Report of a scoping study of forecasting in the national accounts at the Office for National Statistics

Michael P. Clements and David F. Hendry

March 5, 2003

Contents

1 Introduction and remit

This report to the Statistics Commission presents the results of a scoping study on the use of forecasting in the production of the national accounts by the ONS. The study assesses whether, *prima facie*, a useful purpose would be served by a substantive study into the use of forecasting at the ONS.

The specification for the scoping study is set out in a Statistics Commission document, and summarised below under **1.1 Remit**. The report is based on the collection of papers in '*Documentation for the review of forecasting in the UK national accounts*' made available by Amanda Tuke of the ONS for the purposes of the scoping study, and on discussions and interviews with member of ONS on the 10th of January, as recorded in '*Notes of meeting held at Office for National Statistics on January 10th 2003*', also drawn up by Amanda Tuke. The helpful assistance and full cooperation of Amanda Tuke, the National Accounts Group (NAG), and other ONS staff is gratefully acknowledged.

The background for the study, as described by the Statistics Commission Specification of the scoping study, is that the increasing pressure to produce early estimates of GDP has made resort to forecasting techniques and models necessary when components of series are unavailable by the date required. The Statistics Commission's concern is that the production of statistics is not traditionally associated with forecasting, and that details of the forecasts are not published.

1.1 Remit

Our remit is to provide 'a general but authoritative judgement of the issues of interest to the Commission but not a detailed assessment', addressing:

(a) the adequacy of ONS monitoring of its forecasts & forecasting methods;
(b) the strength of interaction with forecasters external to ONS as measured by:
  (1) the extent to which ONS keeps abreast of methods used by other forecasters,
  (2) whether, or in what circumstances, ONS forecasts should be made public;
(c) whether there are seasonal adjustments implicit in the forecasting techniques used, whether these are different from the methods used for adjusting published data, and if so, is the difference in methodology of consequence?
(d) the quality of the available documentation;
(e) whether a knowledge of the statistical properties of forecasting methods would help the ONS and the public quantify the reliability of earlier ONS estimates.

We consider these in turn. First, however, note that we approached the study aware that economic forecasts can be unreliable, and the later use in forecasting of forecast-based preliminary data could lead to contradictions. Also, we do not distinguish explicitly between forecasting a future outcome and a current unknown number, although the latter is often called 'nowcasting'.

2 Assessment of the present situation

(a) While a number of ONS studies bearing on the accuracy of initial estimates of key economic indicators, and more general assessments of accuracy, have been carried out (see e.g., Barkelm, 2000, Richardson, 2002 and Akriditis, 2002, in the '*Documentation for the review of forecasting in the UK national accounts*') there is a case for formal routine monitoring of the forecasts made at a disaggregated level, beyond what currently appears to take place on an *ad hoc* basis. The fact that it is not always the same components on which actual data are unavailable, and therefore for which forecasts are required, could complicate the administration of routine monitoring schemes. Nevertheless, the establishment of such systems would provide useful information on the performance of the forecasts for components which do routinely need to be forecast, by creating a record of tracking performance: in particular, for periods when actual values are available, forecasts could still be produced for immediate evaluation. One might usefully generate forecasts for each period from a number of models or methods. This would enable more effective monitoring, it would allow rapid action to be taken in the event of a deterioration in the performance of any one model or method, and would provide guidance for future choices of models/methods. ONS point to the practical difficulties of setting up such schemes given existing IT systems. We are advised that the National Accounts re-engineering project (NA REP) should provide time for more routine reviewing of the quality of forecasts. The re-engineering of the processes and systems might also usefully allow for the production and monitoring of forecasts from multiple models in some cases. The reasons why ONS might think in terms of using a number of alternative models to forecast a particular number are explained in section 4.

(b) We take this to include whether the level of training and the continuing professional development (CPD) of ONS staff engaged in short-term forecasting is appropriate. The Time Series Analysis (TSA) branch in assisting compilation branches with forecasting problems provides one level of support and training on a case-by-case basis. In addition, more formal internal training of NAG staff includes a one-day course on the time-series functions used in the internal computer package, WinCSDB. This is a fairly recent initiative and fewer than a quarter of staff have attended to date. One suspects that in a short course of this nature, staff learn more about the mechanics of how to implement particular forecasting techniques in the specific software system, rather than developing an understanding of statistical properties of the various techniques.

There might be a role for continuing professional development that seeks to deepen understanding of the techniques, and encourages more exposure to developments in forecasting techniques and methods, over and above that provided by the TSA branch in a problem-solving capacity.

(1) On the extent to which ONS keeps abreast of methods used by other forecasters, we note that there are close links with some other central statistical agencies. Moreover, ONS staff regularly participate in conferences concerned with seasonal adjustment. We also note that the forecasting journal, the *International Journal of Forecasting*, is taken by the ONS, but would suggest encouraging more active participation of staff in the activities of organisations such as the International Institute of Forecasters.[1] Overall, CPD could usefully play a more prominent role in ensuring high quality trained staff, perhaps by commissioning short relevant courses by world experts.

