ECONOMETRIC MODELLING ON AN HP 3000

John Eaton
London Business School

INTRODUCTION

Econometric modelling, particularly of national economies, has become something of a growth industry during the 1970s. The current publically available models of national economies grew out of research projects in universities and other research - based organizations in the late 1960s and early 1970s. This development was made possible by several factors, of which one of undoubted significance was the advent of the electronic computer in the social sciences. For econometric modellers, computers removed the tedium and effort in storing and manipulating the data required, in estimating the relationships that make up the model, and in solving the model for forecasting and simulation purposes.

This paper discussed the implementation of a large econometric model of the UK economy on the HP 3000 system at the London Business School. Though the discussion in this paper is entirely in terms of modelling national economies, modelling techniques such as these are finding increasing application in many large organizations. Thus many large corporations maintain econometric models of the market environment in which they operate, in most cases taking as input the outputs from the large national models of the type discussed in this paper. The purpose of such models from the corporations' view is to enable them to evaluate alternative strategies in the simulated environment, with respect to factors determined by their decisions, and those determined externally by government or their competitors.

The next section provides a general introduction to econometric modelling, and the one following introduces the LBS model of the UK economy. The remaining three sections then discuss the implementation of this model on the HP 3000, in terms of, an on-line database of macro-economic variables;the model estimation process; and the model solution process for forecasting and simulation purposes.

ECONOMETRIC MODELLING OF A NATIONAL ECONOMY

Econometric models, like any other mathematical models, are intended as mathematical repreentations of the significant characteristics of the system being modelled. In the case of a model of a national economy, the relationships that go to make up the model represent the flows or levels of real and financial resources within the economy, and between it and other economies. The relationships may be of three types. Accounting identities, which are of little technical interest, but which are necessary to complete the system; technical relationships, which usually represent institutional or administrative activites (for example, computation of tax yield, given information as to rates, income groups, etc); and finally, behavioural equations, which represent the behaviour of some specific economic activity (for example, the level of imports of fuel, or the rate of inflation). In general, it is these behavioural equations which require the modellers time, effort and skill, though in the majority of current models they comprise less than half of the equations in the model.

Within each behavioural equation in the model the variables appearing on the right-hand side of the equation (the independent or explanatory variables) determine the systematic variation in the variable on the left-hand side of the equation (the dependent variable). It is axiomatic that in a behavioural equation there will be some residual, random variation in the dependent variable which cannot be explained by any combination of systematic variables which have a meaningful economic interpretation. Thus an econometric model is composed of stochastic, statistical relationships; it is not an exact model (though for forecasting and simulation purposes it is usually treated as though it were).

A variable which appears as the dependent variable of one equation may well appear as an independent variable in another equation in the model. Econometric models are usually therefore systems of simultaneous equations, though often they may be broken down into one or more blocks of simultaneous and recursive equations.

Furthermore, it is highly likely that some of the equations will be non-linear in the variables, so that for forecasting and simulation purposes we are faced with the solution of a system of non-linear simultaneous equations. It is impossible to model the entire economic environment; some factors are not modelable (for example, determination of government policy variables), or are infeasible for modelling by an individual group (for example, the world economy, though I will return to this example later). The basic aim is to model those factors which are significant in terms of the movement or management of the national economy. In order to be able to solve the model for future periods, we must have values of these variables determined outside the model, and of the parameters of the equations that go to make up the model. The variables whose values are determined by solution of the model are known as variables endogenous to the system. Those variables whose value must be known in order that the system may be solved are known as exogenous. The exogenous variables may be further sub-divided into those that are purely exogenous (typically government policy variables), and those, known as predetermined variables, which are defined as equal to the value of endogenous variables in periods previous to that for which the model is currently being solved.

The original, and still the major purpose in constructing models of a national economy is to facilitate the evaluation of alternative government policies for the management of the economy. Forecasting initially entered the picture as a means of model validation; that is, to demonstrate that the model is capable of accurately replicating the systematic behaviour of the real economy that it purports to represent. In an on-going context, forecasting has a crucial role to play in keeping the model on track. But the rationale for econometric modelling as opposed to other forecasting methods is that it enables us to consider and assess alternative government policies on a consistent, rigourous basis. Thus, although the public face of the major modelling groups appears in the regular forecasts produced by their models, in practice, much more effort and interest will be found in the alternative strategies for government policy that are evaluated.

