An Historical Perspective on Forecast Errors

Michael P. Clements
Department of Economics,
University of Warwick

and David F. Hendry*
Department of Economics,
University of Oxford.

October 21, 2000

Abstract

Using annual observations on industrial production over the last three centuries, and on GDP over a 100 years, we seek an historical perspective on the forecastability of these UK output measures. The series are dominated by strong upward trends, so we consider various specifications of this, including the local linear trend structural time-series model, which allows the level and slope of the trend to vary. Our results are not unduly sensitive to how the trend in the series is modelled: the average sizes of the forecast errors of all models, and the wide span of prediction intervals, attests to a great deal of uncertainty in the economic environment. It appears that, from an historical perspective, the post-War period has been relatively more forecastable.

1 Introduction

How well or badly forecasters are doing should be measured relative to the predictability of the relevant variables. Little is known about any lower bounds to the predictability of macro-economic time series, and how such bounds change over time, but the historical properties of data is open to analysis. Consequently, we investigate the historical forecastability of UK output measures, using annual observations on UK GDP and industrial production (IP) over the last 100 and 300 years respectively, to gauge changes in predictability relative to some ‘base-line’ models.¹

The forecastability of a variable depends both upon its intrinsic properties and on the model used. The former is not under the investigators’ control, but the epochs in question here witnessed many fundamental technological, legal, social and political changes, most of which had consequences that were not well understood till long afterwards. Part of the apparent change may well be improvements in the measurement system itself. The forecasting model, however, is under an investigators control, and recent research into the sources of forecast failure suggests using univariate time-series forecasting models which are relatively robust to deterministic shifts: section 2 reviews the arguments. A univariate approach obviates the need to model ‘explanatory variables’, which is distinctly difficult over the long historical periods considered (see Hendry, 2000a, for models of several macro time series over 1875–1990), and anyway need not improve forecast accuracy or precision when underlying relationships are

*Financial support from the U.K. Economic and Social Research Council under grant L116251015 is gratefully acknowledged by both authors. All computations were performed using code written in Gauss. Tommaso Proietti kindly provided the Gauss code to estimate the local linear trend model. The data were kindly provided by Charlie Bean. The UK industrial production series is the ‘Output in Industry’ series, and GDP is constant price at factor cost.

¹The data were kindly provided by Charlie Bean. ‘Output in Industry’ was compiled from Crafts and Harley (1992), p.725; Mitchell (1988), p.846; and Central Statistical Office (1993): the data were missing during the second World War so were linearly interpolated between 1938 and 1946. Real GDP in 1985 prices comes from Mitchell (1988), p.836; and *Economic Trends Annual Supplement*, (1993).

changing. Linear time-series models (e.g., of the Box and Jenkins, 1976, form) are easily understood and estimated, and while they may exhibit non-constancy, we allow for that by re-estimating the models' parameters as the estimation and forecast windows move through the data, allowing for gradual parameter evolution. We also consider a class of models that allows the levels and slopes of the trend components of the models to evolve through time. Casual observation of the data (see figure 1) shows that strong upward trends are the dominant features to be modelled, so we focus on this aspect. The growth rates have been volatile, more so for IP than GDP, and the volatility has not been constant over time.

The paper has three objectives. First, we investigate if the 'structural' time-series models employed here, which allow the trend function to change over time, yield either more accurate point forecasts, or better capture the uncertainty surrounding such forecasts, than linear time-series models. To resolve this issue, we report a number of measures of forecast accuracy that assess the performance of models against each other, and consider whether the models' prediction intervals are well calibrated. Secondly, we examine whether the last century was relatively more quiescent (and therefore more forecastable) than the 18th and 19th centuries. Finally, we seek to identify whether periods when forecast performance was particularly poor coincide with periods of turbulence in the economic or legislative fields, including domestic and international conflict. Clements and Hendry (1998b, 1999a) argue that there should be such a match, in so far as turbulence induces deterministic shifts, which are the most pernicious source of forecast failure.

Figure 1 UK output measures and their growth rates.

Section 2 reviews the theoretical background that leads to our choice of forecasting models, and section 3 discusses the models we use. Section 4 provides a brief overview of the measures used to evaluate forecast performance. Section 5 describes the empirical forecasting exercise and our results, and section 6 concludes.

2 Background

Economic forecasting occurs in a non-stationary and evolving world, where the model and the data generation process (DGP) are bound to differ. Long historical time series merely serve to highlight change; and change in turn highlights mis-specifications in the models for the DGP. Key attributes of forecasts are their accuracy and precision, and although other features of the forecast-error distribution may matter, we will focus on bias and variance as measures thereof, despite their well-known drawbacks. Inaccurate or imprecise forecasts may simply reflect the intrinsic difficulty of forecasting the given series. Because there are no absolutes against which to assess the accuracy and/or precision of a forecast, root mean-square forecast errors (which weight the inaccuracy and imprecision together) will be compared across models, and statistical tests used to check if one model is significantly better than the others on this criterion. We also use tests due to Christoffersen (1998) to check whether each model's prediction intervals contain the appropriate number of realized values of the process. However, forecast failure – which occurs when there is a significant deterioration in forecast performance relative to the anticipated outcome – is more easily detected, and can be assessed in a number of ways: see, for example, the tests discussed in Kiviet (1986) and Clements and Hendry (1999b).

Most forecasting models have three main components: deterministic terms (like intercepts) whose future values are known; observed stochastic variables (like prices) with unknown future values; and unobserved errors all of whose values (past, present and future) are unknown. Any, or all, of these components, or the relationships between them, could be inappropriately formulated, inaccurately estimated, or change in unanticipated ways. All nine types of mistake *could* induce poor forecast performance, either from inaccurate or imprecise forecasts. Instead, Clements and Hendry (1998b, 1999a) find that some mistakes have pernicious effects on forecasts, whereas others are relatively less important in most settings.

