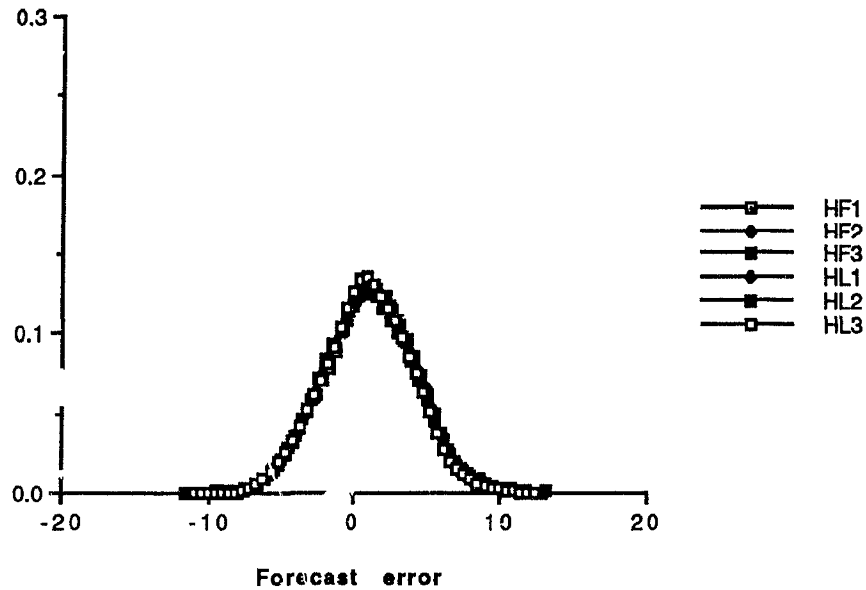
Experiment 3A:3 step ahead forecast for y1



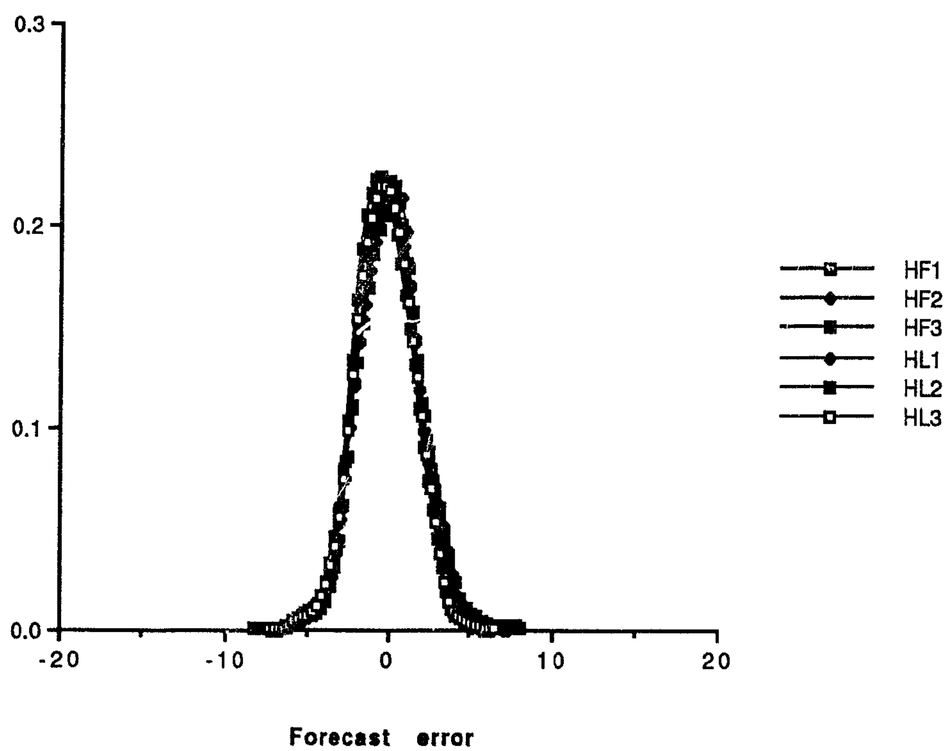
Experiment 4A:3 step ahead forecast for y1



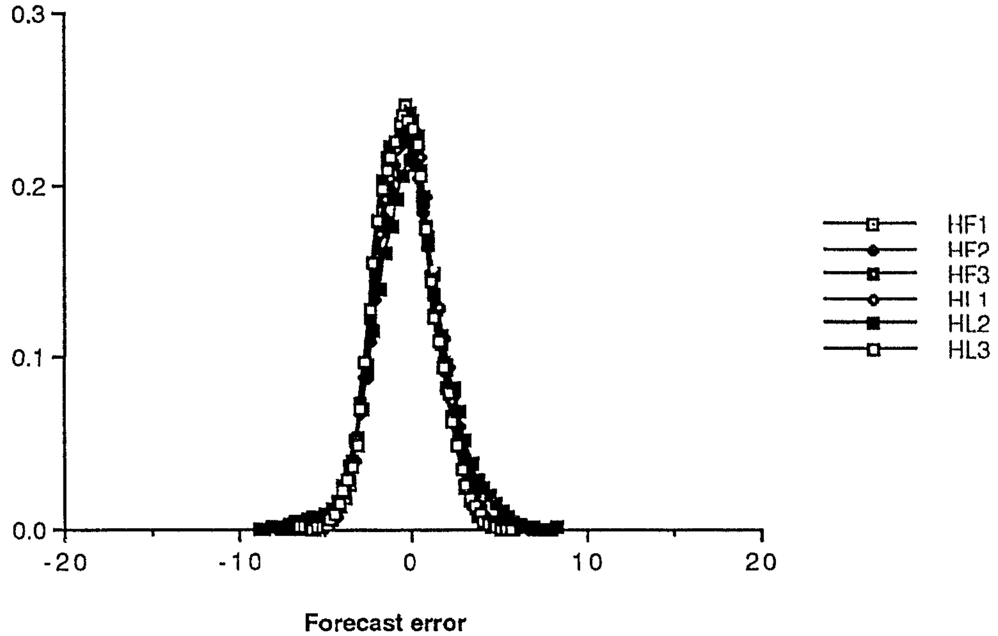
Experiment 1B:3 step ahead forecast for y1



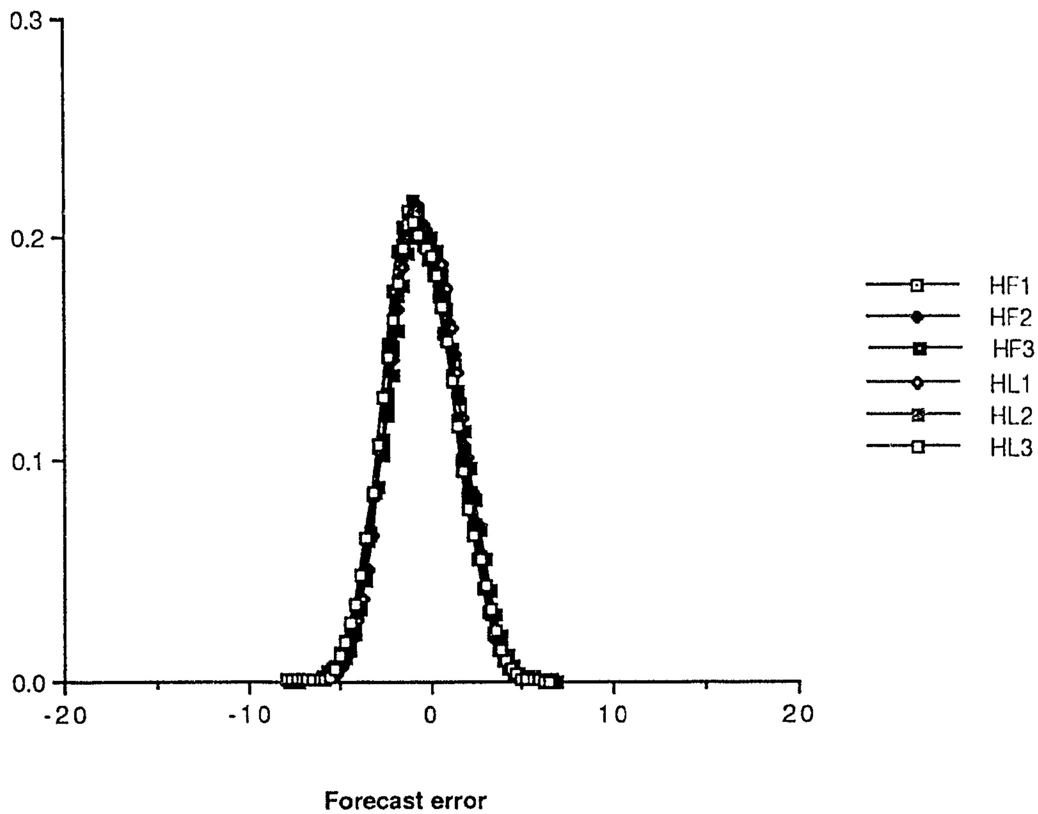
Experiment 2B:3 step ahead forecast for y1



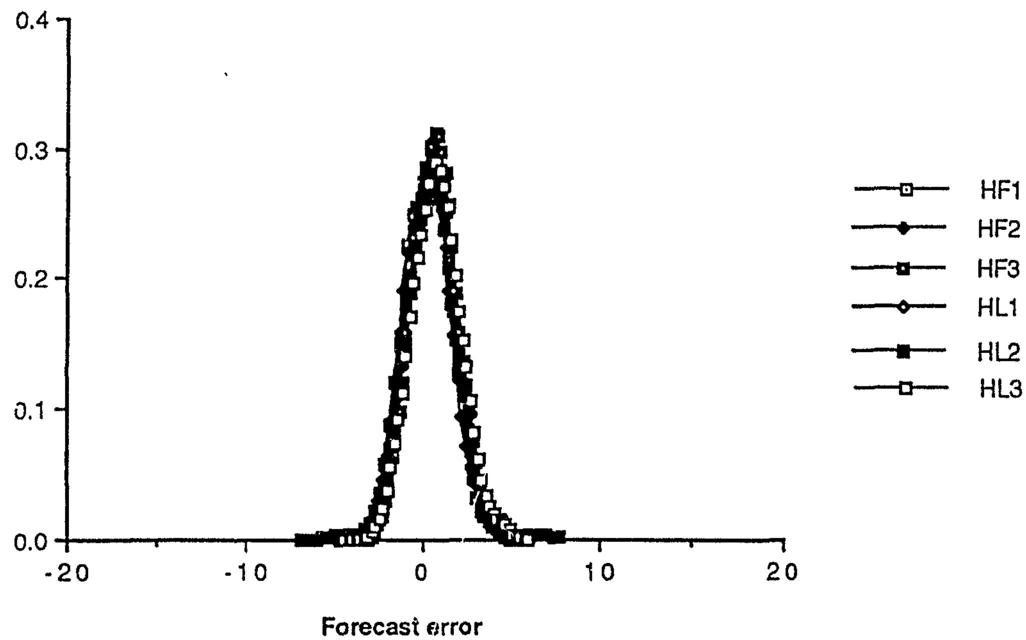
Experiment 3B:3 step ahead forecast for y1