(2) The issue of the circumstances under which ONS forecasts should be made public has two sides. On the one hand, users ought to be in a position to evaluate the adequacy of the preliminary estimates for the purposes to which they wish apply them. Clearly, preliminary (or flash) estimates are bound to be subject to revision as more information accrues, some of which will involve replacing initially-forecast infilling by more factually-based numbers. Users may well find it helpful to know what percentage of the series in the announced figure were forecast-based. On the other hand, publishing the forecasts used for many small dis-aggregated series (internally referred to as ‘plugging the gap of missing data’), often infilled automatically simply to ensure that data can be compiled, seems unlikely to be of more value than the costs of doing so.

(c) Seasonal adjustment of current (and recent past) data itself implicitly entails forecasting, as most SA procedures are long 2-sided moving-average filters, which have to be ‘folded’ at the end points of

[1] The International Institute of Forecasters publishes the *International Journal of Forecasting*, and has as one of its aims making forecasting useful and relevant for decision and policy makers who need forecasts. DFH is an Honorary Fellow of the IIF, and MPC an editor of IJF.

the sample, corresponding to extrapolating future data. Such 'forecasts' are often made by techniques that are different from the methods used for interpolating missing data in the disaggregated preliminary series. Conversely, if seasonally-adjusted disaggregated series underlie the published data, then the missing series must have been forecast in SA form. It is hard to ascertain whether such differences in methodology are of consequence relative to the other difficulties confronting the construction of preliminary data. Certainly, inconsistencies must arise when the forecasts implicit in the SA are not based on the same method as that used to fill in the missing data. Moreover, different outcomes will result if the data are first seasonally adjusted then forecast, or forecast then seasonally adjusted. Nevertheless, such errors are probably small relative to the mistakes in the initial data and any forecast numbers used to fill gaps.

(d) We are advised that the Standards and Guidance (STaG) project will provide a framework for statistical standards to all staff, that will also apply to NA documentation, resulting in an improvement in the documentation of operating procedures, including the forecasting techniques and models used. This will undoubtedly improve the internal functioning of forecasting at the ONS. Standardisation of documentation across compilers of NA will make the system more transparent. ONS staff should be able to more easily and quickly acquire subject-specific knowledge when moving within divisions, and better documentation would seem to be an integral part of an effective forecast-accuracy monitoring scheme.

There is evidence that subject-specific knowledge, or contextual information, can be useful in forecasting, especially when there is a reasonable number of observations available and the patterns in the data are reasonably stable over time. In these circumstances, methods such as autoregressive-integrated moving average (ARIMA) models can outperform exponentially-weighted moving average (EWMA: see e.g., Fildes and Ord, 2002, p.343). Arrangements that encourage the cumulation of subject-specific knowledge, or improve the transfer of such knowledge, would appear to be desirable. We suspect that the ONS gains from the use of 'series owner estimates', where available, for precisely this reason. We note from Reed (2002) (in the '*Documentation for the review of forecasting in the UK national accounts*') that about 40% of the UK preliminary estimates of GDP are attributed to 'series owner estimates', and 14% to the ONS 'internal estimates', where the high proportion coming from the former should allow expert knowledge to be exploited.

(e) We take the two components in turn. A knowledge of the statistical properties of forecasting methods is bound to help the ONS quantify the reliability of all estimates that implicitly or explicitly involve forecasts. Unfortunately, such properties are dependent on the behaviour of the series being forecast: for example, a method that might be near optimal and well behaved for a stationary time series could be highly unreliable in the face of a non-stationary shift. Nevertheless, knowing which methods are robust, and why, and which are not, seems invaluable to any agency making forecasts. For example, we consider that averaging across a range of forecasts can generally improve reliability although that is not current practice, and has obvious resource costs when thousands of series must be handled (less so when appropriate software is available). Much relevant knowledge is available within the ONS, and explicitly underpins their present choice of automatic forecasting technique, but does not exhaust the profession's expertise.

Turning to the public's knowledge base, then generically, the answer is an obvious yes, but it is less obvious how a better understanding could be achieved within current institutional settings. A recent attempt at explaining economic forecasting is provided in Hendry and Ericsson (2001).

A brief general discussion of economic forecasting and what we perceive to be the ONS approach to forecasting is offered to illuminate some of these issues, and provide some background material. Section 3 describes the broad approach to forecasting at the ONS, and section 4 what we perceive to be the nature of the forecasting problem in general terms. Section 4.1 further expands on the use of multiple predictors, and section 4.2 discusses the use made of proxies by the ONS in forecasting. Section 5 summarises our recommendations.

3 Forecasting at the ONS

The ONS views forecasting as being a last resort for 'plugging gaps' when data are unavailable. Forecasting is based on either a version of Holt–Winters or an ARIMA model. The ONS guidelines as to which to use emphasise whether the series is sufficiently important/aggregated to merit going through an ARIMA model fitting exercise, and whether time and expertise are available to do so. The implementation of the chosen method is reviewed on an annual basis. There is little interest in using multivariate approaches; or explanatory variables other than as simple proxies (see section 4.2); or approaches that make use of mixed frequency data (e.g., Montgomery, Zarnowitz, Tsay and Tiao, 1998); or exploiting possible gains that might result from forecasting the aggregate as well as the components (see section 6). There is a clear preference for finding a timely data source that provides a 'measure' of the component that is required, rather than expending resources in an attempt to obtain more accurate forecasts. The pre-eminence given to actual data sources is commendable, and that route should always be fully explored. However, since there will usually be a role for forecasting in the context of preliminary estimates, care is warranted in ensuring that the general unease with the use of forecasting does not lead to the adoption of sub-optimal methods when superior methods exist. We argue in section 4 that the nature of the problem is such that it is largely an empirical question whether the more sophisticated methods noted in this section might yield improvements.

