Monica Ahlstedt

Small Sample Estimation and Stochastic Simulation of an Econometric Model

Suomen Pankki

1986 D:63

Monica Ahlstedt

Small Sample Estimation and Stochastic Simulation of an Econometric Model

Suomen Pankki
Helsinki 1986

D:63

ISBN 951-686-127-X
ISSN 0355-6042

ACKNOWLEDGEMENTS

The subject of this study was proposed to me when I made my first acquaintance with model-building as a member of the Model Group of the Bank of Finland in 1982. Most of the work was done during my second period of secondment to the Research Department in 1984 - 1986, consisting of some 15 months of full-time research. The study was approved as a Licentiate Thesis by the Statistics Department of the University of Helsinki in November 1986.

I am grateful to both the Economics Department and the Research Department for giving me the opportunity to carry through this project. My official examiners Hannu Niemi and Juha Ahtola and my unofficial examiner Ilkka Mellin also deserve thanks for their time and trouble.

I also owe a debt of gratitude to Juhani Hirvonen for his help, especially during the first, empirical part of the study. Equally I wish to thank Matti Virén for his comitment in helping to overcome the difficulties encountered in the final stage of the project and for shared adventures in the world of model-building.

In addition, I had the great privilege of receiving most helpful and encouraging comments from international experts in the field of estimation and simulation. Ray Fair placed his program at my disposal and also helped me in the interpretation of the empirical results. Paolo Corsi and Clive Granger displayed immense patience in helping me to solve some of my technical problems during a course on model-building held in Italy in 1984. John B. Taylor was kind enough to take an interest in my paper during his lecturing at the Bank of Finland.

Many people have contributed to making this study into a finished product. I especially wish to thank Kaisa-Liisa Nordman for her

computational assistance and for her spirit of never considering any task to be too difficult to handle. Päivi Lindqvist did excellent work in typing the manuscript and Malcolm Waters gave my home-made English an Oxford touch. Jukka Syrjänen did a great job in getting the simulation program to work. The staff of the Library of the Bank of Finland was extremely helpful as usual in acquiring literature from all possible sources.

The financial support of Svenska Vetenskapliga Centralrådet is gratefully acknowledged.

Finally, I want to thak my husband, Tor, for providing never- ending support, encouragement and, when needed, a shoulder to cry on.

Helsinki, December 1986

Monica Ahlstedt

CONTENTS

I INTRODUCTION

I.1 The aim of the study

Let us define a macroeconomic model as a system of simultaneous equations describing the behaviour of the economic units that we observe around us and want to explain. Models are the most widely known and used quantitative instruments for economic forecasting and evaluation of the effects of alternative government actions on the economy[1]. When models are used for economic policy purposes, it is important that policy makers should be provided with a measure of reliability along with the forecasts. This study is concerned with the uncertainty inherent in economic models. The aim of the study is twofold: first to investigate how to improve the reliability of the model by minimizing the uncertainty in the estimation phase of the model and then how to calculate the variance-covariance matrix of forecasts so as to measure the reliability of a model.

I.2 Outline of the study

In the construction of an econometric model we need two main tools: economic theory and statistical methods. Economic theory is used to generate the theoretical model of underlying basic relationships. The theoretical model is then transformed into an empirical one by applying statistical estimation methods. The transformation is the outcome of interaction between theory and the information contained in the data. If, for instance, the signs or magnitudes of the estimated parameters do not satisfy the theoretical a priori

---

[1]Since policy evaluations can be interpreted as conditional forecasts, from now on we use the word forecast to also include policy evaluations.

hypothesis we have to reconsider the theory and then go back to the data again. This search procedure of back-and-forth movements between theory and empirical results can be an important part of the empirical work and has to be carried on by means of single equation estimation methods, because they allow a very weak specification of the model as a whole while individual equations are being examined. The method of ordinary least squares is the usual choice, even though it is known not to be theoretically optimal when the model is simultaneous.

During this search procedure the structural form of the model is specified: the functional forms of the equations, the classification of variables into endogenous and exogenous, the restrictions on the parameters and the lag distributions.

Once the model has been estimated the next step is to test and analyze it. There are a great many tests available for examining the properties of both the parameters and the residuals and for detecting misspecification. In addition, multipliers are calculated so as to examine the properties of the model. If the results of this testing and analysis procedure give poor fits, wrong signs or wrong magnitudes of the multipliers in relation to a priori expectations, the search procedure between theory and the results is continued until the model specification is considered to be as correct as possible given the results of the search, test, and analysis procedures.

This study starts from the point where the specification phase has been performed. The point of departure is thus an empirical model the structural form of which has been derived from economic theory and initial estimation from observed data.

When a real-word macromodel is used in forecasting the usual way is to solve it deterministically over the forecast period. The forecasts are then the values of the endogenous variables from a simultaneous solution of the model obtained by replacing the structural disturbances with their expected values, which are all

zero. The deterministic solution has, however, two disadvantages. First, because models are as a rule nonlinear, nonlinearity induces bias into the deterministic solution. Secondly, the deterministic solution takes no account of the random nature of the model, giving only point predictors.

One would not expect the generated point forecasts from the deterministic solution to be perfectly accurate; there will always be some forecast error. The main sources of forecast error are:

- a possible misspecification in the model,
- the uncertainty attaching to the exogenous variables,
- the use of parameter estimates in forecasting as opposed to the unknown true values of the parameters, and
- the presence of random disturbances in the behavioural equations.

The first component is due to the fact that the specification search procedure may yield a model that fits the data well and seems to be quite good, on the basis of the test procedure, when it is in fact a poor approximation to the structure. Another point is that a given model is, even at best, only an approximation of the "real world". The structure of the real economy may be too unstable to be approximated closely by a model with constant parameters. The second source of forecast error is due to the fact that the accuracy of the forecast also depends on the forecaster's ability to anticipate the values of the exogenous variables, i.e. those variables that are used in the model but not predicted by it.

Given the point of departure for our analysis, the first source of error falls outside the scope of the study. Furthermore, we can dispense with the second error component by assuming that the values of the exogenous variables are known with certainty. Consequently, the components of forecast error considered here are those arising from the parameter estimates and from stochastic disturbances. These two components make the model random, and so in

this sense the study deals with measuring the stochastic components of the model. The contribution of these two sources of randomness is measured by means of the variance-covariance matrix of the forecasts. A forecast obtained from a stochastic model is itself a stochastic variable. The calculation of the forecast variance makes it possible to treat the forecasts as random variables. This means that the point forecasts are reported along with an estimate of their second moment to indicate the tolerance interval of the point estimate.

An alternative to the deterministic solution is the stochastic simulation, which takes account of both the nonlinear and the random nature of the model. A stochastic simulation is the solution obtained after assigning a pseudo-random value to the disturbance terms and/or the parameters.

Several stochastic simulation methods have been proposed in the literature for estimating the contribution to forecast errors of these two error sources. All the stochastic simulation methods are based on the assumption of consistency of the estimators of the parameters. However, most macromodels are only estimated using ordinary least squares, which are known to produce estimates which are both biased and inconsistent.

Although the primary aim of this study is to measure the uncertainty present in forecasts, we first have to face the problem of estimating large simultaneous models consistently. In addition to consistency, efficiency is also an important quality of the estimators in this context, because the reliability measure of forecasts, the variance-covariance matrix, is a function of the variance-covariance matrix of the parameter estimates.

The first part of the study is therefore concerned with finding a simultaneous estimation method which is optimal in terms of consistency and efficiency of the parameters. The ideal approach to the comparison of the relative merits of competitive estimation methods would be to rely on analytical formulas for the properties

of the separate estimates. Analytical results have the merit of generality, thus covering the whole parameter space. Unfortunately, general analytical results can be derived only asymptotically. In most cases, however, an economic model has to be specified from finite samples. The reliance on asymptotic results can lead to serious problems of bias and a low level of inferential accuracy when sample sizes are small and asymptotic formulas poorly represent sample behaviour. Analytical small sample properties of estimators can only be derived for limited static models, usually containing no more than two jointly endogenous variables. Therefore the analytical results for finite samples only permit the analysis of the behaviour of estimators in a small part of the parameter space. The severely limited analytical knowledge of small sample distributions has led to the use of sample experiments - that is the Monte Carlo approach, to obtain empirical information about the small sample behaviour of the estimators. The traditional Monte Carlo question is: Are the small sample properties of the estimators close to the large sample properties? Another question is: How serious are effects of departures from the theoretical models on the parameter estimators?

The comparison of estimators in this study is made as complete as possible by using all available approaches for finding the optimal estimation methods. Thus, we use asymptotic theory to select the empirically applicable estimation methods and to provide an analytical background for integrating the empirical results. We then use the available analytical finite sample results and Monte Carlo experiment results as complementary approaches for establishing an a priori ranking order of the estimators. These a priori selected estimation methods are then applied in the empirical estimation of one special model. The empirical ranking order is found by comparing the statistical measures of dispersion along with testing the within sample dynamic behaviour and the post sample one step ahead forecastability of the models built from the various estimates.

One would not expect that a single estimation method performs best on all criteria and all problems. Nevertheless, a method can have

sufficient support to justify its choice as a method to be recommended. In this study, we end up by selecting the iterative instrumental method, which uses the numerical solutions of the model itself as instruments. The method was first introduced by Brundy and Jorgenson (1971) but for some reason it has not been used in reported model estimations. This method is found to be optimal according to the analytical criteria of consistency and relative efficiency. The optimal property in terms of efficiency springs from the utilization of the information embedded in the model. The analytical finite sample properties of the estimators of the method cannot be found. The empirical results of this study show, however, that the estimates compete well with the ordinary least squares estimators in the small sample case. The comparison of simulation behaviour was carried out for a sample size of 20. Monte Carlo expriments have revealed that this sample size favours ordinary least squares against simultaneous estimation methods. When sample size increases, simultaneous methods outweigh ordinary least squares. This gives us reason to believe that, with increasing sample size, the iterative instrumental simultaneous method will be preferable even in simulation behaviour. The method is especially well suited to estimation from undersized samples[2], which is a problem that model builders generally have to face.

Although the iterative instrumental method does not violate any assumption necessary for assuring consistency and efficiency, it does not rely on any subjectively made approximations, which are necessary in all the other applicable estimation methods. It removes the arbitrariness from different sources inherent in the other methods, thus being objective from the methodological point of view, when objectivity is defined to mean that every researcher applies the method equally using no subjective choice.

---

[2]Swamy (1980) gives the following definition: An undersized sample is a sample where the number of the predetermined variables exeeds the number of the observations on which the estimation is performed.

Having dealt with the question of estimation, we then proceed to the second purpose of the study: to obtain an estimate of the variance-covariance matrix of the forecast using different stochastic simulation methods. Comparison of the various methods results in the selection of the residual-based method proposed by Mariano and Brown (1984). A distinctive feature of this method is the absence of approximations, the same feature which characterizes the selected iterative instrumental estimation method. Compared to the other simulation methods, this reduces misspecification sensitivity.

The suitability of the selected iterative instrumental estimation method and the residual-based stochastic simulation method is enhanced by the fact that, compared with other methods, they are both very easy to apply. The small amount of computational work that has to be done for instrumental estimation after the inevitable first ordinary least squares estimation, is more than compensated for by the resulting improvement in the quality of the parameter estimates and in the reliability of the forecasts. In the residual-based stochastic simulation method the simplification derives from the fact that, in the calculation of the variances of the forecast errors, the labourious procedure of generating a new random population becomes unnecessary since the method uses estimated sample period residuals instead of pseudo-generated data.

In the empirical part of this study the methods were applied to the structural annual model of the Bank of Finland. The study is, however, based on the assumption that the basic properties of macroeconomic models are similar enough for the conclusions obtained from the use of a specific model to be generalized to other models. For the iterative instrumental method selected as the best one, the generalization is probably justified for all models with two features typical of most macromodels:

- nonlinearity in variables,
- undersized samples

In addition to the structural model, we also apply the residual-based simulation method to a naive model. This model is a system of completely separate autoregressive equations. The same endogenous variables that are estimated by a behavioural equation in the structural model are simply regressed on a constant, time and the first four lagged values.

II ESTIMATION OF A SIMULTANEOUS MODEL

The standard economic model with additive stochastic disturbance terms can be written as a structural form system of M equations:

$$f(Y,X,d) = u \tag{1}$$

where

Y	is a vector of M endogenous variables
X	is a vector of N exogenous variables
d	is a vector of S structural parameters to be estimated
u	is a vector of M structural stochastic disturbances having zero mean and independently and identically distributed over time, with a finite covariance matrix and independent of all the predetermined variables
f	is a M·1 vector of functional operators, continuously differentiable with respect to the elements of y,x and d.

The stochastic simulation technique, which is introduced in chapter III, is equally well applicable to models where f is linear or nonlinear in its parameters and/or the current endogenous variables. For estimation purposes, in this first part of the study we restrict the analysis to the cases where f is either linear or nonlinear in variables only[1].

If f is either linear or nonlinear in variables, linear single equation estimation is applicable. When the estimation method requires knowledge of the reduced form it is assumed to be linear.

---

[1]The nonlinearity in parameters case is not likely to be used in estimation of macroeconomic models because it makes the search process impossible.

This assumption is an approximation only for the nonlinear case and it is dropped later on in section II.3.

In the first transformation of a theoretical model into an emprical model a simple estimation method has to be used. OLS is preferable because of its simplicity although it is known not to be theoretically optimal. The first question to be dealt with is how the estimators of the parameters can be improved by applying an estimation method that is more appropriate from the point of view of statistical theory.

Statistical estimation methods are frequently justified in econometric work on the basis of certain desirable asymptotic properties of the distribution of the estimators. This is the usual approach because of the relative easiness of deriving asymptotic properties as compared with finding the relevant small sample distributions. In addition, even though the limited degrees of freedom is a small sample problem, the high degree of overidentification of each separate structural equation can be regarded as a justification for the use of asymptotic criteria.

Hendry (1976) has shown that almost all simultaneous estimators can be interpreted as numerical approximations to Full Information Maximum Likelihood estimators and hence one simple formula can be obtained which encompasses all of them. This interpretation leads to asymptotic equivalences between estimators but allows the numerical variants to induce very different finite sample properties. Consequently, using small sample properties is another way of ranking. This can be done in two ways: using analytical results or using Monte Carlo experiments.

The true structural specification of the model is rarely known and hence a valuable criterion for selection between estimators is their robustness to misspecification. A third method of comparison is thus to analyze the dynamic simulation behaviour of the empirical results. Here we refer to the statement of Klein that the

real test of the validity and usefulness of any theory is its ability to predict (Klein, A Textbook of Econometrics, 1956 p. 249). Dynamic simulation comparison will reveal the effect of misspecification on the parameter estimates. Sensitivity against specification errors may be a more important guide in the choice of the estimator than asymptotic properties or formal small sample properties assuming correct specification.

We hope that the findings of these different methods of comparison in this study will also provide us with an answer to the question as to whether there is a contradiction between ranking based on asymptotic properties and the ranking order for small sample estimates.

II.1 Ranking of estimation methods according to their asymptotic properties

When the comparison of estimation methods is based on the asymptotic properties of the estimators, the desirable properties are consistency and asymptotic efficiency. An estimator is said to be consistent if and only if it converges in probability to a point at which the true distribution of observations is yielded (Hatanaka (1976)). An unbiased estimator is said to be efficient if its variance is smaller than the variance of any other estimator. If an estimator cannot be proved to be unbiased, only consistent, which is an asymptotic property, then the estimator can only be asymptotically efficient. An improvement in estimation is thus measured by the probability of concentration about the true parameter values. The most frequently used single equation estimation method, ordinary least squares, OLS, does not take account of the simultaneity between variables. As a consequence OLS estimators are not consistent. Although ranking according to asymptotic properties rules out the OLS method, this does not mean that it is useless. The other ranking criteria presented below will show that OLS estimates should be calculated along with simultaneous method estimates as possible alternatives.

Simultaneous methods have been developed so as to yield consistent estimators under classical statistical conditions that are listed for the disturbance vector in (1). By means of this quality, simultaneous estimators are on an equal footing, so that the comparison has to be based on their relative efficiency.

Simultaneous estimation methods can be divided into system methods and single equation methods. A general view is that the more information is utilized in a method the higher is the efficiency of the estimator. In system methods all the structural equations are estimated simultaneously. These methods make use of all the information in data and the restrictions on the parameters of the full system. The single equation methods make use of only the restriction concerning the particular equation in estimating each of the equations in turn. Single equation estimation methods draw on the rest of the system only to learn which predetermined variables, excluded from this equation, do in fact appear in the other structural equations. No use is made of estimates of parameters of other structural equations; nor do they make use of any a priori restrictions on the other equations. Measured by the utilization of information the class of system estimators is efficient relative to the class of single equation estimators.

The main results on asymptotic efficiency derived for consistent estimates are (Theil, Principles of Econometrics, 1971):

- The full information maximum likelihood (FIML) estimator is asymptotically efficient. The asymptotic variance-covariance matrix[2] converges to the Cramer-Rao lower bound for system estimators. Under certain regularity conditions the asymptotic variance matrix of three stage least squares (3SLS) estimators is identical to that of FIML estimators; thus 3SLS estimators are also asymptotically efficient.

---

[2]From now on the variance matrix.

- Limited information maximum likelihood (LIML) estimators are efficient in single equation estimation and the variance matrix reaches the Cramer-Rao lower bound for this class of estimators. Two stage least squares (2SLS) estimators and LIML estimators converge in distribution to the same limit so that 2SLS estimators are also efficient.

- Instrumental variable (IV) estimators are asymptotically efficient only when all the predetermined variables in the system are used as instruments. In this special case IV-estimators are identical to 2SLS estimators.

When choosing between system estimation and single equation estimation, it also has to be borne in mind that, besides the efficiency criterion, system estimation is more complicated to apply. The optimality properties of system estimation are not valid if there is any misspecification in the model. A misspecification in one equation affects only that particular equation in single equation estimation but is spread over the whole model if system estimation is used.

Another factor that has to be considered when choosing between system and single estimation methods is the question of nonlinearity. Except for some models constructed for pedagogical use macroeconomic models are nonlinear in variables. The use of nominal and volume variables makes for nonlinearity. It is not possible to have a strictly linear economy if one wants to build a system on accepted premises of economic theory. Each individual equation can be linearized so that single equation methods can be used in the ordinary way for linear models. But unless a model is completely linear, both in variables and in parameters, nonlinear methods have to be used in system estimation.

II.2 Limitations for estimation from undersized samples

For models of even moderate size we face the situation where the number of observations does not exeed the number of the exogenous variables in the system; i.e. we are dealing with an undersized sample. This means that system estimation cannot be applied in the original asymptotically efficient form. A condition for the application of FIML and 3SLS estimation is that the number of observations is greater than the sum of the endogenous and exogenous variables. Nor can the above "conventional" 2SLS or the "best" IV single equation estimation technique be used. In estimation from small samples we are faced with the following problem: is it possible to attain efficiency using some modification of the single equation methods that are asymptotically efficient in large samples but are not applicable in small samples?

Since system estimation cannot be applied, all the estimation methods to be discussed here are single equation estimation methods. Therefore it will be sufficient to consider a typical equation in the system specified in (1). Without loss of generality, the first equation may be considered "typical", which allows the use of somewhat simplified notations.

The first equation in a system that can be linearly estimated may be written as:

$$y_1 = Y_1 b + X_1 c + u = z_1 d_1 + u_1$$

where

y_1	is a vector of T observations on the endogenous variable which has been normalized to be "dependent"
Y_1	is a $T \cdot (G_1 - 1)$ matrix of observations on the other included endogenous variables
X_1	is a $T \cdot K_1$ matrix of observations on the included predetermined variables: exogenous and lagged endogenous.

$d_1 = (b\ c)'$, a vector of $(G_1 - 1 + K_1)$ parameters to be estimated

u_1 is a vector of T disturbances

$Z_1 = (Y_1\ X_1)$

In addition $E(uu') = \sigma^2 I_T$. The disturbances are assumed to be mutually independent. The exogenous variables in the system are assumed independent of the disturbance vector while the mutual independence assumption implies independence of the lagged endogenous variables and the current disturbance vector. No assumption about the distribution of the stochastic disturbance vector is needed for estimation. The conventional assumption of normality is necessary only if maximum likelihood estimation is performed. Likewise, normality is required to define the efficiency of the estimators in terms of the Cramer-Rao lower bound. At a later stage we introduce normality to allow hypothesis testing and the construction of confidence intervals for the parameter estimates.

Let X be the T•K matrix of observations on all the exogenous variables and K_2 the number of excluded exogenous variables. Let G_2 be the number of excluded endogenous variables. Then we have $K = K_1 + K_2$ and $G = G_1 + G_2$. The equation is identified only if $G_2 + K_2 \geq G - 1$. We will assume that this condition is satisfied. As a rule $G_2 + K_2$ is far greater than G - 1 so that the equation can be strongly overidentified.

Since Y_1 and the stochastic disturbances are correlated, the OLS estimators

$$(2) \qquad \hat{d}_{OLS} = (Z_1'Z_1)^{-1}Z_1'y_1$$

will be inconsistent. To obtain consistent estimators we use instrumental variables for Y_1. For IV estimation we require a T•K_3 (min $K_3 = G_1 - 1$) matrix W_1 of instruments to estimate

(3) $$\hat{d}_{IV} = (W_1'Z_1)^{-1}W_1y_1$$

The instrumental variables can be chosen in different ways. The statistical requirements for consistency are:

- IV variables are uncorrelated with the structural error terms
- IV variables are correlated with the predetermined variables. From this assumption it follows that the instruments are also correlated with the jointly dependent variables.

In simultaneous equation systems the predetermined variables provide a choice of instrumental variables. Thus, we consider linear combinations of the predetermined variables

$$W_1 = XA_1,$$

where A_1 is a $K \cdot K_3$ matrix, to form the matrix of instruments. A_1 can either be known or estimated as $\hat{A}_1$. It can be shown that among all the instrument matrices $W_1 = X\hat{A}_1$ formed by linear combinations of the predetermined variables, the best choice, that is, the one that minimizes the asymptotic covariance of d_{IV}, is $W_1 = X\hat{A}_1$, where

(4) $$\hat{A}_1 = (X'X)^{-1}X'Y_1$$

Theil (1961) defines a family of estimators called k-class estimators which are a generalization of the expression of $\hat{d}_{IV}$. This family includes all the commonly used simultaneous single equation methods.

Let $\hat{Y}_1$ be the estimated value of Y_1 obtained from the reduced form $\hat{Y}_1 = X(X'X)^{-1}X'Y_1$. Further, let $\hat{V}_1$ be the matrix of estimated residuals residuals obtained from the reduced form so that $Y_1 = \hat{Y}_1 + \hat{V}_1$. We define the instruments as

$$W_1 = [\hat{Y}_1 \; X_1]$$

where we now use $\hat{Y}_1 = Y_1 - k\hat{V}_1$. The k is taken as an arbitrary scalar, which could be stochastic or non-stochastic. For k = 0 we have OLS. For k = 1 we have 2SLS. For LIML k is a stochastic variable > 1.

Nagar (1959) has analytically shown that a sufficient condition for consistency is plim k = 1, which is also the condition for k-class estimators to be IV estimators. Efficiency does not, however, follow from the same conditions as consistency. To have the same asymptotic variance matrix as the efficient optimal IV estimators, k must satisfy the stricter condition

$$\text{plim } \sqrt{T}(k-1) = 0$$

Thus, to be an IV estimator, that is consistent, we require plim(k-1)=0; for efficiency we need plim$\sqrt{T}$(k-1)=0. The choice of

$$k = 1 + (K-K_1)/\sqrt{T}$$

eliminates bias. Given the fact that the conditions for unbiasedness and efficiency are not the same, it follows that the unbiased estimates have larger standard errors than the estimates corresponding to the efficient k.

In the following we only consider 2SLS and IV estimators. We omit the LIML estimators because 2SLS is an improvement over LIML in the sense that the computational burden is lighter, even if the asymptotic properties are the same for both methods. The definition of the k-class family is, however, introduced here because many results in estimation and other inference methods discussed below cover the whole class.

When estimating from a small sample where T < K, the matrix X'X that is used for solving the reduced form residuals is singular

and cannot be inverted. In that case, we cannot apply 2SLS or IV methods in their asymptotically efficient form. Even if $K < T$ the situation may nevertheless be difficult because the low degrees of freedom for estimating the reduced form will diminish the quality of resulting estimates. The usual procedure in the small sample case is to base the estimation on only a subset of the predetermined variables by leaving out of consideration some of the X variables that do not occur in the equation. This leads, however, to loss in efficiency and need not result in consistent estimates. There is no analytically derived unique way of choosing the subset of predetermined variables to be used as first stage regressors. The choice can be made in many ways. Here we consider three approaches:

- the blockdivision approach
- the use of principal components
- the iterative instrumental variable method.

The selection procedures have been considered in the literature mainly as approximations to 2SLS. The procedures can equally well be interpreted as alternative methods of defining $\hat{Y}$ in IV estimation. The essential criterion is the same; maximizing the vector correlation between $\hat{Y}$ and Y while preserving uncorrelatedness with the disturbance term. The following discussion of the selection techniques can therefore be thought to apply equally well to IV and modified 2SLS estimation. Hence we use the concepts of first stage regressors and instrumental variables as synonyms.

II.2.1 Blockdivision approach to 2SLS

The arbitrary choice of first stage regressors can be based on blocks into which the model has been divided. In estimation of the separate equations, all the predetermined variables in the block could be used as first stage regressors. This procedure yields consistent estimators if the X_1 variables are included among the first stage regressors. Asymptotic efficiency depends, as in

block recursive system estimation, on how close to blockdiagonal is the variance covariance matrix of residuals and how close to lower triangular is the matrix of the structural parameters (Theil, 1983). Most economic models are, however, highly interdependent with feedback between blocks and a blockrecursive form is not likely to be found.

II.2.2 Principal components

Kloek and Mennes (1960) have proposed a modified 2SLS method using principal components of the predetermined variables. The method results in consistent estimators if principal components of the predetermined variables are used together with the X_1 variables as first stage regressors. This method should also be efficient relative to other methods using a truncated number of first stage regressors. It is based on the known property that the first k principal components, computed on the basis of the correlation matrix (or variance matrix of normalized variables) of the predetermined variables, explain a larger percentage of the variation of the set of predetermined variables than any other set of k linear combinations of the same variables.

The use of principal components is often suggested as a solution to the multicollinearity problem. The explanatory variables are replaced by a few of their first principal components, which are orthogonal by construction. In this application of the principal component method, some problems arise. First, the principal components corresponding to the largest characteristic roots need not necessarily be the ones that are most correlated with the dependent variable, even though they capture the major part of the total variance of the explanatory variables. The other problem in this pure substitution of exogenous variables is that the substitute variables, the principal components, cannot be given any economic interpretation (Maddala, Econometrics 1977, p. 193). This problem of interpretation is not relevant in our case since we use

the principal components not as substitute variables but rather as auxiliary variables to be used only in the inital regression. In the empirical analysis, we shall see that the first problem is also irrelevant. The first principal components are also the main explanatory variables in the regressions.

Kloek and Mennes (1960) experimented with the characteristic vectors extracted from the correlation matrix of all the predetermined variables X, the excluded predetermined variables X_2 or the excluded predetermined variables orthogonal to X_1. They suggested computing the k principal components on the basis of the characteristic vectors corresponding to the k largest characteristic roots. However, they considered a modification of this rule when the principal components are extracted from the correlation (or variance) matrix of all the predetermined variables X. This modification is to take account of the possibility of high correlation between one or more of the principal components and certain predetermined variables in X_1. Instead of selecting the k components with the largest characteristic roots, they selected in this case the k components with the highest value of the statistic

$$Q_h = v_h(1 - R_h^2)$$

v_h being the characteristic root and R_h the multiple correlation coefficient in regressing the principal component on the X_1 variables.

When selecting k, the number of principal components to be used, there is no unique a priori known optimal number. There are, however, two limits between which k has to be chosen (Kloek and Mennes, 1960): the lower one is the obvious requirement $k > G_1-1$ (see page 21 for the notions). The higher limit called for from the degrees of freedom requirement is $k + K_1 < T$. In choosing k we have to consider the fact that if $k + K_1$ is growing towards T the estimators converge towards the OLS estimators. Usually, the small

sample bias is greater for OLS than for consistent estimates. The number of $k + K_1$ must therefore be less than T. On the other hand, efficiency is growing with growing k because the OLS estimators are known to have the optimal property of minimum variance (Schink, 1971, p. 216).

When the number of k grows towards $T - K_1$, the efficiency of the estimators is affected in two opposite ways. The correlation coefficient of the first stage regression is growing, thus increasing the efficiency of the second stage parameter estimators. On the other hand the falling degrees of freedom make the estimates unreliable, increasing their standard errors. The resulting effect on the parameter efficiency cannot be analytically derived but has to be established empirically. Empirical estimation results obtained by Klein (1969) and Fair (1973) indicate that the best number is not necessarily large relative to the number of predetermined variables.

II.3 The nonlinear case

In applying simultaneous methods in the estimation of the general model in (1) we have to solve two problems. The first problem, which arises from the limited sample size, is how to truncate the set of first stage regressors. This problem was discussed in the previous sections. The other problem stems from the nonlinearity of the model. As we noted above, an economic model is as a rule non-linear in variables. Because the structural form is linear in parameters, we can use linear single estimation methods. However, 2SLS estimation also requires knowledge of the reduced form, which is not necessarily linear.

The steps applied in 2SLS estimation for linear models are:

(i) designate one of the endogenous variables in each equation as the normalized one

(ii) obtain the reduced form

(iii) replace the values of the other endogenous variables by their values predicted from the reduced form

(iv) regress, using least squares, the normalized variable upon the new endogenous variables and the exogenous variables.

In the modified 2SLS methods reported above we have used the analogue of the procedure followed in linear models by assuming a linearization of the unknown reduced form at stage (ii). Some difficulties could arise in doing this (Goldfeld and Quandt 1968). First, we do not know the effect of the nonlinearity upon the error term in the reduced form. Step (iii) above can be justified by the need to eliminate the stochastic component in the other endogenous variables ($\hat{Y}_1 = Y_1 - \hat{V}_1$, see page 22). This may not be successful for nonlinear models because generally in this case the expected value of an endogenous variable calculated from the structure will not be the same as its expected value calculated from the reduced form. There are also other difficulties with estimation of the reduced form. Reduced form will not necessarily have a representation in closed form. Even when the reduced form can be expressed in closed form, it may not be linear in the reduced form parameters.

II.3.1 First stage regressors in polynomial form

To overcome the difficulties connected with the unknown reduced form, Goldfeld and Quandt (1968) proposed an approximation to the nonstochastic part of the reduced form by using an n-th degree polynomial in the predetermined variables representing the first n terms of a Taylor series expansion. Goldfeld and Quandt experimented with two variants of the method: the first, 2SLS1, using both first and second order polynomials, the other, 2SLS2, using a second order polynomial of the predetermined variables of the initial stage. They used Monte Carlo experiments to establish

the small sample properties, which we shall consider later on. Edgerton (1972) proved the consistency of the method by showing that the method is equivalent to the instrumental variable method.