This result follows from a taxonomy of forecast errors which allows for structural change in the forecast period, the model and DGP to differ over the sample period, the parameters of the model to be estimated (possibly inconsistently) from (potentially inaccurate) data, and the forecasts to commence from incorrect initial conditions. The taxonomy reveals the central role in forecast failure of deterministic shifts over the forecast period, whereas other problems, such as poorly-specified models, inaccurate data, inadequate methodology, over-parameterization, or incorrect estimators seem less relevant. Surprisingly, it is even difficult to detect shifts in parameters other than those of deterministic terms (see e.g., Hendry, 2000b). Of course, all inadequacies in models reduce forecast performance relative to the optimum, but the analytics, supported by Monte Carlo and empirical studies, suggest that such effects are 'swamped' by the large errors manifested in forecast failure.

Such a theory reveals that many of the conclusions which can be established formally for correctly-specified forecasting models of constant-parameter processes no longer hold. In weakly stationary processes – or integrated processes reducible to stationarity by differencing and cointegration – a congruent, encompassing model in-sample will dominate in forecasting at all horizons (see Hendry, 1995, for definitions). Causal variables will always improve forecasts relative to non-causal, forecast accuracy will deteriorate as the horizon increases, and there should be no forecast-accuracy gains from pooling forecasts across methods or models: indeed, pooling refutes encompassing. Forecast failure will be a rare event, precisely because the future will be like the present and past. Instead, when the DGP is not stationary and the model does not coincide with the DGP, then new implications are that causal variables need not dominate non-causal in forecasting; forecast failure will occur when the future changes from the present, and is not predictable from in-sample tests, but need not entail changes in estimated parameters nor invalidate a model; and h -step ahead forecasts can 'beat' 1-step forecasts made $h - 1$ periods later

($h > 1$). Moreover, the outcome of forecasting competitions on economic time series will be heavily influenced by the occurrence of unmodelled deterministic shifts that occur prior to forecast-evaluation periods: models which are robust to such shifts will do relatively well.

Given that forecast failure mainly derives from deterministic shifts, there are four potential solutions: differencing, co-breaking, intercept corrections, and updating. Differencing lowers the polynomial degree of deterministic terms: in particular, double differencing usually leads to a mean-zero, trend-free series, because continuous acceleration is rare in economics (even during hyperinflations). The recent study of the Norges Bank model in Eitrheim, Husebø and Nymoen (1999) illustrates the effectiveness of that approach. Next, deterministic non-stationarity can also be removed by co-breaking, namely the cancellation of breaks across linear combinations of variables (see e.g., Clements and Hendry, 1999a, ch. 9). Finally, intercept corrections can be shown to help robustify forecasts against biases due to deterministic shifts, as can the closely-related approach of updating the estimated intercept with an increased weight accorded to more recent data.

These implications help determine the class of model that might minimize systematic forecast failure over the long periods under analysis here.

3 Models

Because of the vexed question of whether output series possess unit roots or are better described as stationary around a linear trend, we estimate both types of model, namely difference stationary (DS) and trend stationary (TS) models. The stochastic-trend model treats the variable $\{y_t\}$ as integrated of order one, $y_t \sim I(1)$, as in the random walk with drift:

$$y_t = y_{t-1} + \mu + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]. \quad (1)$$

The TS model is ostensibly quite different, whereby $\{y_t\}$ is stationary about a deterministic function of time, here taken to be a simple linear trend:

$$y_t = \phi + \gamma t + u_t \quad \text{where} \quad u_t \sim \text{IN}[0, \sigma_u^2]. \quad (2)$$

The disturbances in both models can be stationary Gaussian autoregressive-moving average (ARMA) processes without fundamentally altering the properties of these models (see Clements and Hendry, 2000).

However, both these models can be viewed as special cases of the ‘structural time-series’ class of models of Harvey (1989) (Proietti, 2001, provides a review). Ignoring cyclical and seasonal components, the local linear trend (LLT) structural time-series model can be written as:

$$y_t = \mu_t + \varepsilon_t \quad (3)$$

where:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_{t-1} \\ \beta_t &= \beta_{t-1} + \zeta_{t-1} \end{aligned} \quad (4)$$

and the disturbance terms ε_t , η_t and ζ_t are zero mean, uncorrelated white noise, with variances given by σ_ε^2 , σ_η^2 and σ_ζ^2 . The interpretation of this model is that μ_t is the (unobserved) trend component of the time series, and ε_t is the irregular component. η_t affects the level of the trend, and ζ_t allows its slope (β) to change. We investigate whether this more flexible approach leads to improved forecast accuracy over (1) and (2).

Differencing the equation for μ_t in (4):

$$\Delta^2 \mu_t = \Delta \beta_{t-1} + \Delta \eta_{t-1} = \zeta_{t-2} + \Delta \eta_{t-1},$$

and substituting into (3) differenced twice, we find:

$$\Delta^2 y_t = \zeta_{t-2} + \Delta \eta_{t-1} + \Delta^2 \varepsilon_t,$$

which is a restricted ARIMA(0, 2, 2).²

Estimation of the unknown error variances is by maximum likelihood, most easily achieved after casting the model into state-space form (SSF), e.g.:

$$\begin{aligned} y_t &= \mathbf{Z}\boldsymbol{\alpha}_t + \mathbf{G}\boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}\boldsymbol{\alpha}_t + \mathbf{H}\boldsymbol{\varepsilon}_t, \end{aligned}$$

which define the measurement and state equations respectively, where $\boldsymbol{\alpha}_t = [\mu_t \ \beta_t]'$, $\mathbf{Z} = [1 \ 0]$, $\mathbf{G} = [\sigma_\varepsilon \ 0 \ 0]$, $\boldsymbol{\varepsilon} = [\varepsilon_t/\sigma_\varepsilon \ \eta_t/\sigma_\eta \ \zeta_t/\sigma_\zeta]$ and:

$$\mathbf{T} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 0 & \sigma_\eta & 0 \\ 0 & 0 & \sigma_\zeta \end{bmatrix}.$$

For given values of the hyperparameters $[\sigma_\varepsilon, \sigma_\eta, \sigma_\zeta]$, the Kalman filter is used to obtain the prediction error decomposition, that is, the one-step ahead independent prediction errors, from which the likelihood can be assembled. The likelihood is then maximized over these hyperparameters by a numerical optimization routine.

