MODEL VALIDATION AND FORECAST COMPARISONS :
THEORETICAL AND PRACTICAL CONSIDERATIONS

Mark Salmon and Kenneth F. Wallis

NUMBER 184

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

MODEL VALIDATION AND FORECAST COMPARISONS : THEORETICAL AND PRACTICAL CONSIDERATIONS

Mark Salmon and Kenneth F. Wallis

NUMBER 184

International Symposium on Criteria for Evaluating the
Reliability of Macro-Economic Models

Centro Scientifico IBM, Pisa, December 1980.

December, 1980.

This paper is circulated for discussion purposes only and its contents should be considered preliminary.

MODEL VALIDATION AND FORECAST COMPARISONS: THEORETICAL AND PRACTICAL CONSIDERATIONS

by Mark Salmon and Kenneth F. Wallis

University of Warwick

International Symposium on Criteria for Evaluating the Reliability of Macro-Economic Models

Centro Scientifico IBM, Pisa, December 1980

1. INTRODUCTION

Most macroeconometric models are built with the object, wholly or partly, of providing forecasts. The term "forecast" covers three rather distinct types of exercise:

(a) genuine "ex-ante" forecasts, in which the model user predicts the actual future development of the economy, and for which projected future values of input variables must be supplied;

(b) "ex-post" forecasts, in which the model user eliminates the effects of error in the projections of the input variables by calculating "forecasts" over some period in the recent past, given the actual observed values of the input variables;

(c) hypothetical forecasting or policy analysis exercises, in which the model user estimates the response of the economy to alternative scenarios, that is, to alternative values of policy instruments or to different kinds of exogenous shock.

In each case there is interest in evaluating the results of the forecasting exercise, not only for its own sake but also to provide information that is useful in model validation, that is, in checking the specification of the model. Of course the various forecasting exercises and their respective evaluations are not necessarily independent of one another, for example it is often said that in order to be useful in policy analysis a model should have a good real-world forecasting record over a period that was not part of the estimation period, so that it might also be expected to provide "good" estimates of responses to policy changes.

As elements of the evaluation process we identify for further discussion three main types of forecast comparison:

(a) comparison of the forecasts of a single model with actual outcomes;

(b) comparison of the forecasts of an econometric model with those obtained by ostensibly different methods, such as statistical time series ("Box-Jenkins") methods;

(c) comparison of the forecasts of a number of different macro-econometric models.

Theoretical and practical aspects of these comparisons are discussed in Sections 2-4 of this paper. In general this discussion is set in the context of standard linear textbook models, and in Section 5 we discuss further issues that arise from the non-linearity of practical models. Section 6 deals with structural change, and Section 7 contains concluding comments.

The practical use of model-based forecasts is in decision-making, and in principle the evaluation of a model and its forecasts should be associated with the loss function relevant to the particular objective. In practice this does not feature in the literature, since decision-makers and the users of forecasts seldom discuss such matters, at least not in public. As a result models are generally evaluated by the model-builders themselves and their econometric critics, in rather general terms, using statistical criteria that seem reasonable but are seldom rigorously justified. Model-builders typically report a range of criteria, perhaps in the hope that a particular model-user will find therein the information needed for a particular decision problem. That models are built to achieve different objectives is clear from their different sizes and their coverage of different markets in the economy. Thus a different loss function is implicit in each model, occasionally made more explicit as, for example, when a model is given a specific control orientation (Wall et al., 1975). In the absence of a completely explicit specification of a loss function, our own discussion is concerned with some of the various general criteria that are employed.

2. COMPARING A SINGLE MODEL'S FORECASTS WITH ACTUALITY

We begin by assuming a linear simultaneous equations model, describing the relations between the elements of a vector of current endogenous variables $\underset{\sim}{y}_t$ and a vector of predetermined variables $\underset{\sim}{z}_t$, and of which the reduced form can be written

$$\underset{\sim}{y}_t = \underset{\sim}{\Pi}\underset{\sim}{z}_t + \underset{\sim}{v}_t \, .$$

Given estimates of the coefficients, $\hat{\underset{\sim}{\Pi}}$, based on a sample period $t = 1,\dots,T$, and projections of the predetermined variables, $\hat{\underset{\sim}{z}}_{T+j}$, for a forecast period $t = T+1, T+2,\dots,T+h$, *ex-ante* forecasts are given as

$$\hat{\underset{\sim}{y}}_{T+j} = \hat{\underset{\sim}{\Pi}}\,\hat{\underset{\sim}{z}}_{T+j}, \qquad\qquad j = 1,2,\dots,h.$$

For a correctly-specified model, the corresponding actual values are given as

$$\underset{\sim}{y}_{T+j} = \underset{\sim}{\Pi}\,\underset{\sim}{z}_{T+j} + \underset{\sim}{v}_{T+j} \, ,$$

and in this event the forecast error can be written as

$$\underset{\sim}{y}_{T+j} - \hat{\underset{\sim}{y}}_{T+j} = \hat{\underset{\sim}{\Pi}}(\underset{\sim}{z}_{T+j} - \hat{\underset{\sim}{z}}_{T+j}) - (\hat{\underset{\sim}{\Pi}} - \underset{\sim}{\Pi})\,\underset{\sim}{z}_{T+j} + \underset{\sim}{v}_{T+j} \, .$$

This expression breaks down the forecast error into three components.

First, projections of the predetermined variables are unlikely to be exact, and the error in these projections, $\underset{\sim}{z}_{T+j} - \hat{\underset{\sim}{z}}_{T+j}$, contributes to the overall ex-ante forecast error. Secondly, the coefficient estimates $\hat{\underset{\sim}{\Pi}}$ based on the sample-period information differ from the true values $\underset{\sim}{\Pi}$. Finally the random disturbances $\underset{\sim}{v}_{T+j}$ occurring in the forecast period affect the outcome, the above equation for $\hat{\underset{\sim}{y}}_{T+j}$ having assumed that the optimal forecast of these disturbances is their unconditional mean value of zero. In practice this might be modified either given evidence that the disturbances are autocorrelated, so that earlier residuals are helpful in forecasting later disturbances, or given information about extraneous influences likely to impinge on the model in the forecast period but not explicitly incorporated in the original specification. This last possibility occurs frequently in practical forecasting exercises, since good information about changes in legislation, institutional arrangements both foreign and domestic, labour contracts, and so forth is often available and incorporated into ex-ante forecasts as predictions of residuals, equivalently handled as adjustments to the intercept terms of the appropriate equations. The success of these adjustments reflects the skill and judgment of the model-user, hence their evaluation contributes little to model validation. However a persistent need for modifications to an estimated equation, perhaps with similar justifications being offered from one period to the next, would suggest that an important explanatory variable had been omitted from the original specification. In the absence of a complete respecification, the construction of simple statistical models for the residuals might provide a short-run, but second-best solution (compare Surrey and Ormerod, 1977).

The first component of the forecast error results from errors in the projections of the predetermined variables, and in dynamic models further possibilities arise by distinguishing between lagged endogenous and exogenous variables. We partition $\underset{\sim}{z}_t$, the vector of predetermined variables, into sub-vectors $\underset{\sim}{y}_{t-1}$ and $\underset{\sim}{x}_t$, taking the case of one-period lagged endogenous variables for simplicity. Partitioning the reduced form conformably we have

$$\underset{\sim}{y}_t = \underset{\sim}{\Pi}_1 \underset{\sim}{y}_{t-1} + \underset{\sim}{\Pi}_2 \underset{\sim}{x}_t + \underset{\sim}{v}_t .$$

In ex-ante forecasting one period ahead, only the exogenous variables contribute to the errors in projecting predetermined variables since $\underset{\sim}{y}_T$ is known, and the forecast is given by

$$\hat{\underset{\sim}{y}}_{T+1} = \hat{\underset{\sim}{\Pi}}_1 \underset{\sim}{y}_T + \hat{\underset{\sim}{\Pi}}_2 \hat{\underset{\sim}{x}}_{T+1} .$$

In forecasting further ahead, the model generates its own lagged endogenous variables, and the sequence of forecasts is given by

$$\hat{\underset{\sim}{y}}_{T+j} = \hat{\underset{\sim}{\Pi}}_1 \hat{\underset{\sim}{y}}_{T+j-1} + \hat{\underset{\sim}{\Pi}}_2 \hat{\underset{\sim}{x}}_{T+j} , \qquad j = 2,3,\ldots,h.$$

which reverts to the general expression given above.