II.3.2 Iterative instrumental variable methods

There is, however, another method avoiding the linear approximation of the unknown reduced form. This method also avoids the difficulties associated with the approximation of the set of first stage regressors. The method is consistent and, for linear models, also efficient. For nonlinear models it is efficient relative to other simultaneous methods applicable in undersized samples. It is a method which can be described as an iterative instrumental variable interpretation of the 2SLS. This method (IIV) is based on an iteration process using numerical values of the Gauss-Seidel solutions of the whole model as instruments. The method has been proposed for linear models by Brundy and Jorgensen (1971), Dhrymes (1971) and Dutta and Lyttkens (1974). Hatanaka (1978) has derived the asymptotic properties of the method extended to models nonlinear in variables.

Following Hatanaka the estimation proceeds as follows: In the case where the model is nonlinear in variables the general form of (1) will be:

$$f(Y,X)B+XC = u \tag{5}$$

where Y, X and u are defined as before and f(Y,X) fullfils the requirements of an endogenous function:

- its arguments contain at least one endogenous variable
- it contains no unknown parameters.

Let B,C and f be such that given X and u (5) yields one and only one solution to Y for each period

(6) $y_t = g(x_t, u_t, d)$

where d is the vector of parameter estimates.

The procedure for calculating the IIV estimator with OLS start is as follows:

(i) In the first round of the estimation method, the OLS method is used to construct the first estimates of the vector of the structural parameters, $\hat{d}^{OLS}$.

(ii) Obtain $y_t^{OLS} = g(x_t, 0, \hat{d}^{OLS})$ through (6) as a numerical solution to the model.

(iii) Obtain instrumental variable estimates for each equation separately using $f(y_t^{OLS}, x_t)$ and x_t as instruments. Let $\hat{d}^I$ be this estimate.

(iv) If there are sufficient observations, one more round can be calculated, in which the consistently estimated variance matrix of the previous round is implemented into the estimation. This stage is analogous to the use of a 2SLS estimation error variance matrix to perform 3SLS estimation.

Starting the procedure by using OLS estimators for the Gauss-Seidel solution is well suited to current practice in macro-economic modelbuilding, where the first identification search process is carried out with OLS estimation. Even if estimation starts from inconsistent OLS estimators, the iterative method yields consistent estimators for both linear (Dutta-Lyttkens) and nonlinear models (Hatanaka) because the estimators from (iii), and hence also (iv), fullfil the requirement of instrumental variables. Hatanaka also shows that iteration with OLS start does not result in consistent estimates in the case where lagged endogenous variables and autocorrelated disturbances coexist.

The method is also efficient for linear models. It takes into consideration the full specification of the model and is thus efficient relative to modified 2SLS and other IV methods. In fact, the third stage in the iterative process, the second instrumental variable estimator in the sequence, is asymptotically equivalent to the original efficient 2SLS estimator. In the linear case the estimators are shown to attain the Cramer-Rao lower bound for single equation estimators. Stage (iv) gives iterative estimators that are efficient among system estimators.

In nonlinear models the estimators are asymptotically inefficient owing to

- the discrepancy between the Gauss-Seidel solution, on the one hand, and the expectation of endogenous functions, on the other;
- the failure to take account of the nonlinear effect of the disturbance upon the deviation of endogenous functions from their expectations.

In contrast to the other single estimation methods the IIV method estimator is not strictly speaking a limited information estimator because information on the structure of the entire reduced form parameter matrix is used in computations (when the solved reduced form is used as instruments). This feature has special significance for the estimation method. We pointed out above that if $\hat{Y}$, the estimate of the right-hand endogenous variables, is obtained from the unrestricted reduced form, then $\hat{Y}$ can be used as either regressors (2SLS) or as instrumental variables and the resultant estimators of the parameters are the same. This is no longer true if $\hat{Y}$ is obtained from the Gauss-Seidel solution, which is to be interpreted as the restricted reduced form. In fact, following Maddala (1971) we can talk about the RRF2SLS (restricted reduced form two stage least squares) and RRFIV (restricted reduced form instrumental variables) estimators and the two will be different. The RRFIV is the same as the IIV estimator above. The RRF2SLS is

the estimator that Wold suggests under the name of the fix-point method (Wold, 1965).

The RRFIV method is shown to have the same asymptotic distribution as 2SLS but it is not possible to rank RRF2SLS in relation to 2SLS (Dhrymes & Pandit, 1971). The estimates of some parameters can be more efficient than those of 2SLS, and the estimators of some other parameters can be less efficient than those of 2SLS.

We can very clearly see in this reasoning the contradiction embedded in the use of analytical properties of small sample estimators. These iterative methods have been developed to cope with undersized samples, hence by definition producing small sample estimators. Yet the properties of the estimators are discussed in terms of asymptotic distributions. It is unclear how much the asymptotic arguments can be relied upon. Small sample approximations are required to better evaluate both the RRF2SLS and the RRFIV estimators. Also, knowledge about the effect of the initial estimator being consistent or inconsistent on the small sample properties of the estimators remains to be established.

This section concluded the theoretical comparison of the asymptotic properties of the estimators. Summarizing the results we have seen that all the simultaneous methods, both in the original and in the modified form, are consistent because they can be shown to be IV estimators.

The other desired analytic property is efficiency. If we make the assumption that the error terms have a joint normal distribution we can derive the Cramer-Rao lower bound for the asymptotic variance matrix of the parameter estimators. The methods that yield estimators attaining this lower bound and thus being efficient can only be applied in estimation from samples where the number of observations is greater than the number of the predetermined variables. For undersized samples we have to use modified simultaneous methods the relative efficiency of which cannot be

analytically derived. The only exception is the IIV method. This method yields estimators that are efficient compared to the other methods applicable to undersized samples.

This method is also intuitively appealing for another reason. When estimating from small samples it seems as though one has to pay for shorter data by being forced to use more complicated modifications of the usual estimation methods. Yet there is a contradiction in this payoff because the limited information in short data cannot cope with sophisticated methods. In the iterative method there is no payoff between data and method. The method utilizes in a very simple way all the information available in the data and the model.

II.4 Ranking of estimation methods according to exact finite sample distributions of the estimators

So far we have discussed the asymptotic properties of estimates, this being the usual way of method ranking. A rational choice among alternative methods of estimation from small samples should of course be based on knowledge of the small sample distributions of various types of estimators. The exact finite distributions are difficult to derive, so in comparison based on asymptotic properties the underlying assumption is that the same ordering is also valid in small samples.

Let us anyhow see what is known about the small sample properties of estimators. The finite exact distributions of estimators can be sought in two ways: using analytical derivation or sampling experiments. The analytical approach deals with the derivation of exact expressions for and approximation to the sampling probability distributions and moments of estimators. In the sampling experiment approach, usually covered by the name Monte Carlo procedure, artificial data are generated from which the sampling distributions and moments are calculated. In comparing the estimates from the Monte Carlo data, one estimation method is ranked superior to

another if the estimates of the population parameters produced by it are "closer" to the true value of the parameter than those produced by the others.

In the analytical derivation the formula for the exact or approximately exact value of bias is derived; in Monte Carlo experiments it can be calculated as the difference between the known true parameter value and its estimated value. Both analytical derivation and Monte Carlo simulation have their pros and cons so the results from the different derivations should be used in a complementary way.

The Monte Carlo procedure can only be applied to very small models to keep the sample generating work within manageable size. But these small models can be given complicated structures, which allows the measuring of the effects of different kinds of misspecification on the estimators. The weakness in the Monte Carlo procedure is that the results are tied to the model that is used in the experiment, i.e. they cannot be generalized. Moreover there is always the risk of a substantial sampling error that can be high depending on the number of replications. This defect is mainly due to a cost constraint that restricts the number of replications per experiment.

In the analytical approach the purpose is to attain results which hold for the whole parameter space, thereby giving this approach a generality that is impossible to attain with sampling experiments. The limitation of this approach stems, however, from the simplifications that have to be imposed on the structure of the model to allow the analytical derivation. The stronger the imposed assumptions are, the farther the derivation can be carried. The inclusions of identities, lagged endogenous variables, non-normal errors ets. do not effect the asymptotic results but are impossible to handle in the analytical derivation of finite sample properties.

In the analytical derivation a distinction has been made between the case of two and the case of three or more included endogenous

variables. The latter case, which can be interpreted as the general case, allowing full generalization, leads to very complicated formulas and so most results concern the former case (Mariano (1982)). Even for this limited case the expressions for the sample distributions are fairly complicated, involving a double infinite series closely related to the hypergeometric function. The formulas are quite difficult to compute numerically except in special cases. As a rule, there is no guarantee that they can be expressed in such a way that could enable comparison between different estimators.

One way of handling these complex expressions is to try to obtain approximations of them. The most common method is the Edgeworth approximation (Sargan, 1976). Even though it is an approximation, which by definition should imply a certain simplification, a typical Edgeworth expansion coefficient will be the sum of a large number of terms each of which is a complicated function of the whole set of parameters in the model.

II.4.1 The existence of finite moments

Most results for the case of simultaneous equations are, as pointed out above, derived for models limited to two jointly endogenous variables in the separate equations and for limited-information instrumental methods such as the k-class and the modified two-stage least squares (Mariano 1982, Sargan 1976, Phillips 1983, Handbook of Econometrics, pp. 451-516). In the following we will summarize the main results selected according to their usefulness from the point of view of this study.

Since it is very difficult to work out the actual distributions of the sample functions, a somewhat simpler approach is to attempt to calculate only some of the important properties of the distributions, specifically their moments.

Below we present the available results concerning the existence of moments for various estimators in the general case of a linear

system (following Mariano, 1982). The fact that some estimators from small samples do not possess finite moments means that the mean square error measure does not provide a basis for comparing the point estimates and the forecasts in evaluating the model's performance (Swamy, 1980).

Moments of positive order for estimated structural coefficients for a linear simultaneous system satisfying classical assumptions are finite up to order (see page 15 for the interpretation of the notions underlying the variables):

- $K_2 - G_1$ for 2SLS and 3SLS
- $r - K_1 - G_1$ for modified 2SLS where r is the number of linearly independent first stage regressors
- $N - K_1 - G_1$ for the k-class estimates with k nonstochastic and $0 < k < 1$
- 0 for k-class with nonstochastic $k > 1$
- 0 for instrumental variable estimators with non-stochastic instruments
- 0 for LIML and FIML.

According to these findings only OLS and modified 2SLS with at least one more principal component than joint endogenous variables and modified 2SLS based on block division of the model have finite moments of first and second order. That some estimators do not possess finite moments means that bias and RMSE are not meaningful parameters when the relative merits of alternative estimators are compared. According to Basmann (1961) the comparison should be based on the nonparametric treatment of MAE (or MAPE) and this point has been used in this study in the comparison of the empirically applied estimation methods.

II.4.2 The magnitudes of bias and relative efficiency

The derived expressions for the two first moments are functions of the triple (ρ, μ^2, v), where ρ is the correlation parameter between the structural error and the right-hand-side endogenous variable, μ^2 is the concentration parameter[3] and v is the degree of overidentification. Next, the relevant results for the magnitude of bias and the relative efficiency for the estimation methods applicable in this study (still following Mariano, 1982) are summarized. The results are valid for the limited case where the model includes two jointly endogenous variables:

- The direction of the bias in the k-class estimator (non-stochastic k ε $[0,1]$, that is, OLS and 2SLS) is the same as the direction of correlation, ρ, between the structural error and the endogenous regressors, Y_1. Negative correlation implies a downward bias; positive correlation implies an upward bias.

- The absolute bias is an increasing function of the absolute value of ρ, a decreasing function of the concentration parameter μ^2 and a decreasingly concave function of k. Thus, whenever both exist, OLS bias is always greater in absolute values than 2SLS bias.

- For 2SLS the absolute value of the bias is an increasing function of the degree of overidentification. Since 2SLS probability distributions depend on sample size only through μ^2 and since μ^2 increases with additional observations, the 2SLS bias in absolute values decreases with the inclusion of more observations in the sample.

---

[3]The term concentration parameter derives from the fact that as this parameter increases indefinitely, with sample size staying fixed, for k nonstochastic as well as for LIML, the estimated k-class parameter converges in probability to the true value d. (See page 15).

- The total effect of additional observations on OLS bias, on the other hand, is indeterminate. An increase in the sample size N produces a positive direct effect on the absolute OLS bias through the increase in the degrees of freedom and a negative indirect effect through the increase in μ^2.

- The larger the absolute value of ρ, the larger the OLS bias becomes in relation to 2SLS. The size of OLS bias relative to 2SLS also becomes larger with higher μ^2 or a lower degree of overidentification, v, or a bigger sample size.

- For 2SLS additional sample observations lead to a lower mean squared error. For the other fixed k-class estimators the net effect of increasing sample size is indefinite.

- Measured in relative magnitudes of MSE, larger values of μ^2 and large T favour 2SLS over OLS. In these cases the usual large-sample asymptotics take over and the dominant term is the inconsistency of OLS. In some cases with small values of ρ and T, OLS dominates 2SLS even for large values of μ^2.

- When the degree of overidentification becomes large, the 2SLS and OLS distributions tend to be similar. The higher the degree of overidentification the more regressors can be used in the first stage of 2SLS and the more similar the 2SLS and OLS distributions will be.

According to the above comments the usefulness of the derived results for small distributions of simultaneous estimates is limited. The results hold only for very small correctly specified models containing no lagged endogenous variables and they depend strongly on the unknown population parameters μ^2 and ρ. Some

guidance for the ranking of the estimation methods can, however, be found.

For very small samples OLS is in a strong position compared to the simultaneous methods. Despite its shortcomings in comparison to an asymptotic context, the OLS estimates are worth calculating along with the simultaneous estimates on the basis of small sample argumentation.

II.5 Analytical consequences of approximate specifications

In empirical work we have to make certain approximations; if perfection is the only acceptable objective no empirical work is possible. The asymptotic analytical properties of the estimators presented in the previous sections are valid only under the assumption of a correctly specified model.

In this section the asymptotic implications of misspecifications are discussed. First we discuss the question of whether simultaneous estimation is possible at all when the model size is necessarily always approximate. We then go on to discuss the implications of misspecifications for the estimation results within an approximate model.

II.5.1 Approximation of the size of a model

The consequences of approximation of the size of a model are discussed in the widest sense by Fisher (1963). He refers to two views, each representing a kind of extremity. The first one is that of Liu (1953). His argument goes as follows:

Econometric models are even at best only an approximation of reality. By necessity the modelbuilder abstracts from the real world by limiting the number of equations and variables. In fact, there are

always more variables in the "true" structure than are included in the approximative equations and the equations forming the model are not a complete set of equations. Thus outside equations exist, either relating variables within the model or relating them to unincluded variables. This view has three serious implications:

- first, the usual form of a priori restrictions used for identification - the restriction that some parameters in the structural equations are zero - is likely to be incorrect. The restrictions at best will only hold approximately and that is not good enough. If the variables in question really belong to the equations, then the equations are underidentified and not overidentified as is supposed in empirical work. Hence, the parameters cannot be estimated by any reasonable technique;

- secondly, the exclusion of other variables and equations from the approximate model means that the necessary order condition for identification is not satisfied either;

- thirdly, considering the complete system hardly any variable in the approximate model is truly exogenous. To include only the explicitly stated equations is to treat endogenous variables as exogenous. Once again it follows that underidentification is the usual case and overidentification the exceptional case.

All in all, the current techiques of estimation are entirely misplaced. Structural estimation is generally not possible in simultaneous systems and only reduced forms can be obtained.

The other objection to simultaneous systems referred to by Fisher is that of Wold (1953). According to Wold, the real world is not truly simultaneous at all, but the causation is unilateral and thus true systems are always recursive.

We can summarise these two opposing views as follows: Liu asserts that the world is too simultaneous to be estimated because of underidentification whereas Wold claims there is no simultaneity at all. Fisher takes a stand between these two positions. He does not deny the possibility that either Liu or Wold is right; the real world could be constructed as one of them describes it. But systems do exist in which simultaneity is present but not completely overriding. Thus, the usual kind of simultaneous estimation is entirely possible.

Fisher offers justification for the use of an approximate model by arguing that the "true" system can be interpreted as a self-contained set of blocks for which all the coefficients of omitted variables within the blocks are close to zero. This permits estimation of partial economic models which are in turn parts of models of the socio-physical universe. The proper question is not whether certain parameters are zero or not but whether they are in some sense sufficiently small. The problem is not of having underidentification if the restrictions hold and overidentification if they do not, but rather of having diminishing estimation inconsistency as the restrictions become better and better approximations. Similarly, the problem is not if the omitted variables really have zero coefficients or if variables assumed to be exogenous really are so, but whether these things are sufficiently so in an approximate sense.

Fisher postulates theorems according to which the estimation inconsistency following from the approximate specifications is negligible if

- all a priori restrictions are close approximations,
- omitted variables have small coefficients,
- the endogenous variables have negligible direct effects on the assumed exogenous variables, and
- the endogenous variables have negligible indirect effects on the assumed exogenous variables.

Fisher restricts his analytical derivation to limited information estimators like those belonging to Theil's k-class. The same derivation can be extended to full information maximum likelihood methods under the premise that misspecifications elsewhere in the system must be assumed to go to zero.

Fisher's theorems speak only of the implications of approximations for consistency. They have nothing to say about the efficiency of estimators nor about their small sample properties. It is likely that the asymptotic variances are different for the various estimators and almost certain that their small sample properties are different. The choice among estimators may therefore depend on their behaviour under approximative misspecification, even though negligible inconsistency is assumed for all. This aspect is discussed in section II.6 in the context of the results of Monte Carlo experiments.

II.5.2 Misspecification within an approximate model

Fisher delt with three sources of misspecification connected with the approximation of the size of the model: incorrect a priori restrictions, omitted variables and exogeneity. Sources of misspecifications within the model affecting the results of the estimation appear when the classical assumptions for the stochastic disturbances are violated. Departures from the classical assumptions are heteroscedasticity, autocorrelation and non-normal distribution. All three kinds of departures are known to cause inefficiency. Non-normal distribution affects only maximum likelihood estimation results but for all estimators the nominal significance levels and the powers of the ordinary F- and t-test statistics are incorrect because they are derived under the assumption of normality.

Autocorrelation could from a rigorous point of view always be interpreted as a sign of misspecification in the defined equations,

which should be removed. We should continue to improve the specification until a white noise distribution of the error term is achieved (Spanos, 1984). Serial correlation is caused by omitted variables or wrongly modelled causality between variables. The direction of causality, that is, the division of the variables into exogenous, endogenous and normalized ones, is established in the specification phase of model building and is thus beyond the scope of this study.

If model respecification does not remove the autocorrelation and the modelbuilder accepts the hypothesis of dependence between residuals as being inherent in the nature of the data generating process, there are a great many estimation methods developed for this type of equation. In the nonsimultaneous case there are among others the methods of Cochrane-Orcutt and Hildreth-Lu. In the simultaneous case many methods have been developed which seek to obtain consistent and efficient estimators even when autocorrelation and lagged endogenous variables appear together. (Sargan 1961, Amemiya 1966, Fair 1970, 1972), Dhrymes, Berner and Cummins 1974 and Hatanaka 1976).

So far we have dealt with the effects of approximation in terms of inconsistency and inefficiency. Multicollinearity is another quality affecting the properties of the estimates. If the explanatory variables move together in a sample, this leads to singularity, indeterminacy and nonfinite values of some parameter estimates. If there is perfect correlation between two or more explanatory variables the determinant of the matrix X'X vanishes and we get indeterminate parameter estimates. If the dependence is not complete the estimates can be calculated but it is not possible to detect the separate contributions of related variables to the explanation of the dependent variable. (Klein and Nakamura 1962). In their paper, Klein and Nakamura have analytically shown that 2SLS estimates are more sensitive to the presence of multicollinearity than the OLS estimates. (We also find empirical evidence for that in this study). Similarly, limited information estimates are more sensitive than are 2SLS estimates.

If the parameters are estimable then the problem of multicollinearity in a simultaneous model is not necessarily a major one. Users of econometric models are not often really interested in all the particular structural parameters by themselves; rather they are interested in the solution to the whole system. Correlation between some parameters can, thus, be accepted.

The reasoning in this section has been in very general terms. This is because exact measures of the effects of misspecification are difficult to derive analytically. In the next section, we examine results from Monte Carlo simulations, which is the only method for establishing the sensitivity of various small sample estimators to misspecification.

II.6 Monte Carlo experiments for deriving small sample properties of estimators

Monte Carlo experiments are used as a complementary method to analytical asymptotic and analytical small sample results. In this method the distributions of the estimators are empirically generated, which makes it possible to draw conclusions on the small sample properties in cases for which no analytical small sample results are possible to derive. Cases of interest are those with lagged endogenous variables or different kinds of misspecification. It is also possible to detect the effects of different sample sizes on the estimation results.

The studies of interest in this field are those by Wagner 1958, Nager 1960, Summers 1965, Goldfeld and Quandt 1968, Mikhail 1975 and Mozzami and Buse 1984. The experiments are performed using small models in which various kinds of misspecification can be implemented and their effects analysed. One experiment consists of generating error terms from a known, usually multinormal distribution, and then solving the structural equation for the endogenous variables. Then the resulting data series, conditional

on the generated error terms, known parameter values and known values of the exogenous variables, are used to estimate the coefficients of the equations using the estimation methods to be compared.

Wagner used two linear models, which differed only in the variance matrix of the random disturbances. Each model consisted of two equations; one being over-identified, the other just-identified. The model also contained a one-period lagged endogenous variable. Wagner examined the estimation methods LIML, OLS and IV. The IV method in this case has the same asymptotic properties as 2SLS. Wagner compared only the results for the first overidentified equation. Nagar extended the Wagner study by choosing the same model and same sample size of 20 observations to derive the results for 2SLS and also for the just-identified second equation.

Summers (1965) compared the results of FIML, LIML, OLS and 2SLS estimation of a two over-identified equation linear model. He also investigated the consequences of misspecification for these estimation methods. In the misspecification model, the second equation was just-identified instead of being over-identified. Different degrees of multicollinearity were also introduced. Summers used sample sizes of 20 and 40.

In their Monte Carlo study, Goldfeld and Quandt sought the answer to the question to what extent do the substantive conclusions derived from linear cases hold for nonlinear models. They used two models: one linear in logs, the other containing an endogenous variable of second order. They considered the estimation methods OLS, FIML, 2SLS1[4], in which they used only linear variables in the reduced form, and 2SLS2[4], with both linear and quadratic terms.

---

[4]See page 28 and Appendix III for definition.

Mikhail compared the estimation results of OLS, 2SLS, 3SLS, 3S/OLS, three-stage via OLS, LIML and FIML on a linear two equation model for sample size of 20.

The comparison of the results was made in terms of bias and variance of the estimated structural coefficients and also conditional prediction ability. For most cases it was apparent from the results that the choice of an estimatior depended very little on which of these variables was used. Different kinds of measures were used as criteria in the comparison, although the most frequently used was MMSE[5]. Bearing in mind the argument of Basman (1961) that, since the distributions of the various estimators do not possess finite first and second moments, bias and RMSE are not meaningful parameters, the nonparametric measure MAE[5] was also used. The results indicated, however, that the appraisal of the estimation methods was essentially the same when RMSE was used as a criterion instead of MAE.

In the following sections the results from the above-mentioned studies are put together.

II.6.1 The basic case

The same model was used in several experiments and it is therefore labelled here as the basic case. A sample size of 20 observations was used. This is the sample size that most annual models bases have to deal with. A low contemporaneous correlation coefficient, 0.18, between the residuals was implemented. The following results were reported:

- The small sample bias of OLS exeeds the bias of the consistent estimation methods. Ranking according to the

---

[5]See Appendix III for definition.

biases of estimated parameters puts 3SLS on top followed by LIML, FIML and 2SLS. The two last places were taken by 3S/OLS and OLS.

- OLS has smaller sampling variance around its biased expectation than the consistent methods. Ranking according to the estimated variances gave the order OLS, 3S/OLS, 2SLS, 3SLS and then FIML and LIML.

- Judging by the mean-square error criterion, 2SLS is best and 3SLS in second place. OLS and 3S/OLS ranked third and fourth. Their small variances helped to offset a substantial part of the biases. Due to the small magnitudes of the biases of the other methods, their ranking was identical to the ranking of variances.

- Compared to the true asymptotic standard errors, the asymptotic estimated standard errors of 2SLS, 3SLS and FIML gave a rather satisfactory picture of the variability about the true value. OLS standard errors measure the variability of the estimators about the biased expectation, not about the true value.

The results agree with the conclusion that, for this kind of models, single-equation methods would not be worse than full-information methods. According to asymptotic theory, full-information methods produce a gain in efficiency over the other consistent methods only if the contemporaneous correlations are not close to zero.

II.6.2 The consequences of high contemporaneous correlation

The effects of higher correlation between equations were investigated by generating the structural disturbances from a multinormal distribution with the correlation value 0.76. The following results were obtained:

- The biases of 2SLS, 3SLS, LIML and FIML were so close to each other that no significant differences among them could be detected. As before, OLS bias was greatest of all.

- According to the variance, it was found that in this case of relevant contemporaneous correlation between residuals, the single equation methods, which were well placed in the ranking in the basic model, now lose their good rankings to the theoretically preferable full-information estimators. OLS is relegated from first to fourth place and 2SLS from third to fifth place. The first three positions are taken by 3S/OLS, 3SLS and FIML, respectively.

- According the mean-square error measure, FIML is the best method. Similarly, 2SLS is shown to be best single-equation method followed closely by LIML.

The findings confirmed the expectations based on asymptotic theory. The very small correlation in the error term matrix in the basic case neutralized the effects of joint estimation and stripped the full-information methods of their advantage. The introduction of large covariances brought about the kind of behaviour that the system methods were expected to have in the light of large sample theory.

No attempt was made to determine the smallest value of g, the correlation between the disturbances in the equations which would

make it profitable to use full-information rather than single-equation methods, for the following reasons:

- The value of g favourable to one may not be such for another.

- Even if such a value for g could be worked out for each method, it would be subject to sampling error of considerable magnitude.

- The presumed g would be limited to the use of the particular model simulated in this experiment.

The results show for the high contemporaneous correlation case not only the superiority of FIML and 3SLS known from the theory for their consistency and asymptotic efficiency, but also that 3S/OLS with g's estimated inconsistently by OLS did better than all the consistent single-equation methods, being only marginally less efficient than 3SLS. This would perhaps suggest that a preliminary investigation of the correlation between the disturbances, although inconsistently estimated, would be worthwhile.

II.6.3 The effect of the sample size

The sample size used in the different experiments were 20, 40 and 60 observations. The findings were:

- OLS, compared to itself, measured in terms of bias, behaves best at the smallest sample size considered, 20, but does progressively worse as the sample size increases.

- For the consistent methods the bias and the variance decreases as sample size increases. For large samples the RMSE's are approximately proportional to the inverse of the square root of the sample size.

II.6.4 The effects of multicollinearity

Multicollinearity between the variables was introduced in the experiments to measure the effects on the estimation results:

- OLS fared significantly less badly in the experiments where the multicollinearity was greatest, which indicates less sensitivity for misspecification in the OLS estimation. The closeness to singularity of the matrix of the jointly determined variables may be an important conditional variable in appraising the methods.

- FIML performs well under favourable conditions, but is loses ground when misspecification such as high interdependence between variables or structural misspecification arise. This conclusion stands for the structural parameter estimates. The prediction performance is not much affected.

II.6.5 Non-normal or autocorrelated errors

Misspecification in terms of non-normal or autocorrelated errors was found to have the following implications:

- The effects of departure from normality of the error distribution on the first and second moments of OLS and 2SLS estimates are very slight. This finding illustrates the robustness of these estimates to non-normal error distribution (Knight, 1985).

- In cases of autocorrelated errors Monte Carlo experiments indicate that for small samples the method of least squares is best; OLS for low values of the autocorrelation coefficient r and autoregressive least

squares for large r (Hendry and Srba, 1977), 2SLS for large samples and small r. For large samples and large r an autoregressive instrumental variable estimator performs best.

II.6.6 Nonlinearity in the model

The experiments made with a nonlinear model showed that the conclusions obtained for linear cases also hold for nonlinear models.

- Only for the sample size of 20 did OLS perform well relative to nonlinear FIML and 2SLS2, but could not compete with 2SLS1.[6] The situation changed significantly for larger sample sizes.

- As in the linear case, the RMSE's and also the divergence between different consistent methods decrease as sample size increases.

- Based on the performance of the methods with respect to the RMSE's the following ranking was obtained: 2SLS1, FIML, 2SLS2 and OLS. The ranking between 2SLS1 and FIML is somewhat unclear but both beat 2SLS2 and OLS.

From this experiment we can draw the conclusion that the Taylor approximation of the unknown reduced form of a nonlinear model is better when both first and second order polynomials are used than when only quadratic variables are implemented.

[6]See Appendix III for definition.

III MEASURING THE UNCERTAINTY IN THE SIMULTANEOUS MODEL PREDICTIONS

In the previous sections we have dealt with different aspects of estimation as the first task in the evaluation of a simultaneous model. In the following, we focus on the main question of how to obtain a measure of dispersion for the forecast[1] values of a macroeconomic model. As noted above, the two parts, estimation and evaluation of predictive accuracy, are interrelated; the selection of a particular set of estimators determines the variance for the forecast and, on the other hand, the outcome of the derivation of the variance serves as a guidepost in the selection of the optimal estimator.

No one expects the forecast to be perfectly accurate because any model is, even at best, merely an approximation of reality. Above all, approximation contains the assumption of constant parameters. This implies the concept of no change in the economic structure. This assumption makes already the forecasts, when presented as mean values, deviate from the 'real world'.

The forecasts are thus expected to differ to some extent from the true values but they are also expected to fluctuate around their calculated mean values, the point predictors. There are four main sources of error causing the forecast to fluctuate. They are:

- the presence of random disturbances,
- the use of parameter estimates in forecasting as opposed to the unknown true values of the parameters,

[1] By the forecast we mean the values of the endogenous variables generated from the model solution over a time period not belonging to the sample estimation period (see Appendix III).

- the uncertainty of the values of the exogenous variables, and
- misspecification in the model

The two first sources make the model random. A forecast made with a random model is also a random variable and should also be reported as such. The ideal case would be to present the forecast along with the confidence interval calculated on the basis of the probability distribution and the variance of the forecast. In the following we present a summary of the existing derived results of estimation of the distribution and the variance of the forecast. We proceed according to the type of model examined, starting with the linear static case and concluding with the model for which no analytical results are possible to derive. It will be seen that prediction regions are impossible to derive analytically for dynamic and/or nonlinear systems due to the general non-normality of the endogenous variables. Thus, unlike linear systems, the distribution of the endogenous variables is not characterized by their first and second moments. In the nonlinear case we have to confine ourselves to measuring the dispersion of the endogenous variables by their second moments. Not even for this second moment can analytical results be derived but it is, however, possible to find an estimate for the variance matrix of the forecast error using stochastic simulation techniques which, as it turns out, are applicable equally well to the simpler linear nondynamic single equation models and simultaneous dynamic equation models.

III.1 Analytical results

In all the analytical derivations the exogenous variables are assumed to be known with certainty. The model is also assumed to be correctly specified[2]. Thus, in the analytical derivation the

[2]One of the stochastic simulation methods (Fair, 1980), to be presented later on, also deals with these two sources of forecast error.

components of the forecast error are only those that arise from the presence of the random disturbance in the equations and from the use of random parameter estimates.

III.1.1 Two-variable single equation models

We begin by presenting the following model so as to introduce the notions that will be used throughout the subsequent analysis.