Forecasting takes place in many guises, not always explicitly acknowledged as noted above. However, we view the issue as one of obtaining the 'best estimates of missing data', rather than forecasting *per se*, namely, producing the most reliable preliminary data that is feasible given the available information, the time constraints and the resources that can be deployed. As such, there appears to be a conceptual conflict, where 'forecasting' or even 'nowcasting' are viewed at NAG as necessary evils, rather than an integral part of estimating the final value of a quantity (such as GDP) via a preliminary measure that is inherently subject to revision. In essence the problem is a 'signal extraction' issue for the missing data points, rather than a forecasting exercise, where alternative forecasts deliver signals of varying reliability. If the numbers first announced could be systematically improved by an outside agency (perhaps combining them with other information) then the quality of ONS data is bound to be questioned.

4 The nature of economic forecasting

Given a loss function, such as quadratic loss, for example, the aim of forecasting is to find a forecast $\widehat{y}$ of y which solves:

$$\underset{\widehat{y}}{\operatorname{argmin}} \mathsf{E}\left[y-\widehat{y}\right]^2$$

where $\widehat{y} = g(\widetilde{\mathcal{I}})$, that is, some function $g(\cdot)$ of the available information set, $\widetilde{\mathcal{I}}$. For the preliminary estimate of GDP, for example, the information set is that available up to 24 days after the end of the quarter for which a 'nowcast' is required (Reed, 2000).

There are many different loss functions. If the statistical process generating y were known, the form of the optimal predictor $\widehat{y}$ could be solved for either analytically or by simulation (e.g., Christoffersen and Diebold, 1996, 1997). Often, the loss function may include intangible factors such as 'telling a story about the forecast' which may suggest a causal model, and downplay accuracy. More importantly, the statistical model for y is not known, so neither is the form of g. Moreover, the information set $\mathcal{I}$ is to some extent a choice variable. At issue then, is the type of information to be included in $\widetilde{\mathcal{I}}$, e.g., the history of the series alone, as for a univariate time-series model, survey information, other related series etc.; and the form of the model relating $\widetilde{\mathcal{I}}$ to the object of interest, $g(\cdot)$. In general, the best modelling and forecasting strategy depends upon the properties of the series. This is unknown, so the best strategy in any specific instance is unknown. The properties of a number of commonly-used forecasting models and forecasting strategies have been studied for processes thought to be relevant for economic time series in e.g., Clements and Hendry (1999), and the empirical performance of a wide range of models has been gauged by the M-Competitions reviewed in Fildes and Ord (2002). The forecasting framework in Clements and Hendry (1999) is shown to account for the main findings of the M-Competitions by Clements and Hendry (2001), and leads to better-based explanations for many of the empirical practices of forecasters, such as pooling across a range of methods and models (see e.g., Hendry and Clements, 2002).

The empirical forecasting competitions generally suggest that simple adaptive devices like damped trend and EWMA, such as that embodied in Holt–Winters, work reasonably well, compared to more 'sophisticated' methods such as ARIMA modelling. The extensive use made by the ONS of variants of Holt–Winters would, therefore, appear to be warranted. Some concerns arise over Holt–Winters when the series exhibit trends and seasonals, and whether in such circumstances the 'structural time series' models associated with Andrew Harvey (see e.g., Harvey, 1993, and Koopman, Harvey, Doornik and Shephard, 1999) might generate more reasonable short-term forecasts. Such models allow some flexibility of the specification of the seasonal component. Monitoring the performance of such models alongside the current models with a view to assigning some weight to their predictions in the future might be worth considering.

One of the main messages of Clements and Hendry (1999) is that forecasts are most likely to go seriously awry when there are structural breaks or locations shifts. As we outline in section 6.2, EWMA-type schemes work well when there are 'regular' measurement errors. They can be shown to be optimal for a particular type of process subject to measurement error. As the ONS are aware (see 'NA forecasting review – section 2' in the Documentation provided), Holt–Winters can be problematic when there are outliers, or level-shifts. In some of these circumstances, predicting an unchanged growth rate, as in the WinCSDB function 'Extrap', may be preferable, as we discuss in section 6.2.

The problem of course is the rapid identification of outliers and level shifts in real-time. These effects can perhaps be identified *ex post*, with a time delay, but this will be much more difficult when such events occur close to the present. It is in precisely that case that the distortionary effects on the forecasts would be greatest. This reinforces the case for careful monitoring (see section 4.1 below). Anticipation of imminent unusual events might help avoid large errors, especially if there were historical precedents similar enough in nature to provide useful estimates of the magnitudes of the adjustments required. We formed the impression that ONS staff are skilled in making timely adjustments of this sort.

4.1 Multiple predictors

The importance of multiple predictors is two-fold: revealing potential divergences between methods; and evaluating comparative performance. First, unanticipated shifts in the behaviour of economic time series occur intermittently, and as some methods are less affected by such changes relative to others, divergences between alternative methods show up. Secondly, because the best strategy in any specific instance is unknown, past performance could be used to make the current choice. This requires historical track records of the performance of the methods, and careful monitoring of the recent short-term forecasts and their outcomes to ensure that good past performance does indeed translate into more accurate predictions. A 'forecasting law' noted above is that averaging or pooling across forecasts from a number of models can result in improved accuracy: see, e.g., Newbold and Harvey (2002), Fildes and Ord (2002), and Stock and Watson (1999). This suggests that the performance of such averages should be monitored for any series routinely forecast by the ONS. Although different series are missing on different occasions, suitable software should be able to conduct such analyses relatively automatically.