Errors in the projections of exogenous variables, by definition formed outside the model, also reflect the skill of the model-user rather than that of the model-builder. When the model forecasts are compared to the actual outcomes after the event, as in studies of forecasters' "track records" (for example, Osborn and Teal, 1979), this contribution can be eliminated by considering *ex-post* forecasts. These are based on the actual realized values of $\underset{\sim}{z}_{T+j}$, and are given by

$$\tilde{\underset{\sim}{y}}_{T+j} = \hat{\underset{\sim}{\Pi}} \underset{\sim}{z}_{T+j} = \hat{\underset{\sim}{\Pi}}_1 \underset{\sim}{y}_{T+j-1} + \hat{\underset{\sim}{\Pi}}_2 \underset{\sim}{x}_{T+j}, \qquad j = 1,2,\ldots,h.$$

Ex-post forecasts can be calculated as part of an evaluation of the model's genuine forecasting performance, or as a check on the model's structural stability carried out immediately after specification and estimation by holding back from the estimation period a sub-sample of h observations specifically for this purpose. Assuming that the data are truly generated by the given linear model, the ex-post forecast errors have two components:

$$\underset{\sim}{y}_{T+j} - \tilde{\underset{\sim}{y}}_{T+j} = \underset{\sim}{v}_{T+j} - (\hat{\underset{\sim}{\Pi}} - \underset{\sim}{\Pi})\, \underset{\sim}{z}_{T+j}, \qquad j = 1,2,\ldots,h.$$

Study of the absolute magnitudes of these errors may be informative, in the light of the known development of the economy, but in the absence of a standard of measurement such a study is of limited usefulness. A standard of comparison is provided by an estimate of the variance of these errors, based on the estimated variances of disturbances and coefficient estimates, obtained as a by-product of the sample-period estimation process. Clearly ex-post forecasts can be calculated *within* the estimation period,

$$\tilde{\underset{\sim}{y}}_t = \hat{\underset{\sim}{\Pi}}\, \underset{\sim}{z}_t, \qquad t = 1,\ldots,T,$$

and estimates of the variance of the forecast errors for the period $t = T+1,\ldots,T+h$ are in effect based on the variance of the ex-post forecast errors or residuals $\underset{\sim}{y}_t - \tilde{\underset{\sim}{y}}_t$ in the period $t = 1,\ldots,T$. Formal tests are available to compare this estimated forecast error variance with the actual mean squared error of the forecasts (Christ, 1966, Ch.X.7; Dhrymes et al., 1972; Hendry, 1974), and if the latter is not significantly greater than the former, the model passes its structural stability test. (Using stochastic simulation methods, Fair (1980) provides analogous estimates of the forecast error variance for non-linear models.) Note, however, that there is no statistical theory available for the case in which the forecasts sequentially generate their own lagged values, as follows:

$$\underset{\sim}{y}^*_{T+j} = \hat{\underset{\sim}{\Pi}}_1\, \underset{\sim}{y}^*_{T+j-1} + \hat{\underset{\sim}{\Pi}}_2\, \underset{\sim}{x}_{T+j}, \qquad j = 1,2,\ldots,h.$$

In this case forecast errors typically cumulate, and while various plots can be informative, in particular for checking that the forecasts do not "drift", the theory of standard cusum techniques does not cover models with lagged endogenous variables (Brown, Durbin and Evans, 1975). A further possibility, given forecasts calculated over a suitably long period, is to compare the properties of the forecasts to actuality through the National Bureau of Economic Research business cycle methodology, that is, by considering whether the forecasts reproduce actual behaviour in terms of such features as the length of cycles, the coincidence of turning points, and the lead-lag relations of the variables to one another and to the reference cycle.

If the model fails its structural stability test, then respecification is indicated. However such tests are not very powerful, since specification errors also affect the sample-period estimates of the model's behaviour. For example, omitting a relevant explanatory variable increases the residual variance of an equation and biases its coefficients, hence it is not necessarily the case that forecasts from the misspecified equation have mean squared error greater than that anticipated on the basis of (erroneous) sample-period calculations. Indeed, if the behaviour of the variables is unchanged, the forecast-period estimates will be equally erroneous as the sample-period estimates. Thus forecast periods in which the behaviour of exogenous variables is substantially different from that in the sample period provide particularly informative comparisons, which may lead to improved specification, as in the case of the aggregate consumption function following the high inflation of the early 1970's (Wallis, 1979, Ch.1). In effect, the hypothesis being tested in structural stability tests is that the model's explanation of the variables is equally good in the sample and forecast periods, and no absolute standard is available. These difficulties in directly validating models in an absolute sense have led to econometric forecasts being evaluated relative to forecasts produced by other means, or to forecast comparisons being conducted across different models, in order to gain an impression of their relative merits. These are discussed in turn in the next two sections.

3. COMPARING FORECASTS FROM ECONOMETRIC AND "NON-ECONOMETRIC" REPRESENTATIONS

In the absence of an absolute standard, the forecast performance of an econometric model has often been evaluated relative to the performance of some simpler forecasting rules. In the words of Cooper (1972, pp.828-9):

> Comparing an econometric to a naive method of forecasting supplies a technique for assessing the economic information contained in an econometric model. The defining characteristic of a 'naive' forecasting method is that it depends exclusively on purely statistical properties of economic time series, such as trend, past levels, or past changes. A naive method does not incorporate any economic information ... Forecasts made by naive methods are then compared with forecasts made by other methods. Forecasting methods that cannot do better than a purely mechanical one should be discarded.

That progress has occurred since the early development of econometric models is clear from the fact that the forecasting methods against which the models have been compared have gradually become less naive. Initially "no-change" ($\hat{y}_{T+1} = y_T$) or "same-change" ($\hat{y}_{T+1} - y_T = y_T - y_{T-1}$) forecasting rules were employed. Subsequently autoregressive models of the form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

fitted to individual endogenous variables were used to generate forecasts. Finally there have been applications of the ARIMA models of Box and Jenkins (1970), possibly in their seasonal form, that is,

$$\phi(L)\ \Phi(L^s)\ \Delta^d\ \Delta_s^D\ y_t = \theta(L)\ \Theta(L^s)\ \varepsilon_t$$

where L is the lag operator, the first factors on each side are polynomials in L, the second factors on each side are polynomials in L^s, s is the number of seasons per year, $\Delta = 1 - L$ and $\Delta_s = 1 - L^s$. Nelson (1972) compares the FRB-MIT-PENN model against ARIMA models for certain endogenous variables, and finds that in within-sample comparisons the econometric model was ahead, whereas outside the sample the "pure time-series" models had smaller forecast errors. In this section we challenge the idea that "purely statistical" models for endogenous variables provide an independent check on the econometric model, and more generally consider the relations among the various representations of a model that might be used for forecasting and other purposes (Prothero and Wallis, 1976; Wallis, 1977; Zellner and Palm, 1974).