The general form of this model is:

$$(1) \qquad y_i = a_0 + a_1 x_1 + u_i \qquad \text{for } i = 1,\ldots,N$$

where

y_i is the dependent variable

x_i is the non-stochastic independent variable

u_i are normally distributed random disturbances, terms with mean zero and constant variance σ_u^2. The normality assumption is necessary for the derivation of the variance of the various test statistics

a_0, a_1 are unknown constant parameters.

The best linear unbiased estimate, OLS, of a_0 and a_1 would be used in making forecasts of y_F (F being the forecast period) as follows:

$$(2) \qquad \hat{y}_F = \hat{a}_0 + \hat{a}_1 x_F$$

while the true but unobservable value of y_F is

$$(3) \qquad y_F = a_0 + a_1 x_F + u_F$$

where $var(u_F) = \sigma_u^2$.

The unobservable forecast error in period F is given by:

(4) $$e_F = y_F - \hat{y}_F$$

The variance of e_F is given by:

(5) $$\sigma_F^2 = \sigma_u^2 \left[1 + 1/N + \left((x_F - \bar{x})^2 / \sum (x_i - \bar{x})^2\right)\right]$$

and

$$\bar{x} = 1/N \sum x_i$$

An unbiased estimate of σ_F^2 is given by:

(6) $$\hat{\sigma}_F^2 = \hat{\sigma}_u^2 \left[1 + 1/N + \left((x_F - \bar{x})^2 / \sum (x_i - \bar{x})^2\right)\right]$$

where $\hat{\sigma}_u^2$ is an unbiased estimate of σ_u^2. This expression is a result of straightforward analytical derivation and can be found in any basic econometric textbook.

Expression (5) indicates that as the sample size N increases, σ_F^2 decreases and as $(x_F - \bar{x})^2$ increases, that is when forecasts are made for values of x further away from the mean $\bar{x}$, σ_F^2 increases.

The statistic

(7) $$t = (\hat{y}_F - y_F) / \hat{\sigma}_F$$

is distributed according to the Student "t" distribution with N-2 degrees of freedom. Consider the following probability statement

(8) $$\text{prob } (\hat{y}_F - \hat{\sigma}_F t_{a,N-2} < y_F < \hat{y}_F + \hat{\sigma}_F t_{a,N-2}) = 1-a$$

Since y_F is a random variable, the interval $(\hat{y}_F \pm -\hat{\sigma}_F\ t_{a,N-2})$ cannot be interpreted as a confidence interval in the sense that one would be able to state a confidence interval for an unknown parameter such as a_0 or a_1.

The proper interpretation of the interval defined by expression (8) is that a future value of y_F will be contained in the interval $(\hat{y}_F \pm -\hat{\sigma}_F\ t_{a,N-2})$ with probability (1-a) (Fraser & Guttman, 1956). Such an interval is called a beta-expectation tolerance interval with confidence level 1-a.

A more powerful probability statement would be that a given tolerance interval about y_F contained beta percent of the future values of y_F with probability 1-a. Such a tolerance interval is called a beta-content tolerance interval with confidence level 1-a. It has only been possible to derive this stronger statement for the simple two variable single equation model. Even this derivation obtains only approximative beta-content tolerance intervals because a chi-squared approximation to the normal distribution is employed in the derivation (Wilson A L, 1967). It is, however, possible to derive the beta-expectation tolerance interval for more complicated models as we shall see below.

III.1.2 Nondynamic simultaneous equation models

The general form of this model is given by

(9) $$YB + XC = V$$

where

Y	is an N x G matrix of dependent variables
X	is an N x K matrix of nonstochastic independent variables
B	is a nonsingular G x G matrix of unknown parameters
C	is a K x G matrix of unknown parameters
V	is a N x G matrix of structural disturbance terms.

A typical row of V, V_i, is distributed according to the multivariate normal ditribution with expected value zero and G x G variance matrix Σ_V. The rows of V are assumed independent of one another.

The derived reduced form corresponding to (9) is given by:

$$Y = XA + U \tag{10}$$

where

$$A = -CB^{-1} \quad \text{and} \quad U = VB^{-1}$$

A typical row of U, Ui, is distributed according to the multivariate normal distribution with expected value zero and G x G variance covariance matrix Σ_u defined by:

$$u = (B^{-1})' \; \Sigma_V \; (B^{-1}).$$

While the structural model given by (9) is the one estimated the derived reduced form model (10) is employed in making forecasts. Let the model (9) be estimated using some consistent estimation method yielding the estimates $\hat{B}$, $\hat{C}$ and $\hat{\Sigma}_V$. The corresponding consistent estimates of A and Σ_u are given by

$$\hat{A} = -\hat{C}\hat{B}^{-1} \quad \text{and} \quad \hat{\Sigma}_u = (\hat{B}^{-1})' \; \hat{\Sigma}_V \; (\hat{B}^{-1}). \tag{11}$$

Denoting the forecast period variables as before, the unobservable vector of forecast error, E_F, is given by:

$$E_F' = Y_F' - \hat{Y}_F' = X_F'(A - \hat{A}) + U_F'.$$

Define the K G x 1 vector a* as:

$$a^* = (A.'_1, A.'_2, \ldots, A.'_G)'$$

where the $A._i$ are columns of A. The vector $\hat{a}^*$ would be defined in

the same way as a* except on the basis of the columns of $\hat{A}$. Denote K G x K G variance matrix of the elements $\hat{a}^*$ by $\sum_{\hat{a}}^*$. In addition, define the G x K G matrix Z_F as:

$$Z_F = \begin{bmatrix} X_F' & 0 & & 0 \\ 0 & X_F' & \cdots\cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdot & \cdots & X_F' \end{bmatrix}$$

If a consistent estimate of $\hat{\sum}_{\hat{a}}^*$ can be defined, then a consistent estimate of the variance matrix of forecast error, $\hat{\sum}_F$, is given by:

$$\hat{\sum}_F = Z_F \hat{\sum}_{\hat{a}}^* Z_F' + \hat{\sum} u \qquad (12)$$

The difficulty in evaluating expression (12) is finding a consistent estimate of $\sum \hat{a}^*$. Since $\hat{A}$ is a nonlinear function of $\hat{B}$ and $\hat{C}$, $\sum \hat{a}^*$ is a nonlinear function of the elements in the variance-covariance matrix of the elements of $\hat{B}$ and $\hat{C}$. In the study by Goldberger, Nagar and Odeh (1961) an explicit technique is developed for obtaining a consistent estimate of $\sum \hat{a}^*$, and thereby a consistent estimate of $\hat{\sum}_F$ as in expression (12).

Goldberger et al. make the first-order Taylor expansion of expression (11) and determine an approximate linear relationship between the elements of $\hat{a}^*$ and the elements of $\hat{B}$ and $\hat{C}$. They also show that the approximation becomes more accurate as the sample size increases. Given an approximate linear relationship between the elements of $\hat{a}^*$ and the elements of $\hat{C}$ and $\hat{B}$, it is relatively simple to determine a linear relationship between the elements of $\hat{\sum} \hat{a}^*$ and the consistently estimated variance covariance matrix of the elements of $\hat{C}$ and $\hat{B}$.

Hyman (1969) has used the above cited results of Goldberger et al. to define the statistic

(13) $$T^2 = (\hat{Y}_F - Y_F)\ \hat{\Sigma}_F^{-1}\ (\hat{Y}_F - Y_F)'$$

which is asymptotically distributed as Hotelling's "T^2". Therefore, the statistic $((N-K-G+1)\ T^2/(N \cdot G))$ is asymptotically distributed as Snedecor's F with G and (N-K-G+1) degrees of freedom. Hymans then constructed an asymptotic beta-expectation tolerance ellipsoid for Y_F.

III.1.3 Linear dynamic models

The linear dynamic structural model is of the general form:

(14) $$YB + Y_{-i}\ B_{-i} + XC = V$$

while its corresponding reduced form is given by:

(15) $$Y = Y_{-i}\ A_{-i} + XA + U$$

where

$$A_{-i} = -B_{-i}\ B^{-1},\ A = -CB^{-1} \text{ and } U = VB^{-1}.$$

Y_{-i} is an N x G matrix of endogenous variables lagged i periods

B_{-i} is the corresponding G x G matrix of unknown parameters

The matrices Y, X, B, C, V, A and U are as defined before and the G x G matrix A_{-i} contains the unknown derived reduced form parameters associated with Y_{-i}.

The asymptotic expressions derived by Goldberger et al. presented above are valid for models which are static in the sense that the true observed values are used for the lagged endogenous variables.

Peter Schmidt (1977) has considered the asymptotic distributions of dynamic forecasts where the model itself generates the lagged values of endogenous variables. The results are derived for linear reduced form models only. In the calculation of the variance matrix of the forecast error, the variance matrix of the reduced form parameter estimates is also needed. It is obtained using the formulas developed by Goldberger et al.

The asymptotic covariance of the forecast error $(\hat{y}_F - y_F)$ is of the form $(1/T)\theta + \Delta$. The term $(1/T)\theta$ is due to errors in the estimation of the reduced form parameters: that is, it arises because we do not know the parameters of the model, but only have estimates of them. The term Δ, on the other hand, is due to the random nature of the variable y_F being forecasted and is invariant with respect to the size of the sample used to estimate the parameters of the model.

In the expression $(1/T)\theta + \Delta$, the term $(1/T)\theta$ can be ignored if we are interested only in asymptotic results. Under the null hypothesis of correct specification the test statistic

$$(\hat{y}_F - y_F)((1/T)\theta + \Delta)^{-1} (\hat{y}_F - y_F)'$$

converges in distribution to χ^2_G, where G is the dimension of y. Similarly, the test statistic formed by dividing any element of $(\hat{y}_F - y_F)$ by the square root of the corresponding diagonal element of $((1/T)\theta + \Delta)$ converges in distribution to N(0,1).

III.1.4 Nonlinear dynamic models

The nonlinear dynamic structural model is of the general form

(16) $f(Y_t, Y_{t-k}, X) = V \qquad \text{for } k = 1,\ldots,p$

where f is a G x G matrix valued nonlinear rational function with unknown parameters. In this study we concentrate on the derivation of an estimate of the variance matrix of forecasts made with models of the type (16).

It is assumed that the simultaneous equation system (19) yields one and only one solution for each period for relevant values of the coefficients, the predetermined variables and any value of the disturbance terms. Drawing an analogy with the linear model this solution would be expressed in the reduced form:

(17) $y_t = g(y_{t-k}, x_t, d, u_t) \qquad k = 1,\ldots,p.$

The reduced form may not be expressible in a simple closed form in this nonlinear case. In practice the model is solved iteratively in the normalized form.

Let F be a time period not belonging to the sample estimation period and let the model be used to forecast over the time interval (F+1, F+h). Given the values of the endogenous variables at time F, y_F, and the values of the exogenous variables in the forecast periods, $x_{F+1}, x_{F+2},\ldots,x_{F+h}$, the values of the endogenous variables in the forecast period can be obtained through the normalized form:

$$(18) \qquad y_{F+1} = g(y_{F-k}, x_{F+1}, d, u_{F+1}) \qquad k = 0,\ldots,p$$

$$y_{F+2} = g(y_{F-k+1}, x_{F+2}, d, u_{F+2})$$

$$\vdots$$

$$y_{F+h} = g(y_{F-k+h-1}, x_{F+h}, d, u_{F+h})$$

These are the true forecast values conditional on unchanged structural relationships, certain knowledge of the values of x_F, $x_{F+1},\ldots,x_{F+h}$, and "true" parameters d.

When forecasting with an estimated model we do not know the true parameter vector d. The usual way of forecasting is to insert in the normalized form (18) the values of the predetermined variables x_{F+1}, $x_{F+2},\ldots,x_{F+h}$, the estimated parameter vector $\hat{d}$, the expected values of zero for the disturbance terms u_{F+1}, $u_{F+2},\ldots,u_{F+h}$ and solve the model

$$(19) \qquad \hat{y}_{F+1} = g(y_{F-k}, x_{F+1}, \hat{d}, 0) \qquad k = 0,\ldots,p$$

$$\hat{y}_{F+2} = g(\hat{y}_{F-k+1}, x_{F+2}, \hat{d}, 0)$$

$$\vdots$$

$$\hat{y}_{F+h} = g(\hat{y}_{F-k+h-1}, x_{F+h}, \hat{d}, 0)$$

This forecast is then conditional on the correct specification of the model and no errors in the exogenous variables. The forecasts $\hat{y}_{F+1},\ldots,\hat{y}_{F+h}$ differ from the endogenous variables $y_{F+1},\ldots,y_{F+h}$ because the estimated values are used instead of the unknown vector d, and because of the existence of the random error terms $u_{F+1},\ldots,u_{F+h}$.

Then the observed forecast errors are as follows:

$$e_t = \hat{y}_t - y_t$$

where

e_t is an G x 1 vector (t = F+1,...,F+h)

The asymptotic expected value is expressed as

(20) $$\bar{E}(e_t) = \bar{E}(\hat{y}_t - y_t)$$

while the asymptotic expected value of the variance matrix of forecast errors is given by:

(21) $$\bar{E}(e_t\ e_t') = \bar{E}((\hat{y}_t - y_t)(\hat{y}_t - y_t)')$$

Even in the most straightforward case where the u_t are assumed serially independent and contemporaneously normally distributed, analytical derivation of expressions (20) and (21) is essentially impossible (Schink, 1971).

Even if it were possible to evaluate the expressions for (20) and (21) for the model (16), these asymptotic results would be of limited usefulness. In practical forecasting the forecaster would like to know what the expected forecast bias and variance matrix of forecast error are for sample size T. Moreover, the small sample analysis of (20) and (21) is unfortunately intractable analytically for a model such as (16).

The problem of constructing tolerance intervals for the forecasts of this kind of model must be postponed until more is known about the relevant small sample distributions.

III.2 Simulation techniques

In empirical applications for real world dynamic nonlinear models the analytical derivation of the forecast dispersion has to be replaced by an other method. Simulation techniques are used to generate the empirical outcomes of the unknown joint forecast distribution for the endogenous variables. From these generated data series both small sample and asymptotic estimates of the forecast bias and variance can be computed. The distributions themselves remain unknown because the two moments are not enough to sufficiently characterize the joint distributions. Thus the prediction regions are impossible to construct. The data can, however, be used to get an idea of what the marginal distributions of the separate endogenous variables are like.

There are four methods of simulation to derive a measure of the two first moments. The four methods are:

- Stochastic simulation and re-estimation (Schink 1971)
- Monte Carlo on residuals and coefficients (Fair 1980)
- Analytic simulation (Bianchi and Calzolari 1980)
- Residual based simulation (Brown and Mariano 1984)

The two first methods are based on Monte Carlo techniques, thus involving sampling from a known distribution. The analytic simulation is a combination of analytical methods and numerical simulation. These three stochastic methods have been empirically compared (Bianchi and Calzolari 1982) but so far no empirical results exist for the fourth one, the residual-based simulation method. This study will, to our knowledge be the first one to compare empirical results from this new Brown and Mariano method to the other stochastic simulation methods.

Brown and Mariano (1984) have derived the large sample properties through asymptotic expansions for the prediction in bias and variance of the deterministic predictor (19) and of the predictors

resulting from application of the Monte Carlo and the residual-based methods of stochastic simulation. The results imply that, asymptotically, the deterministic forecast is biased while the predictors from the simulation methods are unbiased. The analytical results also imply that among the simulation methods the residual-based method has asymptotically optimal properties. It seems, however, evident that its strenght lies in its applicability in small samples. In a recent paper (1985) the authors also carry out an analytical investigation of the finite-sample properties of the alternative stochastic predictors for the static non-linear model.

III.2.1 Assumptions and notations

The general form of a dynamic model is given in (16) on page 61. The general model can be either linear or nonlinear in variables. Real world macroeconomic models are as a rule nonlinear.

The simulation results are derived under the assumption of correct specification and consistent estimation.

The assumption of serial independence of the structural disturbances is indispensable for the derivation of the results. If serial correlation is detected then the structural model has to be transformed to leave correlation-free disturbances. The assumption of serially independent error terms enables not only the pseudo-generation of the proxies for the disturbances and the parameters, but also a decomposition of the vector of the forecast error.

When the error term is assumed serially independent, the vector of estimated structural parameters, which is obtained from an estimation procedure applied to the data of the sample period, is independent of the vector of the random error terms in the forecast period, which is again outside the sample estimation period. In

that case, it is possible to decompose the vector of forecast errors in the forecast period F+h, assuming exact knowledge of all the predetermined variables, into two independent components as follows:

$$
\begin{aligned}
(22) \qquad \hat{y}_{F+h} - y_{F+h} &= g(\hat{y}_{F-k+h-1}, x_{F+h}, \hat{d}, 0) - \\
&\quad g(y_{F-k+h-1}, x_{F+h}, d, u_{F+h}) \\
&= [g(y_{F-k+h-1}, x_{F+h}, \hat{d}, 0) - \\
&\quad g(y_{F-k+h-1}, x_{F+h}, d, 0)] \\
&\quad + [g(y_{F-k+h-1}, x_{F+h}, d, 0) - \\
&\quad g(y_{F-k+h-1}, x_{F+h}, d, u_{F+h})].
\end{aligned}
$$

The first component in this decomposition is due to the uncertainty arising from the use of estimated parameters instead of the unknown true ones. The second component is due to the random error term in the stochastic behavioural equations.

The forecast error in this decomposition is the sum of two random vectors: the first is a function of several variables, among which only the vector of estimated coefficients, $\hat{d}$, is random; the second is also a function of several variables, among which only the vector of structural disturbances, u_F, is random. From the assumptions of serial independence of the disturbance terms and of forecasting outside the sample period it follows that the two components of the forecast error are independent. Thus it is possible to analyse the two components separately, and, in particular, an estimate of the variance of the forecast errors can be obtained by summing the estimated variances of the two components (Bianchi and Calzolari, 1982).

The independence between the two components does not hold exactly if lagged endogenous variables are present among the predetermined variables. In this case the two terms are both functions of the

random lagged endogenous variables. The above considerations, however, is still conditional on a given value of the lagged variables (for example, it could be the historical value).

The decomposition of the forecast error can be extended from one-step (or static) simulation to the case of dynamic simulation producing conditional forecasts. The simulations must, however, always be done outside the sample estimation period.

In the following we examine how different simulation methods deal with the estimation of these two components of the forecast error.

III.2.2 Stochastic simulation and re-estimation

The method of stochastic simulation and re-estimation has been used to analyse the small sample behaviour of estimation methods when analytical investigation is difficult or impossible (Hendry and Harrison (1974), Mariano (1980), Mikhail (1972)). Its use for calculting the forecast errors of nonlinear econometric models is proposed and described in Schink (1971). He used it for calculating directly a small sample estimate of the variance of $(\hat{y}_F - y_F)$. In this method, the complete forecast error due to two error sources together is analysed.

The application of this method requires the specification of the distribution of the disturbance terms. It is assumed multivariate normal $u_t \sim N(0,\Sigma)$. Note that this assumption is required in the estimation phase only if maximum likelihood estimation is applied.

A consistent estimate of the variance matrix is calculated in the following way. Let us first look at the linear case. Let

$$Ay_t + Bz_t = u_t \qquad t = 1,2,\ldots,T$$

be a representative equation in a linear model in its structural form, where

y_t is the vector of the endogenous variables at time t

z_t is the vector of the predetermined variables at time t

u_t is the vector of the structural stochastic disturbances at time t.

A,B are matrices of the structural coefficients.

Furthermore the vectors u_t are assumed to be independent and identically distributed with a multivariate normal distribution with zero mean and a covariance matrix, $\sum u_t$, constant over time.

The estimated structural model is

$$\hat{A}y_t + \hat{B}z_t = \hat{u}_t \tag{23}$$

where $\hat{u}_t$ are the estimated residuals and

$$\hat{\sum} u_t = 1/T \left(\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' \right) \tag{24}$$

is a consistent estimate of the variance matrix of the structural equation.

The restricted reduced form, which is used in model forecasting, is

$$y_t = -A^{-1}Bz_t + v_t \tag{25}$$

where

$$v_t = A^{-1} u_t$$

is the vector of the reduced form disturbances at time t.

It is clear that

$$v_t \sim N (0, A^{-1} \sum A'^{-1})$$

so that a consistent estimate of the reduced form covariance matrix, Ω, is available as

$$\hat{\Omega} = \hat{A}^{-1} \hat{\Sigma}_{ut} \hat{A}'^{-1} \tag{26}$$

provided the matrix $\hat{A}$ is nonsingular.

If the model is nonlinear the direct transformation in (26) cannot be applied. In the nonlinear case,

$$f(y_t, z_t, d,) = u_t$$

d is a vector including all the structural parameters, because a clear distinction between the elements of A and B is not possible.

The explicit analytic expression for the reduced form

$$y_t = g(z_t, d, u_t) \tag{27}$$

is, in general, unknown. Nevertheless, the variance matrix of the reduced form (27) can be computed by simulation and the reported methods differ from one another in the way this simulation is performed.

In the stochastic simulation and re-estimation method we make random draws from the distribution $N(0, \Sigma_{u_t})$ and produce via model simulation the elements of the reduced form error matrix of the forecasts.

The method can be summarized as follows;

1. Let a consistent estimate of the variance matrix be denoted by $\hat{\Sigma}_{u_t}$, estimated according to (24). T vectors (one for each sample estimation period) of pseudo-random

error terms, with zero mean and a variance matrix equal to the available $\hat{\Sigma}_{u_t}$, are inserted into the system (16).

2. The system is solved over the sample period, keeping the structural coefficients fixed at their originally estimated values. The simultaneous solution provides pseudo-random values of all the endogenous variables over the sample period.

3. Each set of pseudo-random vector values for the endogenous variables is used, like a new set of data, to re-estimate, along with the values of the predetermined variables, the vector of the structural parameters of the model.

4. The generated coefficients are inserted into the model to produce, via deterministic solution, a vector of pseudo-forecasts over the forecast period (F+1, F+h).

The process is repeated for steps 1 to 4, say, N times. The sample variance of the generated vectors of the N valued pseudoforecasts for each endogenous variable is then the desired small sample estimate of the forecast variance.

This method can be applied with several variants. The re-estimation can be performed with different estimation methods. Moreover, when the model is dynamic, the possibility of different choices arises from the treatment of the lagged endogenous variables in the simulation. The simulation phase can be either static or dynamic and in the re-estimation phase the endogenous variables can be given their historical values or their simulation values can be used.

III.2.3 Monte Carlo on residuals and coefficients

Apart from the assumption of a known distribution of the disturbance term, this method also requires knowledge of the distribution of the estimated structural parameters. As we saw from the previous discussion above the estimation results, it is not possible to derive the small sample distribution of the parameters. For simpler models, as in the linear dynamic and the nonlinear static cases, the distribution of the estimated coefficients can be proved, under sufficiently wide conditions, to follow asymptotically a multinormal distribution. Therefore, the assumption of parameter normality is used for the parameter distributions in the Monte Carlo simulation method.

An estimate of the variance matrix of the asymptotically normal distribution is also required. In system estimation this matrix is a standard by-product supplied by the method. For limited information estimators Theil (1971) has proposed a formula for 2SLS estimates and Brundy and Jorgensen (1971) for IIV estimates. The formulas are naturally valid only asymptotically.

This method has been used by Cooper and Fisher (1974), Haitovsky and Wallace (1972) and Fair (1980). Fair has derived the application to also treat the uncertainty arising from a presumed misspecification of the model. He also stresses the possibility of comparison between empirical models provided by this method (Fair (1980)).

The variance of the second forecast component is simulated as follows:

1. G (the number of the stochastic equations in the model) vectors of pseudo-random numbers for each forecast period from the distribution N $(0, \sum_{u_t})$ are generated and inserted in (16).

2. The system is solved over the forecast period thus producing pseudo-random vector forecasts of all the endogenous variables.

The two steps are repeated, say, N times. The variance of the N pseudo-forecasts for each endogenous variable is then the estimate of the variance of the second component in the error decomposition in (22).

The variance of the first component is calculated as follows:

Let the available estimate of the covariance matrix of the structural parameters $\hat{d}$ be $\hat{\Sigma}_d$. Then the steps in the application are:

1. A vector of pseudo-random numbers from a multivariate normal distribution with mean $\hat{d}$ and covariance $\hat{\Sigma}_d$ is generated. These pseudo-random coefficients replace the original estimates $\hat{d}$.

3. The model is solved over the forecast period, obtaining the vector of pseudo-forecast for the endogenous variables.

The process is repeated, say, N times, and the estimate of the variance matrix of the element of $(\hat{y}_F - y_F)$ is calculated from the N vectors of pseudo-forecasts.

The variances of the two components are summed to give the variance of the whole forecast error $(\hat{y}_F - y_F)$. If we are not interested in the separate error components, only in the total variance, the simulation is made by inserting in the model at the same time random numbers for the error term and for the parameter estimates. The total variance is obtained directly as the variance of the pseudo-random forecasts from the model simulations.

III.2.4 Analytic simulation

This method requires the assumptions of knowledge of the distribution of the random error term, of the distribution of the parameters and an estimate of the variance matrix of the structural parameter estimates. In the previous simulation methods, the methods of Schink and Fair, the errors from the stochastic disturbance term and the random nature of the parameter estimates were calculated together. In the analytic simulation methods the forecast errors of the two components, as presented in (22), are always calculated separately.

In the analytic simulation method the computation of (26) is based on a nonexplicit linearization of the model in the neighbourhood of the solution point corresponding to the period F under consideration. From equations (25) and (26) it is clear that the elements of the matrix A^{-1} are the partial derivatives of the endogenous variables with respect to the elements of the vector u_t. These derivatives can be computed via numerical solution and stored in a matrix $\hat{D}_t$. The reduced form covariance matrix (Ω) at time t can be computed

$$\hat{\Omega}_t = \hat{D}_t \hat{\Sigma}_{u_t} \hat{D}_t' \tag{28}$$

The steps are as follows:

1. A deterministic solution is computed at time t with all u_t set to zero.

2. A value Δu_i is assigned to the disturbance of the first stochastic equation, all the others being still zero, and the model is solved again at time t.

3. The second step is then repeated for all the structural stochastic equations and the differences between the

disturbed solutions and the control solution divided by the values adopted for Δu_i.

$$\Delta y_i / \Delta u_i$$

supply the numerical values of the elements of the matrix of the partial derivatives in D_t. The diagonal elements of the matrix

$$\hat{\Omega}_t = \hat{D}_t \hat{\Sigma}_{u_t} \hat{D}_t'$$

where $\hat{\Sigma} u_t$ is computed as in (24), are the variance of the forecast error due to the random disturbance term in the model.

In the linear case this is an alternative to using equation (26) directly. In the linear case the matrix in (28) is a constant but in the nonlinear case $\hat{D}_t$ will be time-varying. When this method is used for dynamic forecasting, the derivatives in the matrix $\hat{D}_t$ have to be calculated separately for each period t.

The advantage in also using analytic simulation for a linear model as against using (26) directly is that the model is difficult to express in the form of (23), which form is necessary for the inversion of the A matrix. For medium and large-size models, the Gauss-Seidel iterative algorithm for the solution is expressed in a form where each equation is normalized with respect to different endogenous variables. It is then not possible to distinquish the elements of the A matrix from the elements of the B matrix. In such a case it is much easier to compute the $\hat{D}_t$ ($= \hat{A}^{-1}$) matrix by numerical simulation.

As far as the first component is concerned, its variance matrix can be computed by means of a similar linear approximation which is, in many cases, asymptotically exact. If we assume that, as T, the sample size, increases asymptotically

(29) $$\sqrt{T}(\hat{d} - d) \sim N(0, \Sigma_d),$$

and define G_{F+1} as the (M x s) matrix of first order partial derivatives of the vector of the functions g with respect to the elements of d, computed at the point $(y_F, x_{F+1}, d, 0,)$, then, asymptotically,

(30) $$\sqrt{T}\,(g(y_F, x_{F+1}, \hat{d}, 0,) - g(y_F, x_{F+1}, d, 0,))$$

is distributed as $N(0, G_{F+1} \Sigma_d G_{F+1}')$.

An estimated variance matrix of a multivariate distribution, which approximates the variance of the first component of the forecast errors, is obtained by calculating G_{F+1} at the point $(y_F, x_{F+1}, \hat{d}, 0,)$, replacing the Σ_d with the available estimate $\hat{\Sigma}_d$, and dividing $G_{F+h} \hat{\Sigma}_d G_{F+h}'$ by the actual length of the sample period T. This approximation is asymptotically exact if the functions of the vector g are continuously differentiable (Rao, 1973 p. 388) and if the estimated structural coefficients are consistent and asymptotically normally distributed. There is no formal proof for the condition of normality to hold for nonlinear models, so the procedure should be considered approximate, not only for small samples, but even in the large sample case.

Continuity and differentialability of the elements of the unknown vector of the reduced form functional operators y is ensured by the implicit function theorem, which also provides a way of computing the partial derivatives

(31) $$\frac{\partial y}{\partial d'} = - \left(\frac{\partial f}{\partial y'}\right)^{-1} \frac{\partial f}{\partial d'}$$

where the derivatives of the structural form operators, vector f, which is known, can also be analytically computed, once a

deterministic solution of the model at time F+1,...,F+h is known.

For medium- or large-scale models it may be more convenient to perform the above derivations with numerical methods rather than analytically. Finite differences, Δd, are then inserted into the model (16) and the corresponding differences in the endogenous variable, dy, from the model solution over the forecast period are calculated. The ratio $\Delta y/\Delta d$ is an approximative value of the derivative when Δy are the differences in the endogenous variables between a control solution and the disturbance solution with the increments da. The derivatives $\Delta y/\Delta d$ are stored in a matrix $\hat{G}$. In that case the diagonal elements of the matrix

$$\hat{G} \hat{\Sigma}_d \hat{G}'$$

are the variances of the first error component in (22).

III.2.5 Residual based simulation

This method was put forward by Brown and Mariano (1984). The method only takes account of the second component of the decomposed forecast error in (22). Empirical studies have, however, revealed that the second component covers the greater part of the total error (Bianchi & Calzolari, 1982).

The residual-based procedure requires no knowledge of the distribution of the residual term in the model. Only the assumption of no serial correlation is necessary. Nor is the variance matrix of the parameter estimates required. Thus this method avoids the need for approximate specification because it is characterized by low parametrization. In the Monte Carlo methods, presented above, where assumptions about the unknown distribution terms are required, there is a risk that if the wrong distribution is used to generate random draws, bias could be introduced in the predictions.

There is then an asymmetry between the estimation phase and the Monte Carlo prediction phase because the application of estimation methods giving desirable properties for the estimators of models like (16) does not need precise specification of the error distribution.

The Brown-Mariano method reduces the computational burden in the calculations and also the misspecification sensitivity inherent in the Monte Carlo predictor. This new stochastic predictor simply uses the calculated sample period residuals as the stochastic proxies for the disturbance term rather than random draws as in the Monte Carlo predictors above. The method then gives as many pseudo-forecasts as there are observations in the estimation period. The variance of these T forecasts is then the estimate of the forecast error variance.