4.2 Proxies

Here we are concerned with the role of covariate information, namely whether forecasts from models that exploit related series, including survey data, possibly in combination with univariate time-series models, help improve the accuracy with which any missing data are estimated. At present, the percentage changes in proxy variables are often used to provide a direct estimate of the percentage changes in the components for which forecasts are required (after appropriate adjustment for value & volume effects). In general, direct relationship of this sort might need to be tempered by other effects. Formally, this suggests treating the proxy as an explanatory variable in a model for the variable to be forecast. We suspect that ONS staff are skilled at making judgemental adjustments as required, but thinking of the proxy-variable relationship in the context of a model also highlights the need for monitoring and evaluation.

4.3 Surveys

While survey outcomes could also be viewed as covariate information, and so might be used to generate 'forecasts' in which they acted as possible additional 'multiple predictors', they are generally noncausal. Consequently, the reliability of any relationships must always be open to doubt. An alternative role for surveys as an additional source of information is as part of a signal extraction approach to estimating missing data at the forecast origin, namely for ‘plugging the gaps’ by a more sophisticated procedure.

The report by Williamson and Thompson (2002) was examined in detail, and discussed at our ONS meeting, where we were informed that the report had been removed from the NTC web page. That report claimed that the Purchasing Managers Index (PMI) business-survey data could produce more reliable, and more timely, estimates of UK GDP than first estimates produced by the ONS. Their data showed correlations of between 0.3 and 0.78 depending on data frequency and series, and from regression analysis, they claimed more accurate outcomes than the ONS preliminary data. However, it is impossible to judge the import of their claim from the information provided, but it seems subject to the critique of leading indicators in Emerson and Hendry (1996) where the *ex post* performance is often vastly superior to the *ex ante*, reflected in regular revisions of the components of such indicators (also see Diebold and Rudebusch, 1991).

Nevertheless, as we discussed above, monitoring a range of forecasting devices merits attention, and pooling across these can be effective in many situations (Hendry and Clements, 2002), so survey-based forecasts could be one member of that class.

5 Recommendations

Forecasting in the sense of obtaining ‘estimates of missing data’ plays an important role in the compilation of the ‘preliminary’ or ‘first’ estimates of National Accounts. That being the case, one would like to know whether approaches that fully embraced the notion of producing the most reliable preliminary data that is feasible – given available information, time and resource constraints – would yield markedly more accurate estimates than the ‘plugging the gaps’ approach that epitomises current policy.

The size of any possible gains that might be made are largely an empirical matter for the reasons described above. They would become apparent in a real-time forecasting context through the monitoring of multiple predictors for the same numbers, and by adopting schemes that switched between predictors and/or combinations of predictors, on the basis of past performance. Examining past tracking performance – if available at the required level of detail – might allow one to estimate *ex post* the size of potential real-time gains.

We are informed that existing IT systems at ONS would make the real-time application of the multiple predictor–continuous monitoring schema unworkable. The re-designing of the IT systems in the NARP might allow for the incorporation of such features, if these were thought to be desirable. The ONS view is that any further review of forecasting would be especially useful when developments currently in train are nearing completion. We suspect that it might be preferable to decide on how the ONS wishes to conduct its NA forecasting before the software systems are put in place, especially if the systems would rule out, or make prohibitively expensive, certain options.

A small *ex post* pilot study (or studies) focusing on one part of the NA might usefully provide information on the potential gains to adopting a broader, more inclusive approach to forecasting. However, such a study might be a considerable undertaking if suitable software systems for data handling, model estimation, and forecast generation were not in place.

5.1 Responses to the remit

(a) The adequacy of ONS monitoring of its forecasts and forecasting methods *could be enhanced, but that would require greater resource inputs*.

(b) The strength of interaction with forecasters external to ONS as measured by:

  (1) the extent to which ONS keeps abreast of methods used by other forecasters *could be improved by more extensive CPD programs, perhaps by wider use of outside experts*;

  (2) whether, or in what circumstances, ONS forecasts should be made public, *depends on what aspects are under consideration. This is a question that the ONS and Statistics Commission might usefully reflect upon taking into account the views of users of ONS statistics. As a minimum, broad brush estimates of the extent of missing data should be helpful*;

(c) whether there are seasonal adjustments implicit in the forecasting techniques used, whether these are different from the methods used for adjusting published data, and if so, is the difference in methodology of consequence? *Seasonal adjustment procedures are in use, and may entail contradictions with other adjustments, but seem unlikely to be the key problem to improved accuracy of preliminary data*;

(d) the quality of the available documentation *is fair, and is improving*;

(e) whether a knowledge of the statistical properties of forecasting methods would help the ONS and the public quantify the reliability of earlier ONS estimates. *We believe it definitely would, but view this as a long term objective, partly dependent on experts taking the trouble to communicate with the public*.

Overall, we doubt the need for an immediate extensive or in-depth study of the role of forecasting in the preparation of preliminary estimates, but hope that some of the potential improvements suggested here might prove useful. Once the new monitoring systems are in place, however, a return to this problem is merited.