Building a model in the first place is relatively simple as compared with the challenge of keeping it operational on an on-going basis, learning from it, and developing it as the economy itself and economists' understanding of it change. In an important sense, modelling a national economy can be viewed as a continuous exercise in learning by doing. And it is an important part of the make-up of a modeller that he is able to learn from the environment he is modelling, and adapt his perception of the workings of the economy and his model of it accordingly. Some would argue that there was a time at the beginning of the 1970's when econometric modellers were in danger of being swept away by their use of computer technology, and of becoming over confident in the capabilities of themselves and their models. However, the oil-crisis and resultant world inflation provided a salutory shock to modellers everywhere, not only about the tru capabilities of their models, but also about some of the perhaps overlooked characteristics of the environment they were attempting to model.

THE LBS MODEL OF THE UK ECONOMY

The parentage of the LBS model goes back to the mid-1950's, when one of the leading American modellers, Professor Lawrence Klien of the University of Pennsylvania, visited Oxford University. Klien gathered together a small team of British graduate students, with whom he constructed the first econometric model of the UK economy, using annual data. Valuable though this first model was as a research exercise, possibly its greatest contribution lay in highlighting what needed to be done in order to construct a fully operational model of the UK economy. One of Klien's team on this first model was Jim Ball, who in 1965, was appointed Professor of Economics at the then newly founded London Business School (in 1971 becoming Principal of the School). Since that first exercise, he had focussed his research efforts on assembling the basic building blocks for an operational UK model based on quarterly data.

Shortly after joining the LBS he assembled a new model, from which the first forecasts were published towards the end of 1966. Since this time the LBS model has been continuously maintained and developed, and has been used to produce forecasts of the UK economy on a regular basis. The project received a major boost in 1967, when it was funded for the first time by the Social Science Research Council, whose generous support has continued to the present day. In 1976 a consortium of twelve major UK organizations was formed to support the Centre for Economic Forecasting at the LBS, which took over responsibility for the model, with Terry Burns as its first director. In addition to participation in forecast meetings with members of the staff of the Centre, members of the consortium also gained access to the data base, estimation programs and forecasting system maintained by the Centre. Thus, in addition to the ten full-time and four part-time staff of the Centre, who are engaged in maintaining the data-bank and the model, production of the basic control forecasts, and a range of research studies using the model, there is also a group of practising economists and forecasters making use of the system within their own organizations.

The LBS model from which the first forecasts were published in 1966, comprised only 25 equations; the current model is made up of over 300 equations. Although this is not as large as some of the models of the US economy, size is not necessarily a criteria of excellence. An operational model of a national economy does have to grow to a considerable size, simply in order to be able to adequately represent the detail of government policy options (i.e. detailed tax rates, expenditure patterns, financial behaviour, and so on). But increasing size brings with it problems of comprehension and management of the model, as the larger the model, the more unlikely it is that any single person is able to understand all the interactions of the complete model.

The computer system requirements of an econometric model can be considered in three parts. A data base containing the historical data from which the relationships in the model are to be estimated, and two software packages, a statistical estimation program and a model solution/simulation program. Originally, the LBS system was implemented on a large-scale batch-processing machine (an IBM 360/65, located at University College, London University), and accessed by means of an rje terminal. However, this did not provide ideal support for either modellers or forecasters. In addition to the rje link to the 360/65, the LBS has operated an HP 2000 timesharing system since 1972 primarily providing a student computing service. In the mid-1970s a simple forecasting system was implemented on this system, with two objectives in mind. First, to examine to what extent on-line access and interactive processing might better support the modellers and forecasters needs. And second, to aid in the determination of the true computer system needs of the modelling and forecasting system. Nevertheless, although it was clear that the system was more supportive of the modellers and forecasters needs than a batch system, the HP 2000 could not cope with the workload. In 1977, the LBS installed an HP 3000 Series II to provide computing services to all its research activities, including the Centre for Economic Forecasting. The entire modelling and forecasting system has now been implemented on the 3000, exploiting the on-line interactive capabilities of the system.