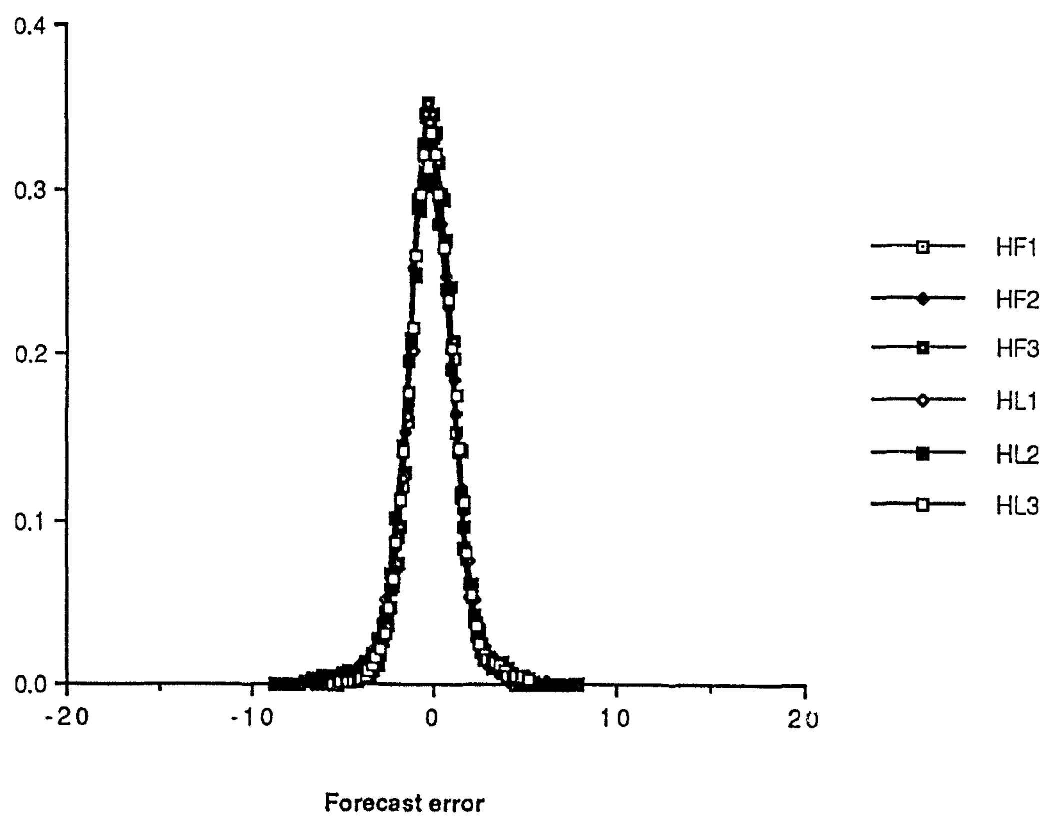
Experiment 4B:3 step ahead forecast for y1

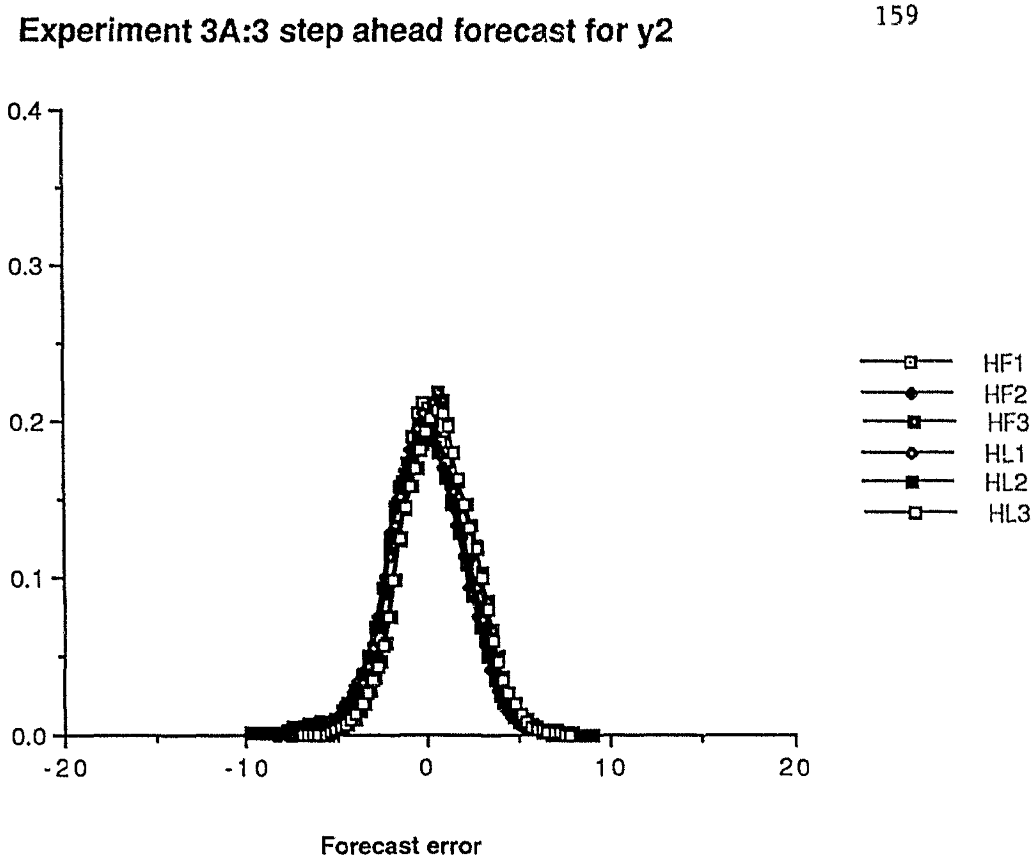


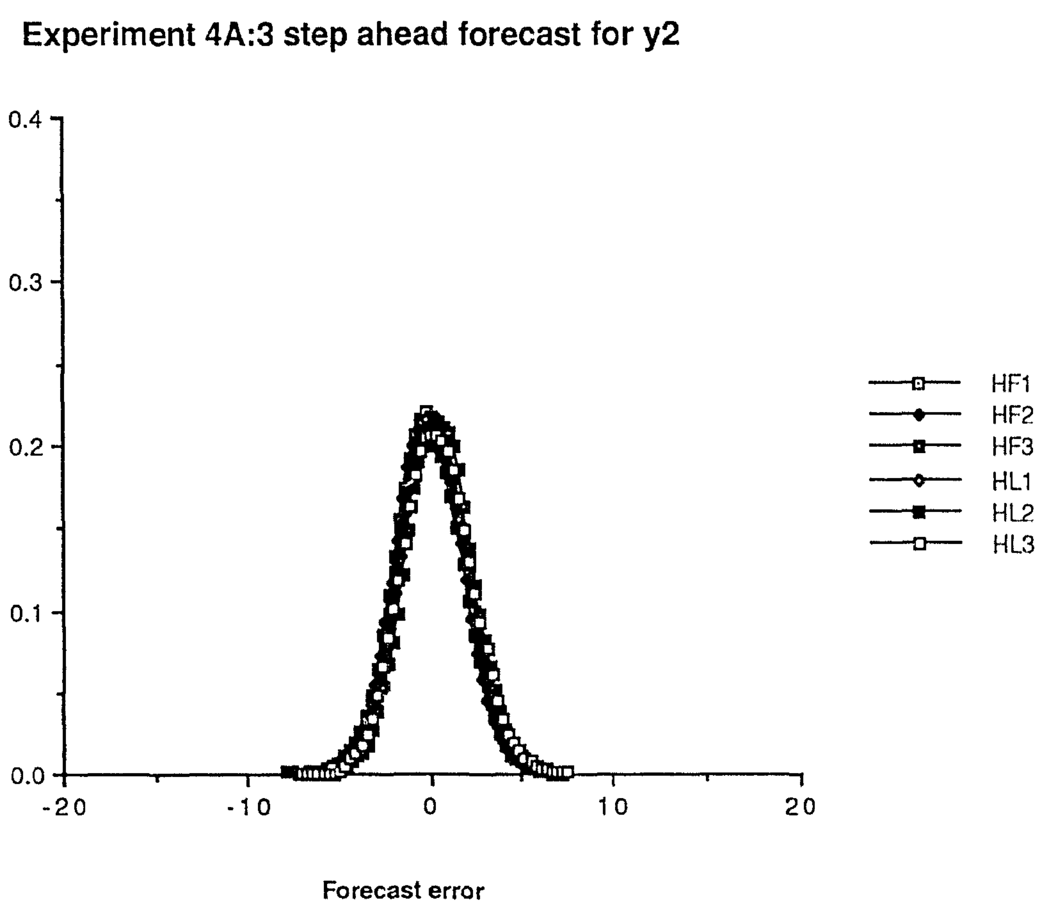
SET 5: Experiment 1A:3 step ahead forecast for y2



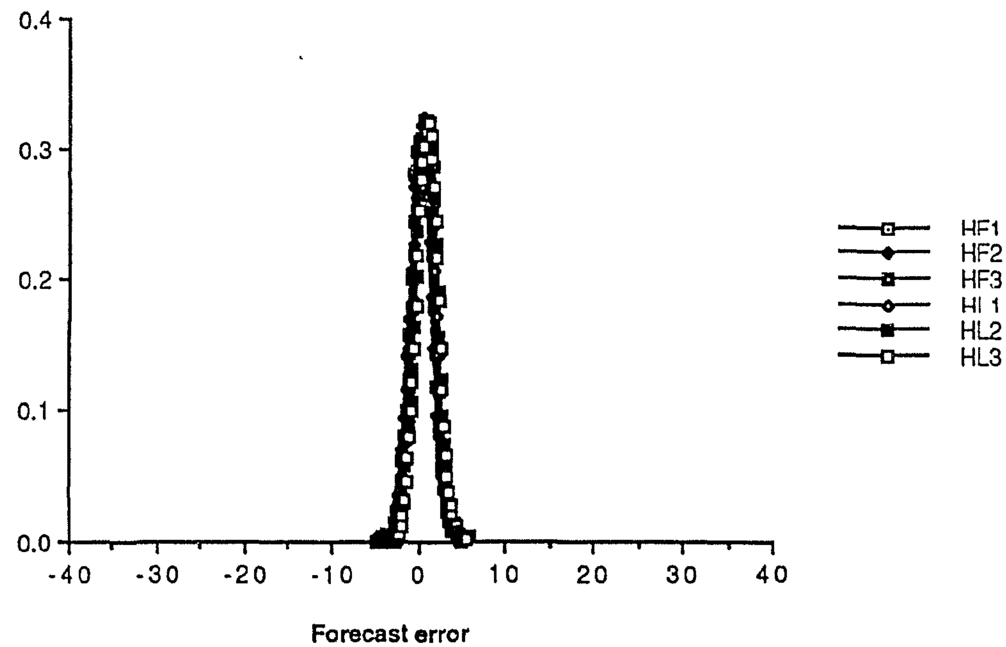
Experiment 2A:3 step ahead forecast for y2