6 Technical Appendix

6.1 Forecasting aggregates and components

Let $\widehat{y}_{i,t|t-1}$ be the forecast of variable i for period t made at time $t-1$, and define:

$$\widehat{\mathbf{Y}}_{t|t-1} = \left[\widehat{y}_{1,t|t-1}\ \widehat{y}_{2,t|t-1} \cdots \widehat{y}_{k,t|t-1}\right]'$$

so $\widehat{\mathbf{Y}} = \left[\widehat{\mathbf{Y}}'_{T+1|T}\ \widehat{\mathbf{Y}}'_{T+2|T+1} \cdots \widehat{\mathbf{Y}}'_{T+P|T+P-1}\right]'$ collects the 1-step forecasts of the k series for $T+1,\ldots,T+p$, and is $kp \times 1$. Then the regression:

$$\begin{aligned} \widehat{\mathbf{Y}} &= \mathbf{Lb} + \mathbf{Y} + \mathbf{e} \\ \widehat{\mathbf{Y}} - \mathbf{Y} &= \mathbf{Lb} + \mathbf{e} \end{aligned}$$

where $\mathbf{Y}$ contains the actual values, and $\mathbf{L} = \mathbf{i}_p \otimes \mathbf{I}_k$, defines $\mathbf{b}$ the $k \times 1$ vector of bias parameters. Here, $\mathbf{i}_p$ is a $p \times 1$ vectors of 1s, and $\mathbf{I}_k$ is a k-dimensional identity matrix, and $\otimes$ is such that $\mathbf{A} \otimes \mathbf{B}$ is $\{a_{ij}\mathbf{B}\}$. The above regression corresponds to averaging the p 1-step errors for each series separately.

The bias-corrected 1-step forecasts are e.g.,

$$\widehat{y}^{\text{bc}}_{i,T+p+1|T+p} = \widehat{y}_{i,T+p+1|T+p} - b_i,$$

where $\widehat{y}_{i,T+p+1|T+p}$ is the original forecast for model i.

Suppose that the k^{th} variable is the sum of the first $k-1$. Then we have forecasts of the components and the aggregate, i.e.,

$$\mathbf{R}\mathbf{y}_t = 0$$

where $\mathbf{R} = [1\ 1\ 1\ 1\ \cdots 1_{k-1}\ \ -1]$ and $\mathbf{y}_t = [y_{1,t} \ldots, y_{k-1,t}, y_{k,,t}]'$. In terms of $\mathbf{Y}$, the kp vector of actual values on the k series for $T+1, \ldots, T+p$, we have:

$$\mathbf{W} = \mathbf{C}\mathbf{Y} = (\mathbf{I}_P \otimes \mathbf{R})\,\mathbf{Y} = \mathbf{0}_{p\times 1}.$$

where $\mathbf{C}$ is $p \times kp$.

With the problem formulated in this way, Guerrero and Peňa (2000, 2001) show that the optimal estimator of the bias, and the minimum mean-squared error predictor of $\mathbf{Y}$ are given by:

$$\widehat{\mathbf{b}} = -\mathbf{\Sigma}_b \mathbf{L}'\mathbf{C}'\mathbf{\Sigma}_d^{-1}\left(\mathbf{W} - \mathbf{C}\widehat{\mathbf{Y}}\right) \tag{1}$$

where $\widehat{\mathbf{Y}}$ is the model-based vector of forecasts, and:

$$\widehat{\mathbf{Y}}_c = \left(\widehat{\mathbf{Y}} - \mathbf{L}\widehat{\mathbf{b}}\right) + \mathbf{A}_z\left[\mathbf{W} - \mathbf{C}\left(\widehat{\mathbf{Y}} - \mathbf{L}\widehat{\mathbf{b}}\right)\right] \tag{2}$$

where:

$$\mathbf{\Sigma}_b = \left(\mathbf{L}'\mathbf{C}'\mathbf{\Sigma}_d^{-1}\mathbf{C}\mathbf{L}\right)^{-1}, \quad \mathbf{\Sigma}_d = \mathbf{C}\mathbf{\Sigma}_e\mathbf{C}',$$

and:

$$\mathbf{A}_z = \mathbf{\Sigma}_e\mathbf{C}'\mathbf{\Sigma}_d^{-1}.$$

$\widehat{\mathbf{Y}}_c$ is the optimal bias-corrected predictor that incorporates the adding-up constraint. $\mathbf{\Sigma}_e$ is $kp \times kp$, and we can assume $\mathbf{\Sigma}_e = \mathbf{I}_p \otimes \mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is the $k \times k$ contemporaneous covariance matrix of the 1-step errors: thus, we also assume this is constant over $T+1$ to $T+p$, and that the autocorrelations are zero.

6.2 Forecast-error adaptation

In any practical forecasting exercise, the forecaster may need to contend with both measurement errors (the variable of interest is measured only imperfectly, e.g., subject to an additive noise component) and breaks or shifts in the underlying generating mechanism of the variable of interest. Given the practical relevance of these two phenomena, we consider the performance of commonly-used forecasting techniques, such as those used by the ONS. EWMA is one of the most famous 'forecast-error correction' mechanisms (FErCMs), and is closely related to Holt–Winters, so will be considered in detail. Setting the forecast of the growth rate to the previous growth rate (or some average of recent growth rates) will be robust to a shift in the level of the process, once that shift has taken place. A forecasting technique that exploits forecast-error based information to achieve a similar aim – correcting for location shifts – is that of intercept corrections (ICs). These add back recent errors (perhaps after smoothing) to offset location shifts, and share many of the characteristics of extrapolating past growth rates forward. ICs can be efficacious in situations in which the form, size, timing and sign of the shift are unknown (see e.g., Clements and Hendry, 1998).