Letting $\mathcal{I}_T$ denote available information, the h -step ahead forecasts are given by:

$$\mathbf{E}[y_{T+h} | \mathcal{I}_T] = \mathbf{Z}\mathbf{E}[\boldsymbol{\alpha}_{T+h} | \mathcal{I}_T] = \mathbf{Z}\mathbf{T}^h \tilde{\boldsymbol{\alpha}}_{T|T}$$

since $\mathbf{E}[\mathbf{G}\boldsymbol{\varepsilon}_{T+h} | \mathcal{I}_T] = \mathbf{0}$, where $\tilde{\boldsymbol{\alpha}}_{T|T}$ is the smoothed estimate of the state vector at time T . For the LLT model, after estimation:

$$\tilde{y}_{T+h|T} = \mathbf{E}[y_{T+h} | \mathcal{I}_T] = \tilde{\mu}_{T|T} + h\tilde{\beta}_{T|T}$$

with an h -step ahead mean-square forecast error (MSFE) of:

$$\text{MSFE}[\tilde{y}_{T+h|T}] = p_{11} + 2hp_{12} + h^2p_{22} + h\sigma_\eta^2 + h(h-1)(2h-1)\sigma_\zeta^2/2 + \sigma_\varepsilon^2 \quad (5)$$

where $\mathbf{P} = \tilde{\mathbf{P}}_{T|T} = \{p_{ij}\}$ is the smoothed estimate of the covariance of the state vector: details are provided in the appendix.

Suppose now that $\sigma_\zeta^2 = 0$, then from (4), $\Delta\mu_t = \beta + \eta_{t-1}$, so that differencing (3) and substituting delivers:

$$\Delta y_t = \beta + \eta_{t-1} + \Delta \xi_t,$$

which is the same as (1) when the disturbance in the latter is an MA(2). Thus, the DS model is a structural time-series model in which the ‘slope’ does not change. The TS model can be obtained as a special case of the structural time-series model by setting $\sigma_\eta^2 = \sigma_\zeta^2 = 0$, so that (4) becomes:

$$\mu_t = \mu_{t-1} + \beta = \mu_0 + \beta t. \quad (6)$$

²See, e.g., Harvey (1989) or Proietti (2001) for the restrictions on the lag 1 and 2 autocorrelations relative to an unrestricted second-order MA process.

(6) indicates that the trend has a constant ‘slope’ ($\sigma_\zeta^2 = 0$), and the ‘level’ is not subject to stochastic shocks ($\sigma_\eta^2 = 0$). Substituting into (3) results in:

$$y_t = \mu_0 + \beta t + \xi_t$$

which is identical to (2) when $\mu_0 = \phi$, $\beta = \gamma$ and $\xi_t = u_t$. Thus, the trend-stationary model is a limiting case of the structural time-series model in which the level and the slope of the trend component are constant over time.

As noted in the previous section, devices are available to robustify forecasts against breaks in means or growth rates. These are discussed at length in Clements and Hendry (1999a), drawing on, *inter alia*, Clements and Hendry (1996, 1998a), where it is shown that bias can often be reduced at the cost of greater imprecision. We experimented with a ‘double-difference’ (DD) predictor, whereby forecasts are generated by solving $\Delta^2 y_t = 0$ for $t = T + 1, \dots, T + h$, to yield predictions of periods $T + 1$ to $T + h$. Here, $\Delta = 1 - L$, $L^n x_t = x_{t-n}$, so $\Delta^2 = (1 - L) - (1 - L)L$. This predictor instantly adapts to changes in the growth rate of y_t : the growth rate observed at period T is predicted into the future. Notice that the DS model incorporates a unit root which will adapt for changes in the level of the series, but not for changes in the growth rate, μ . That is, the growth rate of μ is expected to hold globally. The LLT model allows a local trend, as does a model which estimates γ on a window of observations ending at the forecast origin. We term this model DS-L, and calculate the growth rate from the last five years of data. The DS model uses all the data up to the forecast origin (and may include dynamics, but this is of secondary importance). Finally, we consider ‘intercept-correcting’ the TS model forecasts by adding in the average of the previous two errors (up to and including the forecast origin error). This predictor is denoted TS-IC.

4 Forecast evaluation measures

Until recently, attention in macroeconomics was focused primarily on the production and evaluation of point forecasts (see, e.g., Wallis, 1995), with only summary information regarding the degree of uncertainty associated with forecasts, such as standard errors, being provided. However, Christoffersen (1998) suggests ways of evaluating interval forecasts, as a step toward providing a more complete description of the uncertainty surrounding forecasts.

4.1 Point forecasts

Although there is a large literature on forecast-accuracy comparisons, there have been few formal comparisons of rival forecasts of the same phenomena, where by ‘formal’, we mean comparisons which attempt to assess whether differences between rival forecasts can be attributed to sampling variability, or whether they are ‘significant’. Diebold and Mariano (1995) propose a test of the null of equal forecast accuracy, for an arbitrary ‘loss function’, $g(e_{i,T+h})$, where $e_{i,T+h}$ is the h -step ahead forecast error from using model i commencing at the forecast origin T . The loss differential is defined as $d_{T+h} \equiv [g(e_{i,T+h}) - g(e_{j,T+h})]$ for rival forecasts i and j , so that equal forecast accuracy entails $E[d_{T+h}] = 0$. Given a covariance-stationary sample realization $\{d_t\}$ of n ‘observations’, the asymptotic distribution of the sample mean loss differential $\bar{d}$ ($\bar{d} = n^{-1} \sum_{t=1}^n d_t$) under the null is given by:

$$\sqrt{n} (\bar{d}) \xrightarrow{D} N [0, V (\bar{d})],$$

where we can approximate the variance by:

$$V(\bar{d}) = n^{-1} \left(\gamma_0 + 2 \sum_{i=1}^{h-1} \gamma_i \right), \quad (7)$$

where γ_i is the i^{th} autocovariance (so $V(\bar{d})$ is a weighted sum of autocovariances), and the truncation is determined by the forecast horizon. Diebold and Mariano (1995) discuss the choice of weights (here, rectangular) and the truncation lag.