III.3 Asymptotic properties of the stochastic predictors

The forecasts obtained from estimated real-world macromodels are usually generated as deterministic simulations of models in which structural disturbances are replaced by their expected values. The deterministic solution has, however, two disadvantages. First, because models are as a rule nonlinear, nonlinearity induces bias into the deterministic solution, since expected values of nonlinear functions are not in general equal to the nonlinear functions of the expected values of the random variables:

$$E\ f(x,u) \neq f(x,\ Eu)\ ^{3}$$

---

[3]There are exceptions. For example, if f is an odd function of u, then the deterministic predictor would be asymptotically unbiased. Also, if f is monotonic in u, then the deterministic predictor would be a consistent estimator of the median of y_F. (Brown and Mariano 1983).

This bias is called the simulation bias and does not exist for linear models.

Secondly, the deterministic solution takes no account of the random nature of the model, giving only point predictors.

The use of stochastic simulation procedures in forecasting with nonlinear models takes account of both the nonlinear and the stochastic nature of the models.

In the following, we first examine asymptotic behaviour before going on to consider the finite sample properties of the predictors in a nonlinear simultaneous system.

III.3.1 Asymptotic bias, AMSPE and variance

Brown and Mariano (1983, 1984) have derived the asymptotic expressions of the forecast bias, mean square prediction error and variance for the second component in the decomposition of the error term in (22), for the deterministic predictor, the Monte Carlo predictor (Fair, Bianchi and Calzolari) and their own residual-based predictor. The results are then conditional on the estimated parameters, the exogenous variables and a correct specification of the model. The expressions are composed of the leading terms of the Taylor expansion of the forecast error. The expansion requires the following assumptions:

- The model contains no lagged endogenous variables, an assumption which makes the forecast static so that the results also cover multiperiod forecasts.

- The two first moments of the predicted values are assumed to be finite for relevant parameter estimates, given the the values of the exogenous variables.

- The stochastic disturbances are assumed to be mutually independent and identically distributed with mean zero and a known variance matrix.

Under these assumption the Taylor expansion leads to following conclusions:

1. Asymptotic bias

- The deterministic prediction based on consistent parameter estimates, $\hat{d}$, is in general asymptotically biased due to the nonlinearity in the system. The asymptotic bias is of order $O(1)$ as $T \to \infty$.

- If the model is correctly specified, the Monte Carlo predictor and the residual-based predictor, both based on consistent estimates, have an asymptotic bias of order $O(1/T)$. They thus share the property of being asymptotically unbiased.

If the distribution of the disturbance is incorrectly specified, the Monte Carlo predictor becomes biased but the residual based predictor remains unbiased, provided that the functional form is correctly specified and consistently estimated. The residual-based predictor does not depend on any specific error distribution and is therefore less sensitive to distributional assumptions than the Monte Carlo predictor.

2. Asymptotic mean square prediction error (AMSPE)[4]

When the estimation of d is based on the maximum likelihood method (MLE), we can derive an expression for the lower bound of the AMSPE:

---

[4]See appendix III for definition.

- Because of its nonvanishing asymptotic bias, the deterministic predictor is dominated in terms of AMSPE by both the Monte Carlo and the residual-based predictor in large samples.

- The Monte Carlo predictor approaches the lower bound in the MLE case when S, the number of the stochastic replications, grows large relative to T. However, little is to be gained in increasing replications beyond a moderate number since the consequent reduction in AMSPE is at most a small fraction of the total.

- For S substantially larger than T when the Monte Carlo predictor approaches the lower bound of AMSPE, the Monte Carlo predictor is asymptotically efficient relative to the residual-based predictor.

- For S = T, T being the maximum number of replications possible for the residual-based procedure, the residual-based predictor is asymptotically efficient relative to the Monte Carlo predictor.

But although it is possible to order the Monte Carlo and the residual-based predictors according to their AMSPE in the cases S = T and S >> T, the differences in AMSPE are very small relative to the total AMSPE of either predictor.

3. Asymptotic second moments

As in the comparison of AMSPE, the comparison of the second moments is possible only between the stochastic predictors; the deterministic predictor does not provide estimates of the inherent variance or distribution of the forecasts of the endogenous variables.

As in the comparison of the AMSPE, the comparison of the second moment has to be made in the framework of maximum likelihood

estimators because it provides a measure of the lower bound of the second moment. In the MLE context the following results are obtained:

- The residual-based estimators of the second central moments are consistent.

- When S = T, the residual-based predictor is efficient relative to the Monte Carlo estimator in terms of the second moment, as it was in terms of AMSPE:

- For S sufficiently larger than T the Monte Carlo estimator becomes efficient relative to the residual-based predictor.

- Unlike the results of the comparison of the AMSPE measure, the inefficiency in terms of variance is not necessarily a small fraction of the total.

It would be of major interest to use the estimated variance matrix obtained either by the Monte Carlo method or by the residual based procedure to construct prediction regions centred on the mean of the replicated stochastic simulations. Unfortunately, the probability content of such a prediction region cannot be approximated using asymptotic normality since the leading term in the expansion of the forecast error is not generally normally distributed. Although we could try to use directly the pseudo-forecasts from the stochastic simulations and perhaps accept the hypothesis of their normality, we still lack knowledge of the joint distribution of the predictors from the model as a whole.

III.4 Finite sample properties of the stochastic predictors

The finite sample results are investigated in Mariano and Brown (1985). The results are very limited. They do not cover relative efficiencies of the stochastic predictors with respect to the deterministic predictor, neither do they provide an answer as to the relative magnitude of bias and efficiency between the stochastic predictors. The only analytical results are that the small sample biases and efficiency differ between methods but the relative ordering is ambiguous.

III.5 The simulation error

As we saw above, the deterministic prediction is biased whereas both the Monte Carlo predictor and the residual-based predictor are asymptotically unbiased. Thus we can use the deviation of the deterministic solution from the mean of the replicated stochastic solutions as a measure of the simulation bias. It has also been suggested that this difference could be used as a measure of the stochastic importance of the nonlinearity in the model (Fair, 1980, Bianchi and Calzolari, 1983). A sizable discrepancy has two clear implications: First, forecasts and policy analysis should be conducted via stochastic simulation, rather than deterministic simulation, which of course increases the computional burden. Secondly, many estimation methods rest on the use of conditional expectations of the endogenous variables, and if these are not correctly calculated in the nonlinear model, the properties of the estimation methods will be adversely affected.

Empirical examinations have, in fact, revealed (Fair, 1980, Calzolari and Bianchi, 1983) relatively little difference between the deterministic solution and the mean of replicated stochastic solutions. These findings have led to the conclusion that nonlinearity is not of major concern. Salmon and Wallis (1982) argue, however, that this is a dangerous position to make for a number of reasons:

First, the specification of most models is biased towards linearity through the statistical techniques used in both estimation and specification tests. The estimation method used in the specification phase of the econometric model is mostly OLS. Through its neglect of the simultaneity in the model, the OLS estimator also ignores the stochastic importance of nonlinearity. Further, most of the common summary statistics are in effect linear measures: R^2 is a measure of linear correlation and the Durbin-Watson statistic tests a linear first order autoregression. Hence their use also leads to models that are biased towards linearity.

A second reason for believing that this approach to the measurement of nonlinearity is not correct is pointed out by Mariano and Brown (1983). In the asymptotic expansions they refer to for the first two prediction moments of a nonlinear model, the leading term in the asymptotic prediction bias in the deterministic solution is decomposed into two terms. One is due to possible inconsistencies in the parameter estimates and the other due to nonlinearity. The leading term for the stochastic simulation prediction bias depends only on potentially inconsistent parameter estimates. Thus, from this point of view, the neglect of simultaneity in a deterministic solution is of the same order of magnitude as the neglect of the stochastic importance of nonlinearity.

Thirdly, another point that follows from the work of Mariano and Brown is that, if inconsistent parameter estimates, like OLS, are used when comparing deterministic and stochastic simulations, then it is impossible to separate the effects of nonlinearity from the effects of the use of inconsistent parameter estimates. Thus, the fact that the empirical simulation experiments may have found few differences between deterministic and stochastic simulations may either imply that the stochastic effects of nonlinearity are weak or, as has usually been the case, that inconsistent parameter estimates have been employed.

Brown and Mariano also suggest that the use of nonstochastic simulations in specification searches may lead the model builder away from the true specification to one that performs better in deterministic simulations.

III.6 Stochastic simulation methodology

Stochastic simulation can be performed in many different ways.

III.6.1 Number of replications

The three kinds of methods, Monte Carlo simulation, analytic simulation and residual-based simulation, differ from one another in the number of required simulations. In the residual-based simulation, the maximum number of simulations that can be made is the number of observations in the estimation period. Thus, the number of simulations is for most models something between 20 and 40. The analytic simulation requires for estimation of the second error term as many solutions as there are stochastic equations in the model and for the estimation of the first component of the error term as many solutions as there are structural coefficients to be estimated in the model. By means of the Monte Carlo procedure, either on the residuals or the coefficients, there is no a priori given number for the required simulations. The accuracy of the estimated variances increases with the number of replications, the estimates being asymptotically exact. The analytic simulation procedure, on the contrary, is not exact because it involves, via linearization, a systematic approximation.

Bianchi, Calzolari and Corsi (1979) have made an empirical study to check the effects of the number of solutions chosen for the Monte Carlo simulation against the effect of the linearization in the analytic simulation method. They computed the variance of the second error component of the nonlinear Klein-Goldberger model,

estimated with 2SLS with 4 principal components, using both the analytic simulation procedure and by means of the stochastic simulation approach after 50, 500, 5000 and 50000 replications. The results are displayed in table XXI for the main variables of the model. The results give an idea of the great accuracy of the analytic simulation method. The same accuracy is found for 17 replications in the analytic simulation as for 50000 replications in the Monte Carlo experiment.

III.6.2 Antithetic variates

There are ways of using variance reduction techniques instead of direct sampling in the Monte Carlo methodology. The most common technique is to reuse the known random numbers, either directly or after transformation. This reduces the number of simulations thus economizing on the generation of the random numbers as well as reducing their variability.

Following Hendry (1984) we can describe the technique of using antithetic random variates as follows:

TABLE I

Klein-Goldberger Model: Reduced-form standard errors at 1965

Variable		Standard Errors				
Name	Computed Value	Stochastic simulation Number of replications				Analytic simulation
		50	500	5 000	50 000	
Cd	55.33	2.78	2.48	2.44	2.42	2.42
X	530.1	9.05	8.44	8.52	8.54	8.53
W	310.8	5.24	4.77	4.73	4.77	4.78
Pc	41.97	6.44	6.21	6.16	6.11	6.11
p	1.225	.040	.035	.036	.036	.036

Cd = consumption of durables
X = gross national product
W = wages and salaries and supplements to wages and salaries
Pc = corporate profits including inventory valuation adjustment
p = implicit GNP deflator

Let $\tilde{b}$ and b* be two unbiased estimates for an unknown parameter b such that the "pooled estimator" $\bar{b} = 1/2\ (\tilde{b} + b^*)$ has the expectation $E(\bar{b}) = b$ and variance

$$(32) \qquad V(\bar{b}) = 1/4\ (V(\tilde{b}) + V(b^*)) + 1/2\ \mathrm{Cov}\ (\tilde{b}, b^*)$$

In direct random sampling $\tilde{b}$ and b* are based on independent sets $\{u_i\}$ so that Cov (.) = 0 and $V(\tilde{b}) + V(b^*) = 4V(\bar{b})$. When the random draws u_i are known it may be possible to select pairs which offset each others variability, that is, they are antithetic. For example $\{u_t\} \sim N(0, \Sigma_u)$ and $\{-u_t\}$ are perfectly negatively correlated. Basing $\tilde{b}$ on one and b* on the other of an antithetic pair can

include a negative covariance in many cases thus reducing the variance in (32). In dynamic models it has proved difficult to locate antithetic transformations which generate negative covariances between estimators. In certain cases only the number of simulations is doubled, but nothing is gained in variance reduction.

Calzolari (1980) made use of antithetic variate sampling on the same non-linear Klein-Goldberger model in estimating the variance of the second component. By comparing the standard error for the forecast estimated by means of direct sampling and by means of antithetic sampling, it appeared that the gain in efficiency due to the antithetic variates varied for different endogenous variables in the model from 500 to 50000. That is, it would be necessary to perform between one and one hundred million simple random replications (depending on the variable) in order to obtain the same accuracy in the estimated standard error as is obtained from 1000 pairs of simulations which make use of antithetic variates.

It is, however, noticeable that the findings of Bianchi and Calzolari (1982) show that no significant changes in the expected values of the forecast errors were observed after the first 40 or 50 replications in Monte Carlo simulation. Only the variances of the pseudo-forecasts were reduced by increasing the number of simulations or using antithetic sampling.

III.6.3 Blockdiagonal variance matrix

The use of antithetic variates in the previous section is one way to reduce the number of random sampling and model simulations in the Monte Carlo method. There is another way of simplifying both the analytic simulation and the Monte Carlo method. It involves the construction of the variance matrix of the structural parameter estimates. This matrix is used in the computation of the variance of the first error component.

The variance matrix of the structural disturbances, which is needed in calculations of the second error term, is easily computed from the estimated residuals as (24). But the asymptotic variance matrix of the structural coefficients is directly available only if a full-information estimation method is used. As we have seen above, it is rather difficult to apply system estimation to real-world macroeconomic models because the estimation has to be made from undersized samples. If a consistent single equation estimation method is used, only a blockdiagonal variance matrix is supplied by the method. Additional computations must be performed to obtain covariances between coefficients of different equations as the off-diagonal blocks. Theil (1971) has proposed a formula which can be used to compute these off-diagonal covariances in the case of 2SLS estimation.

The difficulty in applying system methods and the burden involved in the above mentioned additional computations for single-equation methods have raised the question of the importance of the contribution of the covariances to the standard error of the forecasts.

Bianchi, Calzolari and Corsi (1980) have made a comparison for three real-world models to evaluate the effects of the elements in the off-diagonal blocks on the estimated variance of the first error component. As in the case of the study of the effects of using antithetic variates, the study is purely empirical so that the results should not be generalized.

The calculations are made using three models with different degrees of non-linearity and different dimensions. They all give the same empirical evidence in the comparison.

The first model is the linear Klein I model with the complete variance matrix of size 12 x 12. The block-diagonal matrix consists of three blocks of dimension 4 x 4. The second model is the nonlinear Klein-Goldberger model with 54 structural coefficients.

The complete matrix is then 54 x 54. The third model is the nonlinear ISPE model of the Italian economy. The full matrix is of size 75 x 75.

The comparison for all three models between the standard errors computed with the full variance matrix and with the block-diagonal matrix indicate that very minor differences exist between the two cases. The models used cover a wide class of econometric models, which perhaps suggest that the computation of the standard errors could be based on a block-diagonal interpretation in cases when no information is available on the covariances of the structural coefficients between equations, or the computation of them is difficult.

III.7 Comparison of empirical results

Only the first three methods have been applied empirically and the results have been compared for various models. Bianchi and Calzolari (1982) performed some experiments on a set of small, medium and large size real-world models, both linear and nonlinear, comparing the results and performance of these three methods. They based the comparison on the estimated variance of the static forecasts one period ahead. In a later paper (May 1982), the authors extended the comparison to multiperiod dynamic simulations of the same real-world models.

In the static comparison only the first component of the forecast error was considered. The conclusions concerning the equivalence or nonequivalence of the methods are therefore based on the different treatment given to the errors in the estimated coefficients. In estimating those, both the Schink and the Fair methods use the Monte Carlo technique, thus having inherent the risk of non-convergency connected to this method.

The differences between the methods can be summed up as follows:

- The method of stochastic simulation and re-estimation by Schink does not require knowledge of the variance matrix of the structural coefficients. The moments of the coefficients are generated empirically through the simulations.

- The Fair method, Monte Carlo on coefficients, requires knowledge of the distribution of the estimated coefficients and the technical availability of a pseudo-random numbers generator from such a distribution.

- The third method, analytical simulation, requires the assumption of asymptotic normality of the estimated coefficients. The forecast variances are then computed using the assumed distribution of the coefficients, together with partial derivatives which are numerically calculated through successive simulation of the model.

Although the methods differ technically, Bianchi and Calzolari found an approximate equivalence of results for one period forecasts in all cases in which Monte Carlo converged; in other words, no Monte Carlo experiment ever converged to values which differed substantially from those produced from the analytic simulation method. The only cases in which large differences occurred were those in which Monte Carlo did not converge anywhere. The results, therefore, indicate that the nonconvergency of Monte Carlo, due to some matrix to be inverted being close to singular, could be the only source of large differences in the results produced by the three methods for models actually used for forecasting purpose. This is, however, not true for some models when passing from one-period static forecasts to multiperiod dynamic forecasts. Even if the matrix to be inverted is the same as in the converging static case, the dynamic simulation mechanism increases the risk of generating values of the determinant close to zero.

In long dynamic forecasts the analytical simulation method behaves best and the method of re-estimation gives results close to this only in the case of efficient re-estimation methods. The Fair method gives rather different results.

For nonlinear models no univocal results can be established. Three kinds of cases occur in the empirical results:

- All Monte Carlo methods converge and their results are close to each other and to analytic simulation results.

- Some methods, usually those applying efficient estimation, converge to results similar to those produced by analytical simulation, others do not converge.

- Some or all methods converge, but their results are close to those produced by analytic simulation in the first period of forecast, but diverge from each other and from analytical simulation after a few periods.

Nonlinearity is the usual feature of real world models and the empirical results above show that no a priori conclusions are possible in the selection of the proper method of simulating the first component of the forecast error of a special model. The analytical simulation seems to be the most reliable method. It requires, however, the stongest assumption, which is asymptotic normality of the estimated coefficients. In practice the method was also applied when this assumption was violated and the results seemed to be numerically reasonable, even if with unpredictable statistical properties.

IV EMPIRICAL ESTIMATION

In the empirical part of this study the estimation methods and the stochastic simulation methods were applied to the structural annual model of the Bank of Finland. We also applied the simulation technique to an autoregressive naive model.

IV.1 The structural model

The structural model that will be used in the empirical part of this study is the KT-model of the Finnish economy developed at the Bank of Finland. It was originally built by Korkman and Rantala (1980, 1981, 1982), and subsequently re-specified and re-estimated by Ahlstedt and Virén (1984).

The model consists of sixteen stochastic equations plus sixteen identities. In addition to these 32 endogenous variables there are 80 exogenous variables. The number of endogenous variables is large enough to make the model a significant tool in forecasting and policy evaluation. At the same time, the size of the model is such as to make it possible to detect the direction of the causality relationships between its different blocks.

By its structure, the model is a member of the Keynesian family, its main purpose being the explanation of aggregate effective demand and its components, consumption, investment and exports. The supply side is not modelled. The only endogenous component from the supply side is imports but this behavioural equation is specified as a weighted average of the demand components. Apart from the real demand sector, the model consists of a price sector and a monetary sector. The variables and the structural equations are listed in Appendix I and II.

The model is specified from annual data. The lack of longer data series is the main constraint for estimation. Reliable data are available for the period 1960-1983. The lag structure limits the start of the estimation period to 1963 while the last year, 1983, was saved for ex post forecast computations. The estimation period is thus 1963-1982, which makes the sample undersized in relation to the exogenous variables, according to the definition by Swamy (1980). The model is nonlinear in variables but linear in parameters. The empirical specification was carried out using ordinary least squares. This estimation method is known to give biased estimates because it does not take account of the simultaneity present in the model.

IV.2 The naive model

The usual way of assessing the relative performance of a structural model is to compare it with a naive model in which each endogenous variable is simply a function of its own lagged values. The outcome of the naive model then serves as a benchmark for prediction performance of the structural model. In general policy simulations the naive models have no content. If they are modified to include exogenous policy variables they cease to be naive alternatives. But if for prediction purposes the sophisticated models fail to outperform a naive extrapolation rule, then we would be led to conclude that the structural model had underutilized the information available because of statistical and economic errors of specification and sampling errors of parameter estimates.

There were not enough observations to specify an autoregressive moving average model so we had to limit ourselves to the estimation of a purely autoregressive model consisting of a set of completely unrelated equations. A model with 4 lags, a time trend variable and a constant term was estimated for every endogenous variable that had a behavioural equation in the structural model. An autoregressive model was also estimated for the aggregate variable

GDP and its price deflator PQ. The lag length was found in a search procedure where the maximum number of lags was chosen to correspond to approximately the same degrees of freedom in the naive model as in the structural form.

IV.3 The selected methods

As we have seen above the presence of autocorrelation causes inconsistency when conventional estimation methods are used. To cope with autocorrelation we have to use more complicated nonlinear estimation methods. The first thing that it is necessary to do in the empirical estimation is to test the behaviour of the residuals.

To detect autocorrelation we used the asymptotically relevant Durbin-Watson test statistic for equations without lagged endogenous variables and the Durbin h-test for models with lagged endogenous variables. The equations were tested up to the fourth degree of autocorrelation. Taking into account the fact that asymptotic properties do not necessarily hold for small samples we only used the test statistic to indicate the equations where serial correlation might be present. These equations were then re-estimated under the assumption of autocorrelation. The only equation where the estimated parameter of the autocorrelation coefficient reached the significant value 0.3 (Maddala, 1971, p. 283) was that for short-term foreign debt DLUFS. This implies that some degree of inconsistency is incorporated in the estimates of this equation. However, since the autocorrelation was only in one equation, a more complicated method was not necessary. Of course, we know nothing about how autocorrelation affects the small sample properties, nor do we know about the small sample properties in general of the estimators obtained from methods were autocorrelation is taken into consideration. The Monte Carlo results of Moazzami and Buse (1984) indicate that for low values of ρ and small sample size, OLS would have a comparative advantage over the simultaneous methods.

The following methods were used in the estimation of the structural model:

(i) Ordinary least squares (OLS).
(ii) Two stage least squares (2SLS) with instruments selected on the basis of a block-division of the model.
(iii) Two stage least squares using principal components as instruments.
(iv) Iterative instrument estimation.

In the specification of the model ordinary least square estimation was applied to each equation. In the following sections the results of the simultaneous estimation methods are reported.

The analytical finite sample results and the Monte Carlo results reported above indicated that the correlation between stochastic disturbance terms in different equations is the principal reason why a simultaneous-equation system econometric model must be estimated using simultaneous estimation methods. The correlation matrix of the residuals obtained from the stochastic equations thus measures the simultaneity of the whole system.

Let us first examine the correlation matrix of the OLS residuals for the model in question so as to get an idea of the degree of simultaneity. If this matrix were diagonal there would be no contemporaneous correlation between the residuals in the system. Very low off-diagonal elements indicate that no gain is obtained from the use of simultaneous methods.

Table I presents the correlation matrix of the OLS residuals. Only the correlation coefficients from 0.3 upwards are shown. Values lower than that can be regarded as insignificant according to empirical findings for small sample correction (Maddala, 1971, p. 283). Although the matrix is by definition symmetric the whole matrix is shown so as to faicilitate the study of it. As can be seen, less than one third of the correlation coefficients are

The correlation matrix of the structural residuals of OLS estimation

TABLE I

	XQ	MTQ	LHW	CQ	IQC1	KF	VV1Q	PW	PC	PI	PG	IVP	S6230	T123H	LPH	DLU
XQ	1.0							-0.5	0.4		-0.3	0.4			0.4	-0.4
MTQ		1.0			0.4					0.7		0.6				-0.5
LHW			1.0	-0.4		0.4		0.3		-0.3				0.3		
CQ			-0.4	1.0			0.3	-0.5		0.3	-0.4					
IQC1		0.4			1.0					0.3	0.4	0.5				
KF			0.4			1.0	0.3		-0.5							
VV1Q						0.3	1.0		-0.5				0.3			
PW	-0.5		0.3	-0.5				1.0		-0.3	0.3	-0.4				
PC	0.4					-0.5	-0.5		1.0							-0.3
PI		0.7	-0.3	0.3	0.3			-0.3		1.0		0.5				
PG	-0.3		-0.4	0.4				0.3			1.0					
IVP	0.4	0.6			0.5			-0.4		0.5		1.0	-0.3	-0.3	0.5	-0.4
S6230							0.3					-0.3	1.0			0.3
T123H			0.3									0.3		1.0		
LPH	0.4				0.3							0.5			1.0	
DLU	-0.4	-0.5										-0.4	0.3			1.0

greater or equal to 0.3. Only about one tenth are greater than 0.3. Measured in this way correlation between equations does not seem very high but nor is there any indication of diagonality in the matrix either.

The model underlying the correlation matrix consists of a real block, a price variable block and the monetary block. The block of the real variables is formed by the endogenous variables XQ, MTQ, LHW, CQ, IQC1, KF and VV1Q. The price block contains the price variables, PW, PCF, PI, PG and IVP. The variables S623O, T123H, LPH and DLUFS form the monetary block.

On the whole, the correlations are very low. Most of the correlation both in terms of values and frequency is found between the real sector and the price sector, as one would expect. The monetary sector seems to be quite independent. This feature of the model points to a strong position for the OLS estimates.

Since the efficient 2SLS estimation method cannot be applied, we have to choose a subset of first stage regressors from the number of predetermined variables of the whole model. It is possible, without losing consistency, to choose different sets of regressors for different equations. The choice has to be based on own judgement, as there are no objective rules that can be applied. The only necessary requirement for consistency of 2SLS modified estimators is that

- the same set must be used for all of the right hand variables in the same equations; and that

- the predetermined variables, exogenous and lagged endogenous, in the particular equation to be estimated should be included in the set of first stage regressors.

IV.4 2SLS based on blockdivision of the model

In the empirical estimation we have based the selection of the first stage regressors on two different block divisions of the model. All predetermined variables in the block were used as first stage regressors. Nothing can be said on analytical grounds about the efficiency of the estimates. Consistency can be assumed, however, if the error terms are not serially correlated or if there are no lagged endogenous variables among the predetermined variables. This assumption is fulfilled for all the equations except the DLUFS equation.

In the first division we used the three blocks described above: the real block, the price block and the monetary block (model 2TS). Since we have no objective decision rule, this choice can be considered as good as any other arbitrary method of selection.

The blockdivision in the model 2TS seems natural, perhaps because simulation results are always reported according to these sectors. Of course, one could try to measure the efficiency of the estimation based on this division by examining the matrix of the parameters of the endogenous variables. If we could find an order of the equations which would make the matrix of the parameters either triangular as a whole or blocktriangular, the model would be recursive and no simultaneous estimation would be needed. This, however, seems to be impossible and we have to try to find maximal efficiency within a non-recursive system. The efficiency depends on how close to blockdiagonal the parameter matrix is when the equations are in the same order as in the blocks. A pure blockdiagonal representation with only zero values outside the diagonal blocks means that the model can be partioned into self-contained subsystems. Here zero means that the variable does not appear in the equation considered.

The matrix of the parameters of the blockdivision in the 2TS model was printed.[1] It turned out that there were only 14 parameters within the diagonal blocks and 16 outside them. This implied that the division could probably be improved in the sense that we could get more parameters into the diagonal blocks. After a lot of experimentation with the ordering of the equations, we ended up with a partioning in which we had three blocks as before but which now contained 20 parameters inside the diagonal blocks and only 10 outside them. This division was considered to be as close as possible for this model to a division giving a pure blockdiagonal parameter matrix. Most of the improvement resulted from transfering the consumer price variable, PC, from the price block to the monetary block and transferring the short term foreign debt equation DLUFS from the monetary block to the real block. The model was estimated with 2SLS based on this blockdivision (model BL).

IV.5 2SLS using principal components

The idea behind the use of principal components is to reduce the number of first stage regressors by replacing the predetermined variables in the model with a smaller number of constructed variables. When the predetermined variables are strongly correlated, which most certainly is the case for economic variables, the few first principal components usually represent the major part of the variation in the predetermined variables.

---

[1]For space reasons they are not reported but can be provided on request.

TABLE II

Principal components of predetermined variables

COMPONENT	CHARACTERISTIC ROOT	CUMULATIVE FRACTION OF VARIANCE EXPLAINED
1	23.309	0.48561
2	5.436	0.59888
3	4.050	0.68325
4	3.282	0.75162
5	3.019	0.81452
6	2.226	0.86091
7	1.517	0.89251

TABLE III

Ranking oder of principal components

Equation	Order of principal components
CQ	2 3 4 5 6
XQ	1 2 3 4 5
MTQ	1 2 3 4 5
LHW	2 3 5 4 6
PC	1 2 4 3 5
PW	1 4 7 2 8
PI	1 2 3 4 5
PG	1 4 2 5 3
IVP	1 3 2 4 5
IQC1	2 3 1 4 5
S6230	4 3 2 5 6
T123H	2 3 4 6 5
LPH	2 4 3 5 7
KF	1 2 3 4 5
VV1Q	2 5 6 4 7
DLUFS	1 3 5 6 7

In the empirical estimation the principal components were chosen in three different ways.

(i) First, the characteristic roots from the correlation matrix of all the predetermined variables were calculated. Then the principal components corresponding to the largest latent roots were formed. Table II shows the characteristic roots and the cumulative fraction of variance explained by the seven principal components of the correlation matrix of the set of variables of the whole model. When the variables are highly correlated and form a homogenous group, the first principal component explains more than 90 per cent of the total variation. The results shown in table II indicate that the variables are highly heterogenous, and the total variance cannot be concentrated into a few auxiliary variables. The fraction of explained variance starts from 50 per cent for the first principal component and then grows by approximately 5 - 10 per cent for every additional component. The model was estimated using the first 4 and 5 components, 4 being the minimum requirement for identification (model PKE).

(ii) The method of calculating the principal components from the correlation matrix of the whole set of predetermined variables has the disadvantage that high correlation may occur between one or more principal components and the exogenous variables in the particular equations. To reduce multicollinearity, we started from the principal components as in (i), but instead of selecting the principal components with the largest characteristic roots, we selected the components with the highest values of the statistic Q_h.[2] The result of this ranking

---

[2]See page 26.

procedure is given in table III. In the method in (1) the order of the component was 1,2,3,4 and 5, the numbers referring to the order of the corresponding largest characteristic roots. Table III shows that the ranking of the components according to the statistic Q_h does not change the order in all equations because the values of the roots dominate the multiple regression coefficient value. The parameters were estimated using 4 principal components chosen by this ranking criterion (model PKR).

(iii) In the two previous principal component methods one single set of principal components was calculated. In the third version we calculated the principal components separately for every equation. For each equation, the correlation matrix, from which the characteristic roots were calculated, consisted of all the predetermined variables in the model less the variables that occurred as predetermined variables in the equation to be estimated. In the estimation 4,5 and 6 components, calculated separately for each equation, were used as first stage regressors (models YP4, YP5 and YP6).

IV.6 Iterative instrumental methods

The iterative estimation method was applied in two different ways:

(i) In the first iterative estimation the Gauss-Seidel solution for the joint endogenous variables of the OLS-parameter model was used as first stage regressors (model MY). The iterative method with OLS start was shown above to give consistent estimates unless lagged endogenous variables and autocorrelated residuals

coexist.[3] There is one equation, DLUFS, with a significant autocorrelation parameter. In the other single equation methods, error in one equation does not affect the estimation of the other equations. The iterative method, however, uses the information in the whole model when the model solution is used as instruments. The method is more a full information method than the other methods used here. Thus, we might expect that the bias in the whole model caused by serial correlation in one equation is more significant in this method than in the other methods.

(ii) We did one more iterative estimation, this time having a theoretically consistent estimation in the first round. The Gauss-Seidel solution of the 4 principal component parameter model was used as initial stage regressors in the second variant of the iterative method (NY).