THE MACRO-ECONOMIC DATA BANK

The macro-economic data base provides modellers and forecasters with access to the basic raw material for their studies; the time-series recording the detailed movements in the national economy. At its simplest such an on-line data-bank is a source of data for information purposes, providing the answers to queries as to the level or movement in particular variables, or the input for tabular or graphic reports. However, its crucial role is to provide the raw material for the modelling and forecasting processes, for which it must be interfaced with the estimation and solution/simulation systems.

Three macro-economic data-banks are implemented on the HP 3000 at the LBS the major one being that containing quarterly data from the mid-1950s to date on slightly more than 1000 economic variables relating to both the UK and world economy. In addition there are smaller data-banks containing annual and monthly data respectively on the major variables. These data banks are structured as KSAM files with a standard structure applied to other data-banks at the LBS, in particular, a data bank containing stock market data on all UK quoted companies since the mid-1950s. The detailed structure of the LBS KSAM data banks is set out in Appendix A.

The main data bank is updated on-line as new data is released by government and other agencies. Unfortunately, in the UK the government's statistical service is unable to provide data immediately in machine-readable form (only about a month late), so that the majority of the data is entered manually from press releases. A small suite of programs has been developed (in Fortran, using the KSAM intrinsics) to carry out the standard data management functions on-line or in batch as appropriate. Thus, functions to enter update, modify, and list specific variables may be carried out on-line; whilst functions such as to copy, list or transform the entire data bank are carried out as batch jobs. The variables that are used in the equations of an econometric model are rarely published in that form. Instead they are formed as functions of published variables, in some cases very simple ones (such as products or ratios), in other cases much more complex. However, these functions are programmable, so that the updating of a particular published variable, may in turn lead to the automatic updating of several other variables in the data bank which are specified functions of the original variable.

ESTIMATING THE RELATIONSHIPS OF THE MODEL

As discussed earlier, an econometric model is made up of a large number of economic relationships. The process of initially estimating and then maintaining these relationships is a continuous one, since the economy itself, the modellers and other economists' perception of its behaviour, and the statistical tools available to the modeller are continuously changing, albeit slowly in most instances. Given the size and complexity of the national economy, most successful models have begun as top-down exercises, starting simply in order to find out where a more detailed understanding would be most useful. But ultimately the models grow to a considerable size through the necessity to model in detail the impact of government policy variables, so that alternative policy strategies can be evaluated. In general, the modeller will begin with relatively simple relationships, and then incrementally introduce more detail and complexity in order to try to cope with its observed inadequacies; it is a classical example of a man-machine interactive system.

Current economic theory will provide the modeller with initial ideas as to which variables might enter the relationship, but gives much less guidance as to the precise functional form of the relationship, and, in particular, to the distribution of the effects of the independent variables over time. There are relatively few instances where it is believed that the effect of a change in an independent variable on the dependent variable in an equation is totally complete within the current time period. It is usually suggested that there will be carry-over effects over several time periods, and indeed, one of the major reasons for constructing econometric models is to examine in detail the time path of the response of the economy to changes in government policy variables. Thus, for example, the effects on consumption patterns of a change in the rate of a sales tax will not be seen solely in the period in which the change is made, but will be distributed over several subsequent time periods (hence the technical name of distributed lags to refer to this generic problem). Clearly, the significance and magnitude of these carry-over effects dependents on the unit time period of the data used in the model. Thus, if the relationships in the model were estimated using annual data, then we would expect that there would be only relatively small carry-over effects, as compared with a model estimated using monthly data. Most macro-econometric models of national economies are estimated using quarterly data.