The large-sample statistic that Diebold and Mariano (1995) propose for testing the null of equal forecast accuracy is:

$$\frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \overset{app}{\sim} N[0, 1],$$

where $\hat{V}(\bar{d})$ is a consistent estimate of $V(\bar{d})$, using a weighted sum of the sample autocovariances in (7):

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d}).$$

Harvey, Leybourne and Newbold (1997) propose a modified version of this test statistic that corrects for its tendency to be over-sized, and present simulation evidence that attests to the usefulness of their modification. They suggest using:

$$\sqrt{\frac{n+1-2h+n^{-1}h(h-1)}{n}} \times \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}}$$

and comparing this to the Student t distribution with $n-1$ degrees of freedom. To simplify matters, we consider only 1-step forecasts, so that serial correlation is ignored, and use:

$$\hat{\gamma}_0 = n^{-1} \sum_{k=1}^n (d_k - \bar{d})^2,$$

with:

$$\frac{\bar{d}}{\sqrt{\hat{\gamma}_0}} \overset{app}{\sim} N[0, 1],$$

because the correction factor is $\sqrt{n^{-1}(n-1)} \simeq 1$, and for $n > 30$, the normal closely approximates the t distribution.

4.2 Forecast encompassing

For two rival 1-step forecasts, $y_{i,t}$ and $y_{j,t}$, a forecast-encompassing test is implemented by regressing $e_{i,t}$ ($= y_t - y_{i,t}$) on $y_{i,t} - y_{j,t}$:

$$y_t - y_{i,t} = \alpha (y_{i,t} - y_{j,t}) + \zeta_t \quad (8)$$

and using a t-test of $H_0: \alpha = 0$. This is equivalent to the regression:

$$y_t = (1 + \alpha)y_{i,t} - \alpha y_{j,t} + \zeta_t. \quad (9)$$

and is a variant of the test due to Chong and Hendry (1986): see Ericsson (1992) and Clements and Hendry (1998b). Harvey, Leybourne and Newbold (1998) show that this test may be over-sized when

there is a substantial sample of forecast errors available and the forecast errors are not approximately normal. They suggest a number of alternatives, including:

$$R_1 = \frac{1}{\sqrt{n}} \hat{Q}_1^{-\frac{1}{2}} \sum_{t=1}^n (e_{i,t} - e_{j,t})^2 \hat{\alpha}$$

where:

$$\hat{Q}_1 = \frac{1}{n} \sum_{t=1}^n (e_{i,t} - e_{j,t})^2 \hat{\zeta}_t^2$$

and $\hat{\alpha}$ and $\hat{\zeta}_t$ are the OLS estimates of α and the residuals from (8). Harvey *et al.* (1998) also show that the Diebold–Mariano test of equal MSFE is closely related to the notion of forecast encompassing.

4.3 Interval evaluation

Christoffersen (1998) suggests that a ‘good’ interval forecast should have correct conditional coverage, such that the interval is wider in volatile periods than in those of greater tranquility: otherwise, observations outside the interval will be clustered in volatile periods, and absent in tranquil periods. Christoffersen (1998, p.844) develops a ‘unified framework for testing the conditional coverage’ using three likelihood-ratio (LR) tests for unconditional coverage, independence, and conditional coverage respectively.

Let $L_{t|t-1}(p)$ and $U_{t|t-1}(p)$ denote the lower and upper limits of the interval forecast of y_t made at time $t - 1$, with a coverage probability p . Define the indicator function, 1_t , which takes the value unity when y_t lies inside the interval range and zero otherwise. For 1-step ahead forecasts $t = 1, \dots, n$, there is a sequence of outcomes $1_{t=1}^n$.

We wish to test that:

$$E[1_t | 1_{t-1}, 1_{t-2}, \dots, 1_1] = E[1_t] = p.$$

That is, at each point in time, the probability that the actual outcome will fall within the interval is p , and this is independent of the history up to that point. Christoffersen (1998, Lemma 1) establishes that this is equivalent to testing that the sequence $\{1_t\}$ is identically and independently distributed as a Bernoulli distribution with parameter p , denoted: $\{1_t\} \sim \text{IID}_B(p)$.

We can test this hypothesis in two parts. First, the test of correct unconditional coverage: $E[1_t] = p$ versus $E[1_t] = \pi \neq p$. For a sample of n intervals, the likelihood under the null is:

$$L(p; 1_1, 1_2, \dots, 1_n) = (1 - p)^{n_0} p^{n_1}$$

where n_0 is the number of times the actual does not fall within the interval, and n_1 is the number of ‘hits’. The likelihood under the alternative is:

$$L(\pi; 1_1, 1_2, \dots, 1_n) = (1 - \pi)^{n_0} \pi^{n_1},$$

so the standard LR test is:

$$LR_{uc} = -2 \ln \left(\frac{L(p; \cdot)}{L(\hat{\pi}; \cdot)} \right) \underset{\sim}{\sim} \chi_1^2,$$

where $\underset{\sim}{\sim}$ denotes ‘asymptotically distributed as’, and $\hat{\pi} = n_1 / (n_0 + n_1)$.

To test for independence, model the indicator function by a binary first-order Markov chain with transition probability matrix:

$$\mathbf{\Pi}_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix} \quad (10)$$

where $\pi_{ij} = \Pr(1_t = j | 1_{t-1} = i)$. Under independence, $\pi_{ij} = \pi_j$, $i, j = 0, 1$ where $\pi_j = \Pr(1_t = j)$. So under the null, (10) is restricted to:

$$\mathbf{\Pi}_2 = \begin{bmatrix} 1 - \pi_1 & \pi_1 \\ 1 - \pi_1 & \pi_1 \end{bmatrix} \quad (11)$$

The π_{ij} and π_1 are estimated by their sample frequencies, so:

$$\hat{\mathbf{\Pi}}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00} + n_{01}} & \frac{n_{01}}{n_{00} + n_{01}} \\ \frac{n_{10}}{n_{10} + n_{11}} & \frac{n_{11}}{n_{10} + n_{11}} \end{bmatrix},$$

where n_{ij} is the number of times j is followed by i , etc., and:

$$\hat{\pi}_1 = \frac{n_{01} + n_{11}}{n_{00} + n_{10} + n_{01} + n_{11}}.$$

The unrestricted likelihood is:

$$L(\hat{\mathbf{\Pi}}_1) = (1 - \hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1 - \hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}} \quad (12)$$

and the restricted is:

$$L(\hat{\mathbf{\Pi}}_2) = (1 - \hat{\pi}_1)^{(n_{00} + n_{10})} \hat{\pi}_1^{(n_{01} + n_{11})}. \quad (13)$$

The likelihood-ratio test statistic is:

$$LR_{ind} = -2 \ln \left(\frac{L(\hat{\mathbf{\Pi}}_2)}{L(\hat{\mathbf{\Pi}}_1)} \right) \underset{a}{\sim} \chi_1^2,$$

under the null hypothesis of independently-distributed indicator function values.