6.2.1 Exponentially weighted moving average

The EWMA recursive updating formula is, for $\lambda \in (0,1)$ and a scalar time series $\{y_t\}$:

$$\widehat{y}_{T+h|T} = (1-\lambda)\sum_{j=0}^{\infty}\lambda^j y_{T-j}, \tag{3}$$

so (e.g.):

$$\widehat{y}_{T+1|T} = (1-\lambda)\, y_T + \lambda \widehat{y}_{T|T-1} = y_T - \lambda\left(y_T - \widehat{y}_{T|T-1}\right), \tag{4}$$

with start-up value $\widehat{y}_1 = y_1$. Hence, for an origin T, $\widehat{y}_{T+h|T} = \widehat{y}_{T+1|T}$ for all $h > 1$. One can view this method as 'correcting' a random-walk forecast by the latest forecast error $(y_T - \widehat{y}_{T|T-1})$:

$$\widehat{\Delta y}_{T+1|T} = -\lambda\left(y_T - \widehat{y}_{T|T-1}\right). \tag{5}$$

The EWMA forecast function can be seen as approximating the ARIMA(0,1,1):

$$\Delta y_t = \varepsilon_t - \theta\varepsilon_{t-1}, \tag{6}$$

and is the exact forecast function for the ARIMA$(0,1,1)$ when $\theta = \lambda$, so λ can always be set such that EWMA is optimal for an ARIMA$(0,1,1)$.

An ARIMA$(0,1,1)$ might appear to be a restrictive class of data generating process, but it happens to result when the underlying series is a random walk measured with additive error, e.g.,

$$\begin{aligned} y_t^* &= y_{t-1}^* + v_t, \quad \mathsf{E}\left[v_t^2\right] = \sigma_v^2 \\ y_t &= y_t^* + w_t, \quad \mathsf{E}\left[w_t^2\right] = \sigma_w^2 \end{aligned}$$

where y_t^* is the underlying series, and y_t is the measured series, and w_t the measurement error. Combining these two equations yields the single equation:

$$\Delta y_t = \Delta y_t^* + \Delta w_t = v_t + w_t - w_{t-1} \tag{7}$$

where we can write $v_t + w_t - w_{t-1}$ as $\varepsilon_t - \theta\varepsilon_{t-1}$, as in (6) with θ depending on σ_v^2 and σ_w^2: θ lies in the interval $[0,1)$ and decreases monotonically as σ_v^2/σ_w^2 increases. Thus, as the signal (σ_v^2) increases relative to the noise (σ_w^2) the optimal forecast function approaches $\widehat{y}_{T+h|T} = y_T$.

From (6), note that $\widehat{\Delta y}_{T+1|T} = -\theta\varepsilon_T$, and that $\widehat{y}_{T|T-1} = y_{T-1} - \lambda\varepsilon_{T-1}$ implies that $\varepsilon_T = y_T - \widehat{y}_{T|T-1}$, thus:

$$\widehat{\Delta y}_{T+1|T} = -\theta\varepsilon_T = -\theta\left(y_T - \widehat{y}_{T|T-1}\right)$$

establishing the equivalence to (5) when $\lambda = \theta$.

We have established that EWMA is an excellent foil for measurement error when the process to be forecast is approximately a random walk. The structural time series model approach of Andrew Harvey allows one to derive optimal forecast functions for processes with measurement error and more elaborate dynamic structures.

Location shifts cannot be accommodated so easily (at least conceptually) – viewed as 'one-off' deterministic shifts they are not amenable to probabilistic modelling in the way in which recurrent, regular features such as measurement errors are (but see e.g., Hamilton, 1989, for an approach that treats location shifts as being stochastic). Consider the EWMA scheme when there is a shift in the mean of $\{y\}$. The shift will eventually feed through to the forecasts from (4): adding back a damped function of recent forecast errors ought, therefore, to be productive when location shifts are common. The speed with which adjustment occurs depends on the degree of damping, λ, where $\lambda = 0$ corresponds to a random walk forecast. The choice of a large λ prevents the predictor extrapolating the 'noise' in the latest observation, but when there is a shift in mean, the closer λ is to zero the more quickly a break will be assimilated in the forecasts. This argument is formalised below, and the relative costs in terms of forecast accuracy are established.