A linear dynamic econometric model, explaining the behaviour of a vector of endogenous variables $\underset{\sim}{y}_t$ in terms of a vector of exogenous variables $\underset{\sim}{x}_t$ and a vector of disturbances $\underset{\sim}{u}_t$, may be written in *structural form* as

$$\underset{\sim}{B}(L)\underset{\sim}{y}_t + \underset{\sim}{C}(L)\underset{\sim}{x}_t = \underset{\sim}{u}_t$$

where $\underset{\sim}{B}(L)$ and $\underset{\sim}{C}(L)$ are matrices of polynomials in the lag operator, viz.

$$\underset{\sim}{B}(L) = \underset{\sim}{B}_o + \underset{\sim}{B}_1 L + \ldots + \underset{\sim}{B}_r L^r, \quad \underset{\sim}{C}(L) = \underset{\sim}{C}_o + \underset{\sim}{C}_1 L + \ldots + \underset{\sim}{C}_s L^s.$$

Usually $\underset{\sim}{B}_o \neq \underset{\sim}{I}$, hence there is "instantaneous coupling" between the endogenous (output) variables. This is removed in the *reduced form*, which expresses each endogenous variable as a function of predetermined (exogenous plus lagged endogenous) variables, as follows:

$$\underset{\sim}{y}_t = -\underset{\sim}{B}_o^{-1}\{\underset{\sim}{B}_1\underset{\sim}{y}_{t-1} + \ldots + \underset{\sim}{B}_r\underset{\sim}{y}_{t-r} + \underset{\sim}{C}(L)\underset{\sim}{x}_t\} + \underset{\sim}{B}_o^{-1}\underset{\sim}{u}_t.$$

The present formulation generalises the reduced form used in the previous section by including higher-order lags of both endogenous and exogenous variables. It also makes explicit the dependence of reduced form coefficients (the $\underset{\sim}{\Pi}$-matrices, in the previous section) on the structural parameters. In practice estimates of the reduced form coefficients are obtained from estimates of the structural parameters, rather than directly from regressions of each endogenous variable on the predetermined variables, and although the directly estimated reduced form equations have smaller sample-period residual sums of squares, the argument for preferring the "solved out" or restricted reduced form equations is that they forecast better if the (typically *over*-identifying) prior restrictions incorporated in the specification of the $\underset{\sim}{B}$- and $\underset{\sim}{C}$-matrices are correct.

An alternative solution of the structural form is obtained by multiplying through by the inverse, not of $\underset{\sim}{B}_o$, but of $\underset{\sim}{B}(L)$ (under appropriate stability conditions). The result is the *final form*

$$\underset{\sim}{y}_t = -\underset{\sim}{B}(L)^{-1}\,\underset{\sim}{C}(L)\underset{\sim}{x}_t + \underset{\sim}{B}(L)^{-1}\,\underset{\sim}{u}_t,$$

in which each endogenous variable is expressed as an infinite distributed lag function of the exogenous variables, together with an error term that is a moving average of the original structural disturbances. The coefficients in the expansion of $\underset{\sim}{B}(L)^{-1}\,\underset{\sim}{C}(L)$ provide *dynamic multipliers*, describing the response of y_{it} to a unit shock in $x_{j,t-\ell}$. In empirical work the infinite distributed lag is generally approximated by a ratio of finite-degree polynomials in L, but in the present context an explicit expression is obtained by writing $\underset{\sim}{B}(L)^{-1} = \underset{\sim}{b}(L)/|\underset{\sim}{B}(L)|$, where $\underset{\sim}{b}(L)$ is the adjoint matrix of $\underset{\sim}{B}(L)$ and $|\underset{\sim}{B}(L)|$ the determinant. Thus the final form may be written

$$y_t = \frac{-\underset{\sim}{b}(L)\underset{\sim}{C}(L)}{|\underset{\sim}{B}(L)|} \underset{\sim}{x}_t + \frac{\underset{\sim}{b}(L)}{|\underset{\sim}{B}(L)|} \underset{\sim}{u}_t ,$$

giving a set of multi-input "transfer function" equations. Multiplying through by $|\underset{\sim}{B}(L)|$, we obtain a further representation known as the *final equations* (Tinbergen, 1939; Goldberger, 1959):

$$|\underset{\sim}{B}(L)| \underset{\sim}{y}_t = -\underset{\sim}{b}(L)\underset{\sim}{C}(L)\underset{\sim}{x}_t + \underset{\sim}{b}(L)u_t \quad .$$

Since $|\underset{\sim}{B}(L)|$ is a scalar polynomial, each final equation relates a given endogenous variable to its own past values and to the exogenous variables, current and past, but to no other endogenous variable, current or past. In effect, the dynamic interrelations with other endogenous variables have been solved out. The interesting property of this representation is that the autoregressive operator $|\underset{\sim}{B}(L)|$ is common to all endogenous variables (unless, as noted by Goldberger, the model is decomposable, for this results in cancellation of common factors across some of the final equations). Thus the characteristic dynamic behaviour of the endogenous variables could, in principle, be studied by considering a single endogenous variable or, more abstractly, the common characteristic polynomial $|\underset{\sim}{B}(L)|$.

To obtain a further representation, we add an assumption about the data generation process of the exogenous variables, namely that this can be represented as the vector ARMA process

$$\underset{\sim}{G}(L)\underset{\sim}{x}_t = \underset{\sim}{H}(L)\underset{\sim}{\eta}_t ,$$

where $\underset{\sim}{\eta}_t$ is a vector white noise process. The complete model of the data generation process of the observable variables can then be written

$$\begin{bmatrix} \underset{\sim}{G}(L) & \underset{\sim}{0} \\ \underset{\sim}{C}(L) & \underset{\sim}{B}(L) \end{bmatrix} \begin{bmatrix} \underset{\sim}{x}_t \\ \underset{\sim}{y}_t \end{bmatrix} = \begin{bmatrix} \underset{\sim}{H}(L)\underset{\sim}{\eta}_t \\ \underset{\sim}{A}(L)\underset{\sim}{\xi}_t \end{bmatrix}$$

where $\underset{\sim}{u}_t$ is assumed to have a vector moving average representation. (If $\underset{\sim}{u}_t$ is non-autocorrelated, only the leading coefficient matrix $\underset{\sim}{A}_o$ is non-zero. If a vector ARMA representation is alternatively assumed, then it is postulated that the autoregressive operator has been multiplied through and incorporated in $\underset{\sim}{B}(L)$ and $\underset{\sim}{C}(L)$.) On multiplying through by the inverse of the matrix on the left-hand side, defined as its adjoint matrix divided by its determinant, as in the previous paragraph, we obtain "final equations" for $\underset{\sim}{y}_t$ of the form

$$\phi(L)\underset{\sim}{y}_t = \underset{\sim}{F}_1(L)\underset{\sim}{\eta}_t + \underset{\sim}{F}_2(L)\underset{\sim}{\xi}_t ,$$

where $\phi(L) = |\underset{\sim}{G}(L)| \times |\underset{\sim}{B}(L)|$, and $\underset{\sim}{F}_1(L)$ and $\underset{\sim}{F}_2(L)$ are defined in terms of the original matrices and their adjoints. Thus a typical element y_{it} is expressed in terms of a finite number of its own past values and a composite error term which is the sum of a number of separate moving average processes. By a result in time series theory this error term has a representation as a moving average of a single white noise process, ε_{it} say, and so we have a *univariate ARMA model* for each y_{it}, of the form

$$\phi(L)y_{it} = \theta_i(L)\varepsilon_{it} \quad .$$

We can thus interpret this ARMA time series equation for an endogenous variable of an econometric model as an alternative solution form, that is, an implication of the model.

It is interesting to note that a further member of this sequence of alternative representations of the same structural model is the *state space representation* favoured in other disciplines. Thus forecasts generated from the Kalman filter via a state space econometric model may be seen as the application of a common statistical principle to an equivalent representation of the same structure. This equivalence is obtained by rewriting either the structural econometric model or the equivalent ARMA representation in state space form. The system of equations

$$\underset{\sim}{w}_{t+1} = \underset{\sim}{D}\underset{\sim}{w}_t + \underset{\sim}{E}\underset{\sim}{x}_t + \underset{\sim}{\varepsilon}_{1t}$$

$$\underset{\sim}{y}_t = \underset{\sim}{J}\underset{\sim}{w}_t + \underset{\sim}{\varepsilon}_{2t}$$

is known as a *state space model*. The matrices $\underset{\sim}{D}$, $\underset{\sim}{E}$ and $\underset{\sim}{J}$ are constants, not polynomials in L, since the vector of state variables, $\underset{\sim}{w}_t$, incorporates all relevant lags explicitly. The concept of the *state* may be defined either as the minimal set of variables, knowledge of which is necessary for prediction of the future state (excluding exogenous inputs) or alternatively, and more rigorously, as a basis for the predictor space. It can then be seen as a rather fundamental form in which our previous representations may be expressed. A simple transformation allows the above equations to be rewritten as

$$\underset{\sim}{y}_t = \underset{\sim}{J}(\underset{\sim}{I} - \underset{\sim}{D}L)^{-1}\underset{\sim}{E}L\underset{\sim}{x}_t + \underset{\sim}{K}(L)\underset{\sim}{\varepsilon}_t ,$$

which corresponds to our earlier final form representation, but with a different set of parametric restrictions. Many ramifications of this equivalence are developed by Akaike (1974a) and Hannan (1979). It is interesting to note that in general there is not a unique state space representation, and while the structural econometric model leads to one form, the vector ARMA model leads to another. Akaike (1974a) demonstrates the statistical equivalence of the various state space forms.