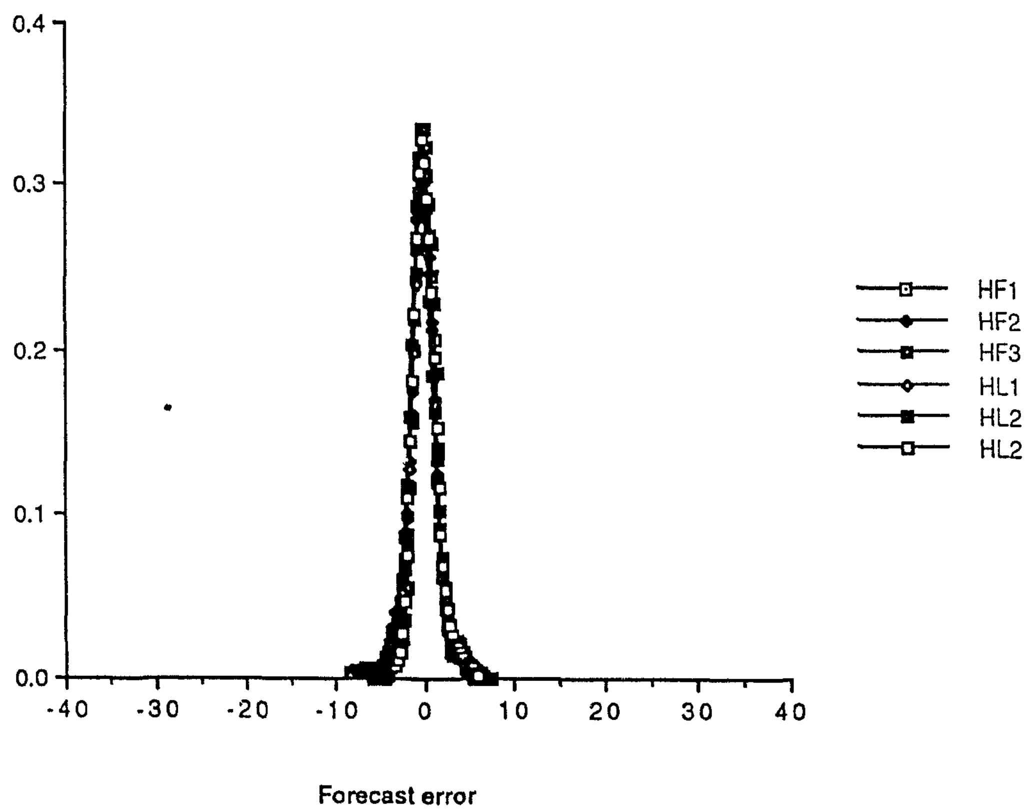

Experiment 3A:3 step ahead forecast for y2
0.4
0.3
0.2
0.1
0.0
-20
-10
0
10
20
Forecast error
HF1
HF2
HF3
HL1
HL2
HL3


Experiment 4A:3 step ahead forecast for y2
0.4
0.3
0.2
0.1
0.0
-20
-10
0
10
20
Forecast error
HF1
HF2
HF3
HL1
HL2
HL3

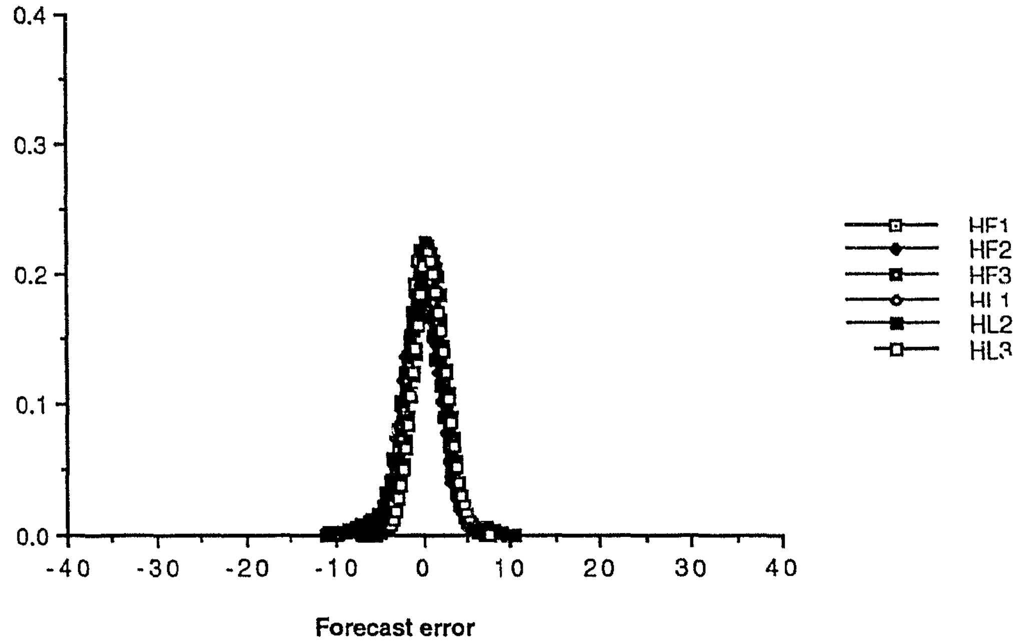
Experiment 1B:3 step ahead forecast for y2



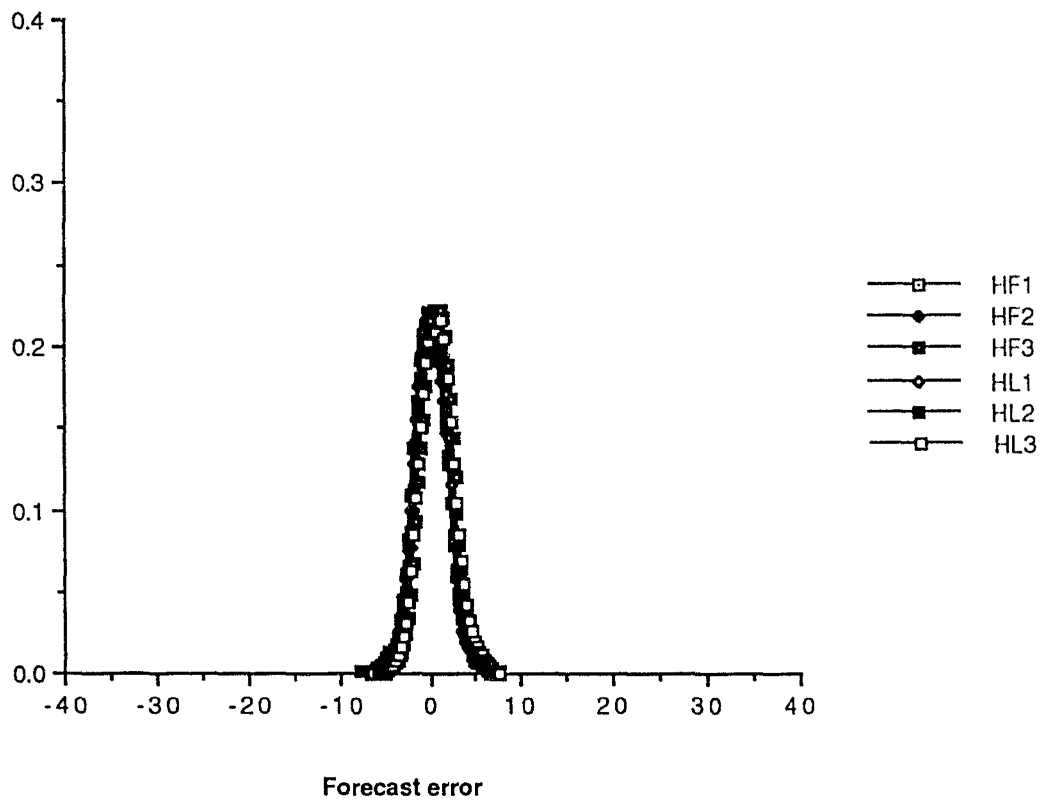
Experiment 2B:3 step ahead forecast for y2



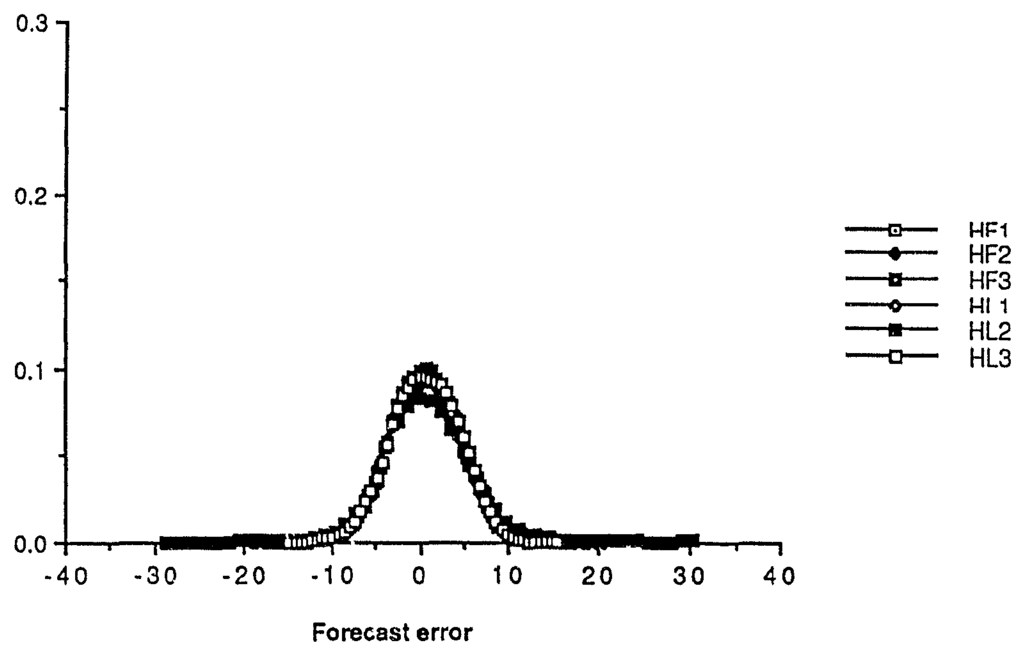
Experiment 3B:3 step ahead forecast for y2



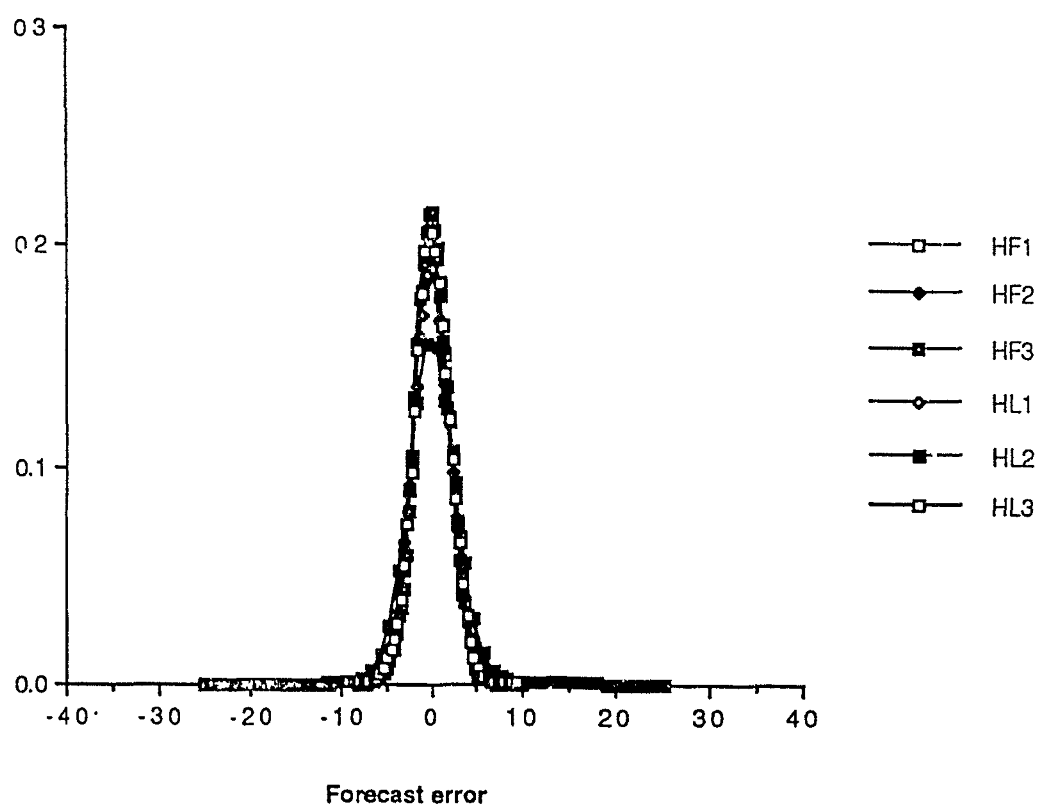
Experiment 4B:3 step ahead forecast for y2



