Economic Forecasting
Lecture 8: Evaluating Forecasts

Richard G. Pierse

1 Introduction

In previous lectures we have looked at many different models for forecasting economic time series, both univariate and multivariate. Often we may have more than one possible model for forecasting a particular series. In this case how can we assess which model produces the best forecast? Is it possible that a forecast can be improved by combining the forecasts from two or more models? These are the main questions addressed in this lecture. Optimal forecasts provide a yardstick by which to measure the performance of a forecast but most forecasts in practice will not be optimal. Another yardstick is the encompassing forecast which can explain the forecasts generated by other (misspecified) models. If we can find an encompassing forecast then this is evidence that we have found a correctly specified model. If not, then combining forecasts may be a way to reduce the variance of forecast errors.

2 Properties of optimal forecasts

Optimal forecasts are the best possible forecasts. Although in practice, most forecasts will not be optimal, it is still useful to examine the properties of optimal forecasts since these are the best forecasts that we can ever hope to achieve. Also it is possible to test with any actual forecast, how far it approximates the properties of an optimal forecast.

2.1 Optimal forecasts

An *optimal forecast* is a forecast that minimises the expected value of the forecaster's loss function

$$L(e_t)$$

given an information set $\mathbf{\Omega}_t$. Under reasonably weak conditions, it can be shown that the optimal forecast of y_{t+h} given the information set at time t, $\mathbf{\Omega}_t$, is given by the *conditional mean*

$$\mathrm{E}(y_{t+h}|\mathbf{\Omega}_t).$$

An optimal forecast uses the available information efficiently so that the optimal forecast error should be *unforecastable* from the information available when the forecast was made. This implies four key properties of an optimal forecast.

2.1.1 Optimal forecasts are unbiased

An optimal forecast will be unbiased so that the forecast error has an expected value of zero. For an h-step ahead forecast of y_{t+h} made in period t, this means that

$$E_t\widehat{y}_{t+h,t} = y_{t+h}$$

so that the expected value of the forecast errors

$$E_t e_{t+h,t} = y_{t+h} - E_t\widehat{y}_{t+h,t} = 0.$$

Note that E_t here denotes that the expectation is formed in period t.

2.1.2 Optimal one-step ahead forecasts have white noise errors

When forecasting one-period ahead, the optimal forecast errors should have no forecastable pattern so that they are distributed as white noise with

$$e_{t+1,t} = \varepsilon_t \sim iid(0, \sigma^2)$$

where *iid* stands for *independently, identically distributed* and means that there is no autocorrelation in the one-step ahead forecast errors.

2.1.3 Optimal h-step ahead forecasts have *MA(h-1)* errors

When forecasting more than one-period ahead, the optimal forecast errors will inevitably be autocorrelated. However, this autocorrelation should cut off at order $h-1$ so that it can be represented as the *MA(h-1)* process

$$\begin{aligned} e_{t+h,t} &= \varepsilon_{t+h} + \theta_1\varepsilon_{t+h-1} + \cdots + \theta_{h-1}\varepsilon_{t+1} \\ \varepsilon_t &\sim iid(0, \sigma^2). \end{aligned}$$

Recall that forecasts in an *ARMA(p,q)* property satisfy this property of optimal forecasts.

2.1.4 Optimal forecast error variances are non-decreasing in h

When forecasting h-periods ahead, the forecast error variance of an optimal forecast should be a *non-decreasing function* of the forecast horizon h. Formally, defining

$$\sigma_h^2 = \text{var}(e_{t+h,t})$$

we require that

$$\sigma_h^2 \geq \sigma_{h-1}^2.$$

For a stationary variable y_t, as h increases, the forecast error variance of an optimal forecast will tend to the unconditional variance of the process $\text{var}(y_t)$.