The basic model-building process is dependent on a high degree of man-machine interaction, which is provided much more successfully on the HP 3000 than on the previous batch system. The model-building, estimation process is both iterative and highly suitable for interactive processing. It is iterative in that the modeller will begin with his initial relationship, which he will then refine, develop and evaluate by estimating many more equations, the precise format of each successive equation being heavily influenced by the results obtained from the estimation of previous equations. It is interactive in that interspersed with the actual estimation of the parameters of an equation, the modeller is likely to manipulate the variables in a variety of ways, and to want varying amounts of detail presented to him. An interactive econometric estimation program (ISEA, Interactive Software for Econometric Analysis) has been implemented on the HP 3000 to provide powerful and detailed support to the modeller. ISEA interfaces with the LBS KSAM data banks to allow the modeller to interactively load data for analysis, to which a wide range of specific data transformations may be applied prior to the estimation process. The relationships in an econometric model are typically estimated using classical regression techniques, which the modeller can apply using ISEA, with interactive control over the specification of the model to be estimated (including specific options tailored to econometric applications, such as estimation subject to polynomial (Almon) distributed lags, first-order autoregressive error terms, linear restrictions on the coefficients, and so on), and over the content and volume of output from the estimation process presented at the terminal. Optionally, large volumes of output may be directed to a lineprinter. It is also possible to interactively estimate the parameters of a Box-Jenkins ARIMA time-series model using ISEA. The objective in the design of ISEA has been to enable the modeller as far as possible to maintain a continuous meaningful man-machine dialogue, in terms that the econometrician can readily comprehend. A brief description of the capabilities of ISEA is given in Appendix B.

MODEL SOLUTION AND SIMULATION

As indicated earlier, in general econometric models will comprise a system of non-linear simultaneous equations, which have to be solved for the endogenous variables in each forecast or simulation period. In order to be able to solve the model in the future, values will be required for those variables which are purely exogenous to the model (values for the predetermined variables will either be the appropriate historical value of the relevant endogenous variable, or its forecast value for a previous period). There are commonly two categories of variables which are likely to be purely exogenous to a model: those whose value is determined either directly or indirectly by the government, central bank, or some other economic regulatory agency; and those measuring the overall economic environment, such as world output, trade, prices and so on. The major models for each country now tend to explicitly model those variables which directly influence their interaction with the economies of other countries (typically, domestic prices, trade prices, money supply, interest and exchange rates). Indeed, there is a project in existence (project LINK, based with Klien at the University of Pennsylvania, and funded by the IMF, NSF and other agencies), to put together on a consistent basis econometric models for individual countries and major regional groupings, and to solve them simultaneously to produce consistent forecasts of the current and capital account trade items, output, inflation interest and exchange rates for the world as a whole. This system is now operational, and the LBS model provides the UK model for the project.

Prior to commencing a forecast or simulation exercise with an econometric model, it is standard practice to adjust it, if necessary, for a variety of potential disturbance factors. First, the individual equations are examined in order to see if the estimated values of the error term (the residuals) over the recent past exhibit any significant systematic behaviour. Although according to statistical theory the expected value of the error term is zero, in practice it is often the case that for a variety of reasons examination of any specific period will demonstrate a significant discrepancy from this expected value. It may well be thought that any observed systematic pattern will continue into the future forecast period, and should be explicitly allowed for. Second, even in the best model, there are some systematic factors whose influence is inadequately accounted for by the structure of the model, but concerning which some reasonable ad-hoc estimate may be made for the future. Finally, it is often the case that the modeller has prior knowledge of some significant event that will occur in the forecast period, but which is not explicitly included in the model. For any or all of these reasons it is usual for the modeller to make adjustments to the structure of the model in the forecast period. This is done by making adjustments to the intercept or constant term of the behavioural equation determining the value of the endogenous variable that it is desired to adjust (hence they are known as constant adjustments). Values for any non-zero constant adjustments must then be specified before forecasting can commence. In practice there tends to be more uncertainty over the appropriate values for the constant adjustments since they are more influenced by the subjective assessment of the modeller.