There were not enough observations nor any computer program available to perform a third round of iteration so as to obtain a full information iterative estimator. To find out whether this third round would have been warranted in terms of achieving maximal efficiency, the correlation matrix of the residuals of the estimation with OLS-start (MY) was printed. The matrix is presented in table IV. As before only the elements >0.3 are shown. If the correlation matrix turns out to be diagonal then there is no need for this third round giving system estimation results. When the correlation matrix is diagonal 2SLS and 3SLS coincide. As we can see only a quarter of the elements outside the diagonal are >0.3. This suggests that the impossibility of carrying out system estimation is not perhaps such a great loss after all.

[3]See page 30.

The correlation matrix of the structural residuals of the IIV model estimation

TABLE IV

	XQ	MTQ	LHW	PW	PC	PI	PG	IVP	IQC1	CQ	S6230	VV1Q	T123H	LPH	KF	DLU
XQ	1.0	0.7		-0.5			-0.5	0.4						0.3		
MTQ		1.0		-0.5		0.7		0.6								
LHW			1.0							-0.6			0.4		0.5	
PW	-0.5	-0.5		1.0			0.5	-0.4		-0.4						
PC					1.0		0.3								-0.4	
PI		0.7				1.0		0.5	0.4	0.3						
PG	-0.5			0.5	0.3		1.0			-0.5						
IVP	0.4	0.6		-0.4		0.5		1.0	0.6					0.5		-0.3
IQC1						0.4		0.6	1.0					0.3		
CQ			-0.6	0.4		0.3	-0.5			1.0	0.4			-0.3		0.3
S6230										0.4	1.0	0.4				0.3
VV1Q								-0.3			0.4	1.0			0.4	
T123H			0.4										1.0			
LPH	0.3							0.5	0.3	-0.3				1.0		
KF			0.5		-0.4							0.4			1.0	
DLU								-0.3		0.3	0.3					1.0

IV.7 Comparison of the various estimates

When estimating a parameter from a very limited sample we have to accept the fact that distribution of the true population parameter remains unknown and no comparison between true value and estimated value is possible. We know that most certainly every estimate has a small sample bias, but we do not know the magnitude of the bias. We also know that the small sample estimates of the standard deviation of the estimates may differ quite appreciably from the corresponding population parameters.

Since the data used in the empirical estimation of the various estimates is the same, the estimates, the forecasts, forecast errors and resulting summary measures are not statistically independent. Hence a formal statistical test cannot be based on a direct comparison of the parameter estimates (Salmon & Wallis, Model validation and forecast comparisons, 1982). It is not therefore possible to find out if the differences between empirical estimates are significant.

Small differences when comparing estimation results parameter by parameter have led a number of investigators to advocate the application of conventional OLS estimation on the grounds that it hardly matters which method is used (Klein, 1960). This is not a correct point of view. Slight differences in the parameter values, even if not statistically significant, can cause noticeable differences in the dynamic simulation path.

A summary statistic applicable to a whole system of equations should be used for the comparison of different methods. Klein suggests the standard error of forecast of endogenous variables. This means that the estimates of the reduced form parameters calculated from the estimated structural form parameters are the relevant subjects of comparison. This choice seems reasonable because the model forecasts are made using the reduced form. However, tiny biases in the estimates of individual structural

parameters can be magnified into sizable biases in the estimates of reduced form parameters (Klein, 1960). Klein also argues that the optimal efficiency properties of OLS structural estimates do not carry over to estimates of the reduced form parameters.

According to this the forecast ability of the estimation methods in simulation behaviour is heavily stressed when we compare the various empirical estimates in terms of

- the statistical measures of the moments of the estimates,
- the accuracy of the within-sample prediction ability, and
- the accuracy of the one-year-ahead outside sample prediction ability.

IV.7.1 Statistical measures of precision

Each of the estimation methods prescribes a formula for computing an estimate of the standard error of the estimated parameters. In table V the point estimates, are presented together with their t-values, for the 18 autoregressive equations. In Appendix II the point estimates for the parameters in corresponding structural equations are presented, together with their standard errors, for OLS, 2SLS (BL) and 2SLS with 4, (YP4) 5 (YP5) and 6 (YP6) principal components and the IIV estimates with OLS (MY) start.[4] We know that the standard errors for OLS estimates are not - even theoretically - the correct ones, whereas the standard errors for the other estimates are at least asymptotically the theoretically, correct estimates. Here we once again face the question of whether there is any point in comparing asymptotic values when estimating from small samples. In relying on asymptotic properties there is implicit the assumption that all we can do is to try to find a method which will be the best possible in the sense that it will have high probability of being correct in the long run, and that requirement is not enough.

[4]See Appendix III for definition.

TABLE V The autoregressive model with 4 lags
The t-values are shown in parentheses below the parameter estimates

Equation	1.LAG	2.LAG	3.LAG	4.LAG	TREND	INTERCEPT	R^2
GDPQ	1.124	-1.030	0.669	-0.513	4146	7821	0.994
	(4.58)	(2.77)	(1.74)	(2.03)	(3.22)	(2.21)	
PQ	1.570	-0.751	-0.015	0.187	0.46	- 6.55	0.998
	(5.67)	(1.34)	(0.03)	(0.61)	(1.91)	(1.85)	
CQ	0.952	-0.636	0.635	-0.606	1941	6447	0.991
	(4.21)	(2.07)	(2.07)	(2.78)	(2.78)	(2.45)	
XQ	0.973	-0.886	0.409	-0.489	1899	-14304	0.976
	(3.98)	(2.56)	(1.18)	(1.83)	(3.42)	(2.85)	
MTQ	0.784	-0.742	0.405	-0.237	1671	-4923	0.961
	(2.95)	(2.14)	(1.14)	(0.84)	(2.04)	(0.94)	
LHW	1.084	-1.011	0.430	-0.322	12.31	2306	0.901
	(4.50)	(2.84)	(1.17)	(1.35)	(2.79)	(3.36)	
PC	1.905	-1.544	0.823	-0.192	0.346	-5.11	0.998
	(6.77)	(2.63)	(1.39)	(0.64)	(1.819	(1.74)	
PW	1.871	-1.522	0.625	0.043	0.410	-6.095	0.999
	(6.80)	(2.63)	(1.07)	(0.14)	(2.06)	(1.92)	
PI	1.202	-0.538	0.541	-0.252	0.686	-9.892	0.997
	(4.54)	(1.32)	(1.22)	(0.86)	(2.07)	(1.90)	
PG	2.027	-1.811	0.793	-0.002	0.379	-5.532	0.999
	(7.38)	(3.03)	(1.31)	(0.01)	(1.89)	(1.81)	
IVP	1.255	-0.691	0.562	-0.141	0.672	-10.12	0.996
	(4.64)	(1.65)	(1.25)	(0.45)	(1.91)	(1.82)	
IQC1	0.871	-0.186	0.167	-0.386	230	146	0.994
	(3.31)	(0.50)	(0.44)	(1.19)	(1.89)	(0.18)	
S6230	0.613	-0.393	-0.163	0.204	0.169	2.413	0.928
	(2.34)	(1.26)	(0.50)	(0.80)	(2.43)	(1.89)	
T123H	1.500	-0.838	-0.376	0.994	-47.04	1155	1.000
	(5.90)	(1.64)	(0.70)	(3.07)	(0.92)	(1.37)	
LPH	1.681	-1.077	0.728	-0.169	-17.00	-51.69	0.999
	(5.75)	(2.01)	(1.36)	(0.49)	(0.31)	(0.05)	
KF	1.936	-1.606	0.786	-0.300	2000	1299	1.000
	(7.20)	(2.90)	(1.45)	(1.21)	(2.25)	(0.62)	
VV1Q	0.877	-0.768	0.273	-0.120	1233	17981	0.955
	(3.19)	(2.13)	(0.73)	(0.45)	(2.39)	(2.24)	
DLUFS	-0.033	0.123	-1.336	0.548	-6.468	445	0.677
	(0.12)	(0.31)	(2.76)	(0.99)	(0.11)	(0.31)	

A look at the estimates and their standard errors shows that they do not differ very much between methods. Surprisingly enough, there are only a few sign differences. They occur when the significance of the parameters is reduced so that the hypothesis of the parameter being zero cannot be rejected. One exception to this rule can be found. The strong multicollinearity between real income and its one period lagged value in the consumption function strongly affected the simultaneous estimation. This led to the use of restrictions on these income elasticity parameters in all the simultaneous estimation methods. They were forced to take the same values that were established in the OLS-estimation, where the income variable and its one period lagged value were estimated without any restrictions. Here we found evidence for the analytical result stated above; OLS is less sensitive to multicollinearity than are the simultaneous methods. The correlation matrices of the explanatory variables in the separate equations reveal, however, that multicollinearity is not a serious problem in this model.[5]

It is not likely that a particular estimation method would behave better than the others when all parameter estimates are compared. An intransitive behaviour in the statistical measures was expected and is also clearly seen. Thus we cannot draw any unambiquous conclusions but instead have to restrict ourselves to talking about tendencies. The estimates of the parameters in the price equations given by the simultaneous methods are close to the OLS-estimates. In the real variable block and the monetary block the 2SLS method with 6 principal components (YP6) gives estimates that are closest to the OLS-estimates. The explanation is that the price block is the less simultaneous part of the model and there are only a few jointly endogenous variables to be replaced in the first stage.

---

[5]The effects of multicollinearity were tested using ridge estimators in the least square estimation (Hoerl and Kennard, 1970). Different values of the ridge constant were compared but the estimates seemed to be highly stable and insensitive to the constant.

Thus the simultaneity bias of the OLS estimates is small. In the other blocks the simultaneity is greater as, too, so is the bias which simultaneous estimation is supposed to remove. In extending the number of principal components we can see that the improvement in the first stage fit leads to a reduction in the standard error of the estimates. The confidence intervals are highest for the 4 principal component case and they systematically decrease as the number of included components increases.

Here we have to remember that the t-values, the standard errors and the confidence intervals are valid only asymptotically and their computations require the assumption of errors being normally distributed.

It is not possible on the basis of the point estimates and their standard errors to make any rigorous conclusions about the priority of the estimators.

IV.7.2 Deterministic prediction: ex post simulation

The next step is to compare the simulation behaviour of the models obtained with the different estimation methods. The model is solved for the estimation period to obtain the dynamic behaviour of the ex post simulation. The model is solved deterministically, which means that stochastic disturbances in the system are set at their expected values of zero. The dynamic deterministic simulation reveals the contradiction in the treatment of lagged endogenous variable in estimation and simulation. In standard estimation techniques the lagged endogenous variables are taken as given data. In simulation the lagged endogenous variables are determined by the system, thus introducing a forecast error that grows with the simulation period.

In addition to this dynamic simulation, we also wish to examine the static behaviour. The model is solved deterministically for the

first year outside the sample period to obtain a one year ex post forecast showing the static prediction performance. This forecast is static because the lagged endogenous variables are given their historical values.

In comparing the within-sample prediction accuracy, the aim is to find the estimates which lead to the most accurate multiperiod prediction. Since this model has many dimensions and the accuracy of explanation may vary among variables, we have chosen to examine all the endogenous variables in some cases and to focus attention on real GDP and its price deflator PQ in other cases. It seems justified to stress these two variables because the model is by nature Keynesian. Accordingly, its main objective is the explanation of aggregate demand. Since GDP and the price deflator are determined in the model as weighted averages of the estimated endogenous demand components, all errors from the demand side are aggregated and accumulated in the multiperiod prediction of the aggregates GDP and PQ.

There are different ways of comparing the outcomes of simulations. We can use graphical interpretations or some statistical measure for the deviations between simulated values and the historical data.

Let us first concentrate on different methods using 4 principal components. In the PKE method we used the same 4 principal components for every equation, corresponding to the 4 largest characteristic roots of the matrix of all the predetermined variables in the whole model. In the PKR method we used the 4 principal components that were selected according to the criterion reported on page 26. In the YP4 method we calculated the principal components separately for each equation and then selected those connected with the largest roots.

Figure I on page 113 shows the ex post simulation of GDP for the different estimation methods using 4 principal components. The

simulation values are divided by the historical values of GDP. The closer to 1 the ratios are, the closer is the fit to the data. As can be seen, there is no such great advantage in the behaviour of the different simulation paths as to make it worthwhile to carry out laborious calculation of different sets of principal components for each equation. Nor does the ranking procedure, which considers the multicollinearity between components and the explanatory variables, noticeably increase the ex post simulation accuracy. It was, however, observed that in the PKE method more iterations were needed to find the Gauss-Seidel solution than in the other methods using 4 principal components.

To measure the dispersion between the estimates of the 4 principal components methods, the trace of the parameter variance matrix for each equation and each method was calculated. The statistics are shown in table VI. The intercept is not reported for the reason that it is not normally important for making inferences about the structural parameter estimates. This comparison also shows that the method which is easiest to compute, PKE, competes well with the more complicated ones.

In figure II on page 113 the ex post behaviour of the methods in which individual sets of principal components were calculated for each equation are compared. Sets of 4,5 and 6 components were used in the estimation. As can clearly be seen, the estimation with only 4 components is superior to the others. The increase in efficiency in the first stage does not compensate for the effects of the loss in the degrees of freedom (see page 26).

Figure III shows the ex post simulation paths of OLS, BL and PKE. In the comparison between these simulations the PKE estimation is clearly inferior to the others. There is major overshooting in the years 1968 and 1977 and major underestimation in the year 1966. These failures are not compensated for by the "good" behaviour in the middle and at the end of the ex post simulation period.

Figure IV shows the simulation accuracy of OLS, IIV with OLS as the starting point (MY) and PKE and IIV with PKE as the starting point (NY). The iterative process leads to a clear improvement in the ex post predictive accuracy in both cases.

After this graphical analysis we next look at a statistical summary of the discrepancy between the simulation values and the historical data. The measure that is most often used is the root mean square error (RMSE). Others include the mean simulation error (MSE), the mean per cent error (MAPE) etc.[6] There is no obvious reason for preferring one measure to the others. In this study we choose to use the MAPE statistics. There could be some doubt about the existence of the first and second moments of the sample statistics, so we want to use a measure that is nonparametric and that is not based on the moments. MAPE is simply the mean absolute difference between the simulated and the historical values in per cent of the historical values. We can then look at the MAPE values for the different estimated models to compare their simulation accuracy.

The model is a simultaneous system which implies that the model as a whole has a dynamic structure which is much richer than that of any one of the individual equations. Thus, even if all the individual equations fit the data well and are statistically significant, there is no quarantee that the model as a whole, when simulated, will reproduce those same data series closely.

---

[6]See Appendix III for definition.

FIGURE I GDP ex post simulation values divided by the actual estimation period values.

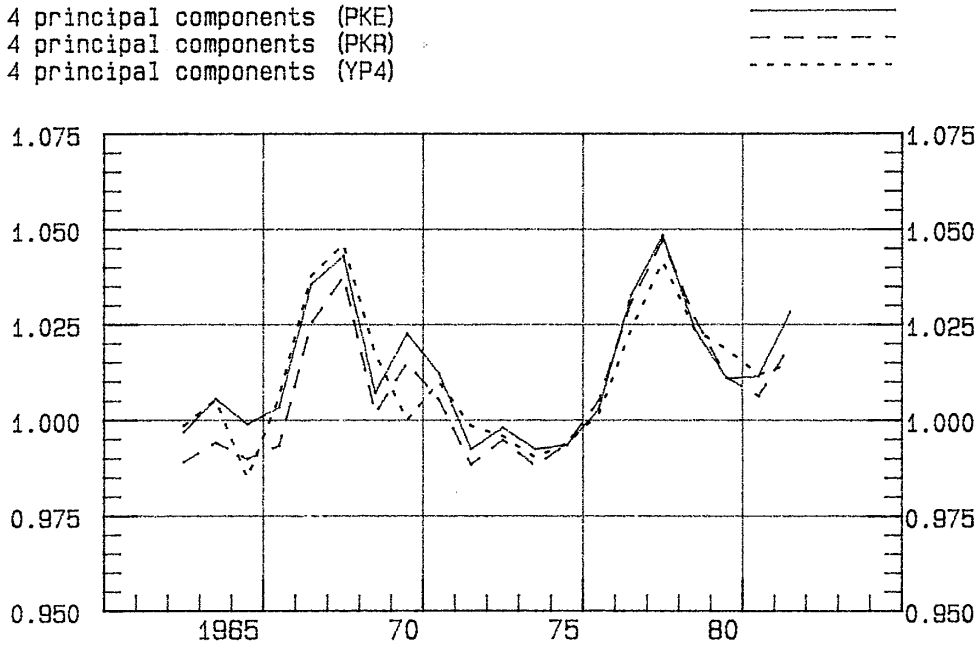


FIGURE II GDP ex post simulation values divided by the actual estimation period values.

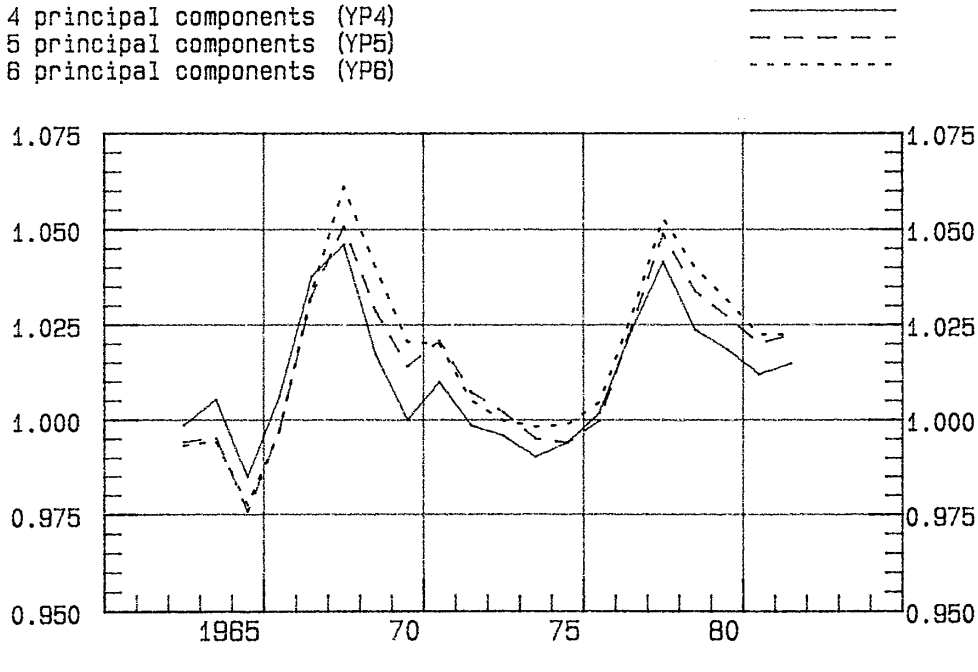


TABLE VI

Comparison of the efficiency of the parameter estimates of the three methods using four principal components

Equation	Method		
	YP4 TRACE	YPR TRACE	YPE TRACE
CQ	0.111	5.680[1]	0.108
XQ	0.332	0.291	0.291
MTQ	3.380	3.380	3.380
LHW	0.296	0.425	0.335
PC	0.054	0.061	0.061
PW	0.383	0.310	0.295
PI	0.005	0.005	0.005
PG	0.084	0.116	0.084
IVP	0.041	0.041	0.041
IQC1	1.434	1.062	1.062
S6230	0.051	0.050	0.050
T123H	0.195	0.242	0.209
LPH	2.352	1.839	2.265
KF	0.033	0.035	0.035
VV1Q	0.780	0.778	0.780
DLUFS	1870.130	345.210	602.970

1 The high value is due to the fact that the ranking criterion ruled out the first principal component from the set of first stage regressors. The exclusion of this largest component, which is the main explanatory variable, resulted in a poor fit of the first stage regression.

FIGURE III GDP ex post simulation values divided by the actual estimation period values.

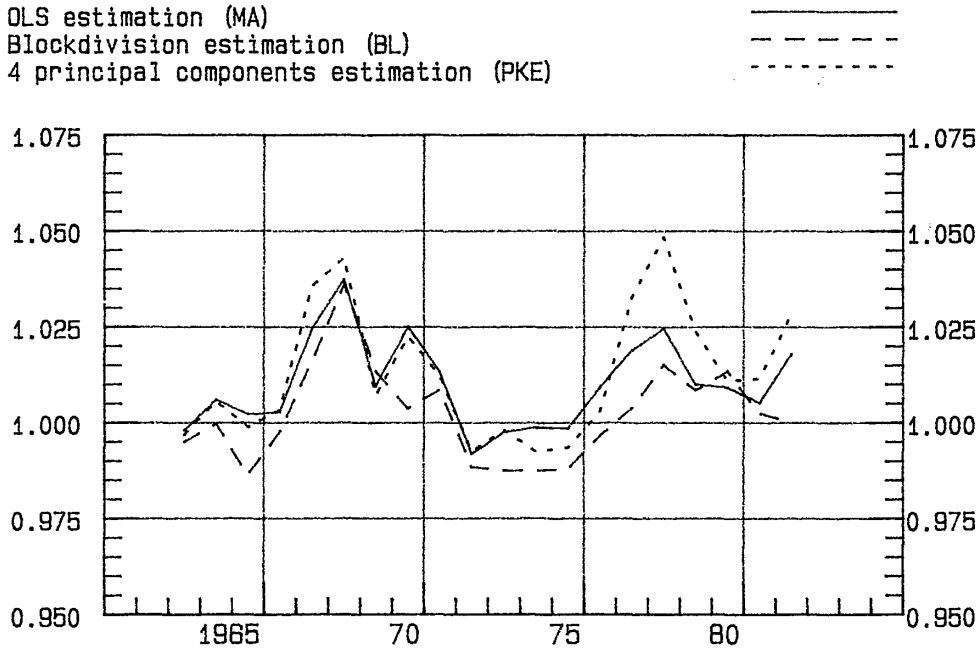
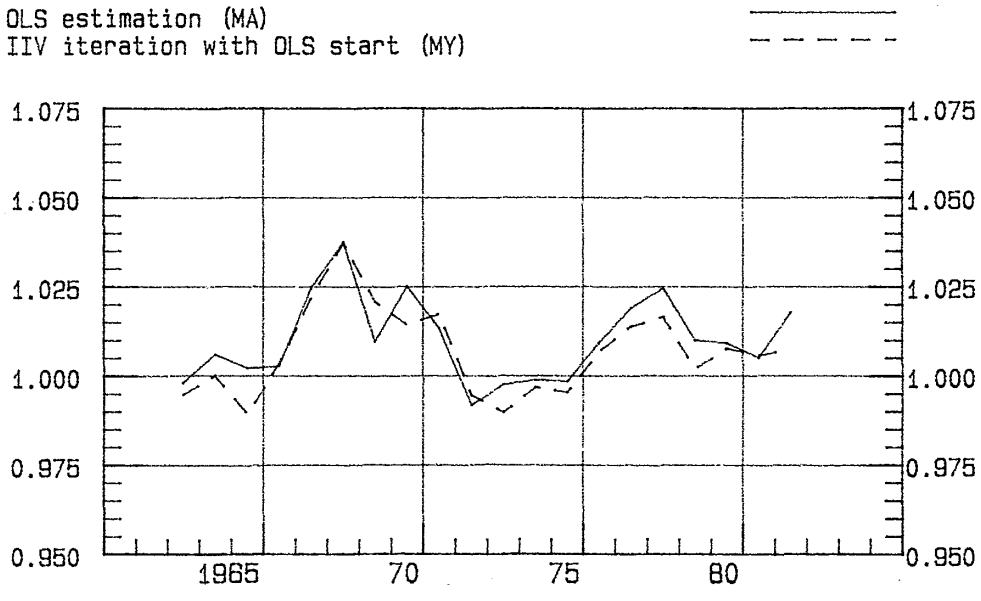
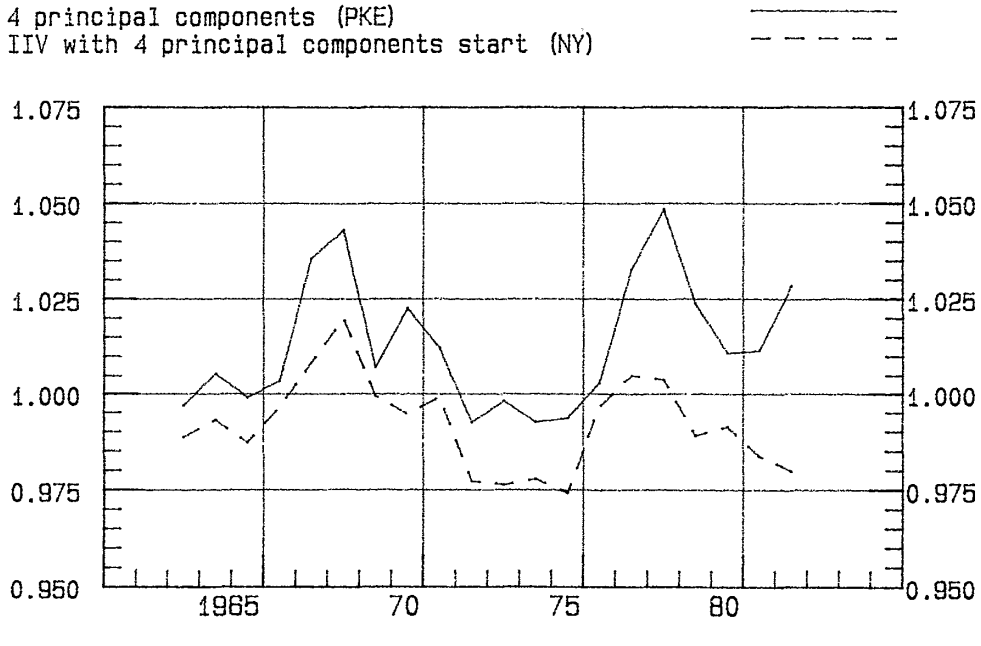


FIGURE IV GDP ex post simulation values divided by the actual estimation period values.





In table VII on page 118 we present three kinds of MAPE values. In the first column are the MAPE values for the individual equations for the model MY for the estimation period 1963-82. In that case, the MAPE is only due to the structural error term of the equation deriving from the fact that stochastic equations do not fit the data perfectly. There is no error from the explanatory variables because they are set to their historical values. The second column shows the same measure for the naive model for the estimation period 1964-82. As can be seen, the individual naive equation errors are greater than those of the individual structural equations. In the third column are the MAPE values of the endogenous variables for the ex post simulation of the whole structural model. These values include both the effect of the individual equation residuals and their simultaneity effect through the right-hand endogenous variables in the model. It can also be seen that the errors attributable to the whole dynamic model solution are far greater than those of the individual structural equations. Since there is no simultaneity in the naive model the MAPE values for the individual equations are also the MAPE values for the whole naive model solution. Thus, a comparison of the second and third columns gives the differences between the MAPE values of the two model solutions. This kind of measure clearly favours the naive model.

Next, we look at the MAPE values obtained from the dynamic ex post simulation of the different estimates. The calculated MAPE values are presented in table VIII. If we examine the aggregate variables GDP and PQ, we can conclude that the OLS estimator and the MY estimators perform best because they have the lowest MAPE values.

TABLE VII

Comparison of MAPE statistics
The MAPE values for the structural equations and
the stuctural model MY are for the estimation period 1963-82.
The values for the naive model are for the estimation period 1964-82.

MAPE			
VARIABLE	INDIVIDUAL EQUATION STRUCTURAL MODEL	INDIVIDUAL NAIVE MODEL	MODEL SIMULATION STRUCTURAL MODEL
CQ	0.87	1.49	2.73
XQ	4.99	4.16	7.21
MTQ	1.01	4.76	4.22
LHW	0.65	0.72	3.02
PC	0.86	2.00	3.12
PW	0.99	2.04	5.53
PI	0.13	2.60	3.68
PG	1.06	1.99	5.49
IVP	1.12	3.20	5.33
IQC1	2.56	4.25	2.58
S6230	2.42	3.50	3.11
T123H	0.99	1.41	4.67
LPH	1.43	1.87	6.92
KF	0.39	0.35	0.79
VV1Q	1.29	2.27	1.45
DLUFS	193.02	244.07	284.00

TABLE VIII

MAPE for ex post simulations for the estimation period 1963-82

$MAPE = (1/N)*100*\Sigma|\hat{Y}_i - Y_i / Y_i$

Equation	Method							
	OLS	BL	YPE	YPR	YP4	YP5	YP6	MY
GDPQ	1.16	0.97	1.58	1.50	1.49	1.92	2.12	1.08
PQ	3.12	2.73	4.03	3.15	3.49	3.81	4.00	3.32
CQ	2.17	2.39	2.42	2.51	2.36	3.27	3.63	2.73
XQ	7.39	7.12	8.89	8.46	8.07	8.28	8.50	7.21
MTQ	3.48	4.02	5.16	5.23	5.07	5.52	5.85	4.22
LHW	2.21	1.34	2.67	1.22	2.12	2.22	2.42	3.02
PC	3.16	3.21	3.86	3.38	3.46	3.91	4.13	3.12
PW	4.67	3.60	5.58	3.27	4.42	4.65	4.91	5.53
PI	3.60	3.60	3.82	2.40	3.18	3.33	3.47	3.68
PG	4.63	3.77	5.50	3.90	3.49	4.68	4.89	5.49
IVP	5.06	3.96	4.93	3.27	4.13	4.43	4.89	5.33
IQC1	2.56	2.59	2.67	2.48	2.56	3.02	3.08	2.58
S6230	2.82	2.94	3.41	3.19	3.55	3.15	3.09	3.11
T123H	2.64	3.57	4.49	3.90	3.66	2.35	3.16	4.67
LPH	6.19	8.87	7.50	8.16	6.99	5.64	6.81	6.92
KF	0.82	0.86	1.03	0.99	0.93	0.90	0.98	0.79
VV1Q	2.06	1.63	2.06	2.06	1.78	2.24	2.47	1.45
DLU	133.80	120.19	263.58	167.08	284.02	222.21	75.78	165.88
MEAN	3.40	3.32	4.09	3.48	3.58	3.72	4.02	3.78

The mean of the MAPE values of the 17 variables[7] over the estimation period for each estimation method indicates the reliability of OLS. As regards the principal component estimates, it can be observed that the ability to generate the historical data decreases when the number of the principal components is increased from 4 to 6.

In all comparisons of the estimates so far we have only presented the 2SLS estimate based on the second blockdivision of the model, i.e. the BL model. In the BL estimation the matrix of the parameters of the endogenous variables was as close as possible to the blockdiagonal interpretation. The reasons for choosing this method instead of the 2TS method, based on division into a real block, a price block and a monetary block, are to be found in table IX.

In the first two columns the standard error of the individual equations are listed. As can be seen, the estimation of the two equations, MTQ and DLUF, which have the greatest standard error are clearly improved in the BL model as compared with the 2TS model. The differences for the other equations are small.