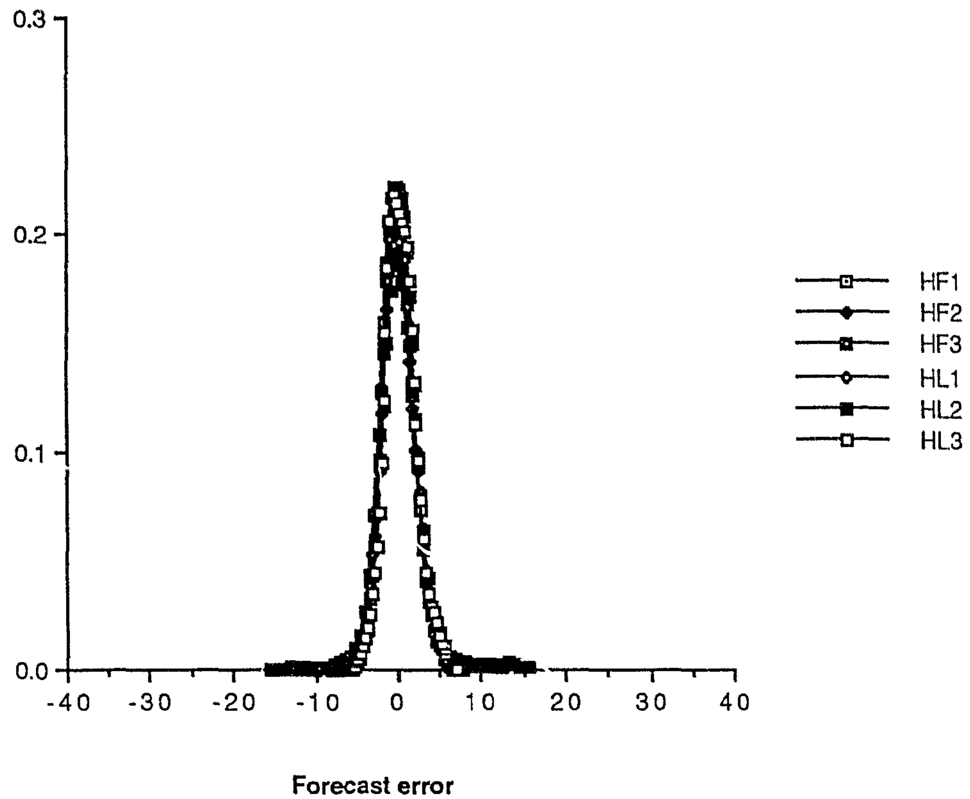
SET 6: Experiment 1A:5 step ahead forecast for y1



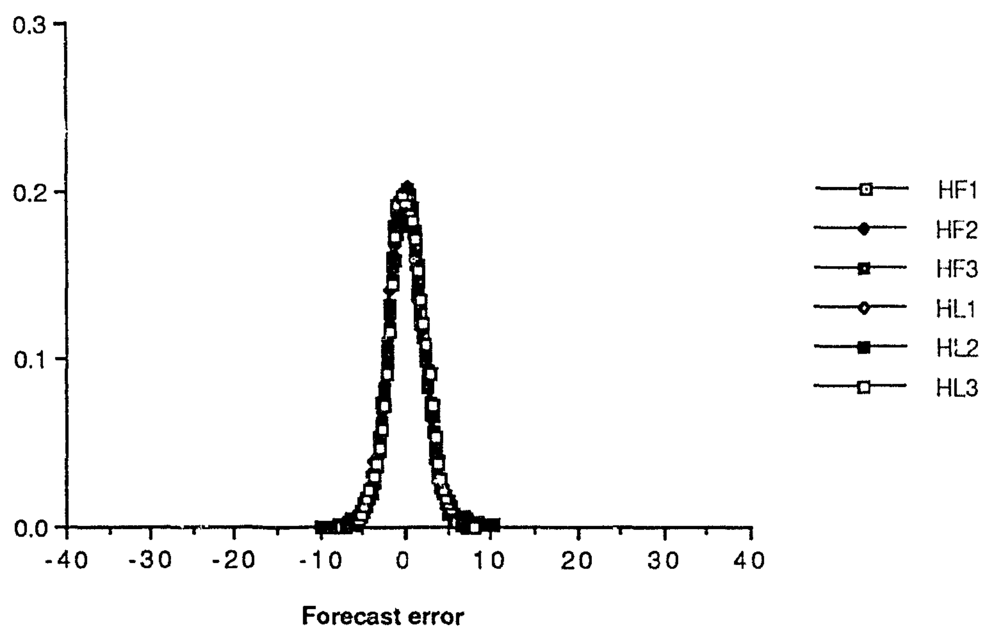
Experiment 2A:5 step ahead forecast for y1



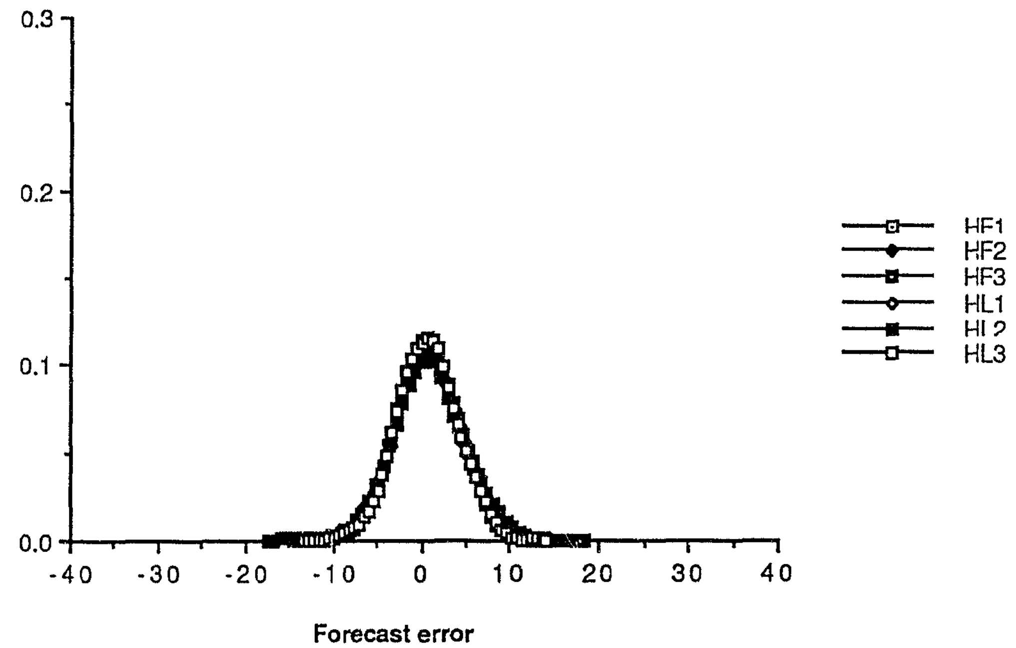
Experiment 3A:5 step ahead forecast for y1



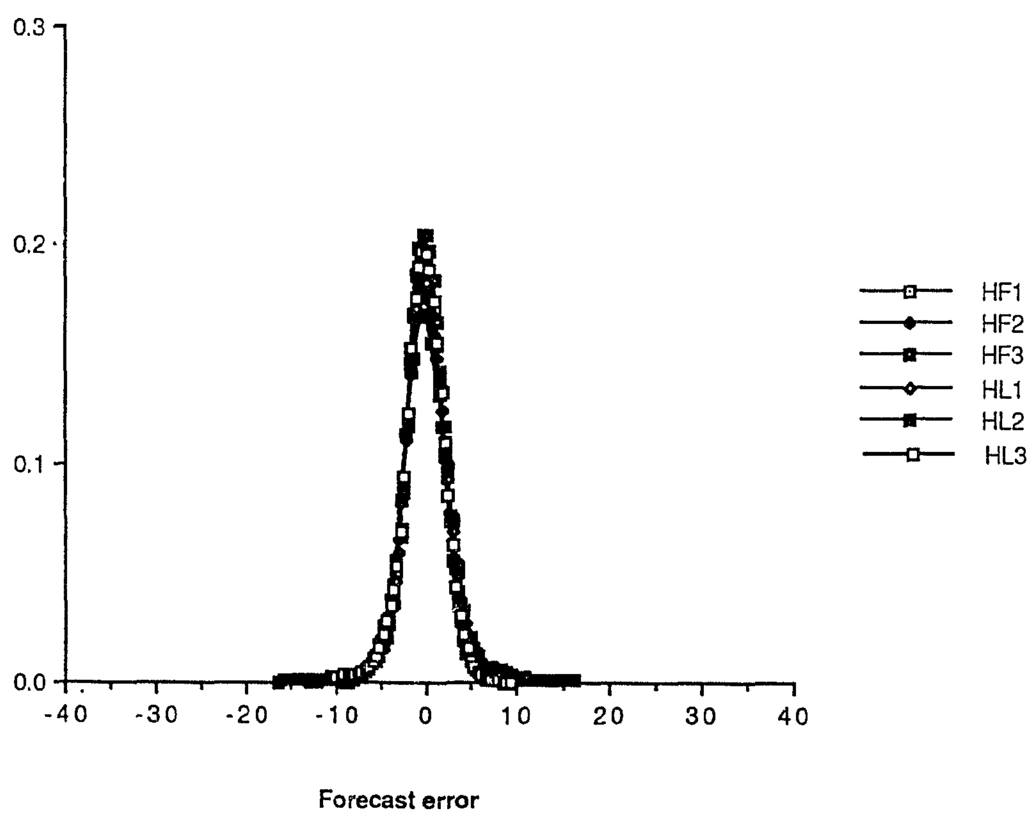
Experiment 4A:5 step ahead forecast for y1



