Department of Economics

Forecasting with Measurement Errors in Dynamic Models

Richard Harrison, George Kapetanios and Tony Yates

Working Paper No. 521 October 2004 ISSN 1473-0278

Queen Mary
University of London

Forecasting with measurement errors in dynamic models

Richard Harrison* and George Kapetanios† and Tony Yates‡

September 2004

Abstract

In this paper we explore the consequences for forecasting of the following two facts: first, that over time statistical agencies revise and improve published data, so that observations on more recent events are those that are least well measured. Second, that economies are such that observations on the most recent events contain the the largest signal about the future. We discuss a variety of forecasting problems in this environment, and present an application using a univariate model of the quarterly growth of UK private consumption expenditure.

Keywords: Forecasting, Data Revisions, Dynamic Models
JEL codes: C32, C53

*Bank of England.
E-mail: Richard.Harrison@Bankofengland.co.uk

†Queen Mary, University of London and Bank of England.
E-mail: G.Kapetanios@qmul.ac.uk

‡Bank of England.
E-mail: Tony.Yates@Bankofengland.co.uk. This paper represents the views and analysis of the authors and should not be thought to represent those of the Bank of England or Monetary Policy Committee members.

1 Introduction

This paper explores the consequences for forecasting of two facts: first, that over time, statistics agencies revise and (presumably) improve observations on economic data, meaning that observations on the most recent data are typically the least well measured. Second, that economies are such that observations on the most recent realisations of economic variables contain the largest signal about future values of those variables.[1]

There is now a large literature that attempts to examine how these facts affect forecasting and monetary policymaking. We will not attempt a survey in this paper, but some key strands of research are these:[2] real-time data sets that enable economists to study the properties of different vintages of data relevant to policymaking have been compiled by Croushore and Stark (2001) for the US, and by Eggington, Pick, and Vahey (2002), Castle and Ellis (2002) and Patterson and Hervai (1991) for the UK. Other studies (examples are Mankiw, Runkle, and Shapiro (1984), Sargent (1989) and Faust, Rogers, and Wright (2001)) have studied whether the statistics agency behaves like a 'rational' forecaster by examining whether early releases of data predict later ones. Still others have studied the implications for monetary policy and inflation forecasts of having to use real-time measures of important indicators like the output gap (Orphanides (2000)and Orphanides and Van-Norden (2001)).

Within this broad literature are papers that study the properties of forecast models in the presence of measurement error, and these are the closest intellectual antecedents of our own. One line of enquiry has been to study a problem of joint model estimation and signal extraction/forecasting. Optimal filters/forecasts are studied in a line of work such as, for example, Howrey (1978) and Harvey, Mckenzie, and Desai (1983). Koenig, Dolmas, and Piger (2003) present informal experiments that reveal the advantages for forecasting of using real-time data for model estimation. Another focus for study has been the idea of evaluating the properties of combinations of forecasts (see, for example, Bates and Granger (1969) and discussions in Hendry and Clements (2001)). Observations on time series at dates leading up to time t are 'forecasts' of sorts of data at time t, so the problem of how best to make use of these data is a problem of combining forecasts.[3]

A brief sketch of our paper makes clear the contribution of our work. Section 2 begins by illustrating and proving how, in an environment when the measurement error does not vary with the vintage, faced with a choice between either using or not using observations on the most recent data, it can be optimal not to use them if the measurement error is sufficiently large. We move on to consider a case when the variance of measurement error is

[1]See Castle and Ellis (2002, page 44) for a discussion of the reasons why data are revised in the UK.

[2]A helpful bibliography can be found at http://phil.frb.org/econ/forecast/reabib.html.

[3] This observation is made in Busetti (2001).

larger, the more recent the data observation. We find that in this case a many-step ahead forecast may be optimal. In Section 2.3 we generalise these results further by assuming that the forecaster can choose the parameters of the forecast model as well as how much of the data to use. We find that it may be optimal to ignore recent data and to use forecasting parameters that differ from the parameters of the data generating process.

Section 3 generalises the results by deriving the optimal forecasting model from the class of linear autoregressive models. This setup allows the forecaster to include many lags of the data to construct the forecast and place different weights (coefficients) on different lags. Unsurprisingly, the optimal weighting scheme differs from the weighting scheme that characterises the data generating process. The greater the signal about the future in a data point, the greater the weight in the optimal forecasting model. More recent and therefore more imprecisely measured data have a smaller weight. The greater the persistence in the data generating process, the greater the signal in older data for the future, and the more extra measurement error in recent data relative to old data makes it optimal to rely on older data.

Throughout, this paper puts to one side the problem of model estimation. We assume that the forecaster/policymaker knows the true model. Taken at face value, this looks like a very unrealistic assumption. But it has two advantages. First, it enables us to isolate the forecasting problem, without any loss of generality. The second advantage is that it also emphasises an aspect of forecasting and policy that is realistic. The forecaster may have a noisy information source that is contaminated with measurement error, but also contains an important signal about shocks. The forecaster may also have an information source that is not contaminated by (at least that source of) measurement error – an economic prior – but that does not contain the same high frequency diagnosis of the state of the economy. The set up we use is just an extreme version of this. We assume that the forecaster's prior about the structure of the economy (the data generating process) is correct.

In Section 4, we present an application to illustrate the gains from exploiting the results in Section 3 using a single-equation forecasting model for the quarterly growth of private consumption in the UK. We use real time data on revisions to national accounts from Castle and Ellis (2002) to estimate how the variance of measurement error declines as we move back in time from the data frontier at T to some $T - n$. We find that the optimal forecasting model does indeed use significantly different weights on the data than those implied by the underlying estimated model, suggesting that the problem we study here may well be quantitatively important.