The relations among these alternative representations raise a number of questions. First, since they are mathematically equivalent, the application of a common statistical approach to the different representations of a given model can only yield different answers if the various representations impose different restrictions on the statistical method, or have different requirements in statistical implementation. The choice of a particular representation that is most appropriate for a given statistical application is often clear, but the choice remains between deriving this representation from an estimated structural form, so that any implied restrictions are imposed and the equivalence of the representation is guaranteed, or obtaining it empirically, perhaps relying on the above development for part of the specification but relying on the data for the remainder.

Economists find the structural form, with its explicit statement of different economic agents' autonomous behaviour, accounting identities, technical relations, and so forth, to be the most convenient framework in which to consider questions of economic theory and on which to impose the resulting restrictions. But to obtain forecasts of the endogenous variables the structural form is transformed to the reduced form as above, and as already noted, given estimates of a correctly specified structure the derived reduced form gives more efficient forecasts than the directly-estimated reduced form (comprising unrestricted regressions of endogenous on predetermined variables). Similarly the implied final form and final equations can be readily derived from an estimated structural form, and in practice dynamic multipliers and/or the properties of the characteristic polynomial are frequently reported. However it is again possible to ignore the structural form and estimate these directly, with the only restriction being a prior classification of endogenous and exogenous variables, the detailed dynamic specification being data-based.

The same arguments apply to the consideration of univariate ARMA representations, except that in this case an additional assumption has been made, namely that the exogenous variables satisfy a vector ARMA process. On incorporating this assumption the "pure time-series" representation of an endogenous variable that is implied by an estimated structure can be derived. However in the last stage of this derivation we construct a moving average error term, with a single innovation process, that is observationally equivalent to the sum of a number of such series, and so the "structural" information on the separate series is lost, and in consequence the implied time-series representation gives less efficient forecasts than the econometric model. In the formal framework of this section, ex-ante forecasts of a given endogenous variable are obtained from the reduced form, with forecasts of the exogenous variables being obtained from the vector ARMA model: it can be shown, following Pierce (1975), that such ex-ante forecasts have smaller mean squared error than forecasts from the implied univariate ARMA models. Again, in contrast to this derivation, a univariate ARMA model for an endogenous variable, to be used in forecast comparisons with the structural model, can be obtained directly from the data, using the "model identification" and diagnostic devices developed by Box and Jenkins and others. It is commonly said that such models "let the data speak for themselves", and so they qualify as "naive" models in the sense of the quotation at the beginning of this section. Usually the degrees of the autoregressive and moving average operators in empirical ARMA models are substantially less than those implied by the formal derivation, since $|\underset{\sim}{B}(L)|$ is typically of high degree, and so in practice there is some conflict between the theoretical and empirical models. Statistical reasons for this are discussed in Wallis (1977), and more general explanations for this disagreement between, in effect, two different approximations to reality are discussed below.

The foregoing analysis shows that there are difficulties in interpreting the univariate ARMA model as an independent check on the econometric model, in terms of forecast or any other comparison. Since the data used in empirical specification and estimation of the two forms are the same, their summary measures are not statistically independent. Moreover the theoretical derivation of the ARMA model from a given structural form does not in general result in restrictions on its parameterisation that could provide the basis of a formal test. The structural form implies the existence of various alternative representations, of which the univariate ARMA model is one, but a given ARMA model is consistent with a number of different structures. Perhaps for this reason, the implied ARMA model is seldom derived (although other difficulties arise in practical models, discussed below) hence in practice the comparison simply rests on the relative forecasting ability of the structural model and the data-based ARMA model. However it is impossible in practice to be sure that the two models are not simply two different representations of the same structure. Different forecast properties may indicate that they *are* two different approximations to reality, or that the two different representations of the same structure are affected in different ways by the requirements of the statistical method in each case. It is difficult to compare models at different points in the sequence of alternative representations, and it cannot be maintained that, from their forecast properties alone, the pure time series models provide an independent check on structural econometric models.

However practical forecast comparisons have often tended to suggest that for a number of variables the empirical time series representations have the better forecasting performance. This result is in conflict with the theoretical result that forecasts based on the implied time series representation have larger mean squared error than those based on the structural econometric model. How can this discrepancy be rationalised? The equivalence of the alternative representations rests on a number of assumptions, few of which may be valid in practice. The implications of the assumption of a *linear* model are examined more generally in Section 5. The impact of parameter estimation errors may be different in different representations, but although little is known about this, it is clear that their contribution to forecast error variance is of relatively small order of magnitude. The vector ARMA process assumed for the exogenous variables may not be a good approximation, since the variables classified as exogenous in macroeconometric models often exhibit jumps and discontinuities. Indeed a number of dummy variables (indicator variables) are often included among the exogenous variables, and while these can be appropriately treated in the model-based forecasts, one would expect that the neglect of such discontinuities in univariate ARMA modelling (unless taken into account by what is termed "intervention analysis" in the time series literature) would work to the disadvantage of the purely statistical forecasts. More generally, however, the model is assumed to be correctly specified, and in practice one doubts this assumption particularly insofar as the dynamic and stochastic specification of large models in concerned. In these areas there is little guidance from economic theory, and one suspects that, taking a large model equation-by-equation, relatively less systematic attention is given to these matters than when a time series model is identified for a single endogenous variable, for it is precisely the dynamic and stochastic aspects of the behaviour of the variables that such models emphasize. Rather than providing an independent post-construction check on an econometric model, comparisons with time series models represent a useful diagnostic device during model-building, moreover attention should not be limited to a comparison of forecast variances, since a comparison of the dynamic specifications of the various representations yields valuable information about their adequacy.

4. COMPARISONS ACROSS MACROECONOMETRIC MODELS

An alternative response to the absence of an absolute standard of forecasting performance against which to assess a single econometric model has been to turn to comparisons across a number of models of a national economy. Three related aspects of the different models have been compared:

(a) pure forecasting ability, based on the mean squared error of forecasts over a post-model-building period;

(b) policy multipliers, through simulation or hypothetical forecasts;

(c) dynamic properties, stability and cyclical behaviour, through stochastic simulation or spectral analysis.

For U.S. examples of (a) and (b) see Christ (1975) and Fromm and Klein (1976), for the U.K. see Laury et al. (1978), and for (c) see Howrey (1972). The validation of a single model as an approximation to reality is not an objective of these exercises, instead models are compared in a rather general, descriptive manner. Thus few strong conclusions have been drawn, although outliers are occasionally detected. We first discuss the statistical foundations, or lack thereof, of these comparisons, in general taking the three types of exercise together, since they have certain essential similarities. These similarities become clear in our subsequent discussion of the experimental design aspects of the comparisons.

It is immediately obvious that the statistical theory needed for a formal evaluation of various large-scale models, seen as non-nested hypotheses, is not available. This was noted by Dhrymes et al. (1972), and although progress has been made, for example by Pesaran and Deaton (1978) and Davidson and MacKinnon (1980), as yet it remains the case.

An alternative approach rests on establishing distance measures for the alternative models from the true probability process: the model that minimises this distance is then considered to be the best available, but again this is a relative comparison and an absolute measure of the "quality" of the chosen model is not provided. These notions underlie recent developments using information criteria (Akaike, 1974b; Sawa, 1978) and the method of support (Edwards, 1972), and relieve objections to the use of significance tests. While these methods are related to estimation of the predictive distribution, and a formal equivalence between Akaike's information criterion and Amemiya's prediction criterion can be shown (Amemiya, 1980), extensions from the regression context to large-scale model validation via forecast comparisons are awaited.