A joint test of correct conditional coverage is $LR_{cc} = LR_{uc} + LR_{ind} \underset{a}{\sim} \chi_2^2$.

5 Empirical forecast comparisons

A number of sampling schemes could be adopted for forecast evaluation (see, e.g., West and McCracken, 1998, pp.818–9). We use a ‘recursive’ scheme, where the parameters are re-estimated in each period as the forecast origin moves forward through the sample. Thus, for an observation vector of length T , the model is first specified and estimated on the data up to period T_0 ($T_0 < T$), and a forecast (point or interval) of $T_0 + 1$ to $T_0 + 4$ is made. Then, the model is re-estimated on data up to and including $T_0 + 1$, and forecasts of $T_0 + 2$ to $T_0 + 6$ are made, and so on.

The specifications of the models are fixed, although we could have optimized lag orders for each estimation sample according to some information criterion, such as BIC (see Schwarz, 1978) or PIC (see Phillips, 1994): Swanson and White (1997) found improved accuracy from such a strategy in some cases. Prediction intervals were calculated using the ‘plug-in’ approach of Box and Jenkins for the DS and TS models. For the LLT model, a $100 \times \alpha\%$ interval was calculated as the central forecast $\pm \Phi^{-1}(\alpha/2) \times \sqrt{\text{MSFE}[\tilde{y}_{T+h|T}]}$, where $\text{MSFE}[\tilde{y}_{T+h|T}]$ is given in (5), and Φ is the standard normal distribution function. This approach ignores parameter-estimation uncertainty (the model parameters are simply replaced by sample estimates), and assumes the model disturbances are Gaussian. Clements and Taylor (2000) survey this literature, and suggest a preferable bootstrap approach, but these refinements would be unlikely to alter the overall picture.

5.1 Industrial production (IP)

We generated sequences of 1 to 4-step ahead forecasts from these univariate time series models for the logarithm of the IP series (y_t) (1700–1988). For the unit root (or difference-stationary, DS) model we specified an ARIMA(1, 1, 0), allowing a non-zero mean; and for the trend-stationary (TS) model, we specified an ARIMA(2, 0, 0) process for $u_t = y_t - \phi - \gamma t$, where ϕ is the intercept and γ , the coefficient on the time trend (t), is the underlying growth rate. Alternative specifications were generally no better in terms of MSFE for the period as a whole, but we did not undertake extensive specification searches. The LLT was as described in section 3, as were the DD, DS-L and TS-IC.

The models were estimated from the period 1700 to 1758, and forecasts were then made of the next four years. The same specifications were re-estimated on data up to 1759, and 1 to 4-step ahead forecasts were again generated. This process was continued up to an estimation sample ending in 1987, with 1 to 4-step ahead forecasts of the observations 1988 to 1991. At each forecast origin, we also calculated prediction intervals, and recorded whether the actual fell within, or outside the associated interval.

For the forecast-evaluation exercise, we split the overall sample of forecasts into three periods: 1759–1850, 1851–1944, 1945–1988 (for 1-step forecasts – for 2-steps ahead, the dates were 1760–1851, etc.). For the whole sample of forecast errors, table 2 shows that the DS and LLT models are more accurate than the TS model, with the LLT having a small advantage at 3 and 4-steps ahead, although the RMSFEs (root MSFEs) suggest considerable uncertainty for all models. For example, an approximate 95% prediction interval for a 1-step ahead forecast using the DS model would be $\pm 2 \times \sqrt{4.82}$, i.e., $\pm 4.4\%$ points of the central projection. We have omitted second-moment results for the DD, DS-L and TS-IC predictors, because these predictors generally fared badly, and in only few cases yielded minor improvements. As is evident from table 1, which records the biases in the forecast errors of IP over the whole period for all the predictors, this is due to higher variability of their forecast errors, because they are manifestly much less biased. For example, the 1-step forecasts of the DS model are on average too low by 1%, those of the LLT are too low by $\frac{1}{2}\%$, while those of DD are virtually unbiased: on average they are too low by only 0.02%. This is consistent with Clements and Hendry (1999a), suggesting that the variance costs offset the bias reductions when weighted by the RMSFE metric: for this reason, our subsequent focus is on the three models of trend, DS, TS and LLT.

Generally, the DS model is the most accurate on RMSFE for the sub-periods as well. The 1945–98 sub-period is associated with less uncertainty, but even here a 95% interval would be $\pm 3.6\%$ points of the central projection, rising to $\pm 5.2\%$ points for a four-year ahead forecast. Table 3 indicates that the DS forecasts at 1-step ahead are often significantly more accurate than those of the TS model on MSFE. This is the case for the whole period and the earliest sub-period, while at conventional significance levels, the LLT is statistically no worse than the DS model at 1-step. For the whole sample period, neither the standard nor modified versions of the forecast-encompassing test reject the null that the DS 1-step forecasts encompass those of the TS forecasts (in line with the findings of the Diebold–Mariano test of equal accuracy), nor do they reject the null that DS encompasses LLT, nor that LLT encompasses TS, but they clearly reject when the tests are reversed, i.e., TS does not encompass DS, LLT does not encompass DS, and TS does not encompass LLT. For the first sub-period, the LLT forecasts are generally preferred on the encompassing criteria, while for 1851–1944, given that we have the DS and TS forecasts, nothing is gained from having the LLT model forecasts as well, though the converse is not true.