To measure the efficiency of the parameter estimates the traces of the variance matrix of the estimates were calculated. The results are shown in the third and fourth columns. The differences between the two methods are negligible. The explanation lies perhaps in the fact that the figures are only approximate and do not measure small sample behaviour, in which greater differences in efficiency would perhaps occur. The last two columns show the MAPE values for ex post simulation over the estimation period. Here we find clear evidence for the superiority of the BL model over the 2TS model. The figure V shows the ex post simulation fit of both models. As before, the criterion is the closer to the value one the figures

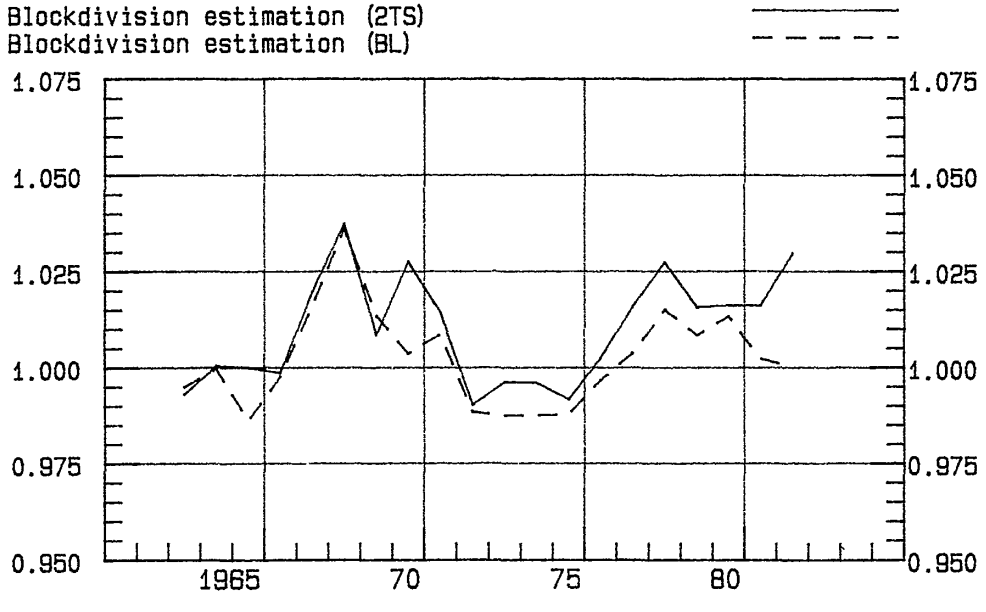
[7]DLUFS was excluded on the basis of its huge values which would otherwise dominate the mean. The high values of DLUFS are due to division by its historical values, which fluctuate around zero.

TABLE IX

Comparison of the efficiency of the parameter estimates in the 2TS and BL models

Egua-tion	Stand.error of equation		Trace		MAPE 1962-83	
	2TS	BL	2TS	BL	2TS	BL
GDPQ					1.33	0.97
PQ					3.98	2.73
CQ	0.008	0.009	0.028	0.050	3.01	2.39
XQ	0.069	0.070	0.142	0.148	7.83	7.12
MTQ	754.156	653.053	0.049	0.034	5.64	4.02
LHW	0.009	0.009	0.135	0.131	2.88	1.34
PC	0.012	0.012	0.038	0.040	3.55	3.21
PW	0.015	0.015	0.247	0.206	6.09	3.60
PI	0.013	0.013	0.004	0.004	4.49	3.60
PG	0.012	0.012	0.058	0.060	5.98	3.77
IVP	0.016	0.016	0.029	0.030	5.67	3.96
IQC1	0.045	0.042	1.459	0.751	2.46	2.59
S6230	0.003	0.003	0.037	0.027	3.67	2.94
T123H	0.016	0.015	0.112	0.077	12.53	3.57
LPH	0.022	0.024	0.059	0.740	18.99	8.87
KF	0.006	0.006	0.011	0.009	0.83	0.86
VV1Q	0.021	0.020	0.813	0.781	1.87	1.63
DLUFS	14.017	12.326	352.980	07.855	37.71	120.19

FIGURE V GDP ex post simulation values of the block-division estimation models divided by the actual estimation period values.



are, the better is the fit. The simulation of the BL model is superior to that of the 2TS model mainly because of the obviously "better" behaviour towards the end of the simulation period. In 2TS simulation, the departure from the value one seems to grow and no return to the historical path can be hoped for.

As in the comparison of the other empirical estimation results, the ex post simulation behaviour, both in terms of MAPE values and graphical analysis, provides the clearest base for comparison. In this case, these two criteria put the BL model before the 2TS model.

In a comparison based on the MAPE values of the ex post simulation, the OLS, BL and MY models perform best. In concession to Klein, we also calculated the means and the standard errors of the forecast errors of the stochastic equations of these three models to obtain one more measure of forecasting ability. The calculated figures are reported in table X. According to this measure the BL model is the best one having the lowest mean for 10 variables and the lowest standard error for 9 variables out of 18. The comparison is not quite valid because the magnitudes of the differences between methods are not taken into account. Only the ordinal ordering is considered.

IV.7.3 Deterministic prediction: ex post forecast

The estimation period is 1963-1982. There is only one year of actual data, the year 1983, which is reserved for use in ex post forecasting.

In table XI, the MAPE values (in fact, only absolute per cent error) for one year outside sample forecasts are presented for the various estimates. In this comparison the MY simulation performs best when we focus on the aggregate variables GDP and PQ. OLS does not behave badly either.

IV.8 Structural versus naive models

In the following we make use of some ideas presented by Nelson (1972) to compare the predictive ability in one period forecasts of the structural models as against the naive model. The structural models, which were chosen according to the findings of the comparisons so far, were the OLS, MY and BL models.

First, the predictive ability was examined in terms of bias. The actual values of the dependent variables in the 16 behavioural equations and in the essential identities GDP and PQ, 18 variables all in all, were regressed on their predicted values from the ex post simulations of the OLS, BL, MY models and the naive model AR. In table XII the estimated slopes and intercepts are presented. The differences between the estimated slopes are small. According to the t-test statistics in the last column of table XII, none of the slope parameters are significantly different from unity. While the deviations from the theoretical values, that is 1, are of rather small magnitude we can conclude that the prediction bias is fairly small for all methods. A ranking order could, however, be established which coincides with the ranking oder given by the MAPE statistics for the structural models. The naive model seems to fulfil the a priori expectation of being the winner when one-period predicitive ability is compared in this way.

The correlation coefficient between structural model errors and naive model errors provide a measure of similarity between the two sets. In table XIII we present the correlation coefficient between the separate structural models and the naive model.

The higher the correlation coefficient, the more of the forecast behaviour is left unexplained by both the structural model and the naive model. The lowest correlation is found between the OLS model and the naive model AR.

TABLE X

Means and standard errors of structural model forecast errors for the estimation period 1963-1982

	OLS model		MY model		BY model	
	mean	stan-dard error	mean	stan-dard error	mean	stan-dard error
GDPQ	1478.20	1652.20	975.53	1586.96	336.62	1722.88
PQ	-1.39	1.05	-1.59	1.37	-1.02	1.07
CQ	-279.00	2323.50	-31.57	2833.69	-653.97	2487.85
XQ	1906.16	2808.44	1753.56	2894.12	1897.72	2753.40
MTQ	698.26	1736.28	1139.65	1925.77	1250.93	1712.09
LHW	56.91	55.10	85.33	65.98	18.39	48.00
PC	0.86	1.53	-0.94	1.51	-0.64	1.55
PW	2.24	1.58	-3.04	2.68	-1.65	1.25
PI	2.14	1.83	-2.28	2.14	-1.69	1.62
PG	1.89	1.04	-2.57	1.48	-1.30	1.04
IVP	-2.59	1.84	-2.93	2.46	-1.96	1.48
OQC1	178.99	402.29	182.78	389.62	160.91	400.80
S6230	0.01	0.35	0.01	0.36	0.01	0.36
T1234	366.93	1776.21	1574.66	1755.05	922.20	2716.88
LPH	-1056.26	2337.27	1115.93	1320.42	-1778.48	3526.68
KF	231.59	2813.51	-508.47	2719.31	-996.58	3095.29
VVTQ	-25.63	1736.40	-317.3	1267.46	-422.73	1256.87
DLUFS	102.80	1782.55	242.21	1950.01	85.15	1738.13

TABLE XI

MAPE for ex post forecasts for the year 1983

Equation	Method							
	OLS	BL	YPE	YPR	YP4	YP5	YP6	MY
GDPQ	0.85	1.17	0.85	1.74	1.50	1.46	1.34	0.17
PQ	0.67	0.58	0.39	0.33	0.67	0.64	0.67	0.79
CQ	3.38	4.52	2.55	4.35	4.61	5.37	4.96	2.32
XQ	1.10	0.38	1.70	2.73	0.12	0.18	0.32	1.67
MTQ	6.00	6.82	3.60	3.40	3.84	5.66	5.86	5.84
LHW	0.95	0.84	1.14	0.70	1.31	1.12	1.05	1.07
PC	0.78	0.95	0.65	0.92	0.77	0.87	0.85	0.51
PW	0.41	0.44	0.45	0.71	0.45	0.47	0.48	0.30
PI	0.48	0.52	0.56	0.73	0.57	0.58	0.59	0.53
PG	0.00	0.03	0.03	0.33	0.02	0.02	0.03	0.18
IVP	1.16	1.04	0.95	0.70	0.91	0.94	1.02	1.16
IQC1	2.92	3.09	0.48	0.46	0.07	3.08	3.10	2.89
S6230	3.16	2.82	2.88	3.19	1.08	1.84	1.98	3.62
T123H	3.27	3.62	4.44	3.53	4.34	3.58	3.22	3.12
LPH	4.20	6.69	3.81	0.87	7.91	7.49	7.43	2.76
KF	0.20	0.10	0.33	0.55	0.43	0.22	0.12	0.21
VV1Q	2.76	2.71	1.96	1.96	2.72	2.73	2.76	2.71
DLU	158.64	145.61	38.47	141.78	57.54	95.61	105.21	225.53
MEAN	1.90	2.14	1.57	1.60	1.84	2.13	2.10	1.76

TABLE XII

Estimated slopes and inter cepts

Model	slope	intercept	t-statistic
AR	0.99551 (0.006)	-385.475 (687.260)	0.736
MY	0.994791 (0.004)	-379.585 (430.252)	1.340
OLS	0.99285 (0.004)	-428.275 (437.750)	1.833
BL	0.99173 (0.005)	-581.970 (544.481)	1.7229

The standard errors are shown in parentheses below the point estimates.

TABLE XIII

Correlation between the structural model and the naive model: one year forecast errors over all endogenous variables

	AR model
OLS model	0.0907
MY model	-0.1092
BL model	-0.1615

TABLE XIV

Composite prediction

	structural model forecast coefficient $\hat{a}_1$	naive model forecast coefficient $\hat{a}_2$
OLS model	0.69819 (6.19)	0.24483 (2.60)
MY model	0.67834 (7.20)	0.31638 (3.35)
BL model	0.60608 (4.66)	0.38646 (2.95)

Student t-values are shown in parentheses below the point estimates.

Next, we consider the concept of composite prediction. If the structural model utilizes the information available, then the naive model, which draws only on a subset of information, should not be able to contribute to the accuracy of a composite prediction which combines both.

A linear composite prediction is of the form

(1) $$E_F = a_1 E_{STR} + a_2 E_{AR} + e$$

where

E_F is the actual value to be predicted
E_{STR} is the prediction obtained from the structural model
E_{AR} is the prediction obtained from the naive model.

If the structural model contains all the information in the naive model, then the OLS estimates a_1 and a_2 of model (1) should be unity and zero, respectively. Since the individual predictions are essentially unbiased, we would expect that in a composite prediction their coefficients would add to approximately unity. In this case we found that the unrestrictedly estimated coefficients were close enough to one so that there was no need to emphasize the constraint $a_2 = 1 - a_1$.

The figures in table XIV on page 128 show that the OLS model contains most of the information in the naive model. The BL model leaves unused part of the information in the naive model. Thus the intersection between the two information sets is greatest for the OLS model and smallest for the BL model.

From this comparison we can conclude that the utilization of information is not complete in the structural models. In forecasting, a composite prediction could lead to an increase in the predictive ability.

In addition to this single period static prediction, the multiperiod dynamic predicition should also be calculated, especially when comparison with a naive model is concerned. The longer period forecasts remove the advantage that the autoregressive scheme has in the short run. The autoregressive models are known to generate rapidly increasing prediction errors as the time horizon of the extrapolation lengthens. For an evaluation of the predictive accuracy of a structural model vis-à-vis a naive model, a multiperiod comparison should be used. For a one period forecast, the naive model is usually the best tool. There is, however, no data at this point for a deterministic multiperiod prediction error comparison but the problem will be tackled in connection with stochastic simulation in chapter V.

According to the ranking order obtained from the comparison of one year prediction between structural and naive models, the MY and OLS are superior to the BL model.

IV.9 MAPE versus RMSE

The comparison of the forecast errors both in ex post simulation and ex post forecasts shows that the MAPE values do not differ very much between methods and the differences are sometimes so vague that the ranking order is not completely clear. There is also a possibility that the outcome of the comparison depends on the choice of the measure of the forecast error. To detect this possibility of the dependence of the outcome on the measure used and to see at the same time if greater differences between methods are to be found, we also calculated another measure for the forecast error both in ex post simulation and in one-period ahead forecasting. The second measure is the squared error (Christ, 1960), which, in fact, is the squared RMSE. This measure was selected as a concession to propriety, RMSE being the most used measure.

In tables XV(a) and XV(b) we present the results of the calculation of the mean squared errors. A comparison of the figures in table XV(a) with the corresponding values of the MAPE statistic in table VIII reveals that the ranking order is exactly the same and has the same degree of vagueness for both measures of simulation errors. It is reassuring to know that at least in this study the ranking is found to be independent of the measure used.

The various methods of comparison above show once more that OLS estimation competes well empirically with simultaneous methods. If we look at the behaviour in ex post simulations and forecasts, OLS and the MY model do equally well. The other empirical comparison measures are too vague to be used as a basis for decision. If we also examine the analytical asymptotic results, then the MY model performs best.

In the following sections, we continue with the valuation of the model using stochastic simulation as a method of achieving a measure of the uncertainty associated with the ex ante forecast error. For the empirical part of the stochastic simulation, we have, on the basis of the estimation results presented above, chosen the OLS and MY models.

TABLE XV(a)

Relative mean squared error of ex post simulations for the period 1963-82

Equation	Method							
	OLS	BL	YPE	YPR	YP4	YP5	YP6	MY
GDPQ	32.6	21.1	69.0	56.9	57.1	86.7	111.4	24.7
PQ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CQ	59.0	70.6	75.7	66.2	72.3	127.1	154.8	89.5
XQ	345.1	327.9	453.6	415.5	399.1	411.8	430.8	344.5
MTQ	78.9	104.1	179.0	176.3	157.8	199.3	219.0	117.7
LHW	1.9	0.8	3.0	0.9	1.8	2.0	2.3	3.5
PC	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
PW	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.2
PI	0.1	0.7	0.1	0.0	0.1	0.1	0.1	0.1
PG	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2
IVP	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.2
IQC1	15.6	15.1	13.3	12.1	12.1	16.6	17.2	15.0
S6230	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T123H	50.1	117.0	188.0	113.8	156.2	33.8	36.7	125.0
LPH	154.0	353.3	271.0	261.5	224.0	80.1	137.7	126.0
KF	25.7	32.5	33.6	31.1	27.8	26.1	32.2	24.6
VV1Q	43.9	25.7	43.9	43.9	32.0	49.5	55.0	25.3
DLU	2403	3140	4531	3212	3756	2737	1213	5013

TABLE XV(b)

Relative mean squared error of ex post forecast for 1983

Equation	Method							
	OLS	BL	YPE	YPR	YP4	YP5	YP6	MY
GDPQ	14.1	28.3	14.8	62.8	46.6	44.5	37.5	0.6
PQ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CQ	127.7	229.4	73.0	211.6	238.2	322.6	275.8	60.5
XQ	6.7	0.8	16.0	41.2	0.1	0.2	0.6	15.4
MTQ	239.5	308.9	85.9	77.0	97.8	213.0	227.8	226.3
LHW	0.3	0.2	0.4	0.2	0.6	0.4	0.4	0.4
PC	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0
PW	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PG	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
IVP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
IQC1	12.5	13.9	0.3	0.3	0.0	13.9	14.0	12.3
S6230	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T123H	96.3	118.2	178.1	112.5	169.5	115.8	93.6	87.8
LPH	131.6	334.1	108.3	5.7	467.1	418.0	411.9	56.6
KF	1.5	0.4	4.2	11.6	6.9	1.8	0.5	1.7
VV1Q	61.2	59.2	30.9	30.9	59.6	60.1	61.5	59.3
DLU	4177	3519	246	3336	549	1517	1837	8441

V EMPIRICAL RESULTS OF THE VARIANCE MATRIX OF THE FORECAST ERROR

The usual way of forecasting with a simultaneous model is to solve the model over the forecasting period deterministically, that is, replacing the structural disturbances by their expected values. An alternative method which takes into account both the nonlinear and stochastic nature of the model is stochastic simulation, in which the structural disturbances are replaced by stochastic proxies. The stochastic simulations are replicated and the mean of the values from the various replications is taken as the predictor. The difference between the mean and the deterministic solution is a measure of the simulation bias in the model. The variance of the various outcomes of the stochastic replications is a measure of the variance matrix of the forecast error originating from the disturbance term in the stochastic equations. This component has in pervious sections been labelled as the second component of the forecast error. If proxies for the structural parameter estimates are also used in the stochastic simulations, then the variance of the various outcomes of the simulation is a measure of both the first and the second component of the forecast error.

There are four methods of simulation for calculation of the variance of the forecast error. In table A, the applicability and necessary assumptions of the various methods are summarised.

The Schink method does not allow separate calculation of the first and second component of the forecast error; only the total effect of both components can be calculated. The methods of Fair and Bianchi-Calzolari allow calculation of both components separately while the method of Mariano-Brown can only be used to calculate the effect of the second component. Table A also shows that the application of the Mariano-Brown method requires less assumptions about unknown distributions and a very limited number of simulations compared to the others. The accuracy of the

calculations in the methods of Schink and Fair grows with the number of simulations performed. In practice 100 is considered to be a suitable number of simulations. The method of Schink requires most computational work, especially since no ready-made computer program is available.

In the following, we report empirical simulation results both for the consistently estimated MY model and the OLS model obtained using the method of Fair, Bianchi-Calzolari and Mariano-Brown in simulation of the variance matrix of the forecast period 1983-86 in the following way:

(i) The stochastic simulation method of Fair was used both in static one-period simulations and in dynamic multiperiod simulations to calculate the variance matrix of the second error component. A computer program developed by Fair also made it possible to compute the variance matrix of the first error component for the static one-period case. In the calculations of the first error component we used a blockdiagonal form of the variance matrix of the parameter estimates. The number of stochastic simulations used was 100.

(ii) The analytic simulation method of Bianchi and Calzolari, the B-C method, was used to calculate the variance matrix of the second error component. The number of required simulations was the number of the stochastic equations, 16, plus one deterministic simulation. The calculation of the matrix of the partial derivates (see page 73) was labourious and the application of this method to a bigger model is questionable.

TABLE A

The applicability of the simulation methods

	Schink[1]	Fair[2]	B-C[3]	M-B[4]
I error component: parameter randomness				
- can be calculated	I and II together	yes	yes	no
- assumption of normality of parameter estimates needed	no	yes	yes	
- number of simulations needed	~ 100	~ 100	Gxs*	
II error component: residual randomness				
- can be calculated	I and II together	yes	yes	yes
- assumption of normality of the residuals	no	yes	yes	no
- number of simulations needed	~ 100	~ 100	G	T*

*) G = number of stochastic equations, s = number of parameters, T = number of estimation period observations.

1) The method of stochastic simulation and re-estimation (Schink)
2) The method of Monte Carlo on residuals and coefficients (Fair)
3) The method of analytic simulation (Bianchi and Calzolari)
4) The residual-based procedure (Brown and Mariano)

(iii) The residual-based method of Brown and Mariano, the B-M method, was the easiest to apply. The variance matrix of the second component of the forecast error was computed both for the static one-period case and for the dynamic multi-period case. The number of simulations was 20, which is the number of the estimation period residuals for each stochastic equation. These residuals were used in the first simulation year 1983. For the following years of the forecast period, the order of the residuals was determined through resampling of these residuals. For each of the 16 stochastic equations, 3 resamplings of the residuals were performed. This procedure is similar to bootstrapping but using all the residuals.

In other words: the residuals for the four prediction years are the same but the order between the residuals in the calculations for separate years has been determined through resampling. The resampling is perhaps not necessary because the autocorrelation in residuals is only a problem in estimation, not in model solution. On the other hand, there is very little likelihood of getting exactly the same residuals for four consecutive periods in a random sampling.

We also tested the use of antithetic variates in this method.

V.1 The simulation bias

First, we examine the simulation bias, which is the difference between the mean of the results of the stochastic solutions and the deterministic solution. In table B, the empirical measures of the simulation bias for models MY and OLS are presented for the aggregate variables GDP and PQ. In these calculations, the method

of Mariano and Brown was used[1]. Column a gives the deterministic solution point predictors. The following columns, b and c, show the mean of the 20 stochastic residual-based simulations and the absolute values of the difference between the mean of the stochastic simulations and the deterministic simulation, that is, the simulation bias, in per cent values of the deterministic solution. The last columns, d and e, present the corresponding results for 40 stochastic replications; a further 20 have been constructed in the form of antithetic variates.

As can be seen the simulation error does not seem noticeable. This result coincides with findings from stochastic simulation with other models (Fair, 1980) but, according to the previously reported findings of Brown and Mariano, we should be careful in drawing too rigorous conclusions.

V.2 The variance matrix in the static case

Let us then look at the results for the variance matrix calculated from the various replications of stochastic simulation.

In table C the results for three different methods applied to the OLS model are compared. The values in the columns are Pearson's coefficient of variation, that is, the ratio between the standard error of the stochastic simulations and the mean of the same simulations. For example, the first row says that the standard error of the forecast deriving from the stochastic residuals for the aggregate variable GDP is 1.3 per cent of the mean value according to the Mariano-Brown method, 1.2 per cent according to the stochastic simulation method with 100 replications and also 1.2 per cent according to the analytic simulation method. The results for the different methods are remarkably similar. According to the

[1]Similar results for the magnitude of the simulation bias were found using the stochastic simulation method of Fair. For space reasons they are not presented here but are available on request.

TABLE B

Simulation errors in MY and OLS models						
			Simulation with sample size 20		Simulation with antithetic sample size of 40	
		Deterministic simulation a*	Stochastic simulation b*	Error in per cent of deterministic simulation c*	Stochastic simulation d*	Error in per cent of deterministic simulation e*
MY model						
GDPQ	1983	207882	207480	0.19	208337	0.21
	1984	211032	211954	0.43	212498	0.68
	1985	215156	216564	0.65	217143	0.91
	1986	218670	220328	0.75	221066	1.09
PQ	1983	133.53	133.67	0.10	133.53	0
	1984	145.48	145.63	0.10	145.67	0.13
	1985	154.48	155.02	0.54	155.18	0.45
	1986	165.52	163.37	0.52	163.54	0.62
OLS model						
GDPQ	1983	209245	209404	0.08	209287	0.02
	1984	212443	212826	0.18	212742	0.14
	1985	216266	216773	0.23	216711	0.21
	1986	219188	219608	0.19	219746	0.25
PQ	1983	133.37	133.48	0.08	133.41	0.03
	1984	146.42	146.76	0.23	146.46	0.03
	1985	156.42	157.10	0.43	156.56	0.09
	1986	165.43	166.59	0.70	165.71	0.17

a) Deterministic simulation: error terms set equal to zero
b) The mean of 20 stochastic simulations with error terms set to the estimation period values
c) The difference between b and a values in per cent of a
d) The mean of 40 stochastic simulations
e) The difference between d and a values and a values in per cent of a.

TABLE C

Comparison of the static case, 1983, the MA model:
Pearson's coefficient of dispersion

	M-B	Stochastic Simulation	Analytic Simulation
GDPQ	0.013	0.012	0.012
PQ	0.017	0.015	0.016
CQ	0.017	0.018	0.017
XQ	0.074	0.077	0.071
MTQ	0.042	0.037	0.035
LHW	0.010	0.010	0.009
PC	0.018	0.018	0.018
PW	0.017	0.016	0.016
PI	0.014	0.015	0.014
PG	0.021	0.019	0.020
IUP	0.016	0.015	0.016
IQC1	0.036	0.034	0.036
S6230	0.024	0.024	0.028
T123H	0.014	0.015	0.014
LPH	0.029	0.032	0.027
KF	0.005	0.005	0.005
W1Q	0.026	0.017	0.018
DLU	0.805	0.903	0.772

Pearson's coefficient of dispersion is the ratio between the standard error of the stochastic simulations and the mean of the some simulations.

findings of previous studies (Bianchi and Calzolari, 1982), the Schink method, the analytic simulation method and the stochastic simulation methods give similar estimates of the variance matrix. The empirical results reported in table C now indicate that the new method of Mariano and Brown gives results similar to the others. Thus, it is possible to get accurate results with a method that requires much less computational work and is therefore also applicable to bigger models.

V.3 The variance matrix in the dynamic case

In table D are reported the results of the stochastic residual-based simulations of 20 observations of all the endogenous variables in the MY model for the period 1983-86. The figures in the columns are: the mean of the 20 pseudo-forecasts and Pearson's coefficient of variation. Table E gives the corresponding figures for the antithetic case.

Tables F and G show the same figures for the OLS model as are shown in tables D and E for the MY model.

In the OLS model the use of antithetic variates seems to make estimates of the moments of the forecasts converge, which does not happen in the case of the MY model. The tendency towards convergency is very slight and could thus be interpreted in two ways:

- either it indicates empirical evidence for the analytical results of the existence of finite moments for the OLS estimators and the absence of moments for the MY estimators,

- or could, according to Hendry and Harrison (1974), indicate that no substantial reduction in variance is gained from antithetic simulation when the model includes a lagged endogenous variable.

TABLE D

Mean values and Pearson's coefficients of dispersion
for residual based simulations with 20 trials for the MY model

MY Model	1983		1984		1985		1986	
	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$
GDPQ	207 480	0.013	211 954	0.020	216 564	0.024	220 328	0.041
PQ	134	0.016	146	0.030	155	0.042	164	0.048
CQ	114 912	0.023	118 696	0.055	119 523	0.043	120 406	0.051
XQ	53 699	0.065	58 533	0.104	62 247	0.145	64 737	0.174
MTQ	70 255	0.034	72 991	0.060	74 405	0.043	75 658	0.075
LHW	3 363	0.010	3 388	0.017	3 391	0.018	3 393	0.028
PC	133	0.010	144	0.031	154	0.039	163	0.055
PW	138	0.017	153	0.032	166	0.047	178	0.062
PI	128	0.014	138	0.028	147	0.039	154	0.050
PG	136	0.020	148	0.034	160	0.043	170	0.047
IVP	131	0.016	141	0.024	151	0.038	159	0.049
IQC1	14 225	0.036	13 721	0.041	13 036	0.052	12 589	0.056
S6230	11	0.025	12	0.026	12	0.035	12	0.036
T1234	93 012	0.014	107 667	0.042	125 034	0.038	144 886	0.061
LPH	76 997	0.030	93 032	0.091	110 042	0.131	128 190	0.162
KF	378 442	0.005	388 887	0.008	398 555	0.010	408 305	0.013
VVIQ	82 741	0.018	84 971	0.025	85 828	0.035	86 560	0.035
DLU(FS)	5 899	0.688	2 933	0.240	-667	8.078	-271	23.618

TABLE E

Mean values and Pearson's coefficients of dispersion
for residual-based simulations with 40 trials with antithetic variates
for the MY model

MY Model	1983		1984		1985		1986	
	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$
GDPQ	208 337	0.013	212 498	0.025	217 143	0.028	221 066	0.041
PQ	134	0.016	146	0.030	155	0.041	164	0.048
CQ	114 644	0.022	118 598	0.054	119 791	0.044	121 005	0.058
XQ	55 098	0.071	59 283	0.117	62 643	0.144	65 136	0.168
MTQ	70 447	0.032	73 182	0.063	74 783	0.056	76 066	0.080
LHW	3 367	0.010	3 392	0.017	3 396	0.020	3 399	0.029
PC	132	0.017	143	0.030	154	0.038	163	0.053
PW	138	0.017	153	0.032	165	0.047	178	0.062
PI	128	0.014	138	0.027	147	0.039	154	0.050
PG	136	0.020	149	0.034	160	0.044	171	0.048
IVP	131	0.016	141	0.024	151	0.039	158	0.051
IQC1	14 222	0.036	13 718	0.041	13 034	0.052	12 589	0.056
S6230	11	0.025	12	0.025	12	0.035	12	0.035
T1234	93 064	0.014	107 697	0.024	124 989	0.037	144 769	0.059
LPH	76 636	0.030	92 592	0.086	109 511	0.124	128 119	0.173
KF	378 354	0.005	388 857	0.008	398 651	0.010	408 485	0.012
VV1Q	82 753	0.018	85 067	0.025	86 091	0.042	86 889	0.051
DLU	5 052	0.863	2 719	2.392	-243	21.327	-138	45.429

TABLE F

Mean values and Pearson's coefficients of dispersion
for residual-based simulations with 20 trials for the OLS model

OLS Model	1983		1984		1985		1986	
	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$
GDPQ	209 404	0.013	212 826	0.018	216 773	0.024	219 608	0.032
PQ	133	0.017	147	0.032	157	0.050	167	0.060
CQ	115 820	0.017	119 529	0.031	121 417	0.042	122 544	0.048
XQ	55 052	0.074	57 734	0.099	60 854	0.127	62 669	0.131
MTQ	70 620	0.042	73 443	0.057	75 453	0.054	76 539	0.071
LHW	3 362	0.010	3 375	0.017	3 363	0.020	3 352	0.022
PC	132	0.018	144	0.036	156	0.054	166	0.071
PW	138	0.017	154	0.029	168	0.046	182	0.063
PI	128	0.014	138	0.020	148	0.030	156	0.039
PG	136	0.021	150	0.034	163	0.052	175	0.068
IVP	131	0.016	142	0.024	153	0.035	161	0.046
IQC1	14 224	0.036	13 720	0.038	13 026	0.039	12 555	0.046
S6230	11	0.024	12	0.026	12	0.035	12	0.033
T1234	93 170	0.014	108 177	0.017	126 772	0.025	148 714	0.036
LPH	77 866	0.029	96 753	0.057	117 655	0.079	140 599	0.104
KF	378 406	0.005	389 480	0.007	399 771	0.008	409 707	0.009
VV1Q	82 810	0.026	85 776	0.041	86 897	0.040	87 587	0.041
DLUFS	4 282	0.805	3 639	1.029	725	6.029	526	9.388

TABLE G

Mean values and Pearson's coefficients of dispersion
for residual-based simulations with 40 trials with antithetic variates
for the OLS model

OLS Model	1983		1984		1985		1986	
	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$	$\bar{y}$	$\hat{\sigma}/\bar{y}$
GDPQ	209 287	0.013	212 742	0.018	216 711	0.026	219 746	0.032
PQ	133	0.017	146	0.032	157	0.049	166	0.060
CQ	115 840	0.016	119 571	0.030	121 556	0.041	119 304	0.046
XQ	54 768	0.068	57 468	0.099	60 334	0.129	62 212	0.127
MTQ	70 439	0.043	73 122	0.057	75 019	0.056	76 099	0.068
LHW	3 361	0.010	3 374	0.017	3 362	0.020	3 352	0.022
PC	132	0.018	144	0.036	155	0.054	165	0.072
PW	138	0.017	153	0.029	168	0.044	181	0.062
PI	128	0.014	138	0.020	148	0.029	155	0.038
PG	136	0.020	150	0.034	163	0.051	175	0.067
IVP	130	0.016	142	0.024	152	0.035	159	0.046
IQC1	14 218	0.036	13 711	0.038	13 019	0.039	12 550	0.045
S6230	11	0.024	12	0.025	12	0.035	12	0.032
T1234	93 176	0.014	108 191	0.017	126 764	0.024	148 690	0.034
LPH	77 480	0.026	96 551	0.051	117 326	0.071	140 087	0.097
KF	378 388	0.005	389 368	0.007	399 565	0.008	409 388	0.009
VV1Q	82 798	0.025	85 676	0.041	86 720	0.040	87 333	0.039
DLUFS	7 333	0.735	3 538	1.061	783	5.696	519	9.560

TABLE H

The dynamic case. Comparison of the Pearson's coefficients for the MY model for residual-based simulation (20 replications) and stochastic simulation (100 replications)

MY model	Residual-based simulation				Stochastic simulation			
	1983	1984	1985	1986	1983	1984	1985	1986
GDPQ	0.013	0.020	0.024	0.041	0.014	0.022	0.030	0.040
PQ	0.016	0.030	0.042	0.048	0.015	0.031	0.043	0.052
CQ	0.023	0.055	0.043	0.051	0.023	0.033	0.041	0.046
XQ	0.065	0.104	0.145	0.174	0.077	0.105	0.133	0.156
MTQ	0.034	0.060	0.043	0.067	0.038	0.057	0.064	0.086
LHW	0.010	0.017	0.018	0.028	0.011	0.018	0.024	0.027
PC	0.017	0.031	0.039	0.055	0.018	0.034	0.048	0.059
PW	0.017	0.032	0.047	0.062	0.017	0.031	0.046	0.055
PI	0.014	0.028	0.039	0.050	0.015	0.026	0.035	0.044
PG	0.020	0.034	0.043	0.047	0.020	0.036	0.050	0.062
IUP	0.016	0.024	0.038	0.049	0.016	0.028	0.036	0.045
IQC1	0.036	0.041	0.052	0.056	0.037	0.044	0.041	0.042
S6230	0.025	0.026	0.035	0.036	0.025	0.029	0.034	0.033
T123H	0.014	0.024	0.038	0.061	0.015	0.021	0.028	0.037
LPH	0.030	0.091	0.131	0.162	0.027	0.054	0.082	0.095
KF	0.005	0.008	0.010	0.013	0.005	0.008	0.011	0.012
VV1Q	0.018	0.025	0.035	0.035	0.017	0.028	0.038	0.054
DLU	0.688	2.240	8.078	3.618	0.963	2.132	9.064	8.048

Table H shows a comparison between the dynamic simulation results of the residual-based method and the stochastic method. In previous section we saw that the results were very similar between methods in the static case. The findings of the figures in table H are that the results for dynamic simulations are also very close to one another. The new Mariano-Brown method also competes well with the more complicated methods in the dynamic case.