2 Some Issues on forecasting under data revisions

We begin, as we described in the introduction, by illustrating how it may be optimal not to use recent data for forecasting, but instead to rely on the model, which we assume is known. We start with a simple model, but we will relax some of our assumptions later.

2.1 Age-invariant measurement error

We start with a model in which the measurement error in an observation on an economic event is homoskedastic and therefore not dependent on the vintage. Assume that the true model is

$$y_t^* = ay_{t-1}^* + e_t \tag{1}$$

where $|a| < 1$ and y_t^* denotes the true series. Data is measured with error, and the relationship between the true and observed series is given by

$$y_t = y_t^* + v_t \tag{2}$$

For this section, we make the following assumptions about the processes for e and v:

$$e_t \sim i.i.d.(0, \sigma_e^2), \quad v_t \sim i.i.d.(0, \sigma_v^2) \tag{3}$$

which encompasses the assumption that the measured data are unbiased estimates of the true data.[4]

We assume that we have a sample from period $t = 1$ to period $t = T$ and we wish to forecast some future realisation y_{T+1}^*. The standard forecast, (when there is no measurement error) for y_{T+1}^* is denoted by $\hat{y}_{T+1}^{(0)}$ and given by $\hat{y}_{T+1}^{(0)} = ay_T$: this is the forecast that simply projects the most recent observation of y_t using the true model coefficient a. We investigate the mean square properties of this forecast compared with the general forecast $\hat{y}_{T+1}^{(n)} = a^{n+1}y_{T-n}$, a class of forecasts that project using data that are older than the most recent outturn.

We begin by finding an expression for the forecast error, and then computing the mean squared error for different forecasts amongst the general class described above. The (true) forecast error (which of course we never observe) is given by $\hat{u}_{T+1}^{(n)} = y_{T+1}^* - \hat{y}_{T+1}^{(n)}$. We know that from (1) we can write:

$$y_{T+1}^* = a^{n+1}y_{T-n}^* + \sum_{i=0}^{n} a^i e_{T+1-i} \tag{4}$$

and from (2) we have:

$$\hat{y}_{T+1}^{(n)} = a^{n+1}y_{T-n} = a^{n+1}y_{T-n}^* + a^{n+1}v_{T-n} \tag{5}$$

[4]The analysis in Castle and Ellis (2002) focuses on the first moment properties of revisions and finds some evidence of bias. But we abstract from that issue here.

So:

$$\hat{u}_{T+1}^{(n)} = \sum_{i=0}^{n} a^i e_{T+1-i} - a^{n+1} v_{T-n} \tag{6}$$

Therefore, the mean square error is simply given by:[5]

$$MSE(n) = a^{2(n+1)}\sigma_v^2 + \left(1 + \frac{a^2 - a^{2(n+1)}}{1-a^2}\right)\sigma_e^2 \tag{7}$$

The next step is to explore the condition that the mean squared error from a forecast using the most recent data is less than the mean squared error that uses some other more restricted information set, or $MSE(0) < MSE(n)$ for some $n > 0$. This will tell us whether there are circumstances under which it is worth forecasting without using the latest data. Doing this gives us:

$$MSE(0) < MSE(n) \Rightarrow a^2\sigma_v^2 + \sigma_e^2 < a^{2(n+1)}\sigma_v^2 + \left(1 + \frac{a^2 - a^{2(n+1)}}{1-a^2}\right)\sigma_e^2 \tag{8}$$

which can be written as:

$$\sigma_v^2 < \frac{\sigma_e^2}{1-a^2} \tag{9}$$

So if $\sigma_v^2 > \frac{\sigma_e^2}{1-a^2}$ it is better in terms of MSE *not* to use the most recent data. The intuition is simply that if the variance of the measurement error σ_v^2 is very large relative to the shocks that hit the data generating process, (σ_e^2), then it is not worth using the data to forecast: the more so the smaller is the parameter that propagates those shocks (a). In fact it follows that if $\sigma_v^2 > \frac{\sigma_e^2}{1-a^2}$ then $MSE(n-1) > MSE(n)$ for all n and therefore we are better off using the unconditional mean of the model to forecast the true series than any other data. The above analysis concentrated on a simple AR(1) model. However, the intuition is clear and is valid for more general dynamic models.

2.2 Age-dependent measurement error

We now investigate a slightly more complex case where the variance of the data measurement error v_t is assumed to tail off over time. This assumption reflects the observation that, in practice, we observe that statistics agencies revise data often many times after the first release. If we assume that successive estimates of a particular data point are subject to less uncertainty (since they are based on more information), then it seems reasonable to assume that the variance of the revision error embodied in the estimate of a particular data point diminishes over time.

The specific assumption we make here is that:

$$Var(v_{T-i}) = \begin{cases} b^i\sigma_v^2, & i = 0, 1, \ldots, N \\ 0, & i = N+1, \ldots \end{cases} \tag{10}$$

for a parameter $0 < b < 1$. We therefore assume that after a finite number of periods $N+1$, there are no further revisions to the data. But for the first $N+1$ periods, the variance of the revision error declines geometrically over time at a constant rate measured by b. This

[5] Notice that this expression requires that the revision errors in (2) are uncorrelated with future shocks to the model (1). This seems like a reasonable assumption.

is a fairly specific assumption which we make here for simplicity and tractability (again the analysis of later sections is more general). Indeed, we know that data are revised for reasons other than new information specific to that series (for example re-basing and methodology changes) so the specification of revision error variance may be more complicated than we have assumed here. But the purpose of the assumption is to be more realistic than the homoskedastic case considered in Section 2.1.