Location shift in the random walk plus noise model Suppose the process (7) is replaced by:

$$\begin{aligned} y_t^* &= \delta 1_{\{t=T\}} + y_{t-1}^* + v_t \\ y_t &= y_t^* + w_t \end{aligned}$$

where we maintain the assumption that $\mathsf{E}\left[v_t^2\right] = \sigma_v^2$ and $\mathsf{E}\left[w_t^2\right] = \sigma_w^2$. Thus, there is an upward shift ($\delta > 0$) in the level of y_t^* (and thus also y_t) at period T (note $1_{\{A\}} = 1$ when A is true and zero otherwise). From (5), $\widehat{y}_{T|T-1} = y_{T-1} - \lambda\varepsilon_{T-1}$, but $y_T = y_{T-1} + \delta + \epsilon_T - \lambda\varepsilon_{T-1}$, so that the actual forecast error is $y_T - \widehat{y}_{T|T-1} = \delta + \varepsilon_T$. Thus the next EWMA prediction is:

$$\widehat{y}_{T+1|T} = y_T - \lambda\left(\delta + \varepsilon_T\right) \tag{8}$$

so that the increase in $\widehat{y}_{T+1|T}$ from y_T incorporating the shift will be partly offset by $-\lambda\delta$. The EWMA is effective at dealing with the measurement error component of $y_T - \widehat{y}_{T|T-1}$, ε_T, but treats the same fraction (λ) of δ as measurement error, whereas the optimal fraction for δ is zero. The EWMA predictor (8) attaches exponentially declining weights to y_T, y_{T-1}, y_{T-2} as in (3), whereas a relatively larger weight on y_T is warranted (y_{T+1} is much more like y_T than earlier pre-level shift values).

The optimal (infeasible) predictor is:

$$\widehat{y}^o_{T+1|T} = y_T - \lambda\varepsilon_T$$

which requires that we are able to separately identify δ and ε_T. Substituting $\varepsilon_T = y_T - \widehat{y}_{T|T-1} - \delta$ gives:

$$\widehat{y}^o_{T+1|T} = \left(1 - \lambda\right) y_T + \lambda\widehat{y}_{T|T-1} + \lambda\delta$$

and proceeding with the EWMA recursion in the usual way, we are able to establish that:

$$\widehat{y}^o_{T+1|T} = \widehat{y}_{T+1|T} + \lambda\delta.$$

λ remains the optimal MA parameter because the signal/noise ratio is unaltered.

Suppose we consider rolling the 1-step EWMA forecasts through the post-shift sample. Because $y_{T+j} = y_{T+j-1} + \epsilon_{T+j} - \lambda\varepsilon_{T+j-1}$, for $j > 0$, and by recursive application of:

$$\widehat{y}_{T+j|T+j-1} = y_{T+j-1} - \lambda \left(y_{T+j-1} - \widehat{y}_{T+j-1|T+j-2} \right)$$

we can show that:

$$\widehat{y}_{T+j|T+j-1} = y_{T+j-1} - \lambda\varepsilon_{T+j-1} - \lambda^j \delta$$

so that in a rolling sequence of 1-step forecasts the impact of δ diminishes.

Differencing (extrapolating the same growth rate into the future if the data are growth rates) can be viewed as an approximation to the infeasible predictor that accounts for location shifts. It works by setting the correction to the random walk predictor to zero: i.e., in order not to implicitly subtract $\lambda\delta$ from y_T, as in (8), we refrain from making an adjustment for $\lambda\varepsilon_T$, as well, by setting $\lambda = 0$, giving the predictor:

$$\widehat{y}^{\Delta}_{T+1|T} = y_T.$$

The associated unconditional mean-squared forecast error (MSFE) is:

$$\mathsf{E}\left[\left(y_{T+1} - \widehat{y}^{\Delta}_{T+1|T} \right)^2 \right] = \left(1 + \lambda^2\right) \sigma^2_{\varepsilon}$$

(the variance of an $\mathrm{MA}(1)$) which is increasing in σ^2_w/σ^2_v, the noise to signal ratio.

By contrast, for the EWMA:

$$\begin{aligned} \mathsf{E}\left[\left(y_{T+1} - \widehat{y}_{T+1|T} \right)^2 \right] &= \mathsf{E}\left[\left(y_T + \varepsilon_{T+1} - \lambda\varepsilon_T \right) - \left(y_T - \lambda\left(\delta + \varepsilon_T\right)\right)^2 \right] \\ &= \lambda^2\delta^2 + \sigma^2_{\varepsilon} \end{aligned}$$

which is also increasing in the noise to signal ratio, as a large noise component translates into a large correction for measurement error and a concomitant (but undesirable) shaving off of the correction to the new level of the process. Differencing will yield a lower MSFE when:

$$1 < \frac{\delta^2}{\sigma^2_{\varepsilon}}$$

so that the choice between EWMA and random walk extrapolation does not depend on λ.

6.2.2 Intercept correction

The role of interception correction can be more easily gauged if we abstract from measurement error, and assume a data generation process of the form:

$$y_t = \mu + \delta 1_{\{t \geq T_1\}} + \rho y_{t-1} + \epsilon_t \;\text{ where }\; \epsilon_t \sim \mathsf{IN}\left[0, \sigma^2_{\epsilon}\right], \tag{9}$$

with $|\rho| < 1$ when $1_{\{t \geq T_1\}}$ is an indicator variable, with the value zero till time $T_1 < T$, after which it is unity. The interesting case is when $T_1 = T - 1$, so the shift has recently occurred. In (9):

$$\mathsf{E}\left[y_{T+1} \mid y_T\right] = \mu + \delta + \rho y_T,$$

for which:

$$\widehat{y}_{T+1} = \widehat{\mu} + \widehat{\rho} y_T \tag{10}$$

(10) is a poor forecast when δ is large, even if the estimated model parameters, $\widehat{\mu}$ and $\widehat{\rho}$, coincide with their population values, μ and ρ, which we now assume for simplicity.