The interval-forecast tests in table 4 indicate that, with the exception of the third sub-period (1945–88), the unconditional coverage of the 1-step prediction intervals is appropriate at the three nominal levels we consider. In the third period there are far too few ‘misses’ (actuals outside the intervals) and

this is signalled by p -values below 5%. Table 5 records the dates of the misses, and confirms their general absence from the post-War period. The whole-period test of the independence of the sequences of hits and misses is more demanding, and rejects for all three models for 50 and 75% intervals (and has a p -value of 0.052 for the TS model even for the 95% interval).³ Table 5 confirms both the clustering of misses for a particular model, and the high correlation of misses between models. Consequently, for the whole period, all three models are rejected at the 10% level on the joint test of correct conditional coverage of 50 and 75% intervals. There is no evidence against the models on the first sub-period, although the second (1851–1944) resembles the whole period. Here again, though, there is less evidence against the LLT model 50% interval, indicating that the LLT model may be able to more accurately characterize, or account for, the uncertainty surrounding the point forecasts at different points in time.

Table 5 also notes any dates where wars (and the oil crises) may have played a role. There is a link between forecast failure and manifest turbulence, but it is not strong: many dates coincide with wars, but there were also many other wars over that period which do not seem to have had marked effects (note that the data were interpolated during the second World War). Other events (financial crises, gold discoveries etc.) have not been investigated, but we note that there are 4 dates between 1929 and 1939.

5.2 Gross domestic product (GDP)

For the logarithm of the historical GDP series (y_t) (1830–1991), we again generated sequences of 1 to 4-steps ahead forecasts. For the DS model, we specified an ARIMA(1, 1, 0), allowing a non-zero mean, and for the TS model, we specified an ARIMA(2, 0, 0) for $u_t = y_t - \phi - \gamma t$. The LLT was as described in section 3, but with $\sigma_\zeta^2 = 0$, because the optimization routines failed to solve when this parameter was allowed to be non-zero. Thus the model is a restricted ARIMA(0, 1, 2).

The models were estimated on the period 1830 to 1858, and forecasts were then made for the next four years. The same specifications were re-estimated on data up to 1859, and 1 to 4-step ahead forecasts were again made. This process was continued up to an estimation sample ending in 1987, with 1 to 4-step ahead forecasts of the observations 1988 to 1991. Because there is only a sample of 100 forecasts, it did not seem prudent to analyze subsets of forecasts.

The RMSEs for GDP over the period 1889–1988 are comparable in magnitude to those recorded for industrial production. The DS model again appears to be the most accurate, followed by the ‘LLT’ (recall $\sigma_\zeta^2 = 0$). But the differences are not statistically significant at 1-step ahead on the Diebold–Mariano test applied to the 1-step forecast errors. The forecast-encompassing tests are able to discriminate between the models. On pairwise comparisons, DS encompasses, but is not encompassed by, TS, and similarly, DS encompasses, but is not encompassed by, LLT. Comparing the two dominated models, TS does not encompass LLT, but at the 5% level we cannot reject that LLT encompasses the TS (although we can at the 10% level). Turning now to the prediction intervals, at the 5% level, there is evidence against the independence of the 75% interval hits and misses for TS, and against the 90% LLT interval, as well as against the 90% DS interval on the joint test.

Table 6 gives the dates of the misses for the 90% intervals. Many of the dates coincide with well-known events, though, as before, there are important events that do not coincide with any forecast failures.

³The power of the test falls in the nominal coverage of the interval at high coverage levels. Intuitively, there are fewer misses from which to deduce whether they are independently distributed amongst the hits.

6 Conclusion

In terms of the first issue – which model appears to offer the most useful characterizations of UK output measures historically – the findings support models with stochastic trends. However, the added flexibility of the LLT model, which allows the slope and level of the trend to change, appears to offer little benefit over the DS model’s unit root and fixed underlying growth rate, except perhaps for forecasts of 3 or 4 years ahead. Nevertheless, the average sizes of the forecast errors, and consequently the wide span of prediction intervals, attest to a great deal of uncertainty in the economic environment.

In relation to the second issue, for IP, the post-War period has been relatively more quiescent, with lower MSFEs, and few (if any) misses using prediction intervals which (because of the way they have been calculated) largely reflect the earlier periods. For GDP, forecast failures occur somewhat less often post War (about once per decade, but bunched in the 1970s) than otherwise, but the 1920s stand out as the hard-to-forecast decade.

Finally, to resolve the third issue, perusal of tables 5 and 6 provides some support for our claim that forecast failure is associated with turbulent periods. For GDP, for example, note the asterisks in the three years following WW1, around the time of WW2, the OPEC oil price hikes, combined in 1980 with the onset of recession. The match seems less close for industrial production, in that several major wars do not appear in the list, albeit that many others do, but as the data were missing during the second World War and were linearly interpolated, they cannot reflect any ‘boom-bust’ that may actually have occurred.

References

- Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day. First published, 1970.
- Central Statistical Office, U. K. (1993). *Economic Trends Annual Supplement*. London: H.M.S.O.
- Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, **39**, 841–862.
- Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, **11**, 475–494.
- Clements, M. P., and Hendry, D. F. (1998a). Forecasting economic processes. *International Journal of Forecasting*, **14**, 111–131.
- Clements, M. P., and Hendry, D. F. (1998b). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Clements, M. P., and Hendry, D. F. (1999a). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
- Clements, M. P., and Hendry, D. F. (1999b). Modelling methodology and forecast failure. Mimeo, Department of Economics, University of Warwick.
- Clements, M. P., and Hendry, D. F. (2000). Forecasting with difference-stationary and trend-stationary models. *Econometrics Journal*. Forthcoming.
- Clements, M. P., and Taylor, N. (2000). Bootstrapping prediction intervals for autoregressive models. mimeo, Department of Economics, University of Warwick.