2.2 Testing for optimality

It is possible to test some of the properties implied by optimality. Consider the regression

$$y_{t+h} = \beta_0 + \beta_1 \widehat{y}_{t+h,t} + \varepsilon_{t+h}.$$

This is called a *Mincer-Zarnowitz* regression after Mincer and Zarnowitz (1969). If the forecast $\widehat{y}_{t+h,t}$ is *unbiased*, then we expect that $\beta_0 = 0$ and $\beta_1 = 1$. We can test this joint hypothesis on the estimated regression parameters $\widehat{\beta}_0$ and $\widehat{\beta}_1$. Furthermore, for the case of one-step ahead forecasts $h = 1$, we expect that the regression residuals are serially uncorrelated. This hypothesis can be tested from the autoregression

$$e_{t+1,t} = \gamma_0 + \gamma_1 e_t + \cdots + \gamma_p e_{t-p+1} + \nu_{t+1}$$

where all the γ parameters should be zero.

An optimal forecast error should be unforecastable from information available at time t. This suggests running the regression

$$e_{t+h,t} = \delta_0 + \delta_1 x_{1t} + \cdots + \delta_{k-1} x_{k-1,t} + \omega_{t+h}$$

where the x_{1t}, x_{2t}, $\ldots$, $x_{k-1,t}$ are a set of $k-1$ variables known at time t. This broadens the information set to test whether the variables x_{1t}, x_{2t}, $\ldots$, $x_{k-1,t}$ help explain any of the forecast error. If so, then the forecast is not optimal.

3 Evaluating forecast accuracy

Several measures have been proposed to evaluate the accuracy of a forecast. Probably the most widely used is the *mean square error* (*MSE*)

$$MSE = \frac{1}{T} \sum_{t=1}^{T} e_{t+h,t}^2$$

or its square root, *root mean square error* (*RMSE*),

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^2}$$

which has the advantage over the *MSE* of being of the same scale as a single forecast error. Note that these measures are averages of forecast errors all forecasting the same number of periods ahead, h.

Another measure is the *mean square percentage error* (*MSPE*)

$$MSPE = \frac{1}{T}\sum_{t=1}^{T}\left(100\frac{y_{t+h} - \widehat{y}_{t+h,t}}{y_{t+h}}\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(100\frac{e_{t+h,t}}{y_{t+h}}\right)^2$$

or its square root, the *root mean square percentage error* (*RMSPE*)

$$RMSPE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(100\frac{e_{t+h,t}}{y_{t+h}}\right)^2}.$$

Percentage errors adjust for the scale of the variable being forecast so that they can be used for comparing forecast errors of different variables.

Measures based on the squared error such as *RMSE* and *RMSPE* penalise large forecast errors more than small forecast errors. They are naturally associated with the quadratic loss function

$$L(e_t) = e_t^2.$$

The *mean absolute error* (*MAE*) defined by

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |e_{t+h,t}|$$

and the *mean absolute percentage error* (*MAPE*) defined by

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|100\frac{e_{t+h,t}}{y_{t+h}}\right|$$

treat large and small errors equally and are naturally associated with the absolute loss function

$$L(e_t) = |e_t|.$$

The *Theil inequality coefficient*, Theil (1961), is defined by

$$U = \frac{RMSE}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} y_{t+h}^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T} \widehat{y}_{t+h,t}^2}}.$$

This statistic always lies between zero and one. When $U = 0$ the forecast is perfect and when $U = 1$ the forecast is as bad as it could be. This statistic is a non-monotonic transformation of the *MSE* and Granger and Newbold (1973) showed that in some circumstances it can fail to select the optimal forecasting model. Theil also suggested a decomposition of forecast error into three components

$$MSE = (\overline{\widehat{y}} - \overline{y})^2 + (\widehat{\sigma}_{\widehat{y}} - \widehat{\sigma}_y)^2 + 2(1 - \widehat{\rho})\widehat{\sigma}_{\widehat{y}}\widehat{\sigma}_y \tag{1}$$

where $\overline{\widehat{y}}$ and $\overline{y}$ are the sample means of the forecast and actual values respectively, $\widehat{\sigma}_{\widehat{y}}$ and $\widehat{\sigma}_y$ are the sample standard deviations of the forecast and actual values respectively and $\widehat{\rho}$ is the sample correlation between forecast and actual values. Dividing both sides of this equation by *MSE*, the three scaled right-hand side terms in (1) sum to unity. The first is the *bias* proportion and is an indication of systematic error, the second is the *variance* proportion and indicates the extent to which the variability of forecast and actual differ and the third is the *covariance* proportion and indicates unsystematic error. Granger and Newbold criticised this decomposition and show that, for the optimal forecast, the last two terms in (1) can be made to take on any values depending on the parameter values in the model. An alternative decomposition, also due to Theil, is