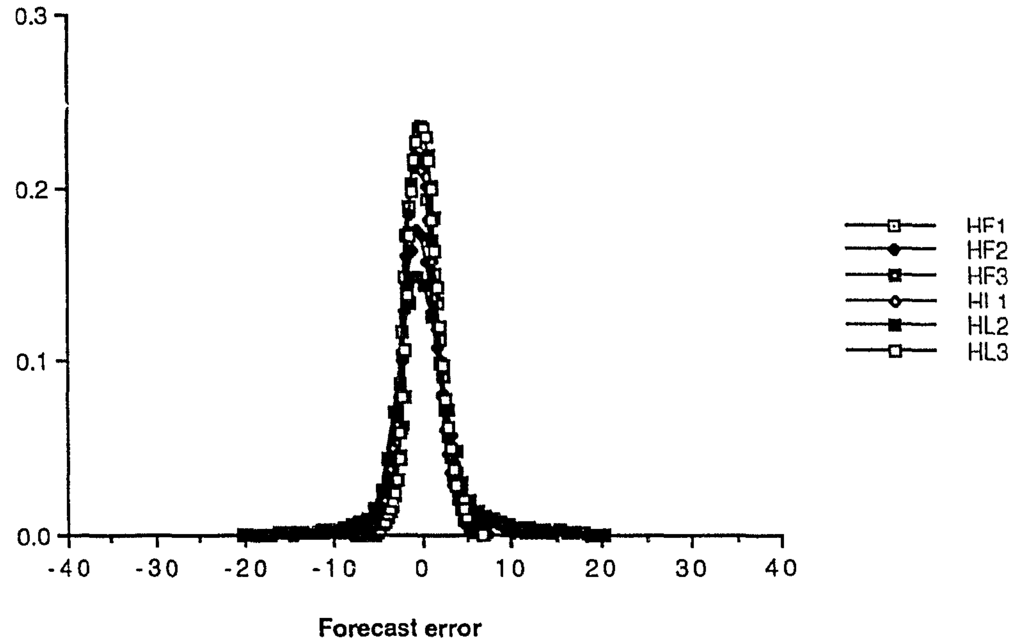
Experiment 1B:5 step ahead forecast for y1



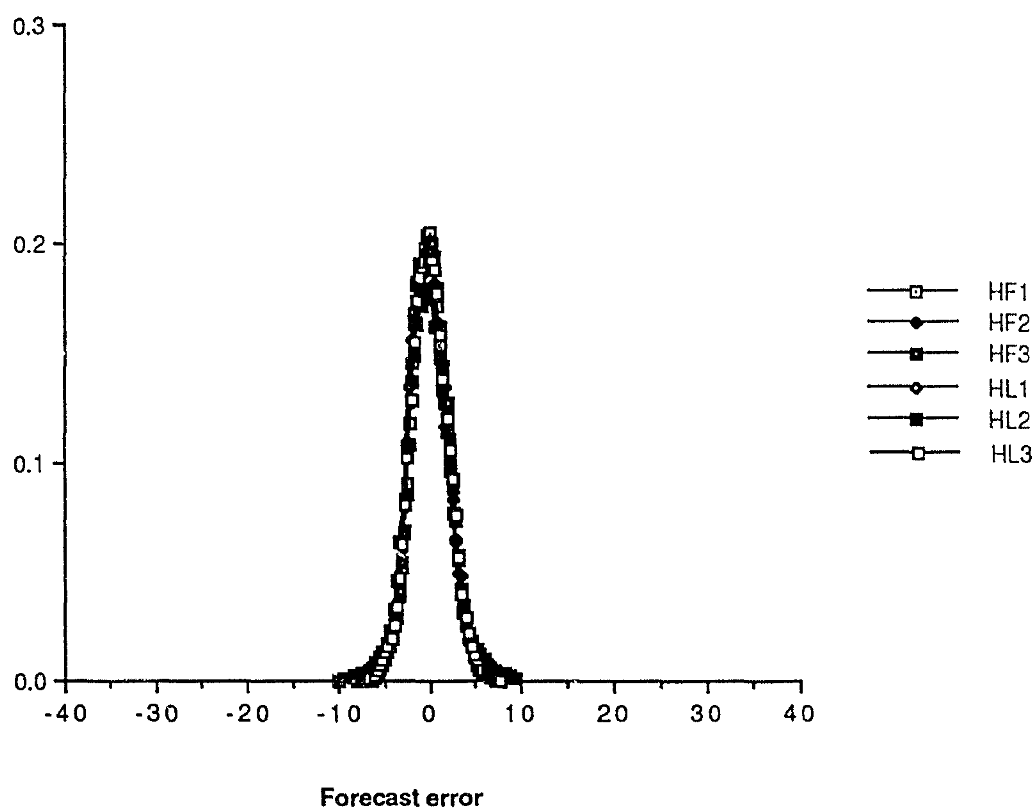
Experiment 2B:5 step ahead forecast for y1



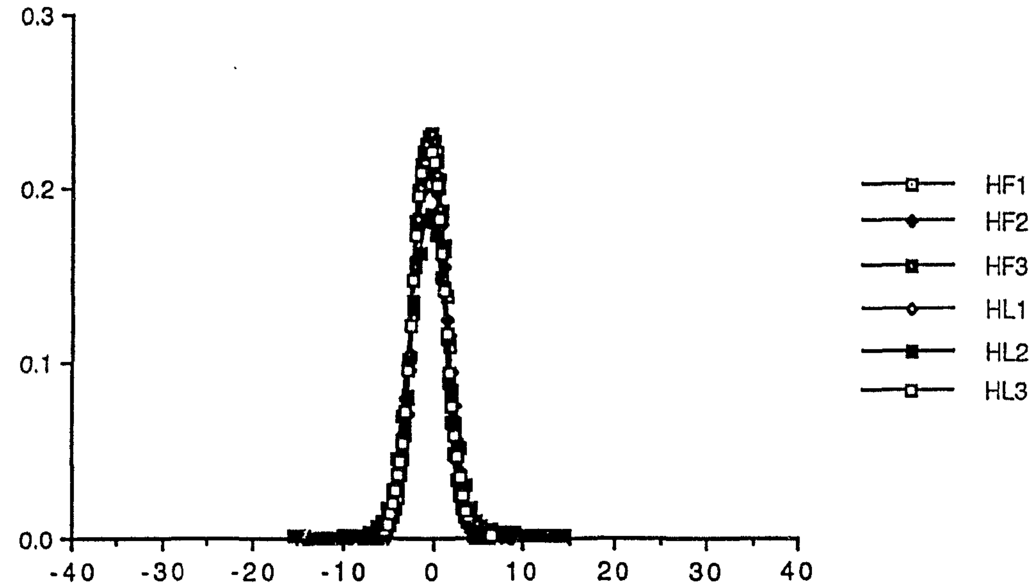
Experiment 3B:5 step ahead forecast for y1