Given values for the exogenous variables and the constant adjustments for the forecast period, then the model may be solved for the values of the endogenous variables for each forecast period. As the model is usually a system of non-linear simultaneous equations it is not possible to compute an analytic solution to the model, and it is necessary to adopt an iterative numerical procedure, starting with an initial estimate of the solution for the forecast period, then hopefully converging to a solution. In general, the solution of systems of non-linear simultaneous equations can pose considerable problems in terms of successful convergance to a consistent solution. However, these problems do not arise in practice in the solution of econometric models, for two reasons. First, most econometric models are only mildly non-linear, with only a few exponential equations being the norm. Second, it is always the case with econometric models that we possess a good initial estimate to the solution for the current period with which to begin the iterative sequence (at worst, we can use the solution for the previous time period; in practice, we often have an existing solution for the present period from a previous forecast run, with only slightly different values for exogenous variables or constant adjustments). Thus it is a general experience of econometric modellers that the computationally simple Gauss-Siedel algorithm will always generate a solution to such a model, normally in less than 20 iterations per period.

Forecasting, or simulating alternative policy strategies, is a highly iterative process. The forecaster will set his initial estimates of the exogenous forecasts and constant adjustments, compute the forecast, evaluate it in some way, modify the exogenous forecasts and/or constant adjustments, compute another forecast, and so on. Because of the simultaneous nature of the models being used, it is normal practice to only make one or two changes each forecast run, so that their effect can be clearly seen (if many changes were made in one run it would not be possible to infer the effects of an individual change). To some extent this process is open-endedand continuous in that the adequacy or otherwise of a particular forecast is ultimately largely dependent upon the subjective assessment of the forecaster. In the context of the LBS model, major forecasts are published three times a year (in January, May and October), with short monthly ones in the intervening months.

The actual implementation of the Gauss-Siedel algorithm for solving the model is as a batch job, as there is nothing to be gained from the forecaster's point of view in interacting with the algorithm. However, from forecast run to forecast run, the forecaster wishes to make a series of small usually incremental changes to the exogenous forecasts and/or constant adjustments. A set of initial exogenous forecasts and constant adjustments is agreed by the staff of the Centre as a 'control' set; the forecaster then creates an Editor file on-line which records his changes to the basic control files (which are structured as KSAM files). He then 'streams' a job which solves the model, using the control exogenous forecasts and constant adjustments modified by the contents of his Editor file. The options available in the processing of the Editor file not only include simple one-for-one changes in value, but also more complex operations such as incrementing all values from a certain date by a constant absolute or percentage factor. The forecast values for the endogenous variables are placed in a KSAM file for future reference; in theory, the forecaster could interrogate this file on-line to see if it was worth printing full forecast details. In practice, the amount of data that most forecasters want to see before they are able to make such a decision is rather large, so the usual option is to print a standard set of national accounts tables on the lineprinter. A typical five year (20 period) forecast takes about 100 cpu seconds on the HP 3000.

In addition to the standard forecasting and simulation operations, the solution system is also used for more research-orientated applications to determine characteristics of the system as a whole. Typical applications of this type would be, solution of the model for within-sample periods to determine the inherent error characteristics; computation of the dynamic multipliers for the major government policy variables, both singly and in combination (the multipliers give the proportionate change in endogenous variables for an appropriate unit change in one or more policy variables, all else being held constant); and computation of stochastic rather than deterministic forecasts, in order to evaluate the likely error bounds for the forecast (this involves setting the error terms in the model to a random value drawn from an appropriate probability distribution). Further, a major exercise is now underway to attempt to compute optimal policy strategies given some desired pattern of activity in the economy for future periods, using optimal control techniques drawn from control engineering.

APPENDIX A

KSAM file structure used at LBS for time-series data-bases:

Bytes	Description	Variable type
1-8	Variable code	character*8
9-44	Variable title	character*72
45-46	Frequency of measurement per year	integer
47-50	Start date	integer*4
51-54	End date	integer*4
55-56	Number of observations	integer
57-62	Variable type	character*6
63-64	Code number	integer
65-66	Lag length	integer
67-68	Transformation number	integer
69-70	1st transformation parameter	integer
71-72	2nd transformation parameter	integer
73-76	3rd transformation parameter	real
77-80	Not used at present	
81-800	Up to 180 observations (or multiple thereof)	real