V.4 The variance matrix of the first and second component of the forecast error

The availability of the Fair computer program made it possible to also calculate the effects of the first error component in static one-period simulations. The OLS and MY models were solved 100 times, each time using proxies for both the random disturbance and the parameter estimates. The variances were calculated from the 100 replications and the Pearson's coefficients for the trials are presented in table I. The first thing to observe is that the results between the two models are again very similar. The figures in columns b show that the standard error grows when the randomness of the parameter estimates is also included in the calculations. The table also shows that the effects of the second component, in column a, are markedly greater than the effects of the first component. This results coincides with previous empirical comparisons (Bianchi and Calzolari, 1982). Thus the measurement of uncertainty in macromodels could be based on only the second component without any great loss of accuracy.

V.5 Stochastic simulation of the naive model

The naive model, unlike the structural model, is linear both in parameters and variables. Consequently there is no simulation bias; for this model the deterministic forecasts coincide with the means of the stochastic simulations. To obtain the variances for the forecasts over the prediction period we used the residual-based method, which according to the findings so far in this study is the easiest to apply although the results are similar to those of the more complicated stochastic simulation methods. The calculated values of Pearson's coefficients are presented in table J. When the figures in table J are compared with the corresponding figures in table D for the MY model, we can conclude that the forecasts from the simultaneous model have greater variances than the forecasts from the naive model. Nor does the variance grow with the time horizon in the naive model as obviously as in the structural model. The non-simultaneous forecasts from the naive model have a smaller dispersion around their mean than the forecasts from the simultaneous structural model.

V.6 The density fuction of the forecasts

In the previous sections we have empirically calculated the mean and the variance of the forecasts of the endogenous variables in the structural and naive models. These two first moments give some information about the distribution of the forecasts but they are not enough to sufficiently characterize the joint distribution. When the distribution remains unknown the prediction regions are impossible to construct.

To get an idea of what the marginal distributions are like we plotted the values of the endogenous variables from the 20 stochastic residual-based simulations for the OLS and MY models as

TABLE I

The effects of the first and second sources of forecast error; static case, stochastic simulation 100 trials

	Model OLS		Model MY	
	a	b	a	b
GDPQ	0.012	0.014	0.014	0.018
PQ	0.015	0.018	0.015	0.020
CQ	0.017	0.022	0.023	0.048
XQ	0.071	0.085	0.077	0.077
MTR	0.035	0.038	0.038	0.048
LHW	0.009	0.022	0.011	0.022
PC	0.018	0.022	0.018	0.021
PW	0.016	0.020	0.017	0.020
PT	0.014	0.015	0.015	0.019
PG	0.020	0.024	0.020	0.024
IUP	0.016	0.019	0.016	0.020
IQC1	0.036	0.042	0.037	0.043
S6230	0.028	0.035	0.025	0.033
TI23H	0.014	0.021	0.015	0.022
LPH	0.027	0.036	0.027	0.052
KF	0.005	0.005	0.005	0.006
VVIQ	0.018	0.021	0.017	0.023
DLUFS	0.772	0.968	0.963	0.986

a) Stochastic simulation with respect to error terms only

b) Stochastic simulation with respect to error terms and coefficient estimates

TABLE J

Pearson's coefficient for the dynamic naive model simulations; residual-based procedure

	1983	1984	1985	1986
GDPQ	0.012	0.019	0.017	0.017
PQ	0.011	0.015	0.017	0.018
CQ	0.014	0.020	0.017	0.018
XQ	0.037	0.050	0.042	0.043
MTQ	0.036	0.044	0.037	0.049
LHW	0.009	0.013	0.010	0.014
PC	0.009	0.022	0.027	0.026
PW	0.008	0.016	0.021	0.017
PI	0.014	0.018	0.017	0.020
PG	0.008	0.017	0.022	0.027
IUP	0.016	0.023	0.022	0.023
IQC1	0.038	0.042	0.056	0.067
S6230	0.037	0.037	0.043	0.025
T123H	0.004	0.005	0.005	0.005
LPH	0.005	0.010	0.012	0.015
KF	0.003	0.007	0.008	0.011
VV1Q	0.026	0.033	0.035	0.025
DLU	0.149	0.401	10.378	0.210

graphical interpretations of the empirical results of an unknown distribution of the aggregate variable GDP. Figure I shows the graphs of the density functions of the 20 observations of GDP for the forecast period 1983-1986 for both models. The graphs are drawn as frequency functions using suitable intervals around the indicated mean.

As can be seen, the functions are really too complicated to be sufficiently characterized by the first two moments. Most of them have several local optimum points.

When the true distributions are not known, the normal distribution could perhaps be used to give an approximation of the tolerance intervals of the forecasts.

V.7 How the forecasts should be reported

The core of this study is the fact that a forecast from a stochastic model is itself a stochastic variable. In that case, it is not enough only to report the point forecast. The dispersion around this point should also be indicated.

Results from several studies, including this one, have shown that the simulation bias in non-linear models is negligible. This means that the common practice of using deterministic simulations to compute the point forecasts is, with some caution, justified. For every model, however, stochastic simulation should be used at least once to provide the dynamic variances of the forecasts. These variances should then be reported along with the deterministic forecast to measure the uncertainty in the point predictions. The finding of this study is that the results of the residual-based simulation method are as accurate as the results of the more complicated ones. The easiness of the use of this new method enables the calculation of the variance for most models.

Here we propose two ways of reporting the forecasts. They are illustrated in tables K and L.

FIGURE 1 Density functions of pseudo-random forecasts of GDP from the residual-based procedure

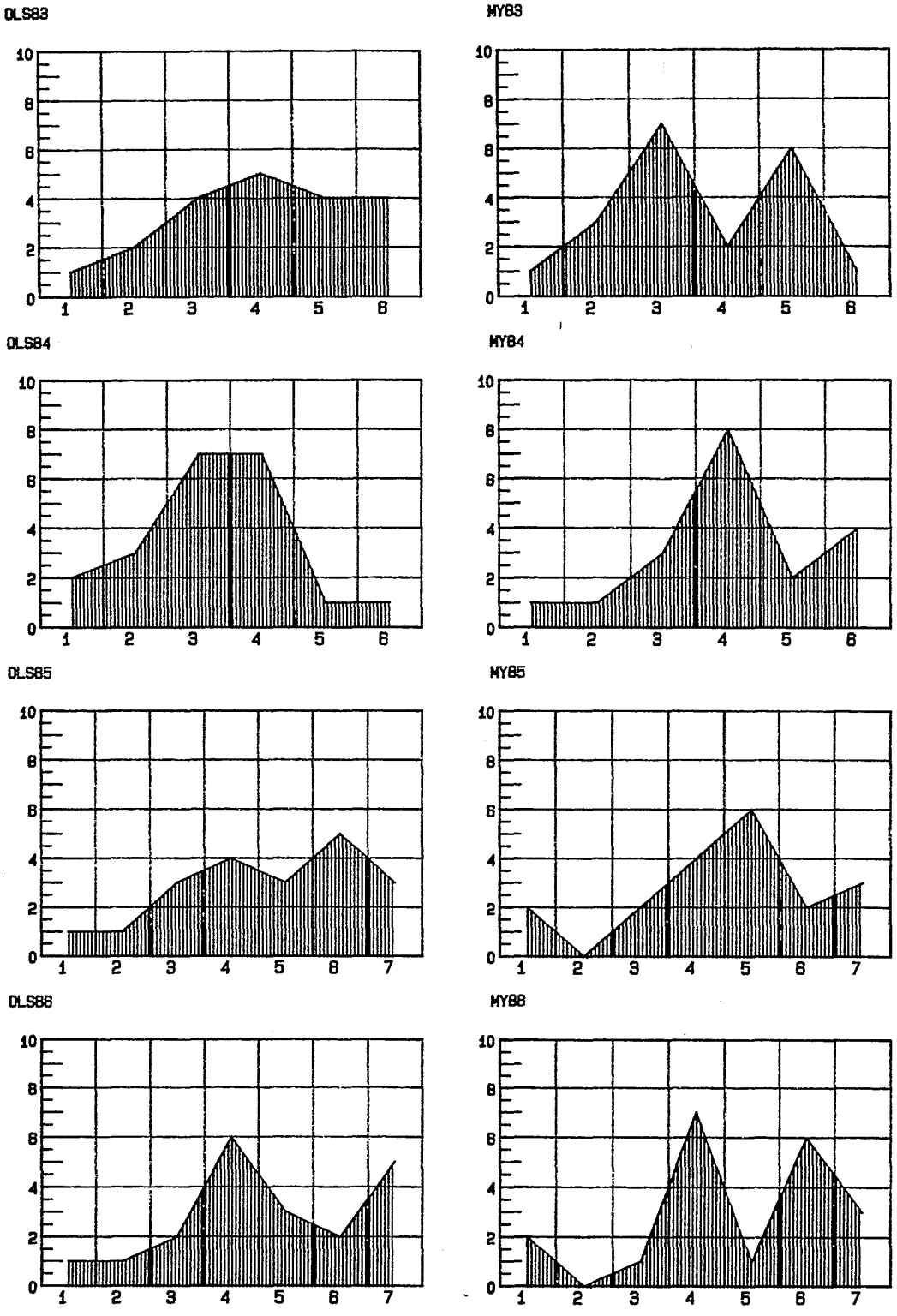


TABLE K

Composite prediction for the year 1983; the MY model and the naive AR4 model, per cent values, standard errors below

	The structural model MY	The AR4 model	Composite prediction
GDPQ	3.1	3.4	3.2
S.E	1.3	1.2	
PQ	9.8	8.5	9.3
S.E	1.6	1.1	
CQ	4.3	3.8	4.1
S.E	2.3	1.4	
XQ	3.2	-3.1	1.2
S.E	6.5	3.7	
MTQ	9.8	7.5	9.0
S.E	3.4	3.6	
LHW	1.2	0.5	1.0
S.E	1.0	0.9	
PC	8.5	7.9	8.3
S.E	1.7	0.9	
PW	10.5	8.9	9.9
S.E	1.7	0.8	
PI	8.7	9.1	8.8
S.E	1.4	1.4	
PG	9.6	8.2	9.1
S.E	2.0	0.8	
IUP	9.5	9.0	9.3
S.E	1.6	1.6	
IQC1	-2.2	7.5	0.9
S.E	3.6	3.8	
S6230	5.4	-1.5	2.7
S.E	2.5	3.7	
T123H	16.8	12.3	15.3
S.E	1.4	0.4	
LPH	19.2	18.3	18.8
S.E	3.0	0.5	
KF	2.9	2.6	2.8
S.E	0.5	0.3	
W1Q	3.0	2.1	2.7
S.E	1.8	2.6	

TABLE L

Forecasts and forecast standard errors for the MY model, per cent values.

MY model		1983	1984	1985	1986
GDPQ	Forecast	3.1	2.1	2.2	1.7
	S.E.	1.3	2.0	2.4	4.1
PQ	Forecast	9.8	9.1	6.5	5.5
	S.E.	1.6	3.0	4.2	4.8
CQ	Forecast	4.3	3.2	0.7	0.7
	S.E.	2.3	5.5	4.3	5.1
XQ	Forecast	3.2	9.0	6.3	4.0
	S.E.	6.5	10.4	14.5	17.4
MTQ	Forecast	9.8	3.9	1.9	1.7
	S.E.	3.4	6.0	4.3	7.5
LHW	Forecast	1.2	0.7	0.1	0.0
	S.E.	1.0	1.7	1.8	2.8
PC	Forecast	8.5	8.4	7.4	5.8
	S.E.	1.7	3.1	3.9	5.5
PW	Forecast	10.5	10.7	8.5	7.7
	S.E.	1.7	3.2	4.7	6.2
PI	Forecast	8.7	7.9	6.9	4.9
	S.E.	1.4	2.8	3.9	5.0
PG	Forecast	9.6	9.1	7.7	6.5
	S.E.	2.0	3.4	4.3	4.7
IVP	Forecast	9.5	8.2	7.1	4.8
	S.E.	1.6	2.4	3.8	4.9
IQC1	Forecast	-2.2	-3.4	-5.0	-3.4
	S.E.	3.6	4.1	5.2	5.6
S6230	Forecast	5.4	6.4	1.6	0.7
	S.E.	2.5	2.6	3.5	3.6
T123H	Forecast	16.8	15.8	16.1	15.9
	S.E.	1.4	2.4	3.8	6.1
LPH	Forecast	19.2	20.8	18.3	16.5
	S.E.	3.0	9.1	13.1	16.2
KF	Forecast	2.9	2.8	2.5	3.4
	S.E.	0.5	0.8	1.0	1.3
VV1Q	Forecast	3.0	2.7	1.0	0.8
	S.E.	1.8	2.5	3.5	3.5

First we use the concept of composite prediction to utilize the summed information in the structural and naive models.

The first column in table K shows the predicted growth values in per cent for the endogenous variables in the structural model MY. Below the predicted growth values are the standard errors in per cent of the mean. In the second column are given the corresponding figures for the naive model AR4. In the third column is the composite prediction, which is the weighted average of the structural forecast and the naive forecast. The estimated weights were presented in table XIV (page 128).

The second way of reporting the forecasts is illustrated in table L. The dynamic forecasts of the endogenous variables (in per cent unity) and their standard errors for the MY model are listed.

The ultimate objective of this study was to calculate the variance matrix of the forecasts, which enables a presentation of simultaneous stochastic model forecasts as in tables K and L.

VI CONCLUSIONS

The main finding of this study is to recommend the IIV iterative instrumental estimation method for consistent estimation of structural parameters in a simultaneous model and the residual-based stochastic simulation method for calculation of the mean and the variance of the forecasts.

The IIV estimation method produces consistent estimators unless lagged endogenous variables and autocorrelated residuals coexist. The estimators are efficient relative to other applicable estimation methods since the method takes into consideration full specification of the model. It solves in a very simple way both the problems associated with the undersized sample and the nonlinearity of the unknown reduced form. In the IIV estimation an initial OLS estimation of the system is used as starting point. The better the OLS-estimators are, in terms of the magnitudes of bias and variance, the better are the IIV estimators. The IIV estimation is also easy to apply and the additional round of estimation that has to be done after the inevitable OLS estimation is well compensated for by the improvement in the quality of the parameter estimates.

Among the stochastic simulation methods we found the new residual-based procedure to be the easiest to apply. It has the drawback of allowing measurement only of the variance in the prediction that comes from the disturbance term. No measurement of the forecast uncertainty deriving from the use of parameter estimates is possible. Empirical comparisons have shown, however, that the major part of the forecast uncertainty is attributable to stochastic disturbances. The drawback may not be of major importance and it is compensated for by the simplicity in the computations and the avoidance of constraining assumptions of unknown distributions which, if incorrect, introduce bias in the calculations. When this method is not based on a Monte Carlo

procedure the risk of sampling error is also avoided. The possibility of nonconvergence of the mean of the replicated stochastic simulations does not show up because the residuals calculated from the sample period are free from outliers. The empirical results of the stochastic simulation methods indicate that this new method of Mariano and Brown gives results similar to the others. Thus, it is possible to get accurate results with a method that requires much less computational work and is therefore also applicable to bigger models.

The estimation and stochastic simulation were applied to the KT-model of the Bank of Finland. We found that the simulation bias was negligible and hence the usual practise of conducting forecasts via deterministic solution is acceptable. The magnitudes of the variance of the static one period forecasts coincide with the findings from stochastic simulation with other models (Fair, 1980). The variances of the dynamic forecasts over the whole forecast period 1983-86 are, however, greater than the dynamic variances for other models. The rapid growth of the variances with time is explained by the inclusion of a lagged endogenous variable in almost every equation. This implies that more attention should be attached to the dynamic behaviour and the long term properties of the model.

In this paper stochastic simulation was used to evaluate the predictive accuracy of the KT-model. A natural extension of this analysis would be to use stochastic simulation to calculate the mean and the variances of the multipliers measuring the effects of policy simulations in the model. In the stochastic simulation of the variance of the forecasts we have to consider both the variance of the parameter estimates and of the stochastic disturbance term. In a stochastic simulation of the variance of the multipliers only the variance of the parameter estimates needs to be considered. According to the findings of this study, and also of other papers, the variance of the parameters is much less than the variance of the stochastic disturbances. This implies that although the forecast level, having a great variance, could be wrong the

multipliers, having small variances, can be assumed to be correct. This means that we can have much greater confidence in the policy simulation properties of the model than in its ability to produce base forecasts. Against this conclusion based on the results of stochastic simulation we can set the Lucas critique. Lucas (1976) argued that model simulations are not reliable for evaluating the effects of policy rules because the constant parameters estimated from historical data change with the policy. Simulations provide no useful information about the effects of alternative economic policies and consequently models should be used only for forecasting. The choice between using modelsolutions in forecasting or in policy evaluations has to be made in another forum.

REFERENCES

AHLSTEDT, M. and VIRÉN, M. (1984), "Kansantalouden osaston vuosimallin uudistettu versio. Discussion paper 21/84. Bank of Finland.

ALMON, S. (1965), "The Distributed Lag between Capital Appropriations and Net Expenditures." Econometrica, Vol. 33. January.

AMEMIYA, T. (1966), "Specification Analysis in the Estimation of Parameters of a Simultaneous Equations Models With Autocorrelated Residuals." Econometrica, 34. April.

BASMANN, R.L. (1961), "A Note on the Exact Finite Sample Frequency Function of Generalized Classical Linear Estimators in Two Leading Overidentified Cases." Journal of the American Statistical Association, Vol. 56. September.

BIANCHI, C., CALZOLARI, G. and CORSI, P. (1979), "A Package for Analytical Simulation of Econometric Models." Optimization Techniques. Proceedings of the 9th IFIP Conference on Optimization Techniques Warsaw, September 4-8.

BIANCHI, C., CALZOLARI, G. and CORSI, P. (1980), "Standard Errors of Multipliers and Forecasts from Structural Coefficients with Block-diagonal Covariance Matrix." Dynamic Modelling and Control of National Economies (IFAC) Warsaw.

BIANCHI, C. and CALZOLARI, G. (1980), "The One-period Forecast Errors in Nonlinear Econometric Models." International Economic Review Vol. 21. February.

BIANCHI, C. and CALZOLARI, G. (1982), "Evaluating Forecast Uncertainty due to Errors in Estimated Coefficients: Empirical Comparison of Alternative Methods." Evaluating the Reliability of Macro-Economic Models. Wiley, New York.

BIANCHI, C. and CALZOLARI, G. (1983), "Standard Errors of Forecasts in Dynamic Simulation of Nonlinear Econometric Models: Some Empirical Results." Time Series Analysis: Theory and Practice 3. O.D. Andersson. North Holland Publishing Company.

BROWN, B.W. and MARIANO, R.S. (1983), "Asymptotic Behaviour of Predictors in a Nonlinear Simultaneous System." International Economic Review, Vol. 24. October.

BROWN, B.W. and MARIANO, R.S. (1984), "Residual-based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System." Econometrica, Vol. 52. March.

BROWN, B.W. and MARIANO, R.S. (1985), "Finite-Sample Properties of Stochastic Predictors in Nonlinear Systems: Some Initial Results." Econometric Society Meeting, Boston.

BRUNDY, J.M. and JORGENSON, D.W. (1971), "Efficient Estimation of Simulataneous Equations by Instrumental Variables." The Review of Economics and Statistics, Vol. 53. August.

CALZOLARI, G. (1979), "Antithetic Variates to Estimate the Simulation Bias in Non-linear Models." Economics Letters, Vol. 4.

COOPER, J.P. and FISCHER, S. (1974), "Monetary and Fiscal Policy in the Fully Stochastic St. Louis Econometric Model." Journal of Money, Credit and Banking 6.

CHRIST, C.F. (1960), "A Symposium of Simultaneous Equation Estimation: Simultaneous Equation Estimation: Any Verdict Yet?." Econometrica, Vol. 28. October.

DHRYMES, P.J. (1971), "A Simplified Structural Estimation Method for Large-Scale Econometric Models." Australian Journal of Economics 13.

DHRYMES, P.J., BERNER, R. and CUMMINS, D. (1974), "A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models with Autocorrelated Errors." Econometrica, Vol. 42, March

DHRYMES, P.J. and PANDIT, V. (1971), "The Asymptotic Distribution of the First Iterate of the Iterated Two Stage Least Squares Estimates." Journal of the American Statistical Association.

DUTTA, M. and LYTTKENS, E. (1974), "Iterative Instrumental Variables Method and Estimation of a Large Simultaneous System." Journal of the American Statistical Association, Vol. 69. December.

EDGERTON, D.L. (1972), "Some Properties of Two Stage Least Squares as Applied to Nonlinear Models." International Economic Review, Vol. 13. February.

FAIR, R. (1970), "The Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and First Order Serially Correlated Errors." Econometrica, Vol. 38. May.

FAIR, R. (1972), "Efficient Estimation of Simultaneous Equations with Autoregressive Errors by Instrumental Variables." Review of Economics and Statistics, Vol. 54.

FAIR, R.C. (1973), "A Comparison of Alternative Estimators of Macroeconomic Models." International Economic Review, Vol. 14. June.

FAIR, R.C. (1980), "Estimating the Expected Predictive Accuracy of Econometric Models." International Economic Review 21. June.

FAIR, R.C. (1984), "Specification, Estimation and Analysis of Macroeconometric Models." Harvard University Press. Cambridge, Massachusetts and London England.

FISHER, F.M. (1961), "On the Cost of Approximate Specification in Simultaneous Equation Estimation." Econometrica, Vol. 29. April.

FISHER, F.M. (1965), "Dynamic Structure and Estimation in an Economy-Wide Econometric Model." The Brooking Quartely Econometric Model of the United States. Edited by J.S. Dusenberry, G. Fromm, L.R. Klein and E. Kuh. Chicago: Rand Mcnall & Co.

FRASER, D.A.S. and GUTTMAN, I. (1956), "Tolerance Regions." Annals of Mathematical Statistics, Vol. 27. March.

GOLDBERGER, A.S., NAGAR, A.L. and ODEH, H.S. (1961), "The Covariance Matrices of Reduced-Form Coefficients and of Forecasts For a Structural Econometric Model." Econometrica, Vol. 29. October.

GOLDFELD, S.M. and QUANDT, R.E. (1968), "Nonlinear Simultaneous Equations: Estimation and Prediction." International Economic Review, Vol. 9. February.

HAITOVSKY, Y. and WALLACW, N. (1972), "A Study of Discretionary and Nondiscretionary Monetary and Fiscal Policies in the Context of Stochastic Macroeconometric Models." The Business Cycle Today. Ed: V. Zarnowitz, NBER, New York.

HATANAKA, M. (1976), "Several Efficient Two-Step Estimators for the Dynamic Simultaneous Equations Model with Autoregressive Disturbances." Journal of Econometrics, Vol. 4. May

HATANAKA, M. (1978), "On the Efficient Estimation Methods for the Macro-Economic Models Nonlinear in Variables." Journal of Econometrics, Vol. 8. December.

HENDRY, D.F. (1976), "The Structure of Simultaneous Equations Estimators." Journal of Econometrics, Vol. 4. February.

HENDRY, D.F. and HARRISON, R.W. (1974), "Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-stage Least Squares." Journal of Econometrics, vol. 2.

HENDRY, D.F. and SRBA F. (1977), "The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems." Econometrica, Vol. 45. May.

HOERL, A.E. and KENNARD, R.W. (1970), "Ridge regression: Biased Estimation for Nonorthogonal problems." Technometrics 12.

HOOPER, J.W. and ZELLNER, A. (1961), "The Error of Forecast for Multivariate Regression Models." Econometrica, Vol. 28. October.

HYMANS, S.H. (1968), "Simultaneous Confidence Intervals in Econometric Forecasting." Econometrica, Vol. 36. January.

KELEJIAN, H.H. (1971), "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables." Journal of the American Statistical Association, Vol. 66. June.

KLOEK, T. and MENNES, L.B.M. (1960), "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables." Econometrica, Vol. 28. January.

KLEIN, L.R. (1956), "A Textbook of Econometrics. Row, Peterson and Company.

KLEIN, L.R. (1960), "The Efficiency of Estimation in Econometric Models." Essays in Economics and Econometrics, 7.

KLEIN, L.R. (1960), "Single Equation vs. Equation System Methods of Estimation in Econometrics." Econometrica, Vol. 28. October.

KLEIN, L.R. and NAKAMURA, M. (1962), "Singularity in the Equation Systems of Econometrics: Some Aspects of the Problem of Multicollinearity." International Economic Review.

KLEIN, L.R. (1969), "Estimation of Interdependent Systems in Macro-econometrics." Econometrica, Vol. 37. April.

KNIGHT, J.L. (1985), "The Moments of OLS and 2SLS when Disturbances are Non-normal." Journal of Econometrics Vol. 27.

KORKMAN, S. (1980), "Malliraportti." Unpublished paper. Bank of Finland.

LIU, T.C. (1955), "A Simple Forecasting Model for the U.S. Economy." International Monetary Fund, Staff Papers, August.

LUCAS, R.E. (1976), "Econometric Policy Evaluation: A Critique". In the Phillips curve and labor markets, ed. K. Brunner and A.H. Meltzer, pp. 14-46. Amsterdam: North Holland Publishing Company.

MADDALA, G.S. (1977), "Econometrics". McGraw-Hill Kogakusha, Ltd. Tokyo.

MARIANO, R.S. (1982), "Analytical Small Sample Distribution Theory in Econometrics: The Simultaneous Equation Case." International Economic Review, Vol. 23. October.

MICHAIL, W.M. (1972), "Simulating the Small-sample Properties of Econometric Estimators." Journal of the American Statistical Association, Vol. 67.

MICHAIL, W.M. (1975), "A Comparative Monte Carlo Study of the Properties of Econometric Estimators." Journal of the American Statistical Association, Vol 70, March.

MOZZAMI, B. and BUSE, A. (1984), "The Small Sample Performance of Some Limited Information Estimators of a Dynamic Structural Equation with Autocorrelated Errors." The European Meeting of the Econometric Society, Madrid. September.

NAGAR, A.L. (1959), "The Bias and Moments Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations." Econometrica, Vol 27. October.

NAGAR, L. (1960), "A Monte Carlo Study of Alternative Simultaneous Equation Estimators." Econometrica, Vol. 28. July.

NELSON, C.R. (1972), "The Prediction of the FRB-MIT-PENN Model of the U.S. Economy." The American Economic Review, Vol. 5.

PHILLIPS, P.C.B. (1983), "Exact Small Sample Theory in the Simultaneous Equation Model." Handbook of Econometrics, Vol I. North Holland Publishing Company.

RANTALA, O. (1982), "KT-osaston vuosimalli." Discussion paper 7/82. Bank of Finland.

SALMON, M. and WALLIS, K.F. (1982), "Model Validation and Forecast Comparisons: Theoretical and Practical Considerations." Evaluating the Reliability of Macro-Economic Models. Wiley, New York.

SARGAN, J.D. (1961), "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive residuals." Econometrica, Vol 29. July.

SARGAN, J.D. (1976), "Econometric Estimators and the Edgeworth Approximation." Econometrica, Vol. 44. May.

SCHINK, G.R. (1971), "Small Sample Estimates of the Variance Covariance Matrix of Forecast Error for Large Econometric Models: The Stochastic Simulation." Ph.D. Dissertation, University of Pennsylvania.

SCHMIDT, P. (1977), "Some Small Sample Evidence of the Distribution of Dynamic Simulation Forecasts." Econometrica Vol 45. May.

SPANOS, A. (1984), Lectures "Recent Advances in Econometric Modelling: Causality, Time Series Methods and Computer Applications." S. Margarita.

SUMMERS, R. (1965), "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators." Econometrica Vol. 33. January.

SWAMY, P.A.V.B. (1980), "A Comparison of Estimators for Undersized Samples." Journal of Econometrics, Vol. 14. October.

THEIL, H. (1971), "Principles of Econometrics." New York: John Wiley.

THEIL, H. (1983), "Linear Algebra and Matrix Methods in Econometrics." Handbook of Econometrics.

WAGNER, H.M. (1958), "A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations." Econometrica, Vol. 26.

WILSON, A.A. (1967), "An Approach to Simultaneous Tolerance Intervals in Regression." Annals of Mathematical Statistics Vol 38. December.

WOLD, H. (1965), "A Fix-Point Theorem with Econometric Background: Parts I and II." Arkiv för Matemativ, pp. 209-240.