$$MSE = (\overline{\widehat{y}} - \overline{y})^2 + (\widehat{\sigma}_{\widehat{y}} - \widehat{\rho}\widehat{\sigma}_y)^2 + (1 - \widehat{\rho}^2)\widehat{\sigma}_y \tag{2}$$

and this has proved to be more useful. In a good forecast the first two terms in (2) should tend to zero so that (after scaling by dividing both sides by *MSE*) the final *unsystematic* proportion is close to unity.

Clements and Hendry (1998) point out that a ranking of *multi-step* forecasts (when $h > 1$) using evaluation methods based on the *MSE* is *not invariant* to linear transformations of the forecast variable such as using Δy_{t+h} in place of y_{t+h}.

4 Evaluating macroeconomic forecasts

In evaluating forecasts from macroeconomic models, one important issue to bear in mind is that the models may be making different assumptions about exogeneity. For example, a variable that is treated as exogenous in one model may be an endogenous variable in another model. This means that the information sets in the two models are different which makes it difficult to compare forecasts with the two models. In order to compare forecasts from models with different exogeneity assumptions, some adjustment must be made such as using actual exogenous values in each model to compare forecasts.

Wallis and Whitley (1991) proposed a decomposition of a published macroeconomic forecast into three components as given by the identity

$$y - f(\widehat{x}, a) = [y - f(x, 0)] + [f(x, a) - f(\widehat{x}, a)] - [f(x, a) - f(x, 0)]. \tag{3}$$

Here x represents the exogenous variables in the model and $\widehat{x}$ are the forecast values of these variables used to produce the model forecasts and a represents judgmental adjustments that are added to the model equations by the forecaster to improve the forecast. Time subscripts have been omitted for clarity. The published forecast is $f(\widehat{x}, a)$ so that the left-hand side of (3) is the forecast error associated with the published forecast. On the right-hand side, this error is decomposed into the sum of three components. The first component, $y - f(x, 0)$, is the pure model error with no judgmental adjustments and the exogenous variables at their actual values. The second term, $f(x, a) - f(\widehat{x}, a)$, is the difference between the forecast (including judgmental adjustments) with actual values of the exogenous variables and with forecast values and so represents the error made in predicting the exogenous variables that feed into the model. The third term, $f(x, a) - f(x, 0)$, is the difference between the forecast (assuming actual values for the exogenous variables) with and without judgmental adjustments and so represents the effect of the judgmental input to the model. Using this decomposition to analyse several different UK forecasts, Wallis and Whitley found that judgmental adjustments served to improve model forecast over pure model forecasts, even though these adjustments were often only intercept corrections based on past single equation residuals, rather than incorporating any extra-model information. Surprisingly, they also found that using actual exogenous values in place of forecast values sometimes made the model forecasts worse. This perverse result that extra information makes the forecast worse is difficult to explain.

4.1 A benchmark model

Sims (1980) is strongly critical of simultaneous macroeconomic models. Sims argued that the restrictions imposed by theory in such models were untested and that the dynamics were primitive so that the models were often dynamically misspecified. As an alternative, he proposed the *vector autoregressive* or *VAR* model in which each model variable is a linear function of unrestricted lags of all the variables in the model up to a certain order. In the *VAR* framework, all model variables are treated as endogenous, as opposed to the simultaneous equation system where there may be exogenous variables. With n endogenous variables, $y_1, y_2, \ldots, y_n$, the first-order *VAR(1)* model is defined by the n equations

$$\begin{aligned} y_{1t} &= c_1 + a_{11} y_{1,t-1} + \cdots + a_{1n} y_{n,t-1} + u_{1t} \\ y_{2t} &= c_2 + a_{21} y_{1,t-1} + \cdots + a_{2n} y_{n,t-1} + u_{2t} \\ &\vdots \\ y_{nt} &= c_n + a_{n1} y_{1,t-1} + \cdots + a_{nn} y_{n,t-1} + u_{nt} \end{aligned} \tag{4}$$

This model is just a vector generalisation of a first-order autoregressive process and can be written in matrix terms as