- Crafts, N. F. R., and Harley, C. K. (1992). Output growth and the British Industrial Revolution: A restatement of the Crafts–Harley view. *Economic History Review*, **45**, 703–730.
- Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.
- Eitrheim, Ø., Husebø, T. A., and Nymoen, R. (1999). Equilibrium-correction versus differencing in macroeconometric forecasting. *Economic Modelling*, **16**, 515–544.
- Ericsson, N. R. (1992). Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. *Journal of Policy Modeling*, **14**, 465–495.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, **13**, 281–291.
- Harvey, D., Leybourne, S., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, **16**, 254–259.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (2000a). Does money determine UK inflation over the long run?. In Backhouse, R., and Salanti, A. (eds.), *Macroeconomics and the Real World*. Oxford: Oxford University Press. Forthcoming.
- Hendry, D. F. (2000b). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65.
- Kiviet, J. F. (1986). On the rigor of some mis-specification tests for modelling dynamic relationships. *Review of Economic Studies*, **53**, 241–261.
- Mitchell, B. R. (1988). *British Historical Statistics*. Cambridge: Cambridge University Press.
- Phillips, P. C. B. (1994). Bayes models and forecasts of Australian macroeconomic time series. In Hargreaves, C. (ed.), *Non-stationary Time-Series Analyses and Cointegration*. Oxford: Oxford University Press.
- Proietti, T. (2001). Forecasting with structural time series models. Mimeo, Dipartimento di Scienze Statistiche, Università di Udine. Forthcoming in M. P. Clements and D. F. Hendry (eds), *A Companion to Economic Forecasting*, (2001), Blackwells.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Swanson, N. R., and White, H. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, **13**, 439–462.
- Wallis, K. F. (1995). Large-scale macroeconometric modelling. In Pesaran, M. H., and Wickens, M. R. (eds.), *Handbook of Applied Econometrics: Macroeconomics*. Basil Blackwell.
- West, K. D., and McCracken, M. W. (1998). Regression-based tests of predictive ability. *International Economic Review*, **39**, 817–840.

7 Appendix: Derivation of MSFE for the LLT

Recall that the model in SSF is given by:

$$\begin{aligned} y_t &= \mathbf{Z}\alpha_t + \mathbf{G}\varepsilon_t \\ \alpha_{t+1} &= \mathbf{T}\alpha_t + \mathbf{H}\varepsilon_t \end{aligned}$$

where $\alpha_t = [\mu_t \ \beta_t]'$, $\mathbf{Z} = [1 \ 0]$, $\mathbf{G} = [\sigma_\varepsilon \ 0 \ 0]$, $\varepsilon = [\varepsilon_t/\sigma_\varepsilon \ \eta_t/\sigma_\eta \ \zeta_t/\sigma_\zeta]$ and:

$$\mathbf{T} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 0 & \sigma_\eta & 0 \\ 0 & 0 & \sigma_\zeta \end{bmatrix}.$$

Below, the matrix $\mathbf{P} = \{p_{ij}\}$ is the updated smoothed MSE matrix of the state vector. The h -step ahead forecasts are given by:

$$\mathbb{E}[y_{T+h} \mid I_T] = \mathbf{Z}\mathbf{T}^h \tilde{\alpha}_{T|T}.$$

For the LLT model:

$$\tilde{y}_{T+h|T} = \mathbb{E}[y_{T+h} \mid I_T] = \tilde{\mu}_{T|T} + h\tilde{\beta}_{T|T}.$$

The h -step forecast error is:

$$\begin{aligned} e_{T+h|T} &= y_{T+h} - \tilde{y}_{T+h|T} \\ &= \mathbf{Z} \left(\alpha_{T+h} - \mathbf{T}^h \tilde{\alpha}_{T|T} \right) + \mathbf{G}\varepsilon_{T+h} \\ &= \mathbf{Z}\mathbf{T}^h (\alpha_T - \tilde{\alpha}_{T|T}) + \mathbf{Z} \sum_{i=0}^{h-1} \mathbf{T}^i \mathbf{H} \varepsilon_{T+h-1-i} + \mathbf{G}\varepsilon_{T+h} \end{aligned} \quad (14)$$

because:

$$\alpha_{T+h} = \mathbf{T}^h \alpha_T + \sum_{i=0}^{h-1} \mathbf{T}^i \mathbf{H} \varepsilon_{T+h-1-i}.$$

The three terms in (14) are uncorrelated, so:

$$\mathbb{E}[e_{T+h|T}^2] = \mathbf{Z}\mathbf{T}^h \mathbf{P} \mathbf{T}^{h'} \mathbf{Z}' + \mathbf{Z} \sum_{i=0}^{h-1} \mathbf{T}^i \mathbf{H} \mathbf{H}' \mathbf{T}^{i'} \mathbf{Z}' + \mathbf{G}\mathbf{G}'$$

which for the LLT specializes to:

$$\text{MSFE} = p_{11} + 2hp_{12} + h^2 p_{22} + h\sigma_\eta^2 + h(h-1)(2h-1)\sigma_\zeta^2/6 + \sigma_\varepsilon^2.$$

Table 1 Biases of multi-step forecast errors for Industrial production, 1759–1988 .

h	DS	TS	LLT	DD	TS-IC	DS-L
1	0.93	1.57	0.46	0.03	0.21	0.05
2	1.60	2.64	0.71	0.05	0.40	0.10
3	2.34	3.75	0.96	0.07	0.61	0.15
4	3.03	4.76	1.18	0.06	0.80	0.17

All figures have been multiplied by a hundred.

Table 2 RMSE and MAE of multi-step forecast errors.

	RMSE			MAE		
h	DS	TS	LLT	DS	TS	LLT
Industrial production forecast errors: 1759–1988						
1	4.82	5.08	4.95	3.82	4.07	3.97
2	6.56	7.07	6.55	5.20	5.70	5.18
3	7.72	8.60	7.63	6.18	7.12	6.01
4	8.60	9.84	8.49	6.89	8.20	6.70
Industrial production forecast errors: 1759–1850						
1	4.52	5.15	4.41	3.64	4.25	3.55
2	5.67	6.98	5.43	4.44	5.63	4.27
3	6.44	8.68	6.16	5.16	7.40	4.87
4	7.66	10.60	7.28	6.01	9.10	5.66
Industrial production forecast errors: 1851–1944						
1	5.56	5.58	5.87	4.46	4.47	4.87
2	7.75	7.80	7.91	6.39	6.42	6.54
3	9.32	9.39	9.34	7.73	7.78	7.72
4	10.08	10.19	10.14	8.24	8.32	8.22
Industrial production forecast errors: 1945–1988						
1	3.58	3.60	3.68	2.82	2.84	2.93
2	5.41	5.46	5.42	4.25	4.32	4.21
3	6.27	6.37	6.23	4.98	5.15	4.76
4	6.86	7.01	6.85	5.86	6.05	5.60
GDP forecast errors: 1889–1988						
1	3.25	3.49	3.39	2.30	2.43	2.41
2	5.50	6.08	5.58	3.94	4.22	4.00
3	7.31	8.05	7.30	5.25	5.73	5.27
4	8.58	9.41	8.48	6.19	6.70	6.21

All MSFE and MAE figures have been multiplied by a hundred.