It is also clear that the choice of economic structure differs between models, with a different loss function implicit in each one, as already noted. The most obvious manifestation is the differing size of models, which results in certain variables being variously treated as endogenous or exogenous, although differences in the classification of variables can arise for other reasons. The type of policy objective to be considered clearly influences the level of disaggregation and size of the model. The main reason for the increasing size of models, the Treasury model of the U.K. economy having some 500 equations, is the need to model the channels through which various policy instruments affect the economy, and in practice the number of available instruments is large. The tax system involves a large number of different rates and allowances, and monetary and interest rate policy works through many channels. To describe the differential impact of particular measures on particular groups or sectors of the economy or particular types of expenditure a full account of these impacts is required. Of course to forecast broad aggregates a smaller, more aggregated model might suffice. However in forecast comparisons the fact that the larger model provides information on a number of matters about which the smaller model has nothing to say is usually neglected, there being no loss function specified. Ex-post forecast comparisons clearly given an advantage to models that treat a variable that is difficult to forecast as exogenous, and this can be eliminated to some extent by basing the comparison on ex-ante forecasts. But comparing models augmented by their forecasting rules for exogenous variables may distort the appraisal of different models that are designed to serve different ends.

The specification of a particular model may also be influenced by the statistical method employed, the sample period, the level of temporal aggregation, any pre-filtering of the data, and an emphasis on one or another of the alternative representations discussed in the previous section. The particular objective will lead a model-builder to make a careful choice under each of these headings, and Cooper's (1972) attempt to compare various models on a standardised basis by reestimating them mechanically on a consistent body of data has been criticised by Howrey et al. (1974) for its neglect of this ("tender loving") care.

An important feature both in model specification and in elucidation of the model's properties is the nature of the input characteristics, given not only by the choice of variables to be treated as exogenous but also by their intrinsic behaviour: together we term these the *input topology*. This then becomes the distinguishing feature of the three types of comparison outlined at the beginning of this section. In principle, comparisons across models of forecasts, policy multiplier paths, or cyclical behaviour in stochastic simulation, provided that there is no unequal injection of the skill and judgment of the model-user, yield partial information on essentially the same characteristic of the models, namely the map of their dynamic properties. The question is then one of experimental design, since the three exercises differ only in input topology. Thus a forecast comparison is based on those inputs that are economically meaningful in the particular forecast period, a multiplier comparison is typically based on unit impulse or unit step change inputs, and stochastic simulation comparisons usually study the response to a white noise, random input. It is potentially confusing that different conclusions about the relative merits of various models may be drawn from these different exercises, and there is room for an analysis that fully explores the dynamic response of alternative models in an objective manner, the crucial aspect being the design of the input signal.

The characteristic polynomial $|\underset{\sim}{B}(L)|$ determines the dynamic behaviour of the model, as noted in the previous section. In transfer function analysis, the positioning of the roots of the characteristic polynomial, or poles, determines the stability, speed of response and cyclical behaviour of the system. The eigenvalues of the system provide precisely the same information. Writing the polynomial $|\underset{\sim}{B}(z)|$ as

$$(z - \lambda_1)(z - \lambda_2)(z - \lambda_3)\ldots(z - \lambda_n)$$

where the poles are $\lambda_1, \lambda_2, \ldots, \lambda_n$, the dynamic behaviour of the system may be seen in the frequency domain to be composed of the product of terms, each resonant at a particular frequency determined by the position of the particular pole or eigenvalue. In designing an input signal we would wish to ensure that the input was "rich enough" in frequency terms to draw behaviour in the output from each of these poles. A badly designed input will not "excite" all the possible modes of behaviour in a model. In comparative studies the possibility exists that an input signal, apparently well justified on economic criteria, may excite different modes of behaviour in different models, or a particular mode of behaviour to a different extent. Of the three types of comparative exercise, the pure forecast comparison is potentially the most deficient in this respect. A full frequency response analysis avoids these difficulties. While multiplier paths generated as step or impulse response functions provide essentially equivalent information in the time and frequency domains, in other areas different information is obtained. For example, it is difficult in time domain analysis to detect frequency-dependent differences between models such as might arise, during model-building, from the varying use of seasonally filtered data or varying emphasis on short-term and medium-term responses.

As the final form or final equations make clear, an input signal may stimulate the system either via the exogenous variables or via the disturbances, which often have different economic interpretations. However these relations also show that for dynamic analysis the only distinction lies in the different lag polynomials applied to the input, which alter the characteristics of an otherwise standard input, and may result in the observation of apparently different dynamic responses. For example, in the classic study by the Adelmans (1959), white noise disturbances generated cycles at business-cycle frequencies, while white noise innovations in exogenous variables did not. Moreover, the specification of these lag polynomials is typically based on the sample-period data, and so may reflect a particular input topology rather than a prior theoretical view of the dynamics. However almost no systematic examination of the nature of the inputs that characterise the model during the sample period is carried out. Such an examination would provide useful preliminary information in a model comparison exercise, providing guidelines as to possible uses of the model and an indication of the frequency range over which the model may be more reliable. If no evidence of cycles is found in the "unforced" behaviour of a system, but cycles are present in the input topology and these induce cycles in the endogenous variables, it is important to say why.

A final question is whether the white noise input commonly used in stochastic simulation, which is potentially exciting at each frequency of the dynamic response, is the optimal input signal. The recent literature on the experimental design of dynamic systems (Goodwin and Payne, 1977; Zarrop, 1980) suggests that it may not be. If an optimum is defined as a maximum of some scalar function of the Fisher information matrix, it appears that a finite number of discrete frequency components is preferable to white noise when estimating the parameters of a dynamic regression. The problem with a white noise input is that it enhances noise at irrelevant frequencies. A physical analogy suggested to us is that of determining the resonant frequency of an organ pipe: a sequence of tuning forks, each of a particular pitch, offers a better prospect of finding the frequency than white noise forced into the other end of the pipe! Of course this is essentially an argument about efficiency in estimation, which may not necessarily override other advantages of white noise, such as ease of interpretation. With a white noise input, the transfer function of the dynamic response is directly proportional to the output spectrum, although in practice output spectra from different models are not readily distinguishable from one another - the "typical spectral shape" predominates. The cross-spectrum yields information about the separate gain and phase shift components of the frequency response function, again easily calculated and giving information across the full frequency range if a white noise input is used. The information on gain and phase may be combined in the Nyquist diagram, which has proved a useful tool in the physical sciences and might be fruitfully employed in econometrics in general, and in model comparisons in particular.

5. NONLINEARITY

Most practical econometric models are nonlinear, and in this section we examine how this affects the various procedures for comparing models, discussed in the preceding sections in an entirely linear context. We first consider statistical consequences of nonlinearity, and then consider its impact on the forecast comparison exercises discussed in Sections 2-4.

5.1 Statistical implications of nonlinearity

The nonlinear system may be written in *structural form* as

$$\underset{\sim}{f}(\underset{\sim}{y}_t, \underset{\sim}{z}_t ; \underset{\sim}{\theta}) = \underset{\sim}{u}_t ,$$

where $\underset{\sim}{f}$ is a vector of functions, having as many elements as the vector $\underset{\sim}{y}_t$, and $\underset{\sim}{\theta}$ is a vector of parameters. We assume that this implicitly defines a single inverse relationship giving the solution

$$\underset{\sim}{y}_t = \underset{\sim}{g}(\underset{\sim}{u}_t, \underset{\sim}{z}_t ; \underset{\sim}{\theta}),$$

valid for relevant z-values, analogous to the reduced form in the linear case. Typically an explicit analytic solution does not exist, and the structural equations are solved numerically. The *conditional expectation* of the endogenous variables is written

$$E(\underset{\sim}{y}_t) = E\{\underset{\sim}{g}(\underset{\sim}{u}_t, \underset{\sim}{z}_t ; \underset{\sim}{\theta})\} = \underset{\sim}{h}(\underset{\sim}{z}_t, \underset{\sim}{\theta}) .$$

Since the expected value of a nonlinear function of a random variable is not in general equal to the nonlinear function of the expected value of the random variable, it is generally the case that

$$h(z_t, \theta) \neq g(0, z_t ; \theta) .$$

The right-hand side is termed the *deterministic solution*. It is the solution for y_t of the equations $f(y_t, z_t ; \theta) = 0$, but unlike the linear case it is not equal to the conditional expectation of y_t, which is what is required in forecasting exercises. However the conditional expectation can be estimated by stochastic simulation, and a common measure of the (stochastic) importance of nonlinearity is given by the deviation of the deterministic solution from the mean of replicated stochastic simulations. A sizeable discrepancy may have serious implications. First, it implies that, to avoid serious biases, policy analysis should be conducted via stochastic simulation rather than, as is commonplace, deterministic solution, and this considerably increases the computational burden. Secondly, many estimation methods rest on the use of the conditional expectations of the endogenous variables, and if these are not correctly calculated in the nonlinear model the properties of the estimation method will be adversely affected. Notice that it is only nonlinearity between the endogenous variables that is our present concern, and a nonlinear relation between exogenous variables is of no statistical importance.