Under our assumptions, the MSE as a function of n is given by

$$MSE(n) = a^{2n+2}b^n\sigma_v^2 + \sum_{i=0}^{n} a^{2i}\sigma_e^2, \quad n = 0, 1, \ldots, N \tag{11}$$

$$MSE(n) = \sum_{i=0}^{n} a^{2i}\sigma_e^2 = \sum_{i=0}^{N} a^{2i}\sigma_e^2 + \sum_{i=N+1}^{n} a^{2i}\sigma_e^2, \quad n = N+1, \ldots \tag{12}$$

We want to examine when $MSE(n) > MSE(N+1)$, $n = 0, 1, \ldots, N$. It is clear that $MSE(n) > MSE(N+1)$, $n = N+2, \ldots,$. So, for $n = 0, 1, \ldots, N$

$$MSE(n) > MSE(N+1) \Rightarrow a^{2n+2}b^n\sigma_v^2 + \sum_{i=0}^{n} a^{2i}\sigma_e^2 > \sum_{i=0}^{n} a^{2i}\sigma_e^2 + \sum_{i=n+1}^{N+1} a^{2i}\sigma_e^2 \tag{13}$$

or, in terms of the signal-noise ratio, $\sigma = \sigma_e^2/\sigma_v^2$:

$$\frac{b^n(1-a^2)}{1-a^{2(N-n+1)}} > \sigma^2 \tag{14}$$

Figure 1: b=0.99

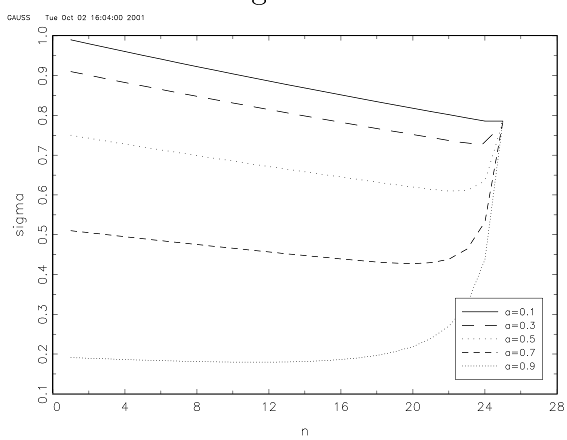
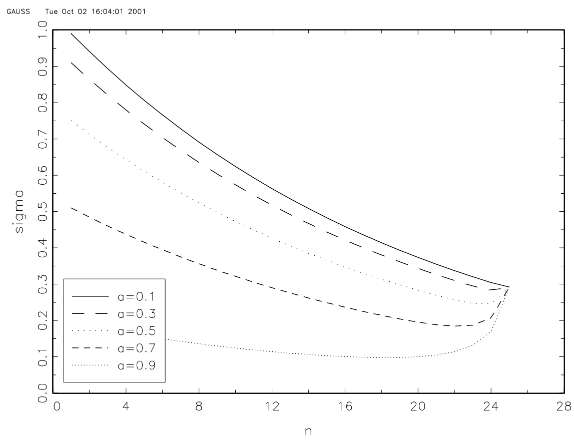


Figure 2: b=0.95



So if $\sigma^2 < \frac{b^n(1-a^2)}{1-a^{2(N-n+1)}}$ for all n then the best forecast for y_{t+1} is $\hat{y}_{t+1}^{(N+1)}$. To clarify the range of relevant values for σ we graph the quantity $\frac{b^n(1-a^2)}{1-a^{2(N-n+1)}}$ over n for $N = 24$, $b = 0.99, 0.95, 0.9, 0.5$ and $a = 0.1, 0.3, 0.5, 0.7, 0.9$ in Figures 1-4. If each period corresponds to one quarter, then our assumption $N = 24$ corresponds to the situation in which data are unrevised after six years. While this is naturally an approximation (since rebasing and methodological changes can imply changes to official figures over the entire length of the data series) it seems a plausible one.

Clearly, the more persistent the process is (the larger the a) the lower σ^2 has to be for $\hat{y}_{t+1}^{(N+1)}$ to be the best forecast. Also, the more slowly the revision error dies out (the larger the b), the lower σ^2 has to be for $\hat{y}_{t+1}^{(N+1)}$ to be the best forecast. Note that some of the curves in the figures are not monotonic. This indicates that although $\hat{y}_{t+1}^{(N+1)}$ is a better forecast than $\hat{y}_{t+1}^{(0)}$, there exists some $N+1 > n > 0$ such that $\hat{y}_{t+1}^{(n)}$ is better than $\hat{y}_{t+1}^{(N+1)}$.

Figure 3: b=0.9

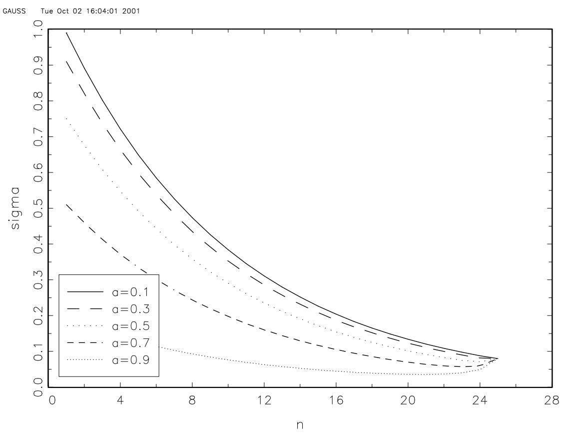
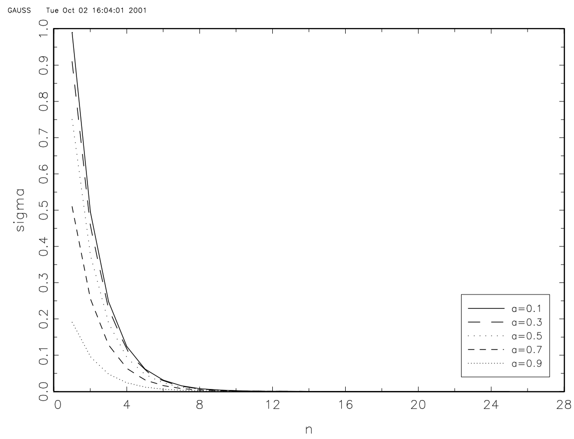