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}\mathbf{y}_{t-1} + \mathbf{u}_t$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of observations on the endogenous variables at time t, $\mathbf{A}$ is an $n \times n$ square matrix of coefficients and $\mathbf{c}$ an $n \times 1$ vector of intercepts. The process can easily be generalised to higher order as in the p-th order *VAR(p)* process, defined by

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-1} + \mathbf{u}_t.$$

The *VAR(p)* model can be regarded as a benchmark model against which to test the forecasts from a simultaneous equations model. The *VAR* model is *atheoretical* and incorporates no restrictions from economic theory. If such restrictions are important for forecasting, then we would expect the simultaneous equation model to out-forecast the *VAR* model. In practice the reverse has often proved to be the case. McNees (1986) found that *VAR* models outperformed macroeconomic models in forecasting many macroeconomic variables in the US and similar results have been found for the UK. However, these results have been criticised by Hendry and Clements who argue that although *VAR* models may outperform structural models in times of structural change, they will be subject to unstable parameters and will not outperform structural models in times of structural stability.

5 Forecast encompassing

One forecast is said to *encompass* another forecast when the forecasts of the first model can 'explain' the forecasts of the other model. In this case the first forecast incorporates all the information in the other forecast and there is nothing to be gained by combining them. In the regression

$$y_{t+h} = \beta_1\widehat{y}^a_{t+h,t} + \beta_2\widehat{y}^b_{t+h,t} + \varepsilon_{t+h}$$

$\widehat{y}^a_{t+h,t}$ and $\widehat{y}^b_{t+h,t}$ are the forecasts for y_{t+h} from models a and b respectively. If the forecasts from model a encompass the forecasts from model b, then we expect that $\beta_1 = 1$ and $\beta_2 = 0$. since the forecasts from model b contain no additional information not already included in the forecasts from model a. Conversely, if the forecasts from model b encompass the forecasts from model a, then we expect that $\beta_1 = 0$ and $\beta_2 = 1$. In all other cases, neither forecast encompasses the other and there is a gain to combining the two forecasts to form a composite forecast. Chong and Hendry (1986) argue that if neither forecast encompasses the other forecast, then both forecasting models must be misspecified.

6 Combining forecasts

Suppose that we have several competing forecasts. How do we proceed? One possibility is simply to choose the 'best' forecast using the measures of forecast accuracy discussed above. Another possibility is to try and combine the forecasts together to produce a *composite* forecast. It turns out that this is often a fruitful approach. Two methods for forecast combination have been suggested in the literature: the covariance method first put forward by Bates and Granger (1969) and the regression approach of Granger and Ramanathan (1984). It turns out that both methods give the same results.

6.1 Covariance method

Suppose that we have two unbiased forecasts $\widehat{y}^a_{t+h,t}$ and $\widehat{y}^b_{t+h,t}$ with associated forecast errors $e^a_{t+h,t}$ and $e^b_{t+h,t}$. Each of these forecast errors has a variance

$$\text{var}(e^a_{t+h,t}) = \sigma^2_a$$

$$\text{var}(e^b_{t+h,t}) = \sigma^2_b$$

and the two forecast errors have a covariance defined by

$$\text{cov}(e^a_{t+h,t}, e^b_{t+h,t}) = \sigma_{ab}.$$

Consider the composite forecast defined by the linear combination

$$\widehat{y}^c_{t+h,t} = \alpha\widehat{y}^a_{t+h,t} + (1-\alpha)\widehat{y}^b_{t+h,t} \tag{5}$$

where $0 \leq a \leq 1$ is the weight associated with forecast $\widehat{y}^a_{t+h,t}$ and $0 \leq 1-a \leq 1$ is the weight associated with forecast $\widehat{y}^b_{t+h,t}$. Note that we impose the restriction that the weights sum to unity.