Most examinations of the stochastic importance of nonlinearity carried out as described above have in fact found relatively little difference between the deterministic solution and the mean of replicated stochastic simulations. This has provided some confidence in the existing practice of deterministic forecasting and a general impression that nonlinearity is not of major concern. However we argue below that this is a dangerous position to take 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 to determine the specification of most large models is still ordinary least squares. Through its neglect of the simultaneity problem the OLS estimator also ignores the stochastic importance of nonlinearity. Moreover the linear-in-parameter constraint can in fact be a serious restriction when a conditional expectation is a general nonlinear function of the parameters and the conditioning variables. This point is best seen by viewing the question of model estimation *and specification* as one of optimal signal extraction. In this case the optimal estimate of both functional form and parameters would be obtained by orthogonal projection on the conditioning variable set. Thus the conditional expectation of the variable defines both the optimal functional form and then, perhaps in a separate exercise, the optimal parameter estimates. However in practice the functional form is invariably predetermined and only the parameter values are determined by an orthogonal projection operation, on the assumption that the functional form is true. The restriction imposed by ordinary least squares and most other minimum distance methods in econometrics is that the functional form be a linear combination of possibly nonlinear functions of the variables. It is unlikely in general that this linear-in-parameter model will coincide with the nonlinear expression for the conditional expectation, $\underset{\sim}{h}(\underset{\sim}{z}_t, \underset{\sim}{\theta})$, although it may perhaps be justified as a linearization via a Taylor series expansion. Notice that this point is equally applicable to the nonlinear simultaneous equations estimators such as nonlinear two-stage least squares when applied to models where the linear-in-parameter assumption is maintained. Thus in the present context even though the models are found through the simulation comparisons to be relatively linear this does not necessarily imply that the underlying economic structure *is* linear but it may indicate that the models are biased towards linearity.

A second reason for believing that this approach to the measurement of nonlinearity is flawed is provided by Mariano and Brown (1980). In a theoretical examination, through asymptotic expansions for the first two prediction moments of a nonlinear model, they find that the leading term in the asymptotic prediction bias in the deterministic solution may be decomposed into two terms, one due to possible inconsistencies in the parameter estimates and the other due to the nonlinearity. The leading term for a stochastic simulation 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. A second important observation that follows from their work is that if inconsistent parameter estimates are used when comparing deterministic and stochastic simulations then it is impossible to separate the effects of nonlinearity from the use of inconsistent parameter estimates. Thus the fact that the simulation exercises may have found little difference 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. They also conclude that the use of nonstochastic simulations in both estimation and validation may lead the model specification consistently away from the true specification to one that performs better in terms of deterministic simulations.

Finally we note the results of Fair and Parke (1980). In their examination of alternative estimators in the context of a nonlinear model they find that certain policy multipliers that follow from the use of ordinary least squares deviate from a relatively uniform set of multipliers obtained from nonlinear simultaneous estimators. As our preceding discussion would show this could either be due to a neglect of simultaneity or to a neglect of the stochastic importance of nonlinearity on behalf of the OLS estimator.

In section 3 we discussed a hierarchy of alternative representations of a structure within the linear framework. When the original structure is actually nonlinear then it is difficult to exploit these equivalences in practice, hence the importance in the nonlinear case of our earlier statement, that a common statistical principle may not give equivalent results when applied to apparently equivalent mathematical representations if they impose different restrictions on the statistical method. This not only applies to our foregoing discussion of the difficulties of estimating a conditional expectation correctly in a nonlinear model, but becomes particularly important when we attempt to compare different representations of what may be the same structure, such as with the forecasts of a structural econometric model and the forecasts from an ARMA model. In this case the most obvious distinction is that these comparisons are with a *nonlinear* model(although typically examined through deterministic simulation) and a *linear* time series model. Thus we are immediately assured that the two competing representations must be supporting different approximations to reality, although it is possible that the linear time series model is representing a linearization of some underlying nonlinear time series representation. Evidence to this effect may already exist given that the forecast performance of the linear time series representations tends to decay faster than that from the nonlinear structural models. Thus the linear ARMA representations are perhaps only *locally* justified in reality, because the Jacobian of the transformation from the unobserved disturbances to the observed endogenous variables is not constant in a nonlinear system.

5.2 Forecast comparisons in the context of nonlinearity

Nonlinearity also invalidates many of the alternative methods of forecast comparison discussed in section 4. The reason has already been mentioned: the Jacobian of a nonlinear transformation is not a constant function throughout the sample. Thus policy multipliers and empirical spectra are not uniquely defined and eigenvalues do not exist in the same sense, since they are aspects of a linear algebra. It is surprising that this fact has been so readily disregarded, particularly in comparisons of policy multiplier paths. Although such exercises still do provide comparative information their generality is substantially weakened. Perhaps in an attempt to avoid this problem various linearizations have been adopted. The question then is how relevant are these linearized models to policy analysis and forecast comparisons.

The first type of linearization follows from a retention of only the first terms in a Taylor series expansion of the nonlinear model around some point, perhaps the sample mean values of the variables. This could of course be generalised to successive linearizations around a number of points, which would in effect give a number of models, one for each linearization point. Once linearized the properties of the submodels could be analysed in exactly the manner discussed in the first part of this paper. However since the act of linearization has different effects on the different models it is virtually impossible to gain unambiguous comparisons of the underlying models in this way. A second form of linearization that has been discussed in the literature (Bowden, 1974; Aoki, 1980) but does not appear to have been applied in practice is a linearization around a particular form of behaviour or state, rather than around a point. This approach suggests that rather than somewhat arbitrarily choosing a point about which to linearize, a better choice is to ensure that the resulting set of linearized submodels at least uniformly coincided with common behaviour, for example, by collecting all the upswings of a business cycle into one model and all the downswings into another. A third type of linearization arises from our preceding discussion of the conditional expectation of nonlinear models. By stochastically simulating the nonlinear model and averaging the results of successive replications we are assured of a data series that conforms both in the conditional expectations implied by the nonlinear model and also exactly reflecting the dynamic structure of the model. A linearization may then be constructed from these data by modelling this simulation output series using linear time series techniques. Since the resulting model captures the linear component of the total response of the nonlinear model we may now directly observe the degree of nonlinearity by measuring the relative size of the residual in the time series model, perhaps through R^2, an opportunity not directly available in the preceding linearizations. Thus this exercise provides a linearization around the conditional expectation which could again be analysed using the linear techniques discussed above and further a qualitative analysis of the properties of the nonlinear component in the entire residual could also be made. This approach to separating the linear and nonlinear responses is well founded mathematically in terms of a Volterra series expansion (Subba Rao, 1979). If x_t and y_t denote scalar input and output variables related through a nonlinear relation

$$y_t = f(y_{t-j}, x_{t-\ell}) \qquad \text{for } j,\ell \geq 0$$

then, under certain conditions, we may write the relation in the form:

$$y_t = \sum_{\tau_1=0}^{\infty} g_1(\tau)\, x_{t-\tau_1} + \sum_{\tau_1=0}^{\infty} \sum_{\tau_2=0}^{\infty} g_2(\tau_1,\tau_2)\, x_{t-\tau_1}\, x_{t-\tau_2}$$

$+$ higher-order terms.

This is known as a Volterra expansion and the functions g_i corresponding to higher-order transfer functions are known as the Volterra Kernals. The linearization we have suggested may then be seen as an attempt to estimate the linear response through the linear transfer function, and to allow the study of the higher-order nonlinearities that occur in the resulting residual.