Experiment 4B:5 step ahead forecast for y1

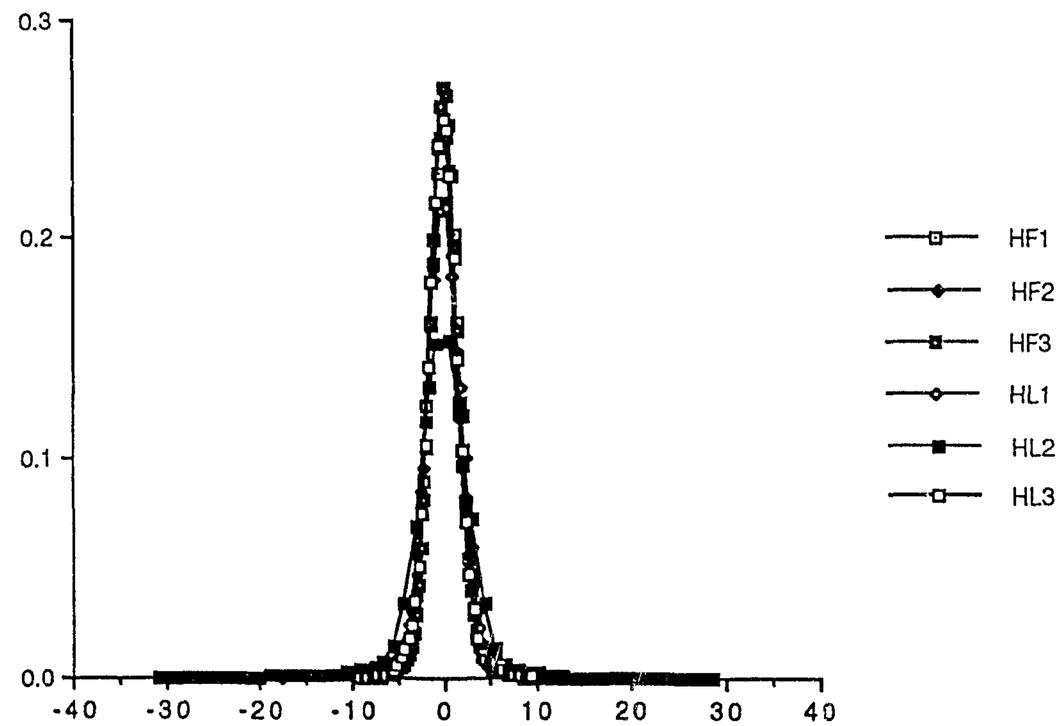


SET 7: Experiment 1A:5 step ahead forecast for y2



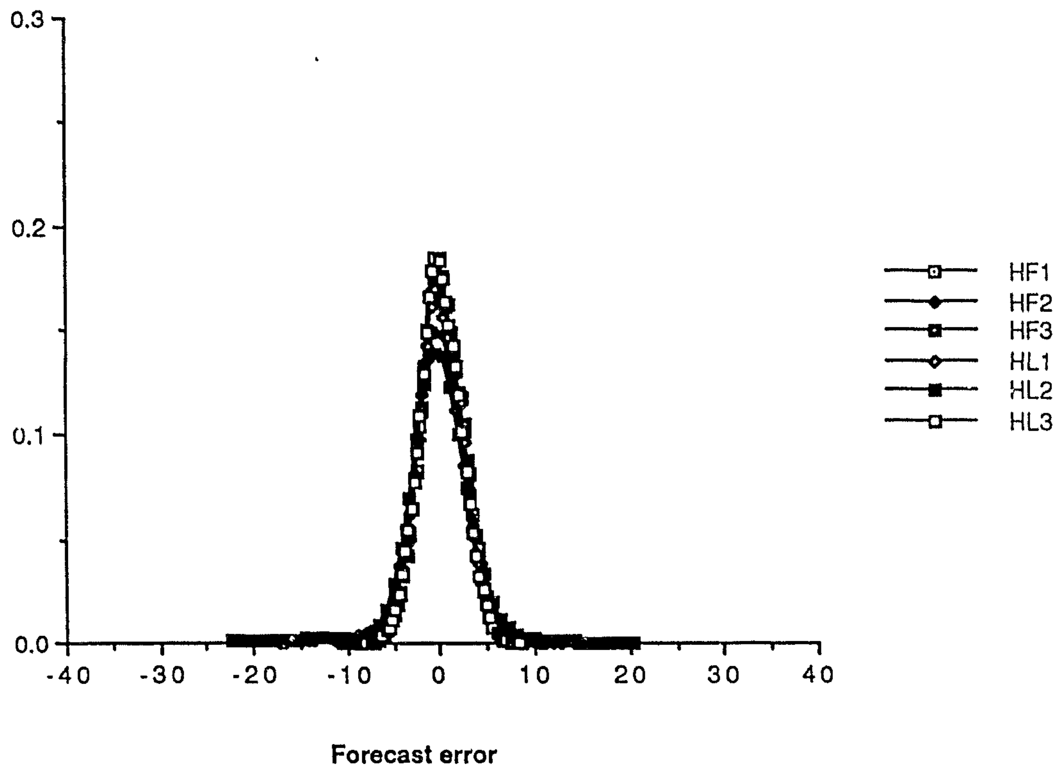
Forecast error

Experiment 2A:5 step ahead forecast for y2

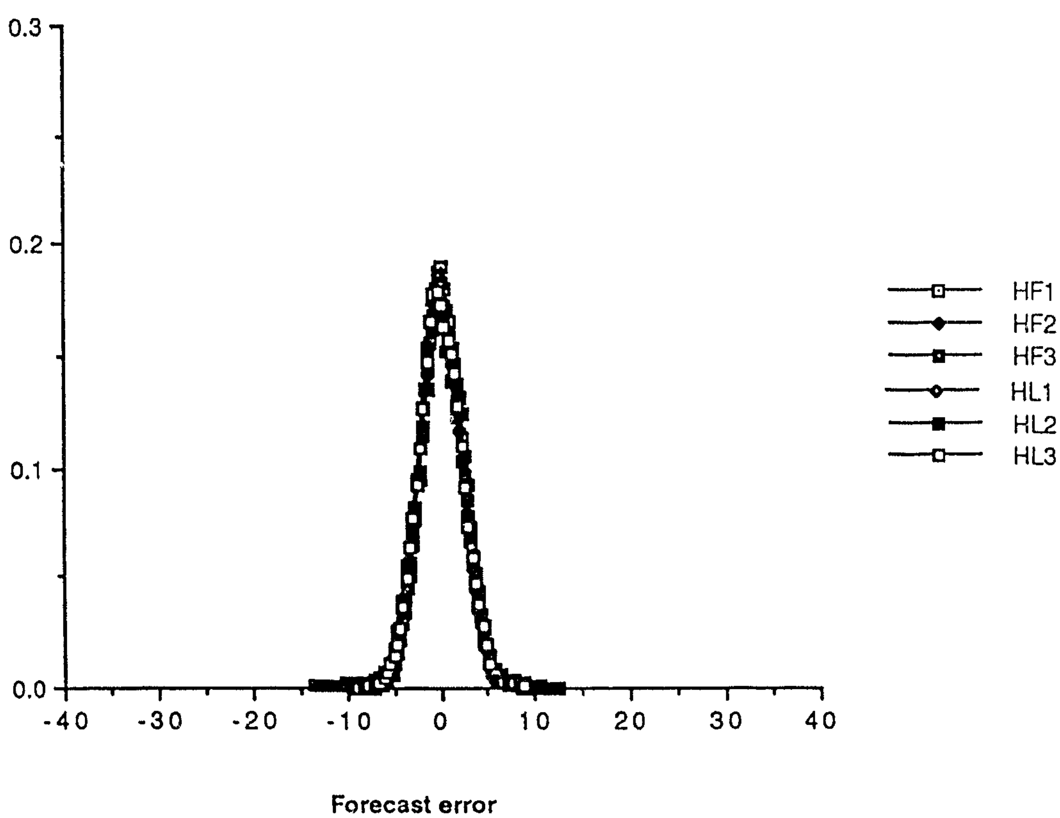


Forecast error

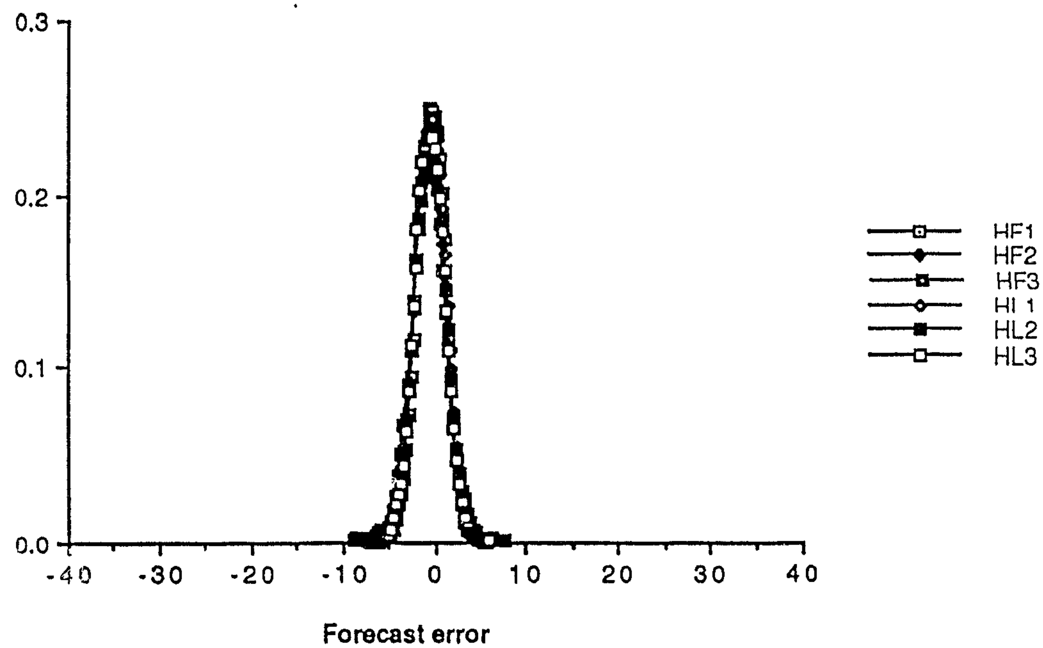
Experiment 3A:5 step ahead forecast for y2



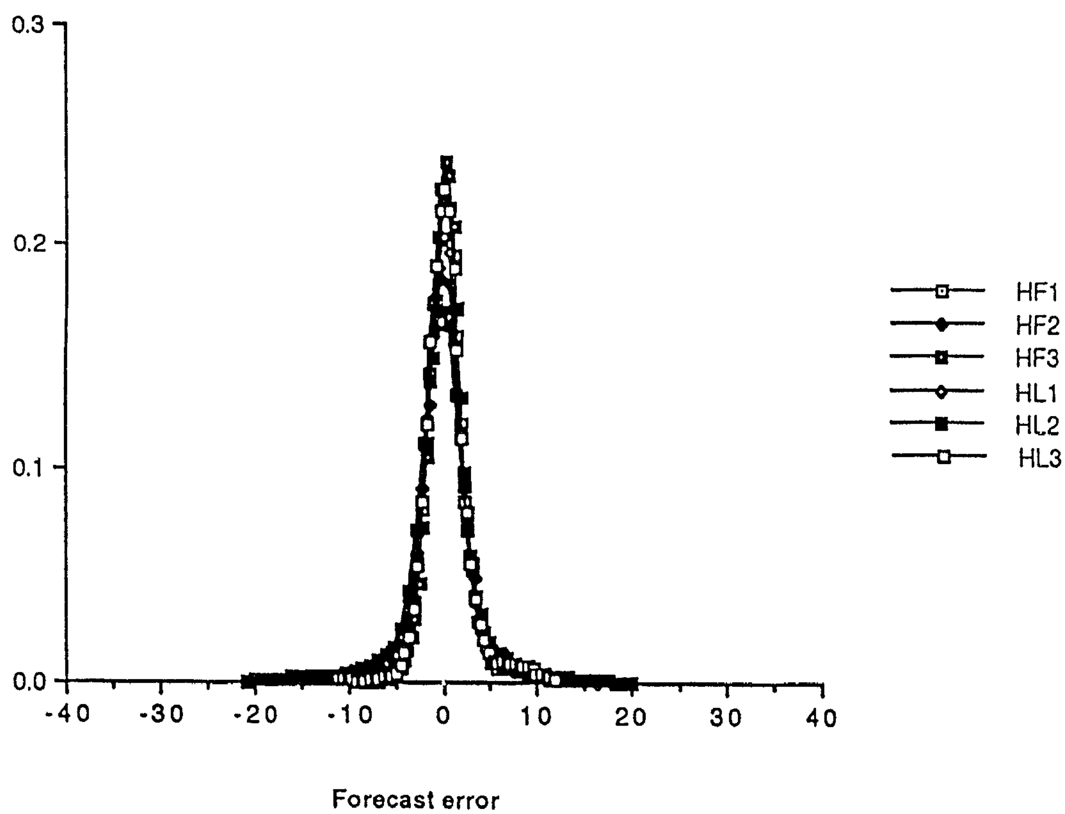
Experiment 4A:5 step ahead forecast for y2



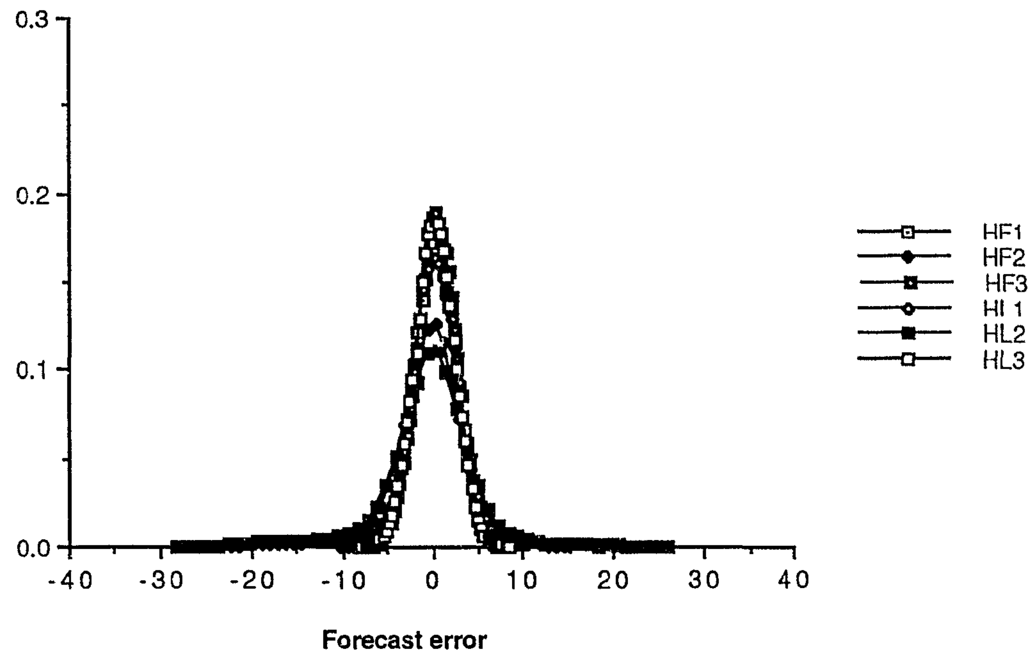
Experiment 1B:
Five-step a head forecast error for y2