The forecast error of the composite forecast is given by

$$e^c_{t+h,t} = \alpha e^a_{t+h,t} + (1-\alpha)e^b_{t+h,t}$$

and will have variance defined by

$$\text{var}(e^c_{t+h,t}) = \sigma^2_c = \alpha^2\sigma^2_a + (1-\alpha)^2\sigma^2_b + 2\alpha(1-\alpha)\sigma_{ab}. \tag{6}$$

We can choose the optimal weight α^* to minimise the variance of the composite forecast. Differentiating (6) with respect to α and setting to zero we have

$$\frac{\partial\,\text{var}(e^c_{t+h,t})}{\partial\alpha} = 2\alpha\sigma^2_a - 2(1-\alpha)\sigma^2_b + 2(1-\alpha)\sigma_{ab} - 2\alpha\sigma_{ab} = 0$$

which can be solved for α to give

$$\alpha^* = \frac{\sigma_b^2 - \sigma_{ab}}{\sigma_a^2 + \sigma_b^2 - 2\sigma_{ab}}.$$

It can be shown that the optimally combined forecast $\widehat{y}^c_{t+h,t}$ will have a forecast error variance that is no larger than the smaller of the variances of the two component forecasts or

$$\sigma_c^2 \leq \min(\sigma_a^2, \sigma_b^2).$$

Thus we cannot lose by combining forecasts and will generally gain.

In the special case where the two component forecasts are uncorrelated, $\sigma_{ab} = 0$ and we have

$$\alpha^* = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2}.$$

As $\sigma_b^2 \to 0$, the second forecast becomes more precise and, from the formula, we see that $\alpha^* \to 0$ so that all the weight is put on forecast b and $\widehat{y}^c_{t+h,t} \to \widehat{y}^b_{t+h,t}$. Conversely, as $\sigma_a^2 \to 0$, the first forecast becomes more precise and, from the formula, $\alpha^* \to 1$ so that $\widehat{y}^c_{t+h,t} \to \widehat{y}^a_{t+h,t}$.

In practice we do not know the population variances σ_a^2, σ_b^2 and σ_{ab} so we replace them by consistent estimates $\widehat{\sigma}_a^2$, $\widehat{\sigma}_b^2$ and $\widehat{\sigma}_{ab}$ and use the formula

$$\widehat{\alpha}^* = \frac{\widehat{\sigma}_b^2 - \widehat{\sigma}_{ab}}{\widehat{\sigma}_a^2 + \widehat{\sigma}_b^2 - 2\widehat{\sigma}_{ab}}.$$

6.2 Regression method

Consider the regression

$$y_{t+h} = \beta_0 + \beta_1 \widehat{y}^a_{t+h,t} + \beta_2 \widehat{y}^b_{t+h,t} + \varepsilon_{t+h}. \tag{7}$$

This can thought of as a way of estimating the optimal parameters in the linear combination of forecasts (5) in which case we expect $\beta_0 = 0$, $\beta_1 = \alpha$ and $\beta_2 = (1 - \alpha)$. Including an intercept β_0 in the regression allows for the possibility that the forecasts are *biased* in which case β_0 estimates the extent of the bias. Estimating the parameters β_1 and β_2 separately allows for the possibility that $\beta_1 + \beta_2 \neq 1$. To impose the condition that $\beta_1 + \beta_2 = 1$, we can rewrite the regression (7) as

$$y_{t+h} - \widehat{y}^b_{t+h,t} = \beta_0 + \beta_1(\widehat{y}^a_{t+h,t} - \widehat{y}^b_{t+h,t}) + \varepsilon_{t+h}. \tag{8}$$

From the definition of regression, the *OLS* estimate of β_1 in (8) is given by

$$\widehat{\beta}_1 = \frac{\widehat{\text{cov}}(\widehat{y}^a_{t+h,t} - \widehat{y}^b_{t+h,t})(y_{t+h} - \widehat{y}^b_{t+h,t})}{\widehat{\text{var}}(\widehat{y}^a_{t+h,t} - \widehat{y}^b_{t+h,t})} = \frac{\widehat{\sigma}_b^2 - \widehat{\sigma}_{ab}}{\widehat{\sigma}_a^2 + \widehat{\sigma}_b^2 - 2\widehat{\sigma}_{ab}} = \widehat{\alpha}^*$$

so the regression method will estimate the optimal value of α. The regression method can easily be generalised to the case where there are more than two component forecasts. It is also possible to allow for serial correlation in the regression model (7) by assuming that ε_{t+h} follows an *ARMA(p,q)* process. This is especially important in the case of multi-step forecasts with $h > 1$ where we know that even optimal forecasts will be serially correlated.