The following table indicates the practicality of this expansion. In this case the GDP-Government expenditure link of the Australian NIF7 model has been examined through a white noise input applied to government expenditure. The observed GDP output is then modelled using linear techniques at different levels of variance in the white noise input.

Thus we attempt to measure scale change effects of nonlinearity as considered by Zellner and Peck (1973). Two points are worth noting, first the constancy of the linear representation under the increasing scale of the government expenditure level, and secondly the decreasing R^2 indicating as we would expect that the nonlinear effect is becoming more important.

TABLE (1)

Model: $y_t = a_1 y_{t-1} + b_o u_t$

Scale change $\alpha(\sigma)$	a_1	b_o	Long Run Multiplier	Time Constant	R^2
10	-.542	.739	1.61	1.63	.71
5	-.534	.762	1.59	1.63	.889
2	-.537	.761	1.61	1.64	.897
1	-.529	.764	1.57	1.62	.901
.5	-.538	.752	1.61	1.63	.904
.1	-.525	.761	1.55	1.60	.909

The foregoing "statistical" linearization has certain attractions over the two preceding mathematically oriented linearizations but is restricted in that the importance of the numerator polynomials for the exogenous variables have been ignored. Hence in terms of practical utility for policy analysis rather than model comparison this approach is limited. Alternatively, as in the example, we may shock the exogenous variables in the nonlinear model directly and whilst this may give some guidance as to the degree of nonlinearity in particular channels it does not provide a linearization around the conditional expectation. One solution would be to shock simultaneously both the exogenous channels of interest and the stochastic terms in the nonlinear equations. The resulting output from the stochastic simulations could then be modelled as a system of linear multi-input multi-output transfer functions and would again provide a linearization around the conditional expectation. This latter suggestion seems to provide a number of practical advantages for model comparison, policy calculations and the systematic study of nonlinearity in the residuals.

6. STRUCTURAL CHANGE

Different linearizations of a nonlinear model, applicable over different ranges of behaviour as discussed in the previous section, may appear to exhibit structural change whether or not the underlying model is stable. More generally a model, linear or not, that is characterised by a particular sample period may appear to break down when the range of behaviour over which it is expected to operate changes from that over which it was specified. We have assumed throughout this paper that there is a unique and stable structure that the model-builder is attempting first to capture and then to validate through forecast comparisons. In the absence of such a stable structure most econometric analysis, including forecast comparisons, breaks down. A number of attacks have recently been made on this concept of a stable economic structure, and they are examined in this section. We consider first the question of structural change in the absence of rational expectations, and then the argument derived from the "Lucas critique" concerning the stability, in the face of rational expectations, of the "laws of motion" that govern economic behaviour. Salmon (1980) examines the impact of these arguments on modelling practice.

We begin by distinguishing two types of structural shift, namely a change in the behavioural basis for economic decisions, and a change in the institutional structure. Clearly over an historical period changing opportunity sets alter observed consumption behaviour, however it is most probable that this occurs within the framework of a constant structure of tastes. Again, over time this structure of preferences, defined by basic tastes, is itself subject to change as completely new economic environments evolve. However this type of fundamental change in the economy and consequently tastes will occur over a substantially longer period than our econometric models are covering. Thus we may be assured in applied econometrics of a fundamental constancy in the parameters that define the tastes and technology of the observed economic system and hence in the constancy of the fundamental "laws of motion" of the economy. However if we now consider how econometric models appear to break down and fail to predict as the institutional structure in which they are defined changes, we are at first sight led to question the preceding argument for constancy. However the reason why we fail to observe constancy is simply a reflection of the level of approximation to reality of the macro model. Given our constancy premise, there would exist a meta model defined by the constant parameters of taste, technology etc. that would also define uniquely the reaction of the economic agent to any outside stimulus. This model of course is bound to be highly complicated and well beyond the reach of the economist given his limited knowledge of economic behaviour and restricted data sets. So the reason that we observe variation following an institutional change is that the approximate model has been specified at a level such as to prevent it from accommodating the change. Alternatively in estimation based on a particular sample any information that may previously have been available on behaviour in the new circumstances may not have been sufficient or persistently exciting to enable the specification of a more general model to have been detected, or even thought necessary. Thus structural shift in this sense provides prima facia evidence of misspecification and an opportunity to use the new information to determine a better specification.

When the autonomous economic structure is supposed to have changed in the forecast period, compared to the sample specification, then we may only have been able to accommodate it beforehand if a suitable signal was present in the estimation period. The fact that we do observe apparent structural breakdowns does not necessarily imply that the underlying structure is not constant. It may however mean that we should develop new techniques for examining the information within an available sample (Salmon, 1980).

Let us now turn to the question of the effect of rational expectations on the preceding arguments. Lucas (1976) criticizes the standard use of policy simulation to compare the effects of alternative policy rules on the grounds that the "structure" of econometric models is not invariant to changes in policy. The models contain behavioural relations derived from the optimal decision rules of economic agents, based in part on their expectations of the future movements of relevant variables. Changes in the nature of these movements are said to cause changes in the optimal decision rules, hence "any change in policy will systematically alter the structure of econometric models" (p.41). Prescott (1977) also supports the view that the "laws of motion" of the economy cannot be "policy invariant". It is important to distinguish in this debate the role of a change in government policy, which really is just a structural shift in the sense of the preceding discussion, from the presence of rational expectations in the model.

Consider first a change in government policy that will impinge on the economic agent's consumption choice. Economic theory postulates that the consumption decision is based on the optimisation of a utility function, the form of which is determined by tastes and other basic motives for economic behaviour. These basic or instinctive motives could be seen to determine the functional form and any parameters in this utility function. The premise of constancy argues that these are intrinsically fixed (for ever) by definition of the uniqueness (or more correctly the existence) of an individual. When the utility function is then optimised there will be a one-to-one relation, in the sense of structural constancy, between the utility function and the derived reaction function, which the econometrician observes in the aggregate as the consumption function. The situation is exactly that of any other, perhaps exogenous, structural change in the economy. To suggest that government policy causes structural variation in the agent's *true* reaction function must imply a change in, and hence previous inadequacy of, the utility function of the economic agent. However this effectively denies the existence of such an agent given the premise of constancy. Once again the reasons why the econometrician may observe variation are either because the approximate model is too naive, in that it fails to model appropriately the true reaction function, or that the signal lacks persistent excitation and thus fails immediately to identify that particular mode of behaviour.

The Lucas critique can then be seen as a criticism of the commonplace confounding of two sources of dynamics: the economic structure, incorporating decision rules in which expectations appear, and secondly the expectation formation process, which in many cases will reflect the stochastic structure of exogenous variables. Wallis (1980) argues that these should be, and can be, kept separate, and has suggested ways of achieving this. If this is done, the traditional view of econometric policy evaluation can be reasserted. What is not traditional is the resulting emphasis on the modelling of information flows, decision sequences, and the expectation formation process in general. Nevertheless the notion of an *economic* structure that results is entirely consistent with preceding arguments for stability.

When we move on to consider expectations about government policy instruments we must first determine how these obviously varying expectations are formed. If we endogenise government behaviour and then use the model's forecasts as proxies for the expectations we may apparently once again confront an argument concerning structural variation, since it is frequently suggested that different political parties have different basic motives, leading to variation in endogenised government reaction functions. Notice that this variation should only be observed in the equations explaining government behaviour and *not* elsewhere in the model. However even here it is possible to make a wider case regarding uniformity in the role and motives of all governments and the population they serve, although clearly the successful modelling of government behaviour is much further from our grasp. Indeed the basic premise of constancy may well not carry over so readily in this case given the particular temporal structure of governmental power.

Thus in conclusion we suggest that the weight of the Lucas critique bears simply on the suggestion that, to be good approximations to reality, models should include rational expectation terms. There seems to be little strength in the argument that the economic environment is not constant; what may be required however are more informative statistical techniques to indicate the type of model we may (or may not) have specified.

7. CONCLUSION

Many of our suggestions and conclusions have been mentioned in passing in our discussion of theoretical and practical aspects of model validation and forecast comparisons. Here we restate briefly some points that symposium participants might find the most controversial.