Figure 4: b=0.5



2.3 Optimising over the choice of data frontier and the projection parameters

The analysis in the previous section constrained the forecaster to use the true model when forecasting future outturns. The only choice variable was therefore the horizon n upon which to base the forecast $\hat{y}_{T+1}^{(n)} = a^{n+1} y_{T-n}$. This section generalises the problem of the policymaker/forecaster so that it is possible to construct a forecast that does not use the true model parameter. Specifically, we allow the policymaker/forecaster to use the forecast $\hat{y}_{t+1}^{(n)}(\tilde{a}) = \tilde{a}^{n+1} y_{t-n}$ where $\tilde{a}$ may differ from a. In this setting, there are two choice variables ($\tilde{a}$ and n) and so it might be more appropriate to use a different parameter value, $\tilde{a}$, and the most recent data as opposed to older data and the true model parameter, a.

This section therefore extends the setup and views the mean square error as a function of n and $\tilde{a}$ where the forecast is given by $\hat{y}_{t+1}^{(n)}(\tilde{a}) = \tilde{a}^{n+1} y_{t-n}$ and $\tilde{a}$, n are to be jointly determined given the structural parameters a, σ_v^2 and σ_e^2. We extend the analysis along these lines assuming that the revision error variance is given by $Var(v_{t-i}) = b^i \sigma_v^2, i = 0, 1, \dots, N$ and $Var(v_{t-i}) = 0, i = N+1, \dots$ as before.

Now the mean square error is a joint function of n and $\tilde{a}$ given by

$$MSE(n, \tilde{a}) = (a^{n+1} - \tilde{a}^{n+1})^2 \frac{\sigma_e^2}{1-a^2} + \frac{(1 - a^{2(n+1)})\sigma_e^2}{1-a^2} + \tilde{a}^{2(n+1)} b^n \sigma_v^2, \quad n = 0, 1, \dots, N \quad (15)$$

$$MSE(n, \tilde{a}) = (a^{n+1} - \tilde{a}^{n+1})^2 \frac{\sigma_e^2}{1-a^2} + \frac{(1 - a^{2(n+1)})\sigma_e^2}{1-a^2}, \quad n = N+1, \dots \quad (16)$$

and we wish to find the optimal values for n and $\tilde{a}$. To do so, we analyse a two-step minisimisation problem. First we will minimise the mean squared error with respect to the forecasting parameter $\tilde{a}$. This allows us to write down a set of mean-squared errors that use the optimal forecasting parameters as n changes. To find the best forecast simply requires choosing the n that gives the overall smallest mean-squared error.

We therefore begin by minimising $MSE(n, \tilde{a})$ with respect to $\tilde{a}$. The first order necessary conditions are:

$$\frac{\partial MSE(n,\tilde{a})}{\partial \tilde{a}} = \begin{cases} -2(a^{n+1}-\tilde{a}^{n+1})(n+1)\tilde{a}^n \frac{\sigma_e^2}{1-a^2} + 2(n+1)\tilde{a}^{2n+1}b^n\sigma_v^2 = 0 & n = 0,1,\ldots,N \\ -2(a^{n+1}-\tilde{a}^{n+1})^2(n+1)\tilde{a}^n \frac{\sigma_e^2}{1-a^2} = 0 & n = N+1,\ldots \end{cases} \tag{17}$$

Rearranging gives

$$-\tilde{a}^n \left[a^{n+1} - \left(1 - \frac{b^n(1-a^2)}{\sigma^2}\right)\tilde{a}^{n+1}\right] = 0, \quad n = 1,\ldots,N \tag{18}$$

$$-\tilde{a}^n \left(a^{n+1} - \tilde{a}^{n+1}\right) = 0, \quad n = N+1,\ldots \tag{19}$$

For (18), disregarding complex roots and under the convention of square roots being positive numbers, the solutions are $\tilde{a} = 0$ and $\tilde{a} = \sqrt[n+1]{\theta}$ where $\theta = \frac{\sigma^2 a^{n+1}}{\sigma^2 + b^n(1-a^2)}$. For (19), they are, intuitively, $\tilde{a} = 0$ and $\tilde{a} = a$. Note that for positive a, $\theta \geq 0$ making sure than the second solution of (18) is real. It is easy to verify that the non-zero solutions are minima.

We can now incorporate the solutions of this minimisation into the expression for the mean squared error. We define $M\hat{S}E(n) = \min_{\tilde{a}} MSE(n, \tilde{a})$:

$$M\hat{S}E(n) = \left[(a^{n+1}-\theta)^2 + (1-a^{2(n+1)})\right]\frac{\sigma_e^2}{1-a^2} + \theta^2 b^n \sigma_v^2 \quad n = 0,1,\ldots,N \tag{20}$$

$$M\hat{S}E(n) = \frac{(1-a^{2(n+1)})\sigma_e^2}{1-a^2} \quad n = N+1,\ldots \tag{21}$$

which has to be minimised over n. Standard methods do not apply as n takes only discrete values[6]. Nevertheless, we again see that it is not neccesary that the best forecast uses the most recent data.