Table 3 Point forecast evaluation: tests of equal forecast accuracy and of forecast encompassing.

	DS,TS		DS,LLT		TS,LLT	
Industrial production forecast errors: 1759–1988						
DM	–		.147		.859	
FE(1)	0.166	–	0.249	–	–	0.114
FE(2)	0.094	–	0.325	0.002	–	0.143
Industrial production forecast errors: 1759–1850						
DM	–		.740		1.0	
FE(1)	0.180	–	0.023	0.438	–	0.072
FE(2)	0.127	–	0.050	0.505	–	0.081
Industrial production forecast errors: 1851–1944						
DM	.092		.058		.076	
FE(1)	0.305	0.192	0.962	0.002	0.820	0.002
FE(2)	0.237	0.133	0.963	0.002	0.823	0.003
Industrial production forecast errors: 1945–1988						
DM	.182		.169		.258	
FE(1)	0.450	0.303	0.679	0.112	0.979	0.172
FE(2)	0.451	0.306	0.690	0.125	0.979	0.185
GDP forecast errors: 1889–1988						
DM	.147		.122		.735	
FE(1)	0.397	–	0.618	0.005	0.002	0.079
FE(2)	0.528	0.004	0.748	0.064	0.003	0.082

Rows headed DM report p -values of Diebold–Mariano tests of equal forecast accuracy on MSFE of x, y . A value of 0.04, say, indicates that there is only a 4% chance that y is no less accurate than x . A value of 0.80, say, indicates that there is a 20% chance that x is no less accurate than y .

FE(1) is the standard forecast encompassing test, where for x, y , the first entry is the p -value of the null that x forecast encompasses y , and the second entry is the p -value of the hypothesis that y forecast encompasses x . FE(2) is the modified form of the test given in Harvey *et al.* (1998). ‘–’ denotes zero to three decimal places.

Table 4 Interval forecast evaluation.

	DS			TS			LLT		
	Uncond.	Ind.	Cond.	Uncond.	Ind.	Cond.	Uncond.	Ind.	Cond.
Industrial production: 1759–1988									
50%	0.235	–	–	0.692	–	0.001	0.147	0.071	0.068
75%	0.591	–	0.001	0.407	0.002	0.005	0.939	–	–
90%	0.105	0.412	0.193	0.825	0.052	0.149	0.827	0.112	0.275
Industrial production: 1759–1850									
50%	0.210	0.304	0.269	0.300	0.546	0.484	0.060	0.781	0.163
75%	0.325	0.214	0.284	0.239	0.246	0.255	1.000	0.086	0.229
90%	0.427	0.280	0.407	0.350	0.710	0.603	0.238	0.375	0.336
Industrial production: 1851–1944									
50%	0.302	0.003	0.008	0.215	0.001	0.002	0.148	0.121	0.105
75%	0.132	0.003	0.004	0.132	0.013	0.014	0.132	0.003	0.004
90%	0.890	0.235	0.489	0.838	0.071	0.192	0.037	0.383	0.078
Industrial production: 1945–1988									
50%	0.015	0.094	0.013	0.015	0.094	0.013	0.006	0.445	0.017
75%	0.023	0.563	0.065	0.023	0.563	0.065	0.007	0.326	0.016
90%	0.002	NA	NA	0.002	NA	NA	0.181	0.659	0.371
GDP: 1889–1988									
50%	0.071	0.857	0.193	0.317	0.557	0.510	0.161	0.191	0.159
75%	0.817	0.190	0.413	0.494	0.033	0.081	0.482	0.440	0.580
90%	0.118	0.051	0.044	0.206	0.422	0.325	0.063	0.021	0.012

The elements are p -values of the nulls of the Christoffersen tests of correct unconditional coverage, independence, and correct conditional coverage of 50, 75 and 90% nominal intervals. ‘–’ indicates zero to three decimal places. ‘NA’ indicates that there were no ‘misses’, so the LR_{ind} (and therefore LR_{cc}) tests are inapplicable.

Table 5 Dates of forecast failure.

Industrial production: 1759–1988				
	DS	TS	LLT	War
1774	*	-	*	w?
1792	-	*	-	w
1799	*	*	-	w
1810	-	*	-	w
1811	-	*	-	w
1815	-	*	-	w
1817	-	*	*	-
1825	*	*	*	-
1827	-	*	*	w?
1828	*	*	*	-
1836	*	*	-	-
1839	*	*	-	w?
1842	-	-	*	-
1844	*	*	-	-
1867	-	-	*	-
1869	-	-	*	-
1870	*	*	-	w?
1871	-	*	-	-
1879	-	-	*	w?
1885	-	-	*	w?
1886	-	-	*	-
1906	*	*	*	-
1916	-	-	*	w
1917	*	*	*	w
1918	-	-	*	w
1920	*	*	*	a
1921	*	*	*	a
1922	*	*	-	a
1924	*	*	*	-
1929	*	*	*	-
1931	*	*	*	-
1934	-	-	*	-
1936	-	-	*	-
1975	-	-	*	o
1980	-	-	*	o
No.	16	22	24	
%	0.70	0.10	0.10	

The dates record when 1-step forecasts lay outside their 90% prediction intervals. *w* denotes a war involving the UK directly, *w?* when there was a war ongoing, *a* denotes the aftermath of World War I, and *o* the oil crises.

Table 6 Dates of forecast failure.

GDP: 1889–1988			
	DS	TS	LLT
1892	*	*	*
1894	*	-	*
1898	-	*	-
1903	-	*	-
1908	*	*	*
1915	*	-	*
1919	*	*	*
1920	*	*	*
1921	*	*	*
1922	*	-	-
1923	*	-	-
1926	*	*	*
1927	*	*	*
1931	*	*	*
1940	*	*	*
1941	-	-	*
1944	*	*	*
1945	-	-	*
1946	-	-	*
1973	*	*	*
1974	*	-	-
1975	-	-	*
1980	-	-	*
No.	15	14	16
%	0.15	0.14	0.16

The dates record when 1-step forecasts lay outside their 90% prediction intervals.