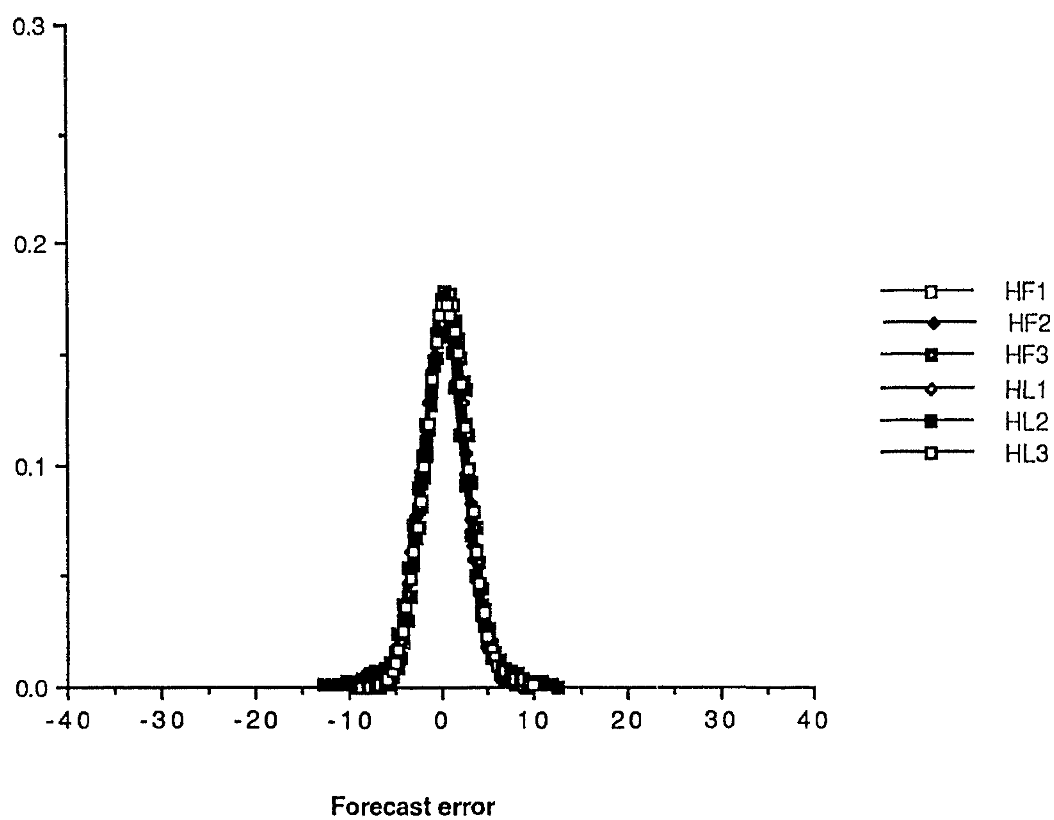
Experiment 2B:
Five-step a head forecast error for y2



Experiment 3B:5 step ahead forecast for y2



Experiment 4B:5 step ahead forecast for y2



CHAPTER 6: SUMMARY AND CONCLUSIONS

In this study we sought to understand the relative small sample properties of several two-step estimators of dynamic simultaneous equations models with autocorrelated errors using the Monte Carlo approach. The six estimators studied were proposed by Hatanaka and include three full information estimators denoted by HF1, HF2 and HF3 and three limited information estimators denoted by HL1, HL2 and HL3. All the six estimators have the desirable asymptotic properties of consistency and asymptotic efficiency. If asymptotic properties of the estimators reflect their performance in small samples, we would expect the full information estimators to perform relatively better than their limited information counterparts. However, since asymptotic properties of the estimators are not necessarily a reflection of their performance in small samples, we examined the relative performances of these estimators using typical samples of sizes 30 and 60. For purposes of experimentation, a two-equation model was chosen which roughly reflects the characteristics of a real-world model and, in addition, satisfied the standard assumptions usually made for models of this type. The changes in the parameters were made to reflect varying degrees of autocorrelation and changes in the coefficients of the lagged endogenous variables. To this end, four different structures depicting these changes were considered. As it turned out the performance of the estimators were sensitive to changes in the magnitudes of the autocorrelation coefficients as well as the coefficients of the lagged endogenous variables.

The study covered the econometric issues of structural estimation, prediction and hypothesis testing. In particular, the following questions were addressed:

1. How do the small sample properties of these estimators compare for purposes of structural estimation and for dynamic simulation forecasting?

2. How reliable are the formulae for the asymptotic covariance m ·ices of the estimators of structural coefficients and the dynamic simulation forecasts for purposes of testing hypotheses in small sample situations?

3. How do the kernel density estimates of the sampling distributions of the structural parameters and the forecasts compare when small samples are used and what are the effects of changing the sample size on the shapes of these distributions?

The first issue was analysed using the basic measures of bias, dispersion, skewness and kurtosis.

We present below a summary of the conclusions arrived at in this study. We categorize the conclusions into specific conclusions and general conclusions.

Specific conclusions

a) Structural estimation

Biases:

All the six estimators were significantly biased. The absolute values of the biases were very large in cases where the degree of autocorrelation in the model was also large. In such cases the biases were mainly positive. However, as the degree of autocorrelation was reduced the biases became mostly negative. In general, the limited information estimators had lower absolute values of the biases than the full information estimators in estimating the parameters of the equation with high autocorrelation and the reverse was true in cases of low absolute values of the autocorrelation coefficient. The most consistent rankings of the estimators according to the bias criterion seemed to occur only if the difference between the autocorrelation coefficients in the two equations was relatively large.

Referring to the absolute values of the biases, the ranking of the estimators according to this criterion was almost impossible, except for the fact that HF2 and HL2 tended to have larger biases. The biases generally decreased as the absolute values of the magnitudes of the autocorrelation coefficients were decreased and in this case the differences among the estimators became less pronounced. The biases of the autoregressive coefficients were generally negative and tended to decrease as the sample size was increased. Furthermore, the absolute values of the biases of the elements of Σ were relatively larger than those of r_{11} and r_{22}.

Mean Square Errors:

In general the limited information estimators tended to have larger MSE's than the full information estimators when autocorrelation was high. However, the reverse was true when autocorrelation was low. Also the MSE's of the elements of Σ were generally higher than those of the non-zero elements of R. The MSE's generally decreased as the sample size was increased. In terms of overall ranking of the estimators according to the MSE criterion, the full information estimators HF1 and HF3 and their limited information counterparts performed marginally better than HF2 and HL2. Perhaps, the only useful conclusion that could be reached is that HL2 performed consistently worse. This suggests that HL2 should possibly not be considered as a serious candidate for structural estimation. The ranking became extremely weak except in cases where there were no significant differences in the autocorrelation coefficients of the two equations. The rankings of the estimators according to the MSE's were only strong with sample size 30. The rankings became weaker as the sample size was increased. In fact, most of the time the estimators tended to perform better with an increase in sample size. Thus, as the sample size was increased the differences amongst the estimators became negligible. Except for HF2 and HL2, the differences among the estimators were not all that pronounced.

Hypothesis testing:

The asymptotic covariance matrices of the estimates of the structural parameters produced anomalous results for all the six estimators. The failure to find any rejections indicates the need to make significant small sample adjustments to the computed covariance matrix of errors, which in turn, might increase the power of the test. The exact nature of the small sample adjustments required needs to be explored.

Density estimates, skewness and kurtosis:

The ranking of the six estimators according to the density estimates is almost impossible. However, an inspection of the density estimates reveals that these are more or less symmetric. This observation remained invariant to the model used and was also supported by the skewness statistics. In relative terms, the full-information estimators tended to be more skewed than their limited information counterparts when autocorrelation was high and the reverse was true when autocorrelation was low. Also if the autocorrelation coefficient for a particular equation was high, the distribution of the estimates of the parameters of that particular equation tended to be more skewed than those of the other equation, for any given estimator. The kurtosis statistics show that the density estimates were leptokurtic. The value of the kurtosis statistic was greater than 3 for all coefficient parameters in all models suggesting that the distributions were more peaked than the standard normal distribution. The skewness and kurtosis increased as the sample size was increased. However the density estimates provided us with visual descriptions of the distributions of the two-step estimators. It must also be mentioned that the skewness and kurtosis statistics of the non-zero elements of Σ were relatively larger than the non-zero elements of R and these values tended to increase as the sample size was increased.

In relative terms the full information estimators tended to be more skewed than their limited information analogues when autocorrelation was high and the reverse was true when autocorrelation was low. If autocorrelation was high for a particular equation the distribution of the estimates of the coefficients of that equation tended to be more skewed.

b) Prediction

The conclusions that have emerged from our analysis of the sampling distributions of the dynamic simulation forecasts of estimators are given below.

Like the structural parameters the biases of the prediction errors of all the six estimators were significant at the 5 percent level. Although the direction of the biases was not clear, the biases of the prediction errors became generally negative as the gap between the autocorrelation coefficients of the two equations was widened. It is impossible to make a meaningful ranking of the estimators according to the magnitude of the prediction biases. Generally HF2 and HL2 were the most biased though the results were rather mixed. The biases were smallest for model 4 compared to the corresponding forecast periods in models 1, 2 and 3 suggesting that decreasing the autocorrelation coefficients might reduce the prediction biases. However, the rankings of the estimators became fairly consistent when the difference between the autocorrelation coefficients of the two equations was large.

Significant differences amongst the estimators emerged only if there were large differences in the autocorrelation coefficients and when the sample size was small. In general, the limited information estimators tended to outperform the full information estimators according to the MSEP criterion when autocorrelation was low. Overall, there seemed to be no substantial differences among the predictors HF1, HL1, HF3 and HL3. However, HF2 and HL2 performed consistently worse. Thus, for forecasting purposes, HF2 and HL2 should not be used. The MSEP behaved as expected from theory (i.e., decreased as the sample size was increased), except for HF2 and HL2 which tended to behave erratically as the sample size was increased. In general, the full information estimators dominated the limited information analogues when autocorrelation was low and imited information estimators performed marginally better when autocorrelation was high. Most of the times the estimators performed better with an increase in sample size and the differences among the estimators became less pronounced.

Hypothesis testing:

A major conclusion with respect to the use of the asymptotic standard errors is that inferences concerning predictions is not very reliable unless the sample sizes were fairly large. As noted above, the number of type I errors were generally larger for T = 30 than for T = 60. When T = 30, the asymptotic formulae resulted in too many rejections as indicated by the number of type I errors. As the sample size was increased the number of type I errors generally decreased. For these formulae to be reliable, samples of at least size 60 are required for valid inferences to be made. This is likely to be a stringent requirement since applied researchers work typically with sample sizes that are, in most cases, considerably less than 60, except in cases where the data are either quarterly or monthly. In particular, the asymptotic formulae were extremely unreliable when the absolute values of the autocorrelation coefficients are close to unity. Thus, if the autocorrelation coefficients are high, fairly large samples are required.

Returning to the relative performances of the six estimators considered in this study we note that when autocorrelation was very high, the limited information estimators HL1, HL2 and HL3 tended to perform relatively better than the full information estimators HF1, HF2 and HF3 in tests of hypotheses and construction of confidence intervals for predictions. This result is not surprising as the computation of the asymptotic covariance matrices of the full-information estimators involve a greater number of matrix inversions than the limited information estimators. This enhances the possibility of rounding errors to accumulate which might result in near singularity of the matrices involved. As it turned out the singularity problem mentioned above was relatively more frequent with the full-information estimators than with the limited information estimators. Even in cases where the autocorrelation was high, the ranking of the estimators was not consistent for all the forecast periods. However, as the absolute values of the autocorrelation coefficients were reduced, the full information estimators regained their dominance over the limited information estimators and that the predictors in small samples behaved as suggested by asymptotic theory.

Though the differences between HF1 and HF3 and their limited information counterparts HL1 and HL3 were marginal, the limited information estimators seem preferable for purposes of testing hypotheses about predictions and for constructing confidence intervals of forecasts when autocorrelation was high. The fact that the limited information estimators performed quite well when autocorrelation was high is of practical significance, because they are computationally easier to apply.

Density estimates, skewness and kurtosis

The kernel estimates of the sampling distributions suggest that the distributions of the prediction errors were almost symmetric and that they remain invariant to changes in the autoregressive coefficient and/or the coefficient of the lagged endogenous variable. It should be emphasized that the density estimates become extremely useful when used in conjunction with the skewness and the kurtosis statistics. It turned out that the density estimates of the six two-step estimators considered here were quite similar. The most significant differences amongst the estimators occurred only in cases where the differences between the autocorrelation coefficients of the two equations were quite high, i.e., in model 3 where r_{11} and r_{22} were 0.9 and -0.6, respectively.