3 A general approach to forecasting with dynamic models under data revisions

In this section we propose a general method of forecasting in autoregressive models under a general known form of data revisions. The extension from the previous sections is that we optimise the forecasting model from within the linear class of models. Specifically, we allow the forecaster to choose the optimal weights and lags on all past data. The method described here can be easily extended to multivariate models. In particular VAR models could easily be accommodated (including of course vector error correction models).

We assume a univariate $AR(p)$ model to illustrate the method. The true process is given by

$$y_t^* = \sum_{i=1}^{p} a_i y_{t-i}^* + e_t \tag{22}$$

[6] As the minimum of any function is also the minimum of a positive monotonic transformation of that function, we can take logs of the mean-squared-error expressions to show that, for real n, the minimum is obtained for

$$n = \frac{\log\left[\frac{(1-a^2)\theta\log(b)}{2\sigma^2 a\log(a)}\right]}{\log(a/b)}$$

This can be written as a VAR(1) model of the form

$$\mathbf{y}_t^* = \mathbf{A}\mathbf{y}_{t-1}^* + \epsilon_t \tag{23}$$

where $\mathbf{y}_t^* = (y_t^*, y_{t-1}^*, \ldots, y_{t-p+1}^*)'$, $\mathbf{A} = (\mathbf{A}_1', \mathbf{e}_1, \ldots, \mathbf{e}_{p-1})'$; $\mathbf{e}_i$ is a $p \times 1$ vector with an element of 1 at the i-th place and zeroes everywhere else; $\mathbf{A}_1$ is a $p \times 1$ vector of the autoregressive coefficients a_i; $\epsilon_t = (e_t, 0 \ldots, 0)'$. Now the observed data are given by

$$\mathbf{y}_t = \mathbf{y}_t^* + \mathbf{v}_t \tag{24}$$

where $\mathbf{v}_t = (v_t, v_{t-1}, \ldots, v_{t-p+1})$. At time T we wish to determine the optimal forecast for y_{T+1}^*. We assume that the revision error $\mathbf{v}_T$ has a variance matrix which is given by $\mathbf{\Sigma}_v^T$. Our aim is to determine the optimal forecasting model of the form $\hat{y}_{T+1} = \tilde{\mathbf{A}}_1\mathbf{y}_T$. in terms of mean square error, where $\tilde{\mathbf{A}}_1$ is a $1 \times p$ vector. Note that the restriction on the dimension of $\tilde{\mathbf{A}}_1$ to be the same as that of the order of the true process is not problematic because we can simply increase the order of the process by setting the higher order a's equal to zero. This means that the true data generating process might be an AR(1) even though we can write it as an AR(p) with the coefficients on lags $2, \ldots, p$ set equal to zero.

The forecast error for the forecast of the above form is given by

$$y_{T+1}^* - \hat{y}_{T+1} = \mathbf{A}_1\mathbf{y}_T^* + \epsilon_T - \tilde{\mathbf{A}}_1\mathbf{y}_T^* + \tilde{\mathbf{A}}_1\mathbf{v}_T = (\mathbf{A}_1 - \tilde{\mathbf{A}}_1)\mathbf{y}_T^* + \tilde{\mathbf{A}}_1\mathbf{v}_T + \epsilon_{T+1} \tag{25}$$

where $\mathbf{A}_1$ is the first row of $\mathbf{A}$. The mean square error is given by

$$(\mathbf{A}_1 - \tilde{\mathbf{A}}_1)\mathbf{\Gamma}(\mathbf{A}_1 - \tilde{\mathbf{A}}_1)' + \tilde{\mathbf{A}}_1\mathbf{\Sigma}_v^T\tilde{\mathbf{A}}_1' + \sigma_\epsilon^2 \tag{26}$$

where $\mathbf{\Gamma} = E(\mathbf{y}_T^*\mathbf{y}_T^{*'})$. The covariances of an AR(p) process are given by the first p elements of the first column of the matrix $\sigma_\epsilon^2[\mathbf{I}_{p^2} - \mathbf{A} \otimes \mathbf{A}]^{-1}$. We have assumed that the error process is uncorrelated with the true process of the data. In the data revision literature this is referred to as the error-in-variables model. This assumption is not crucial to our analysis and could be relaxed as long as the covariances between the true process and the data revision errors could be estimated. We want to minimise the mean square error in terms of $\tilde{\mathbf{A}}_1$. We will use matrix optimisation calculus to solve this problem. We rewrite the expression for the mean square error using only terms involving $\tilde{\mathbf{A}}_1$ since the rest of the terms will not affect the minimisation. We have that the mean square error is given by

$$\tilde{\mathbf{A}}_1\mathbf{\Gamma}\tilde{\mathbf{A}}_1' + \tilde{\mathbf{A}}_1\mathbf{\Sigma}_v^T\tilde{\mathbf{A}}_1' - \mathbf{A}_1\mathbf{\Gamma}\tilde{\mathbf{A}}_1' - \tilde{\mathbf{A}}_1\mathbf{\Gamma}\mathbf{A}_1' = \tilde{\mathbf{A}}_1(\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)\tilde{\mathbf{A}}_1' - 2\tilde{\mathbf{A}}_1\mathbf{\Gamma}\mathbf{A}_1' \tag{27}$$

We differentiate with respect to $\tilde{\mathbf{A}}_1$ and set to zero to get

$$(\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)\tilde{\mathbf{A}}_1' - \mathbf{\Gamma}\mathbf{A}_1' = 0 \tag{28}$$

giving

$$\tilde{\mathbf{A}}_1^{opt\ '} = (\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)^{-1}\mathbf{\Gamma}\mathbf{A}_1' \tag{29}$$