APPENDIX B

ISEA commands (version 6.4, 2 October 1978):

AUTO	Computes autocorrelation coefficients		
BOXJ:	Computes the parameters of a seasonal Box-Jenkins ARIMA model:		
	Output options:		
	AUTO:	autocorrelation coefficients of residuals	
	PAUT:	partial autocorrelation coefficients of residuals	
	PRES:	print actual, fitted, residuals and percent error	
	SRES:	save, in the ISEA data matrix, the fitted, ratio of actual/fitted, or the fitted + and - one or two standard errors	
	GRAF:	plots residuals or actual and fitted at terminal	
	VCOV:	variance-covariance matrix of estimated coefficients	
	BCOR:	correlation matrix of the estimated coefficients	
	FORE:	compute a forecast, and (optionally) save in the ISEA data matrix	
	MODL:	prints standard output on the lineprinter	
CORR:	Computes simple correlation coefficients		
DATA:	Enters data into the ISEA data matrix,		
	T:	from the terminal	
	K:	from an LBS KSAM file	
	F:	from a file 'kept' by ISEA	
	E:	from an Editor file	
DIM:	Re-dimensions the ISEA data matrix (the default is 300 observations on 20 variables; this may be increased to 500 observations and 50 variables, so long as the product of the two does not exceed 8000).		
EDIT:	Edits the data in the ISEA data-matrix:		
	ACOL:	adds a column	
	CCOL:	changes a column	
	AROW:	adds a row	
	CROW:	changes a row	
	COBS:	changes an observation	
	AFOR:	adds date from a CEF KSAM forecast file	

GRAF: Plots a variable at the terminal, either across or down the page.

KEEP: Keeps the ISEA data matrix in a private sequential file

LPNT: Prints the contents of the ISEA data matrix on the lineprinter

MEAN: Computes means and standard deviations of variables

MISS: Specifies a value to indicate missing observations

NAME: Prints the names of the variables currently in the ISEA data matrix

PAUT: Computes the partial-autocorrelation coefficients

PLOT: Plots up to three series on an HP 7203 graph-plotter

PRNT: Prints the contents of the ISEA data matrix

REG: Specifies a regression equation (default is OLS), options:

- OLS: Ordinary Least Squares
- TSLS: Two-stage Least Squares
- RSLS: OLS subject to linear restrictions
- ALMN: Almon distributed lags
- AUTO: Include 1st order-autoregressive term
- NCON: Supress constant term
- LOG: Print true statistics for a log equation
- WGHT: Weighted least squares
- DSCT: Discounted least squares
- STRT: Start date for sub-sample
- END: End date for sub-sample

Output options:

- AUTO: autocorrelation coefficients of residuals
- PAUT: partial autocorrelation coefficients of residuals
- PRES: print actual, fitted, residuals and percent error
- SRES: save, in the ISEA data matrix, the fitted, ratio of actual/fitted, or the fitted + and - one or two standard errors
- GRAF: plots residuals or actual and fitted at terminal
- VCOV: variance-covariance matrix of estimated coefficients
- BCOR: correlation matrix of the estimated coefficients
- FORE: compute a forecast, and (optionally) save in the ISEA data matrix
- MODL: prints standard output on the lineprinter

SMTH: Exponential smoothing (Winters method)

TRAN: Compute any of the following transformations:

LAG:	lag the observations on a variable
LOG:	natural logs
DIFF:	differencing
MSUM:	moving sum
MAVE:	moving average
ADD:	sum two variables
SUBT:	subtract one variable from another
MULT:	multiply two variables
DIVI:	divide one variable by another
RECP:	reciprocal of a variable
POWR:	raise to a power
EXP:	e to the power of a variable
CSUM:	cumulative sum
TIME:	time trend
MULC:	multiply by a constant
ADDC:	add a constant
SUBC:	subtract from a constant
ABS:	absolute value
SUMV:	sum several variables
EXS:	simple exponential smoothing
DEV:	positive, negative (or both) deviations from zero, mean or a specified value
DUMY:	form a dummy variable
SDUM:	form seasonal dummy variables