The regression method can also be generalised to produce optimal non-linear combinations of forecasts such as the quadratic model

$$y_{t+h} = \beta_0 + \beta_1 \widehat{y}^a_{t+h,t} + \beta_2 \widehat{y}^b_{t+h,t} + \beta_3 (\widehat{y}^a_{t+h,t})^2 + \beta_4 (\widehat{y}^b_{t+h,t})^2 + \beta_5 (\widehat{y}^a_{t+h,t})(\widehat{y}^b_{t+h,t}) + \varepsilon_{t+h}.$$

In this model squares and cross-products of the component forecasts are included in the regression. Higher order powers could also be included and the linear model (7) is simply the special case where $\beta_3 = \beta_4 = \beta_5 = 0$.

6.3 Averaging forecasts

Rather than estimate an optimal combination of forecasts, a simpler method is to impose an equal weighting and take the simple average of the forecasts. In the case of two forecasts, this results in the composite forecast

$$\widehat{y}^m_{t+h,t} = 0.5 \widehat{y}^a_{t+h,t} + 0.5 \widehat{y}^b_{t+h,t}$$

with variance

$$\operatorname{var}(e^m_{t+h,t}) = 0.25\sigma^2_a + 0.25\sigma^2_b + 0.5\sigma_{ab}.$$

Since in general, for any two random variables a and b,

$$\operatorname{var}(a-b) = \operatorname{var}(a) + \operatorname{var}(b) - 2\operatorname{cov}(a,b) = \sigma^2_a + \sigma^2_b - 2\sigma_{ab} \geq 0$$

(with equality only in the case of perfect correlation of a and b), it follows that

$$2\sigma_{ab} \leq \sigma^2_a + \sigma^2_b \tag{9}$$

so that averaging will normally reduce the forecast variance. For example, with $\sigma^2_a = \sigma^2_b = 1$, the inequality (9) implies that $\sigma_{ab} \leq 1$ so that

$$\operatorname{var}(e^m_{t+h,t}) = 0.5 + 0.5\sigma_{ab} \leq 1 = \sigma^2_a = \sigma^2_b$$

and the composite forecast has smaller variance than either of its component forecasts. The average forecast will be unbiased as long as all the component forecasts are unbiased. However, because no intercept is estimated, the average forecast will in general be biased if any of the component forecasts are biased (unless by fluke the weighted sum of the biases is zero).

The UK Treasury publishes summary forecasts based on averages of many different forecasts made by city firms and other independent forecasters. In the absence of information about the variances and covariances of the individual forecasts, the composite summary forecast should have smaller variance than the individual forecasts from which it is constructed.

References

[1] Bates, J.M. and C.W.J. Granger (1969), 'The combination of forecasts', *Operations Research Quarterly*, 20, 451–468.

[2] Chong, Y.Y. and D.F. Hendry (1986), 'Econometric evaluation of linear macroeconomic models', *Review of Economic Studies*, 53, 671–690.

[3] Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.

[4] Granger, C.W.J. and P. Newbold (1973), 'Some comments on the evaluation of economic forecasts', *Applied Economics*, 5, 35–47.

[5] Granger, C.W.J. and R. Ramanathan (1984), 'Improved methods of combining forecasts', *Journal of Forecasting*, 3, 197–204.

[6] McNees, S.K. (1986), 'Forecasting accuracy of alternative techniques: a comparison of US macroeconomic forecasts', *Journal of Business and Economic Statistics*, 4, 5–23.

[7] Mincer, J. and V. Zarnowitz (1969), 'The evaluation of economic forecasts', in J. Mincer (ed.), *Economic Forecasts and Expectations*, National Bureau of Economic Research, New York.

[8] Sims, C.A. (1980), 'Macroeconomics and reality', *Econometrica*, 48, 1–48.

[9] Theil, H. (1961), *Economic Forecasts and Policy*, North-Holland, Amsterdam.

[10] Wallis, K.F. and J.D. Whitley (1991), 'Sources of error in forecasts and expectations: UK economic models, 1984–1988', *Journal of Forecasting*, 10, 231–253.