The second derivative is given by $(\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)$ and by the positive definiteness of this matrix the second order condition for minimisation of the mean square error is satisfied. This result is of some interest because it may be viewed as analogous to similar results in other literatures. Note first the similarity between this result and the standard signal extraction result which says that the optimal filter for distinguishing between signal and noise is equal to the autocovariance of the signal ($\mathbf{\Gamma}$ in our case) divided by the sum of the signal and noise autocovariances. Note that if $p = 1$ then $\tilde{\mathbf{A}}_1^2 \leq \mathbf{A}_1^2$. By the positive-definiteness of $\mathbf{\Gamma}$

and $\mathbf{\Sigma}_v^T$ one might conjecture that this result would extend to the multivariate case where $\tilde{\mathbf{A}}_1\tilde{\mathbf{A}}_1' \leq \mathbf{A}_1\mathbf{A}_1'$. Unfortunately, this is not the case. Although it is likely that this result will hold it is by no means certain. Another interesting corollary of the above result is that the method applies equally to measurement error. The only assumption we have made is that there exist an error in the measurement of the true data whose covariance is given by $\mathbf{\Sigma}_v^T$. This clearly covers cases of data measurement error.

The above analysis concentrated on one-step ahead forecasts. The general problem of h-step ahead forecasting can be dealt with similarly by minimising the sum of the 1 to h-step ahead forecast errors with respect to a suitably defined set of coefficients $\tilde{\boldsymbol{A}}$ just as we did above. We analyse this case in what follows: We want to minimise the variance of the forecast errors of the 1-step to n-step ahead forecasts. As we need to minimise a scalar function we choose to minimise the trace of the forecast error variance-covariance matrix of the 1 to h step forecasts. We assume for simplicity that $p > h$. If this is not case it can always be made the case by increasing p. Using the previous notation we know that

$$\boldsymbol{y}^*_{T+h} = \boldsymbol{A}^n\boldsymbol{y}^*_T + \boldsymbol{A}^{h-1}\boldsymbol{\epsilon}_T + \ldots + \boldsymbol{\epsilon}_{T+n} \tag{30}$$

So

$$\boldsymbol{y}^*_{T+h,h} = (y^*_{T+1}, \ldots, y^*_{T+h}) = \boldsymbol{A}^{(n)}\boldsymbol{y}^*_T + \boldsymbol{A}^{(h-1)}\boldsymbol{\epsilon}_T + \ldots + \boldsymbol{\epsilon}_{T+h,h} \tag{31}$$

where $\boldsymbol{A}^{(h)}$ denote the first h rows of $\boldsymbol{A}^h$ and $\boldsymbol{\epsilon}_{T+h,h}$ is a vector of the first h of the vector $\boldsymbol{\epsilon}_{T+h}$. So the forecast error is given by

$$\boldsymbol{y}^*_{T+h,h} - \hat{\boldsymbol{y}}_{T+h,h} = \boldsymbol{A}^{(h)}\boldsymbol{y}^*_T + \boldsymbol{A}^{(h-1)}\boldsymbol{\epsilon}_T + \ldots + \boldsymbol{\epsilon}_{T+h,h} - \tilde{\boldsymbol{A}}\boldsymbol{y}^*_T - \tilde{\boldsymbol{A}}\boldsymbol{v}_T \tag{32}$$

The part of the variance of the forecast error, depending on $\tilde{\boldsymbol{A}}$, which is relevant for the minimisation problem, is given as before by

$$\tilde{\boldsymbol{A}}(\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)\tilde{\boldsymbol{A}}' - 2\tilde{\boldsymbol{A}}\mathbf{\Gamma}\boldsymbol{A}^{(h)\prime} \tag{33}$$

Differentiating and noting that the derivative of the trace of the above matrix is the trace of the derivative gives

$$tr((\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)\tilde{\boldsymbol{A}}' - \mathbf{\Gamma}\boldsymbol{A}^{(h)\prime}) = 0 \tag{34}$$

If the matrix is equal to zero then the trace is equal to zero and so if

$$(\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)\tilde{\boldsymbol{A}}' - \mathbf{\Gamma}\boldsymbol{A}^{(h)\prime} = 0 \tag{35}$$

the first order condition is satisfied. But the above equality implies that

$$\tilde{\boldsymbol{A}}^{opt\ \prime} = (\mathbf{\Gamma} + \mathbf{\Sigma}_v^T)^{-1}\mathbf{\Gamma}\boldsymbol{A}^{(h)\prime} \tag{36}$$

Finally, the variance of the optimal coefficients is easily obtained using the Delta method.

As we see from the above exposition our method essentially adjusts the coefficients of the AR model to reflect the existence of the measurement error. Of course, using the 'wrong' coefficients to forecast introduces bias. Often, we restrict attention to best linear unbiased (BLU) estimates. But this means that if we have two models, one unbiased but with high variance, and the other with a small bias but low variance, we always pick the former. This would be true even if the low-variance estimator had a lower forecast MSE than the unbiased estimator. So the reduction in variance induced by 'aiming off' the true coefficients outweighs the loss from bias.