(1) To evaluate a model in an absolute sense requires the specification of a loss function. In its absence, only relative statements can be made.

(2) Acceptance of the null hypothesis in a post-sample parameter stability test implies that the degree of misspecification does not differ in the sample and forecast periods, and not that it is small.

(3) The theoretical equivalence between different representations of a given model may be difficult to realise in practice. Evidence that different representations have different forecast properties is *by itself* difficult to interpret.

(4) Most of the methods of comparison we have discussed are attempts to map the dynamic properties of a model. Frequency response analysis, perhaps using Nyquist plots, appears to have some advantages. The experimental design of inputs requires further attention, and in the meantime the use of discrete sinusoidal inputs appears to merit investigation. (In discussing Howrey (1972), Holt makes a similar suggestion, without any argument.)

(5) Nonlinearity disrupts many of the techniques developed for linear models. Practical models may be seriously biased towards linearity, and the degree of nonlinearity underestimated. The use of linearization around the conditional expectation should be explored.

(6) The "Lucas critique" does not halt the search for constant economic structures, nor does it require fundamental change in simulation philosophy, once rational expectations terms have been correctly incorporated where necessary.

REFERENCES

ADELMAN, I. and ADELMAN, F.L. (1959). The dynamic properties of the Klein-Goldberger model, *Econometrica*, 27, 597-625.

AKAIKE, H. (1974a). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, 26, 363-387.

______ (1974b). A new look at the statistical model identification. *I.E.E.E. Transactions on Automatic Control*, AC-19, 716-723.

AMEMIYA, T. (1980). Selection of regressors. *International Economic Review*, 21, 331-354.

AOKI, M. (1980). Note on comparative dynamic analysis. *Econometrica*, 48, 1319-1325.

BOWDEN, R.J. (1974). Non-linearity and non-stationarity in dynamic econometric models. *Review of Economic Studies*, 41, 173-179.

BOX, G.E.P. and JENKINS, G.M. (1970). *Time Series Analysis, Forecasting and Control.* San Francisco : Holden-Day.

BROWN, R.L., DURBIN, J. and EVANS, J.M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society* B, 37, 149-192.

CHRIST, C.F. (1966). *Econometric Models and Methods.* New York : Wiley.

______ (1975). Judging the performance of econometric models of the U.S. economy. *International Economic Review*, 16, 54-74.

COOPER, R.L. (1972). The predictive performance of quarterly econometric models of the United States. In *Econometric Models of Cyclical Behavior* (B.G. Hickman, ed.), pp. 813-926. New York : Columbia University Press for NBER.

DAVIDSON, R. and MACKINNON, J.G. (1980). Relations among some non-nested hypothesis tests. Department of Economics Discussion Paper 369, Queens University.

DHRYMES, P.J. et al. (1972). Criteria for evaluation of econometric models. *Annals of Economic and Social Measurement*, 1, 291-324.

EDWARDS, A.W.F. (1972). *Likelihood.* Cambridge : University Press.

FAIR, R.C. (1980). Estimating the expected predictive accuracy of econometric models. *International Economic Review*, 21, 355-378.

FAIR, R.C. and PARKE, W.R. (1980). Full information estimates of a non-linear macroeconometric model. *Journal of Econometrics*, 13, 269-291.

FROMM, G. and KLEIN, L.R. (1976). The NBER/NSF model comparison seminar : an analysis of results. *Annals of Economic and Social Measurement*, 5, 1-28.

GOLDBERGER, A.S. (1959). *Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model.* Amsterdam : North-Holland.

GOODWIN, G.C. and PAYNE, R.L. (1977). *Dynamic System Identification, Experiment Design and Data Analysis.* New York : Academic Press.

GRANGER, C.W.J. and ANDERSEN, A.P. (1978). *An Introduction to Bilinear Time Series Models.* Gottingen : Vandenhoeck and Ruprecht.

HANNAN, E.J. (1979). The statistical theory of linear systems. In *Developments in Statistics, Vol. 2.* (P.N. Krishnaiah, ed.), pp. 83-121. New York : Academic Press.

HENDRY, D.F. (1974). Stochastic specification in an aggregate demand model of the United Kingdom. *Econometrica*, 42, 559-578.

HOWREY, E.P. (1972). Dynamic properties of a condensed version of the Wharton model. In *Econometric Models of Cyclical Behavior* (B.G. Hickman, ed.), pp. 601-663. New York : Columbia University Press for NBER.

HOWREY, E.P.,KLEIN, L.R. and MCCARTHY, M.D. (1974). Notes on testing the predictive performance of econometric models. *International Economic Review*, 15, 366-383.

LAURY, J.S.E., LEWIS, G.R. and ORMEROD, P.A. (1978). Properties of macroeconomic models of the U.K. economy : a comparative study. *National Institute Economic Review*, No. 83, 52-72.

LUCAS, R.E. (1976). Econometric policy evaluation : a critique. In *The Phillips Curve and Labor Markets* (K. Brunner and A.H. Meltzer, eds.), pp. 19-46. Amsterdam : North-Holland.

MARIANO, R.S. and BROWN, B.W. (1980). Asymptotic behavior of predictors in a nonlinear simultaneous system. Unpublished paper, Wharton School, University of Pennsylvania.

NELSON, C.R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review*, 62, 902-917.

OSBORN, D.R. and TEAL, F. (1979). An assessment and comparison of two NIESR econometric model forecasts. *National Institute Economic Review*, No. 88, 50-62.

PESARAN, M.H. and DEATON, A.S. (1978). Testing non-nested nonlinear regression models. *Econometrica*, 46, 677-694.

PIERCE, D.A. (1975). Forecasting in dynamic models with stochastic regressors. *Journal of Econometrics*, 3, 349-374.

PRESCOTT, E.C. (1977). Should control theory be used for economic stabilization.In *Optimal Policies, Control Theory and Technological Exports*, eds. K. Brunner and A.H. Meltzer, Carnegie-Rochester Conference Series on Public Policy, Vol. 7., North-Holland.

PRIESTLEY, M.B. (1979). Nonlinear models in time series analysis. *The Statistician*, 27, 159-176.

PROTHERO, D.L. and WALLIS, K.F. (1976). Modelling macroeconomic time series (with discussion). *Journal of the Royal Statistical Society* A, 139, 468-500.

SALMON, M. (1980). Structural constancy and model selection. Unpublished paper, Department of Economics, University of Warwick.

SAWA, T. (1978). Information criteria for discriminating among alternative regression models. *Econometrica*, 46, 1273-1292.

SUBBA RAO, T. (1979). On the theory of bilinear time series models-II. Unpublished paper, Department of Mathematics, The University of Manchester Institute of Science and Technology.

SURREY, M.J.C. and ORMEROD, P.A. (1977). Formal and informal aspects of forecasting with an econometric model. *National Institute Economic Review*, No. 81, 67-71.

TINBERGEN, J. (1939). *Statistical Testing of Business-Cycle Theories, II : Business Cycles in the U.S.A., 1919-1932.* Geneva : League of Nations.

WALL, K.D., PRESTON, A.J., BRAY, J.W. and PESTON, M.H. (1975). Estimates of a simple control model of the U.K. economy. In *Modelling the Economy* (G.A. Renton, ed.), pp. 398-446. London : Heinemann Educational Books.

WALLIS, K.F. (1977). Multiple time series analysis and the final form of econometric models. *Econometrica*, 45, 1481-1497.

______ (1979). *Topics in Applied Econometrics* (2nd. ed.). Oxford : Basil Blackwell.

______ (1980). Econometric implications of the rational expectations hypothesis. *Econometrica*, 48, 49-73.

ZARROP, M.B. (1980). *Optimal Experiment Design for Dynamic System Identification.* Berlin : Springer-Verlag.

ZELLNER, A. and PALM, F. (1974). Time series analysis and simultaneous equation econometric models. *Journal of Econometrics*, 2, 17-54.

ZELLNER, A. and PECK, S.C. (1973). Simulation experiments with a quarterly macroeconometric model of the U.S. economy. In *Econometric Studies of Macro and Monetary Relations* (A.A. Powell and R.A. Williams eds.). Amsterdam : North-Holland.