WOLD, H. in association with JUREEN, L. (1953), "Demand Analysis." New York, John Wiley and Sons.

APPENDIX I

Variables in the KT-model

CQ	private consumption expenditure
CR	cash reserve deposit ratio
D6368	dummy for years 1963 - 1968
D78	dummy for years 1978 - 1982
DLUFS	short-term foreign debt of firms and banks
GBH	central government bonds held by public
GDP	gross domestic product
GQ	central government consumption expenditure
GROWTH	growth of production, lagged five-year-moving average
IF	investment, firms
IQ	fixed investment, firms
IQC	residential investment
IVP	residential investment prices
IVQ	inventory investment and statistical discrepancy
KF	stock of inventories
LHW	paid labour input
LN	employment, 1000 persons
LPF	bank loans to firms
LPH	bank loans to households
LTF	loans of firms
LU	unemployment rate
MC	call money rate
MSVT	imports of Finland's major export countries
MTQ	imports of goods and services
MTV	imports of goods and services
PC	private consumption prices
PG	central government consumption prices
PI	fixed investment prices
PM	import prices

PMR	import prices of oil
POV	bank deposits of the public
PQ	GDP-deflator
PW	wage rate
PX	export prices
R	bank lending rate
RO	interest rate on central government bonds
RU	interest rate on 3 month eurodollar deposits
SVTUL	unit labor costs of Finland's major export countries
TFF	long-term foreign debt, net
TH	time deposits of households
THR	household tax rate
TREND	linear trend
VVQ	stock of inventories
XQ	exports of multilateral goods and all exports of services
XTQ	exports of goods and services
XTV	exports of goods and services
YF	disposable income of firms
YH	disposable income of households

APPENDIX II

Parameter estimates OLS, BLS, YP4, YP5, YP6 and MY.

DLOGXQ=	1.234 DLOG MSVT	-0.513 DLOG(PWX/SVTUL)
	(0.232)	(0.261)
	1.246	-0.661
	(0.234)	(0.295)
	1.265	-0.904
	(0.248)	(0.519)
	1.258	-0.820
	(0.242)	(0.402)
	1.255	-0.782
	(0.240)	(0.388)
	1.219	-0.319
	(0.236)	(0.289)

DMTQ=	0.638DIQ+	0.668DIVQ+	0.288DCQ+	0.420DXTQ+	6245(PQ/PM)
	(0.095)	(0.048)	(0.062)	(0.054)	(2537)
	0.626	0.635	0.367	0.354	5638
	(0.125)	(0.058)	(0.090)	(0.080)	(3068)
	0.738	0.642	0.182	0.554	8966
	(0.821)	(0.123)	(0.650)	(0.704)	(17869)
	0.673	0.661	0.258	0.474	9553
	(0.694)	(0.093)	(0.526)	(0.574)	(15516)
	0.673	0.661	0.258	0.474	9553
	(0.694)	(0.093)	(0.526)	(0.574)	(15516)
	0.650	0.660	0.422	0.263	3875
	(0.168)	(0.114)	(0.141)	(0.152)	(4527)

```
LOG(LHW/GDPQ!0.5!)= -0.304  +0.923LOG(LHW/GDPQ!0.5!)1!!
                    (0.266) (0.071)

                    -0.157   0.962
                    (0.293) (0.079)

                    -0.367   0.906
                    (0.459) (0.124)

                    -0.271   0.931
                    (0.444) (0.120)

                    -0.255   0.936
                    (0.328) (0.088)

                    -0.534   0.860
                    (0.306) (0.082)

-0.098LOG(PWS/PQ) -0.205DLOG(GDPQ)!0.5! -0.232DLOG(GDPQ)!1.5!
(0.098)           (0.112)               (0.111)

-0.043            -0.232                -0.220
(0.109)           (0.120)               (0.114)

-0.121            -0.212                -0.227
(0.172)           (0.159)               (0.125)

-0.086            -0.252                -0.209
(0.167)           (0.152)               (0.123)

-0.081            -0.248                -0.211
(0.122)           (0.130)               (0.116)

-0.186            -0.256                -0.190
(0.114)           (0.126)               (0.119)

DLOGPW= 0.772DLOG(PC!0.5!)+0.055DLOGPX
       (0.074)             (0.042)

        0.755               0.054
       (0.076)             (0.042)

        0.727               0.052
       (0.153)             (0.043)

        0.740               0.053
       (0.078)             (0.042)

        0.755               0.054
       (0.077)             (0.042)

        0.752               0.054
       (0.081)             (0.042)
```

```
 0.532D(YF/GDPV)!1! +1.241GROWTH +1.524DTHR
(0.204)             (0.203)      (0.395)

 0.520               1.282        1.531
(0.204)             (0.207)      (0.395)

 0.501               1.345        1.515
(0.226)             (0.375)      (0.408)

 0.509               1.316        1.522
(0.206)             (0.211)      (0.397)

 0.520               1.280        1.531
(0.204)             (0.209)      (0.396)

 0.518               1.287        1.530
(0.205)             (0.216)      (0.396)

DLOGPC= 0.337DLOG(PC!1!) -0.468DLOG(GDPQ/LHW)
       (0.076)           (0.135)

        0.349            -0.438
       (0.080)           (0.157)

        0.418            -0.276
       (0.094)           (0.199)

        0.403            -0.349
       (0.089)           (0.183)

        0.404            -0.350
       (0.089)           (0.182)

        0.378            -0.332
       (0.087)           (0.179)

 0.503DLOGPWS +0.187DLOGPM -0.036(MC-R)!2!
(0.076)       (0.039)      (0.030)

 0.483         0.191       -0.037
(0.086)       (0.040)      (0.031)

 0.367         0.215       -0.037
(0.1130)      (0.045)      (0.030)

 0.400         0.210       -0.035
(0.105)       (0.0433)     (0.033),

 0.399         0.210       -0.035
(0.105)       (0.043)      (0.030)

 0.426         0.202       -0.038
(0.101)       (0.042)      (0.030)
```

DLOGPI= 0.607DLOGPWS+ 0.391DLOGPM -0.172(MC-R)!1.5!

(0.052)	(0.042)	(0.045)
0.613	0.388	-0.002
(0.052)	(0.043)	(0.045)
0.621	0.383	-0.179
(0.057)	(0.045)	(0.050)
0.621	0.383	-0.179
(0.054)	(0.044)	(0.046)
0.622	0.382	-0.180
(0.054)	(0.044)	(0.050)
0.632	0.377	-0.185
(0.053)	(0.043)	(0.050)

DLOGPG= 0.088DLOG(PG!1!) + 0.864DLOGPW + 0.110DLOGPM -0.281DLOG(GQ)

(0.084)	(0.113)	(0.040)	(0.162)
0.078	0.884	0.107	-0.297
(0.094)	(0.138)	(0.043)	(0.175)
0.100	0.842	0.115	-0.262
(0.109)	(0.176)	(0.048)	(0.198)
0.112	0.818	0.119	-0.241
(0.109)	(0.175)	(0.048)	(0.198)
0.112	0.818	0.119	-0.241
(0.108)	(0.174)	(0.048)	(0.197)
0.220	0.611	0.160	-0.067
(0.122)	(0.194)	(0.054)	(0.222)

```
DLOGIVP=-0.259DLOG(IVP)-1 + 0.450DLOG  +0.850DLOGPWS
        (0.103)             (0.050)     (0.107)

        -0.291               0.446       0.892
        (0.113)             (0.050)     (0.122)

        -0.331               0.442       0.942
        (0.129)             (0.052)     (0.144)

        -0.315               0.444       0.921
        (0.123)             (0.051)     (0.135)

        -0.288               0.447       0.888
        (0.118)             (0.050)     (0.129)

        -0.290               0.446       0.891
        (0.114)             (0.050)     (0.122)

 -0.109(MC-R)!1! -0.142(MC-R)!2!
 (0.042)         (0.039)

 -0.116          -0.148
 (0.042)         (0.039)

 -0.124          -0.153
 (0.040)         (0.040)

 -0.121          -0.151
 (0.040)         (0.040)

 -0.115          -0.147
 (0.040)         (0.040)

 -0.115          -0.147
 (0.040)         (0.040)

LOGIQC1= 3.038 + 0.656LOG(IQC1)!1! + 0.086LOG(GLH/IVP)
        (0.764) (0.091)              (0.034)

         3.131   0.665                0.095
        (0.816) (0.097)              (0.038)

         2.264   0.751                0.059
        (1.139) (0.137)              (0.055)

         3.200   0.636                0.105
        (0.905) (0.108)              (0.044)

         3.204   0.636                0.105
        (0.889) (0.106)              (0.043)

         3.054   0.654                0.090
        (0.769) (0.091)              (0.035)
```

```
 0.312DLOGIVP -0.277(MC-R)!1!  -0.379(MC-R)!2!
(0.231)       (0.100)          (0.100)

 0.240        -0.271           -0.387
(0.273)       (0.115)          (0.119)

-0.062        -0.284           -0.412
(0.340)       (0.130)          (0.130)

 0.083        -0.262           -0.404
(0.310)       (0.120)          (0.120)

 0.086        -0.262           -0.403
(0.283)       (0.120)          (0.120)

 0.258        -0.275           -0.385
(0.244)       (0.110)          (0.120)

LOGCQ= 0.884 + 0.667LOG(YH/PC) + 0.245LOG(YH/PC)!1!
      (0.155) (0.149)            (0.141)

        0.948   0.600              0.300
       (0.214) (0.022)            (0.022)

        1.109   0.594              0.297
       (0.218) (0.020)            (0.020)

        1.139   0.592              0.296
       (0.229) (0.021)            (0.021)

        1.149   0.592              0.296
       (0.227) (0.021)            (0.021)

        1.194   0.589              0.294
       (0.273) (0.025)            (0.025)

 0.350DLOG(LPH/PC) -0.202(R-DLOGPC)
(0.141)            (0.100)

 0.270             -0.558
(0.100)            (0.506)

 0.479             -0.445
(0.250)            (0.250)

 0.611             -0.513
(0.191)            (0.250)

 0.580             -0.523
(0.179)            (0.250)

 0.706             -0.624
(0.262)            (0.330)
```

RO= 0.646RO!1! +0.059DLOGPC +0.573R +0.007LOG(GBH/YH)
(0.078) (0.023) (0.129) (0.002)

0.655 0.064 0.550 0.006
(0.080) (0.025) (0.140) (0.002)

0.693 0.075 0.429 0.005
(0.097) (0.034) (0.201) (0.003)

0.670 0.065 0.492 0.005
(0.093) (0.032) (0.191) (0.003)

0.666 0.063 0.502 0.005
(0.086) (0.027) (0.167) (0.003)

0.651 0.068 0.576 0.007
(0.081) (0.031) (0.131) (0.002)

LOG(TH/PC)= 0.805LOG(TH/PC)!1! +0.398DLOG(YH/PC)
(0.092) (0.159)

0.804 0.478
(0.100) (0.186)

0.761 0.623
(0.133) (0.337)

0.811 0.477
(0.121) (0.301)

0.829 0.422
(0.107) (0.249)

0.790 0.376
(0.108) (0.199)

-0.808DLOGPC +0.211LOG(LPH/PC)!1! -0.921(RO-R)
(0.129) (0.095) (0.640)

-0.794 0.212 -0.722
(0.146) (0.104) (0.711)

-0.774 0.256 -1.004
(0.211) (0.139) (0.960)

-0.769 0.205 -0.768
(0.198) (0.126) (0.885)

-0.753 0.186 -0.754
(0.192) (0.111) (0.880)

-0.826 0.227 -1.134
(0.154) (0.112) (0.760)

LOG(LPH/PC)=	-0.479	+0.607LOG(LPH/PC)!1!	-0.257(MC-R)!1.5!	
	(0.303)	(0.123)	(0.100)	
	-0.281	0.743	-0.373	
	(0.391)	(0.158)	(0.176)	
	-1.639	0.155	0.163	
	(0.942)	(0.373)	(0.370)	
	-1.434	0.243	0.083	
	(0.803)	(0.315)	(0.320)	
	-1.408	0.256	0.071	
	(0.793)	(0.310)	(0.310)	
	-0.377	0.617	-0.272	
	(0.350)	(0.138)	(0.160)	

+0.614LOG(POV/PC)	-0.203LOG(LPF/PC)	-0.997DCR!1!
(0.173)	(0.059)	(0.628)
0.509	-0.260	-1.203
(0.231)	(0.090)	(0.708)
1.345	-0.373	-1.301
(0.580)	(0.150)	(0.982)
1.218	-0.353	-1.279
(0.494)	(0.134)	(0.893)
1.203	-0.355	-1.289
(0.488)	(0.133)	(0.885)
0.540	-0.144	-0.841
(0.205)	(0.075)	(0.659)

LOG(KF/GDPQ)=	0.120	+0.857LOG(KF/GDPQ)!1!	-1.065DLOGGDPQ
	(0.025)	(0.041)	(0.063)
	0.116	0.864	-1.108
	(0.027)	(0.043)	(0.070)
	0.088	0.918	-1.270
	(0.038	(0.065)	(0.154)
	0.099	0.900	-1.235
	(0.034)	(0.058)	(0.138)
	0.103	0.893	-1.209
	(0.032)	(0.054)	(0.128)
	0.115	0.868	-1.117
	(0.026)	(0.043)	(0.078)

-0.013LOG(PMR/PI)	-0.118(R-DLOGPI)	-0.053(MC-R)!1.5!
(0.004)	(0.034)	(0.026)
-0.014	-0.132	-0.055
(0.004)	(0.037)	(0.028)
-0.021	-0.168	-0.100
(0.006)	(0.055)	(0.050)
-0.019	-0.139	-0.093
(0.006)	(0.047)	(0.040)
-0.018	-0.128	-0.088
(0.005)	(0.043)	(0.040)
-0.015	-0.124	-0.065
(0.004)	(0.038)	(0.030)

LOGVV1Q=	-0.408	+0.251LOGVV1Q!1!	+0.763LOGGDPQ!1!
	(0.852)	(0.126)	(0.087)
	-0.410	0.253	0.764
	(0.855)	(0.126)	(0.089)
	-0.410	0.253	0.764
	(0.855)	(0.126)	(0.087)
	-0.410	0.253	0.764
	(0.854)	(0.126)	(0.087)
	-0.408	0.254	0.763
	(0.853)	(0.126)	(0.087)
	-0.410	0.253	0.764
	(0.855)	(0.126)	(0.087)

-0.223(RU-DLOGPI)	-0.738(LFT/GDPV)-1
(0.090)	(0.165)
-0.248	-0.739
(0.095)	(0.166)
-0.249	-0.739
(0.115)	(0.166)
-0.244	-0.739
(0.110)	(0.166)
-0.231	-0.738
(0.100)	(0.087)
-0.248	-0.739
(0.090)	(0.166)

DLUFS/PI=	-20.429	+ 0.270(DLUFS/PI)-1	-0.402D(XTV/PI)
	(9.978)	(0.171)	(0.127)
	-22.695	0.305	-0.379
	(11.680)	(0.178)	(0.160)
	-18.931	-0.046	-0.822
	(38.127)	(0.0583)	(0.812)
	-28.052	0.014	-0.610
	(19.353)	(0.449)	(0.344)
	-33.803	0.206	-0.439
	(14.445)	(0.273)	(0.169)
	-31.189	0.340	-0.456
	(15.200)	(0.261)	(0.590)

+0.548D(MTV/PI)	-0.264(TFF/PI)-1	+0.307(IF-YF)/PI	+13.299D6368
(0.118)	(0.203)	(0.105)	(7.902)
+0.580	-0.240	+0.309	+14.549
(0.207)	(0.209)	(0.178)	(8.444)
+0.560	-0.271	+0.425	+14.431
(0.304)	(0.413)	(0.324)	
+0.557	-0.299	0.474	18.468
(0.248)	(0.328)	(0.228)	(12.663)
0.644	-0.228	0.437	21.147
(0.173)	(0.262)	(0.188)	(10.166)
0.729	-0.135	0.355	20.069
(0.340)	(0.352)	(0.183)	(11.596)

APPENDIX III

GLOSSARY OF TERMS AND DEFINITIONS

1. List of estimation methods mentioned in the text

OLS	Ordinary least squares
2SLS	Two stage least squares
2SLS1	Two stage least squares using first and second order polynomials of the predetermined variables as first stage regressors
2SLS2	Two stage least squares using second order polynomials of the predetermined variables as first stage regressors
3SLS	Three stage least squares
3S/OLS	Three stage least squares with OLS start
FILM	Full information maximum likelihood
LIML	Limited information maximum likelihood
IV	Instrumental variable estimation
IIV	Iterative instrumental estimation
RRF2SLS	Restricted reduced form two stage least squares
RRFIV	Restricted reduced form instrumental variable estimation
MLE	Maximum likelihood estimation

2. List of estimated models in this study

The names of the models refer to the estimation methods that have been used. The structure of the models is the same, that is the KT-model of the Bank of Finland.

OLS model	Ordinary least squares estimation
2TS model	2SLS estimation based on a first division of the model into a real block, a price block and a monetary block
2BL model	2SLS estimation based on a second division of the model in which the matrix of the parameters of the endogenous variables was as close as possible to the blockdiagonal interpretation
PKE model	2SLS estimation with 4 principal components corresponding to the largest characteristic roots, the same for each equation
PKR model	2SLS estimation with 4 principal components selected on the bases of size of the corresponding characteristic roots and the correlation coefficients in regressing the principal components on the predetermined variables in each equation
YP4 model YP5 model YP6 model	2 SLS estimation with 4, 5 and 6 principal komponents calculated separately for each equation
MY model	2SLS estimation using OLS start
NY model	2SLS estimation using PKE start

3. The simulation time horizon

Forecast and prediction are used to mean the same thing. They also mean the same thing as solution and prediction.

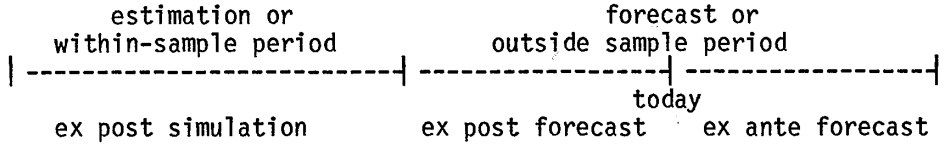
The time horizon over which the simulation is performed is defined in this study as follows:

- an ex post simulation is a model solution over the estimation period.

- an ex post forecast is a model solution beyond the estimation period using actual existing values of the exogenous variables.

- an ex ante forecast is a model solution beyond the priod for which data exist.

The diagram below illustrates the time horizon:



4. Statistical measures

MAPE mean absolute per cent erro

$$1/T \sum | (Y_t^s - Y_t^a) / Y_t^a$$

MAE mean absolute error

$$1/T \sum | Y_t^s - Y_t^a |$$

RMSE root mean square error

$$\sqrt{1/T \sum (Y_t^s - Y_t^a)^2}$$

AMSPE asymptotic mean square prediction error

$$\operatorname*{plim}_{t\to\infty} \; 1/T \sum (Y_t^s - Y_t^a)^2$$

where

- Y_t^s is the simulation (forecast, prediction) value.

- Y_t^a is the actual value.

- T is the number of observations.

RMSE and MAPE are the usual measures. When using RMSE we have, contrary to the MAPE statistics, to assume the existence of the first and second moments of the model forecasts.

Let the simulated value for method k be denoted by Y_k and the historical value by Y. The bias of the simulation value for method k is then $EY_k - Y$ and the corresponding variance $E(Y-EY_k)^2$. Taken alone as a measure of goodness of the simulation both bias and the variance are poor measures. The expected squared error

$$E(Y_k-Y)^2$$

takes account of both variance and bias because it can be shown to be the sum of the variance and the squared bias:

$$E(Y_k-Y)^2 = E(Y-EY_k)^2 + (EY_k-Y)^2$$

This measure is used as the loss function in statistical decision theory. The MAPE measure gives equal weights to the errors but when using the expected squared error the loss associated with the errors is held propotional to the square size of the error. Of course, it depends on the purpose of the comparison whether or not it is preferred to penalise large errors more heavily.

PUBLICATIONS OF THE BANK OF FINLAND

Series D (ISSN 0355-6042)
(Nos 1 - 30, Publications of the Bank of Finland Institute for Economic Research, ISSN 0081-9506)

1. PERTTI KUKKONEN On the Measurement of Seasonal Variations. 1963. 11 p.

2. The Index Clause System in the Finnish Money and Capital Markets. 1964, Rev. 1969. 15 p.

3. J.J. PAUNIO Adjustment of Prices to Wages. 1964. 15 p.

4. HEIKKI VALVANNE - JAAKKO LASSILA The Taxation of Business Enterprises and the Development of Financial Markets in Finland. 1965. 26 p.

5. MARKKU PUNTILA The Demand for Liquid Assets and the Development of the Liquidity of the Public in Finland during 1948-1962. 1965. 110 p. (In Finnish)

6. J.J. PAUNIO A Theoretical Analysis of Growth and Cycles. 1965. 117 p. (In Finnish)

7. AHTI MOLANDER The Determination of Aggregate Price and Wage Levels in Finland 1949-1962. 1965. 159 p. (In Finnish)

8. ERKKI PIHKALA The Permanent Commissions of COMECON as Realizers of the Distribution of Labour. 1965. 35 p. (In Finnish)

9. KARI NARS Price Policy Parameters of the State. 1965. 118 p. (In Swedish)

10. HEIKKI VALVANNE The Framework of the Bank of Finland's Monetary Policy. 1965. 34 p.

11. JOUKO SIVANDER On the Theory and Measurement of Elasticities of Substitution in International Trade. 1965. 91 p. (In Finnish)

12. TIMO HELELÄ - PAAVO GRÖNLUND - AHTI MOLANDER Memorandum for Wage Negotiations. 1965. 56 p. (In Finnish)

13. ERKKI LAATTO Quarterly Volume Series of Finland's External Trade in Goods between 1949 and 1964 - Adjusted for Certain Short-term Fluctuations. 1965. 24 p. (In Finnish, Summary in English)

14. DOLAT PATEL The Shares of the Developing Countries in Finnish Foreign Trade. 1966. 31 p.

15. PEKKA LAHIKAINEN On Variations in the Relationship between Output and Labour Input. 1966. 25 p. (In Finnish)

16. HEIKKI U. ELONEN On Demand for and Supply of Finance for an Enterprise. 1966. 88 p. (In Finnish)

17. TIMO HELELÄ - J.J. PAUNIO Memorandum on Incomes Policy. 1967. 10 p.

18. KARI NARS Study in the Pressure of Demand. 1967. 118 p. (In Swedish)

19. KARI PUUMANEN Debt Instruments tied to Price Indices as Objects of Choice. 1968. 186 p. (In Finnish)

20. RICHARD ALAND The Investment Banking Function in the United States. 1968. 31 p.

21. TIMO HELELÄ Strikes and Industrial Relations in Finland, 1919-1939). 1969. 341 p. (In two volumes.) (In Finnish)

22. SIRKKA HÄMÄLÄINEN On Subjective Factors Affecting Household Savings and the Possibility of Quantifying Them. 1969. 177 p. (In Finnish)

23. HEIKKI KOSKENKYLÄ An Evaluation of the Predictive Value of the Investment Survey of the Bank of Finland Institute for Economic Research. 1969. 12 p.

24. HEIKKI KOSKENKYLÄ On the Statistical Problems of the Investment Survey of the Bank of Finland. 1970. 71 p. (In Finnish)

25. PERTTI KUKKONEN - ESKO TIKKANEN Icebreakers and Winter Traffic. 1970. 136 p. (In Finnish)

26. HEIKKI U. ELONEN - ANTERO ARIMO A Study of the Finances of the Lutheran Church in Finland. 1970. 73 p. (In Finnish)

27. JUHANI HIRVONEN A Simultaneous Econometric Model for International Economy. 1971. 64 p. (In Finnish)

28. HEIKKI KOSKENKYLÄ On Problems of the Theoretical and Empirical Analysis of Investment - A Study of Manufacturing Investment in Finland in 1948-1970. 1972. 182 + 58 p. (In Finnish) (ISBN 951-686-001-X).

29. A Quarterly Model of the Finnish Economy by the Model Project Team of the Research Department. 1972; 1973 Sec. ed. 103 p. (ISBN 951-686-007-9).

30. HANNU HALTTUNEN Production, Prices and Incomes in the Quarterly Model of the Finnish Economy. 1972; 1973 Sec. ed. 120 p. (In Finnish) (ISBN 951-686-003-6, ISBN 951-686-013-3)

31. SIMO LAHTINEN Demand for Labour in the Quarterly Model of the Finnish Economy. 1973. 171 p. (In Finnish, Summary in English) (ISBN 951-686-008-7).

32. MAURI JAAKONAHO An Empirical Study of the Consumption of Electricity in Finland. 1973. 144 p. (In Finnish) (ISBN 951-686-009-5)

33. ESKO AURIKKO The Foreign Trade Block in the Quarterly Model of the Finnish Economy. 1973. 100 p. (In Finnish, Summary in English.) (ISBN 951-686-011-7)

34. HEIKKI KOSKENKYLÄ - ILMO PYYHTIÄ Problems of Resource Allocation in Finland. 1974. 59 p. (In Finnish) (ISBN 951-686-014-1)

35. IMMO POHJOLA An Econometric Study of the Finnish Money Market. 1974. 120 p. (In Finnish) (ISBN 951-686-016-8)

36. JUHANI HIRVONEN On the Use of Two Stage Least with Principal Squares with Principal Components. 1975. 91 p. (ISBN 951-686-023-0)

37. HEIKKI KOSKENKYLÄ - ILMO PYYHTIÄ The Use of the Capital-Output as an Investment Criterion. 1975. 65 p. (In Finnish, Summary in English.) (ISBN 951-686-024-9)

38. ALPO WILLMAN An Econometric Study of the Effects of Fiscal Policy Measures. 1976. 217 p. (In Finnish) (ISBN 951-686-028-1)

39. JORMA HILPINEN Migration, the Labour Participation Rate and Cyclical Variations in Employment. 1976. 69 p. (In Finnish) (ISBN 951-686-030-3)

40. OLAVI RANTALA Factors Affecting the Choice of Saving Objects in Finland. 1976. 115 p. (In Finnish) (ISBN 951-686-031-1)

41. The Flow-of-Funds Framework as an Analytical Tool (AHTI HUOMO The Flow-of-Funds Approach; TAPIO KORHONEN The Main Linkages between State Finances, Balance of Payments and Financial Markets; IMMO POHJOLA State Finances in the Framework of Flow of Funds; OLAVI RANTALA The Use of Flow of Funds and its Limitations in Quantitative Analysis). 1976. 98 p. (ISBN 951-686-033-8)

42. ILMO PYYHTIÄ Shadow Prices and the Allocation of Factors of Production in Finnish Manufacturing Industry in 1948 - 1975. 1976. 176 p. (In Finnish) (ISBN 951-686-035-4)

43. PETER NYBERG Fluctuations in the Supply of Labour in Finland. 1978. 64 p. (In Finnish) (ISBN 951-686-046-X)

44. MARJA TUOVINEN On Formation of Inflation Expectations and the Optimality of an Inflation-Expectation Series. 1979. 154 p. (In Finnish) (ISBN 951-686-056-7)

45. KALEVI TOURUNEN Finnish Industrial Inventory Investment in 1961 - 1975. 1980. 70 p. (In Finnish) (ISBN 951-686-059-1)

46. URHO LEMPINEN Rational Expectations in Macroeconomic Theory. 1980. 82 p. (In Finnish) (ISBN 951-686-060-5)

47. HANNU HALTTUNEN - SIXTEN KORKMAN Central Bank Policy and Domestic Stability in a Small Open Economy. 1981. 79 p. (In Finnish) (ISBN 951-686-066-4)

48. SEPPO KOSTIAINEN Transmission mechanisms in Finnish financial markets and industry's investment decisions. 1981. 126 p. (In Finnish, Summary in English) (ISBN 951-686-067-2)

49. URHO LEMPINEN A Theoretical Study of Central Bank Financing and Foreign Finance as Substitutes for Each Other. 1981. 131 p. (In Finnish) (ISBN 951-686-069-9)

50. ILMO PYYHTIÄ The Bank of Finland Investment Inquiry as an Instrument for Forecasting Industrial Investment. 1981. 93 p. (In Finnish) (ISBN 951-686-071-0)

51. ILKKA SALONEN On the Measurement of Technological Development with the Aid of a Production Function; an Application to the Finnish Economy. 1981. 88 p. (In Finnish) (ISBN 951-686-073-7)

52. ALPO WILLMAN The Effects of Monetary and Fiscal Policy in an Economy with Credit Rationing. 1981. 64 p. (ISBN 951-686-075-3)

53. JOHNNY ÅKERHOLM The Macro-economic Effects of Fiscal Policy in the Short Run. 1982. 173 p. (In Swedish) (ISBN 951-686-079-6)

54. HANNELE LUUKKAINEN An Integrated Model of Consumer, Investment and Financial Decision in the Household Sector. 1983. 128 p. (Summary in English.) (ISBN 951-686-085-0)

55. Inflation and Economic Policy (TAPIO PEURA Inflation in Finland; JOHNNY ÅKERHOLM Different Explanations of Inflation and Economic Policy; JUKKA PEKKARINEN On Finland's Wage Inflation: Rigidity of Real Wages or of Income Distribution?; ALPO WILLMAN Dependence of Domestic Inflation on Foreign Inflation in the Light of Finnish Research on Inflation; PENTTI FORSMAN On the Long-Term Cost of Inflation; P. SCHELDE ANDERSEN Inflation: Theories, Evidence and Policy Implications; GAVIN BINGHAM Inflation: an Overview). 1983. 204 p. (All except last two in Finnish) (ISBN 951-686-088-5)

56. PETER JOHANSSON The Effect of Interest Rate Policy on Output and the Price Level. 1984. 91 p. (In Finnish, Summary in English.) (ISBN 951-686-091-5)

57. PENTTI PIKKARAINEN The Demand for Energy in Finnish Industry in 1960 - 1982. 1984. 86 p. (In Finnish, Summary in English.) (ISBN 951-686-096-6)

58. ILKKA LYYTIKÄINEN An Econometric Model of the Finnish Labour Market: En Empirical Investigation Based on Data for the Period 1960 - 1982. 1984. 157 p. (In Finnish, Summary in English.) (ISBN 951-686-098-2)

59. The BOF3 Quarterly Model of the Finnish Economy (edited by Juha Tarkka and Alpo Willman.) 1985. 465 p. (ISBN 951-686-108-3)

60. JARMO PESOLA Cyclical Behaviour of Inventory Investment in the Finnish Corporate Sector, An Econometric Study on the Basis of Quarterly Time Series Data for the Period 1963 - 1981. 1985. 178 p. (In Finnish, Summary in English.) (ISBN 951-686-109-1)

61. JUHA TARKKA The Finnish Bank Loan Market and Neoclassical Monetary Theory. 1986. 162 p. (In Finnish, Summary in English.) (ISBN 951-686-114-8)

62. PENTTI PIKKARAINEN The Choice of the Optimal Currency Basket for a Central Bank. 1986. 77 p. (In Finnish, Summary in English.) (ISBN 951-686-121-0)

63. MONICA AHLSTEDT Small Sample Estimation and Stochastic Simulation of an Econometric Model. 1986. 181 p. (ISBN 951-686-127-X)

ISBN 951-686-127-X
ISSN 0355-6042