Clearly the method we suggest is optimal in terms of mean square forecasting error conditional on being restricted to use p periods of past data, where $p = T$ is a possibility. It is therefore equivalent to using the Kalman filter on a state space model[7] once $p = T$. Nevertheless, the method we suggest may have advantages over the Kalman filter in many cases. Firstly, the method we suggest is transparent and easy to interpret structurally. For example, one can say something about the coefficients entering the regression and how they change when revisions occur. It is also possible to carry out inference on the new coefficients. We can obtain the standard errors of the modified coefficients from the standard errors of the original coefficients. So in forecasting one can say something about the importance (weight) of given variables and the statistical significance of those weights. From a practical point of view where a large model with many equations is being used for forecasting, and one which must bear the weight of economic story-telling, one may want to fix the coefficients for a few periods and not reestimate the whole model. Our method has some advantages over the Kalman Filter in uses of this sort, since it just uses the same coefficients rather than applying a full Kalman filter every period. Finally, the method we have is nonparametric as far as variances for the revision error are concerned. We have a $T \times 1$ vector of errors at time T. In the most general case, these errors can have any $T \times T$ covariance matrix that represents all possibilities for how the variance of measurement error varies by vintage, over time, (and, in a multivariate setting, across variables). In other words, our procedure allows for time variation in the covariances, heteroscedasticity and serial correlation. The state space cannot easily attain that sort of generality. In fact a standard state space imposes rather strict forms of covariance to the errors that are unappealing in the context we are envisaging. These can be relaxed with great difficulty only and by experienced state space modellers.

Another point worth making in this context concerns the distinction between the 'news' and 'noise' alternative representations of measurement error as discussed by, e.g., Mankiw, Runkle, and Shapiro (1984) and Sargent (1989). In this paper we have adopted the 'noise' interpretation. It worth noting that adopting the extreme 'news' representation where the agency publishing data knows the correct economic model and uses it to optimally filter data prior to release would make our suggested methodology redundant for linear models. If the published data are not equal to the truth but are an optimal projection conditional on all available information obtained via, say, the state space representation of the economic model then the optimal forecast is obviously obtained by using the published data in the economic model. In the context of autoregressive models, the best one can do is simply use the available data together with the model, as this is equivalent to using the state space representation.

Finally, we note that a number of practical complications have been assumed away in the discussion of our method. For example, the variance of revision errors often depends not just on how long it has been since the first release of the data, but also on the time that data were first released. For example, in the UK, large revisions occur once a year with the publication of the 'Blue Book' by the Office of National Statistics.

[7]For more details on the state space representation of the case we consider see Harvey, Mckenzie, and Desai (1983). There the authors develop a state space methodology for irregular revisions for univariate models. Multivariate models are briefly discussed as well.

4 Empirical illustration

We apply the general method of optimising a forecast model to an consumption forecasting equation based on a simple AR model. Such models, however, have been found to have very good forecasting performance in a variety of settings. The model is given by

$$\Delta c_t = a_0 + \sum_{i=1}^{p} a_i \Delta c_{t-i} + e_t \qquad (37)$$

where c_t is the (log of) consumption. Many equations of this general form include an error correction term. However, there is significant evidence to indicate that error correction terms may not be very helpful in a forecasting context. Evidence presented by Hoffman and Rasche (1996) demonstrates that the forecasting performance of VAR models may be better than that of error correction models over the short forecasting horizons which concern us. Only over long horizons are error correction models shown to have an advantage. Christoffersen and Diebold (1998) cast doubt on the notion that error correction models are better forecasting tools even at long horizons, at least with respect to the standard root mean square forecasting error criterion. They also argue that although unit roots are estimated consistently, modelling nonstationary series in (log) levels is likely to produce forecasts which are suboptimal in finite samples relative to a procedure that imposes unit roots, such as differencing, a phenomenon exacerbated by small sample estimation bias.

We use real time data from 1955Q1-1998Q2 for forecasting. We use the revision data available to provide estimates of the revision error variances. We assume that revisions do not occur in general after 24 revision rounds. More specifically we estimate the data revision variances as follows: We use real time data to estimate the variance of the revision error between release i and release $i+1$, $i = 1, \ldots, 24$. Denote these estimates of the variance by ζ_i. Then, the variance of the revision error of the i-th release is equal to $\sum_{j=i}^{24} \zeta_i$. The standard deviations for ζ_i estimated from the data revisions using data from 1983Q1-2001Q1 are given in Table 1. These are also plotted in Figure 5.

We want to investigate the out-of-sample performance of the above equation. We consider four variants of it. The four variants reflect the number of lags of consumption growth considered which varies from 1 to 4. We assume that the revision error becomes smaller and eventually disappears after 24 rounds of revisions.

We compare the forecasting performance of the models using optimal parameter estimates, obtained using the method of the previous section and as many lags as those in each model, and standard parameter estimates. The out-of-sample forecast evaluation exercise is carried out as follows. Starting at 1984Q3 the model is estimated over the period 1955Q1-1978Q3 to get parameter estimates. Then data up to 1984Q2 and the estimated parameters are used to forecast consumption at 1985Q3. The reason for not using the period 1978Q4-1984Q3 data for estimating the coefficients is to ensure (within the assumptions of the experiment) that the original parameter estimate reflects the true parameter rather than be contaminated by revision errors in the data. We continue producing forecasts until 1998Q2. So the whole forecast evaluation period is 1984Q3-1998Q2 (14 years). The reason we stop at 1998Q2 is because we need to use the most recently available data for the evaluation period as proxies for the true data (uncontaminated by noise). Although, for the initial periods, it may be the case that the number of observations used to estimate the coefficients is small this is rectified rapildy as more data accumulate for parameter estimation in later periods.