The skewness and the kurtosis of the sampling distributions of the dynamic simulation forecast errors generally decreased as the sample size was increased.

General conclusions

1. No estimator emerged as the universal winner in all the

three aspects, i.e., structural estimation, prediction and hypothesis testing. This is partly because the relative small sample properties of the estimators depended to some extent on whether the models were used for structural estimation and hypothesis testing or prediction and hypothesis testing. More specifically, the relative small sample properties of the estimators for structural estimation tended to differ from their properties in prediction. Furthermore, the relative performances of the estimators were sensitive to the degree of autocorrelation, as well as to changes in the coefficients of the lagged endogenous variables. However, if the sample sizes are fairly large, there is no reason to choose one estimator over the other as the rankings became extremely weak, except in cases where the differences in the autocorrelation coefficients across equations were large.

2. The rankings of the estimators based on prediction performances are much stronger than their rankings in structural estimation as measured by Kendall's W. This is not surprising since prediction involves more steps than structural estimation thus leaving more room for differences amongst the estimators to emerge.

3. Density estimates should be used in conjunction with other descriptive statistics like skewness and kurtosis. A plot of the density estimates enables us to examine visually the shape of the sampling distributions and their sensitivity to changes in the sample size. In particular, the information

about the sampling distributions of the estimators, which are conventionally provided by the skewness and the kurtosis statistics, are also apparent from the density estimates. However, the density estimates may not facilitate the ranking of estimators in cases where the sampling distributions of these estimators are very similar in shape.

An important criterion for evaluating the alternative estimators is their relative robustness to changes in the values of the parameters. The ideal requirement is for an estimator to be able to perform quite well in the parameter space that is most likely to be encountered in actual empirical work. Although Monte Carlo studies are often criticized for concentrating on a few selected points in the appropriate parameter space, nevertheless, the results of such studies provide us with some useful clues regarding the shapes of the small sample distributions of the estimators.

Clearly, the foregoing discussion suggests that the choice among different estimators is not an easy one as their performances were similar in many respects. There were no significant differences among the estimators when large sample sizes were used, except in cases where there are large differences in the autocorrelation coefficients of the two equations. Nevertheless, the results of the sampling experiments presented here provide useful information regarding the relative small sample performances of the estimators for models of the type used in this study. We hope that these results will be of some use to applied researchers as well as a guide to those who might wish to pursue this problem using analytical techniques. Furthermore, we were able to corroborate some of the results of Moazzami and Buse (1986). The computational convenience of Hatanaka's estimators render them quite attractive in empirical work and any effort to understand their relative performances in small samples is fruitful. Finally, it must be noted that the results presented in this study are subject to correct specification. The effect of specification errors on the relative performances of these estimators is an area for future research.

BIBLIOGRAPHY

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

Baillie, R.T.(1982),"Prediction from the Dynamic Simultaneous Equation Model with Vector Autoregressive Errors," Econometrica,49,1331-1337.

Basmann, R.L., D.H.Richardson, and R.J. Rohr (1974),"An Experimental Study of Structural Estimators and Test Statistics Associated With Dynamical Econometric Models," Econometrica,42, 717-730.

Binkley J.K. and G.Nelson (1984), "Impacts of Alternative Degrees of Freedom Corrections in Two and Three Stage Least Squares," Journal of Econometrics, 24,223-233.

Box, G.E.P, and M.E. Mueller(1958),"A Note on the Generation of Random Normal Deviates," Annals of Mathematical Statistics, 28,610-611.

Brown, T.M. (1960),"A Distribution Free Method of Equation System Structure Estimation,"International Economic Review, 3,173-191.

Brundy, J.M., and D.W. Jorgenson(1971),"Efficient Estimation of Simultaneous Equations by Instrumental Variables," The Review of Economics and Statistics, LIII,207-224.

Buse, A.(1989),"Efficient Estimation of Structural Equation with First Order Autocorrelation,"Journal of Quantitative Economics, 5,59-72.

Cacoullos, T.(1966),"Estimation of a Multivariate Density," Annals of the Institute of Mathematical Statistics, 18,179-189.

Chow, G.C., and R.C. Fair(1973), "Maximum Likelihood Estimation of Linear Equation Systems with Autoregressive Residuals," Annals of Economic and Social Measurement, 2,17-28.

Dhrymes, P.J.(1970),Econometrics:Statistical Foundations and Applications, New York:Harper & Row.

______,R.Berner, and D.Cummins(1974),"A Comparison of Some Limited Information Estimators For Dynamic Simultaneous Equation Models With Autocorrelated Errors," Econometrica, 42,311-322.

______, and J.B. Taylor(1976),"On an Efficient Two-Step Estimator For Dynamic Simultaneous Equation Models With Autoregressive Errors,"International Economic Review, 17,362-376.

______(1974),"A Note on An Efficient Two-Step Estimator," Journal of Econometrics ,2, 301-304.

______,and H.E. Erlat(1974),"Asymptotic Properties of Full Information Estimators of Dynamic Autoregressive Simultaneous Equation Models," Journal of Econometrics, 2, 247-259.

Evans, G.B.A. and N.E. Savin(1982), "Conflict Among the Criteria Revisited; The W, LR and LM Tests," Econometrica, 50, 737-748.

______ (1987), "Conflict Among Testing Procedures in A Linear Regression Model With Lagged Dependent Variables," in Advances in Econometrics Fifth World Congress (vol.1), W.Hildenbrand, ed., Oxford University Press, 263-283.

Fair, R.C.(1970),"The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica,38,507-516.

______(1972),"Efficient Estimation of Simultaneous Equations With Autoregressive Errors by Instrumental Variables," The Review of Economics and Statistics,54,444-449.

______(1984),Specification,Estimation and Analysis of Macroeconometric Models,Cambridge,Massachusetts:Harvard University Press.

Fix, E., and J.L. Hodges(1951),"Discriminatory Analysis, Nonparametric Estimation: Consistency Properties," Report No 4, Project No. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

Fomby,T.B.,R.C. Hill, and S.R. Johnson(1984),Advanced Econometric Methods,New York:Springer Verlag.

Godfrey, L.(1976),"Testing for Serial Correlation in Dynamic Simultaneous Equation Models," Econometrica,44, 1077-1084.

______ (1978a),"Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," Econometrica, 46, 1293-1301.

_______(1978b),"Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables,"Econometrica,46,1303-1310.

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

Guilkey, D. (1975),"A Test for the Presence of First Order Vector Autoregressive Errors when Lagged Endogenous Variables are Present,"Econometrica, 43, 711-717.

Harris, P. (1985) An Asymptotic Expansion for the Null Distribution of Efficient Score Statistic, Biometrica, 72, 653-659.

Hatanaka, M.(1974),"An Efficient Two Step Estimator for the Dynamic Adjustment Model with Autoregressive Errors," Journal of Econometrics,2,199-220.

________(1976),"Several Efficient Two Step Estimators for the Dynamic Simultaneous Equation Model with Autoregressive Disturbances,"Journal of Econometrics,4, 189-204.

Hendry, D.F.(1972),"Maximum Likelihood Estimation of Simultaneous Regression Equations with Errors generated by a Vector Autoregressive Process," International Economic Review,12,257-272.

______(1974),"Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process: A correction," International Economic Review,15,260.

______,and R.W. Harrison(1974),"Monte Carlo Methodology and the Small Sample behaviour of Ordinary and Two-Stage Least Squares,"Journal of Econometrics,2,151-174.

______(1976),"The Structure of Simultaneous Equations Estimators," Journal of Econometrics,4,51-88.

______,and F. Srba(1977),"The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems," Econometrica,45,969-990.

______(1984),"Monte Carlo Methods In Econometrics," in Handbook of Econometrics (vol. 2), Z.Grilches and M.D. Intriligator, eds., North Holland, Amsterdam, 937-976.

Hurd, M.D.(1972),"Small Sample Estimation of a Structural Equation with Autocorrelated Errors,"Journal of the American Statistical Association, 67, 567-573.

Intriligator, M.D.(1978), Econometric Models Techniques and Applications, New Jersey, Prentice Hall.

Knight, J.L. (1982), "Asymptotic distribution of Restricted Reduced Forms and Dynamic Multipliers in a Linear Dynamic Model with Vector Autoregressive Errors," International Economic Review, 23, 553-563.

Maddala, G.S. (1974), "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," Econometrica, 42, 841-851.

Magnus, J.R. and H.Neudecker, "Symmetry, 0-1 Matrices and Jacobians: A Review," Econometric Theory, 2, 157-190

Maritz, A.(1978), "A Note of Correction to Guilkey's Test For Serial Independence in Simultaneous Equation Models," Econometrica, 46, 471.

Moazzami, B. and A. Buse(1986), "The Small Sample Performance of Some Limited Information Estimators of a Dynamic Structural Equation with Autocorrelated Errors," Journal of Statistical Computation and Simulation, 26, 55-78.

Neudecker, H.(1969), "Some Theorems of Matrix Differentiation with Special Reference to Kronecker Matrix Products," Journal of the American Statistical Association, 65, 953-963.

Parzen, E.(1962), "On the Estimation of Probability Density and Mode," Annals of Mathematical Statistics, 33, 1065-1076.

Power, S., and A.Ullah (1986), "Nonparametric Monte Carlo Density Estimates of Rational Expectations Estimators and Their t-ratios," University of Regina Discussion Paper No. 16.

Raj, B.(1980), "A Monte Carlo Study of Simultaneous Equation Estimators with Normal and Non-normal Disturbances," Journal of the American Statistical Association, 75, 221-229.

Rao, U.L.G.(1986), "Generalized Least Squares Estimation of Simultaneous Equations Models with First Order Autocorrelated Errors," Metron, XLIV, 263-276.

Rosenblatt, M.(1956), "Remarks on some Nonparametric Estimates of the Density Function," Annals of Mathematical Statistics, 27, 832-837.

Rothenberg,T.J., and C.T. Leenders(1964),"Efficient Estimation of Simultaneous Equation Systems," Econometrica,32,57-76.

Rothenberg,T.J.(1984),"Hypothesis Testing in Linear Models When the Error Covariance Matrix is Non Scalar," Econometrica, 52, 827-842.

_______(1987), "Comparing Alternative Asymptotically Equivalent Tests," in Advances in Econometrics Fifth World Congress (vol. 1), W. Hildenbrand, ed.,Oxford University Press, 255-262.

Sargan, J.D.(1961),"The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals," Econometrica,29, 414-426.

Schmidt, P.(1973),"The Asymptotic Distribution of Dynamic Multipliers," Econometrica, 41,161-164.

Schmidt, P.(1974),"The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model," Econometrica, 42, 303-309.

______(1976),Econometrics,New York,Dekker.

Siegel, S.(1956),Nonparametric Statistics for the Behavioural Sciences, New York, McGraw-Hill.

Silverman, B.W.(1986),Density Estimation for Statistics and Data Analysis, New York, Chapman and Hall.

Smith, V.K. (1973),Monte Carlo Methods: Their Role in Econometrics, D.C. Heath, Massachusetts.

Theil, H.(1958),Economic Forecasts and Policy, North Holland, Amsterdam.

Ullah, A.(1988),"Nonparametric Estimation of Econometric Functionals," Canadian Journal of Economics, 21, 625-658.

Wang, G.H.K., and W.A. Fuller(1982),"Estimators for a Simultaneous Equation Model with Lagged Endogenous Variables and Autocorrelated Errors," Communications in Statistics, Simulation and Computation,11,123-142.