At time T, there was a residual of $\widehat{u}_T = y_T - \mu - \rho y_{T-1}$, where from (9) at time T:

$$\widehat{u}_T = \mu + \delta + \rho y_{T-1} + \epsilon_T - \mu - \rho y_{T-1} = \delta + \epsilon_T. \tag{11}$$

To set the model 'back on track' (i.e., fit the last observation perfectly), the IC $\widehat{u}_T$ is often added to (10) to yield:

$$\widehat{y}_{\iota,T+1} = \mu + \rho y_T + \widehat{u}_T \tag{12}$$

with the forecast error:

$$\begin{aligned} y_{T+1} - \widehat{y}_{\iota,T+1} &= \mu + \delta + \rho y_T + \epsilon_{T+1} - (\mu + \rho y_T + \widehat{u}_T) \\ &= \delta + \epsilon_{T+1} - \widehat{u}_T = \delta + \epsilon_{T+1} - (\delta + \epsilon_T) \\ &= \Delta\epsilon_{T+1} \end{aligned}$$

Thus, the IC forecast changes the forecasting error to the difference thereof, thereby removing the impact of the shift δ in the equilibrium mean. The unconditional mean squared forecast error is:

$$\mathsf{E}\left[(y_{T+1} - \widehat{y}_{\iota,T+1})^2 \right] = 2\sigma_\epsilon^2, \tag{13}$$

as against the minimum obtainable (for known parameters) of σ_ϵ^2. When $\delta > \sigma_\epsilon^2$, ICs such as $\widehat{u}_T$ have excellent properties in this setting.

6.2.3 The relation of EWMA to IC

Four components contribute to the forecasting success of EWMA:

- adapting the next forecast by the previous forecast error ($\lambda \neq 0$);
- differencing to adjust to location shifts (y_{T+1} is set to the last observed level y_T: at least when $\lambda = 0$);
- the absence of deterministic terms which could go awry;
- rapid adaptivity when λ is small.

As is apparent, some of these aspects pull in different directions. The correction of a forecast by a previous forecast error is reminiscent of intercept correction. However, EWMA differs from IC by both the sign and size of the damping factor, $-\lambda$ in place of unity, so may outperform when there are large measurement errors at the forecast origin, but not when there are unanticipated shifts. The sign change is not due to IC being an autoregressive, rather than a moving-average, correction: rather, the aim of the IC is to offset a location shift, whereas EWMA seeks to offset a previous measurement error, using differencing to remove location shifts. Thus, measurement errors are an important *caveat* to the explanations for the empirical success of ICs discussed in Clements and Hendry (1999, Ch.6): some of the potential roles of an FErCM conflict. In particular, to offset previous mis-specifications or measurement errors requires the opposite sign for an FErCM to that needed for offsetting breaks.

References

Christoffersen, P. F., and Diebold, F. X. (1996). Further results on forecasting and model selection under aymmetric loss. *Journal of Applied Econometrics*, **11**, 561–572.

Christoffersen, P. F., and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, **13**, 808–817.

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

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (2001). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.

Diebold, F. X., and Rudebusch, G. D. (1991). Turning point prediction with the composite leading index: An ex ante analysis. In Lahiri, K., and Moore, G. H. (eds.) , *Leading economic indicators: New approaches and forecasting records*, pp. 231–256. Cambridge: Cambridge University Press.

Emerson, R. A., and Hendry, D. F. (1996). An evaluation of forecasting using leading indicators. *Journal of Forecasting*, **15**, 271–91.

Fildes, R., and Ord, K. (2002). Forecasting competitions – their role in improving forecasting practice and research. In Clements, M. P., and Hendry, D. F. (eds.) , *A Companion to Economic Forecasting*, pp. 322–353: Oxford: Blackwells.

Guerrero, V., and Peňa, D. (2000). Linear combination of restrictions and forecasts in time series analysis. *Journal of Forecasting*, **19**, 103–122.

Guerrero, V., and Peňa, D. (2001). Combining multiple time series predictors: a useful inferential procedure. Mimeo, Universidad Carlos III de Madrid.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.

Harvey, A. C. (1993). *Time Series Models* 2nd (first edition 1981) edn. Hemel Hempstead: Harvester Wheatsheaf.

Hendry, D. F., and Clements, M. P. (2002). Pooling of forecasts. Forthcoming, *Econometrics Journal*.

Hendry, D. F., and Ericsson, N. R. (eds.) (2001). *Understanding Economic Forecasts*. Cambridge, Mass.: MIT Press.

Koopman, S. J., Harvey, A. C., Doornik, J. A., and Shephard, N. (1999). *Structural Time Series Analysis, Modelling, and Prediction using STAMP* 2nd edn. London: Timberlake Consultants Press.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association*, **93**, 478–493.

Newbold, P., and Harvey, D. I. (2002). Forecasting combination and encompassing. In Clements, M. P., and Hendry, D. F. (eds.) , *A Companion to Economic Forecasting*, pp. 268–283: Oxford: Blackwells.

Stock, J. H., and Watson, M. W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In Engle, R. F., and White, H. (eds.) , *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger*, pp. 1–44. Oxford: Oxford University Press.

Williamson, C., and Thompson, L. (2002). Early, accurate, monthly GDP estimates: a PMI-based model of gross domestic product. Economic research monograph no. 3, NTC Research Limited, Henley-on-Thames.