We look at the RMSE ratios of the forecasts coming from optimal and standard parameter

Figure 5: Revision error std. deviations for Consumption

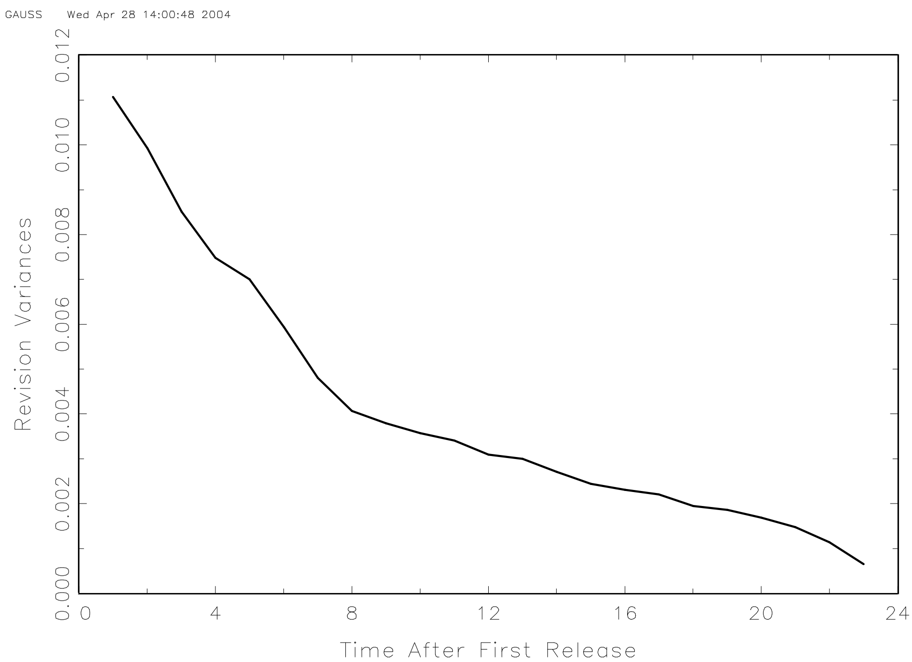


Table 1: Consumption revision error standard Deviations

Horizon	Standard Deviation
1	0.0111
2	0.0099
3	0.0085
4	0.0075
5	0.0070
6	0.0059
7	0.0048
8	0.0041
9	0.0038
10	0.0036
11	0.0034
12	0.0031
13	0.0030
14	0.0027
15	0.0024
16	0.0023
17	0.0022
18	0.0019
19	0.0019
20	0.0017
21	0.0015
22	0.0011
23	0.0007

estimates and we also look at the Diebold-Mariano tests (see Diebold and Mariano (1995)) looking at the null hypothesis that the two forecast are equally good in terms of RMSE. The test simply compares the means of the squared forecast errors of the two forecasts[8]. Results are also considered for the two seven year subperiods within the whole evaluation period. Results are presented in Table 2. . Figure 6 presents the one-step ahea forecasts for the whole period compared to the realised data.

Clearly the forecasts using the optimal coefficients outperform the standard forecasts for all models for the whole period and the two subperiods. The Diebold Mariano statistics indicate statistically significant superiority (at the 5% significance level) in the whole period and the second subperiod. Figure 6 illustrates the advantage of using the optimal coefficients. The optimal forecasts weighs less fluctuations in past data which may be considered to be partly arising out of measurement noise. Consequently, the optimal forecasts are smoother.

5 Summary

In this paper we have explored the effects of data revision on forecasting models. We have shown that in the presence of data revisions it is possible that forecasting with older data may provide superior forecasts in terms of mean square error compared to forecasts which

[8]Positive test statistics indicate superiority of the forecasts based on the optimal forecasting coefficients and vice versa.

Figure 6: One-Step Ahead Forecasts

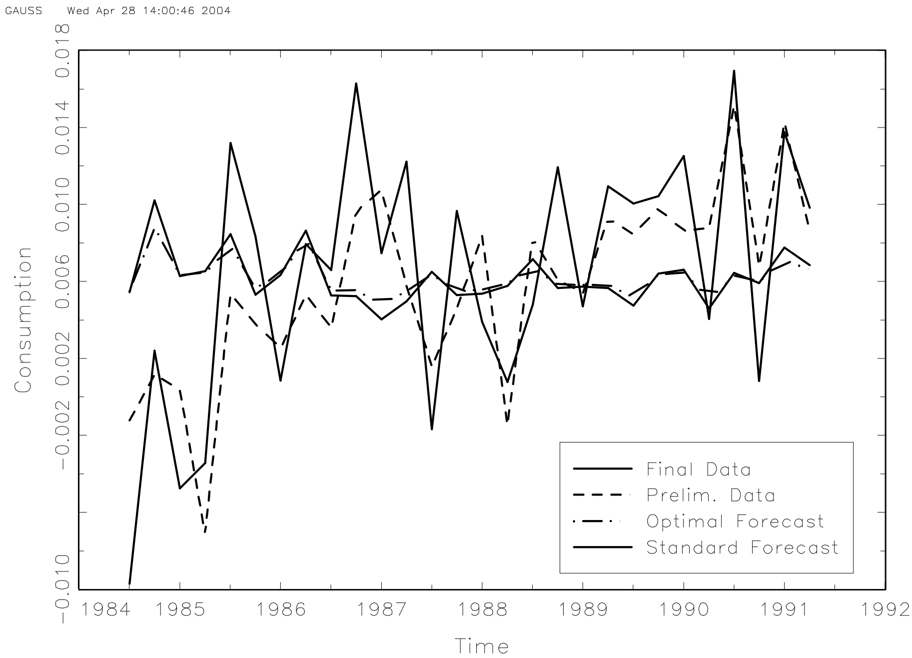


Table 2: MSE ratios and Diebold-Mariano tests[a]

	Whole period		First subperiod		Second subperiod	
Model	MSE Ratio	D-M Test	MSE Ratio	D-M Test	MSE Ratio	D-M Test
AR(1)	0.958	2.914*	0.956	2.261*	0.963	2.845*
AR(2)	0.945	2.771*	0.937	2.384*	0.966	2.318*
AR(3)	0.954	1.991*	0.941	1.916	0.988	0.739
AR(4)	0.929	2.740*	0.911	2.676*	0.976	1.09

[a]* denotes significance at the 5% level

Table 3: Whole Period Coefficients in AR forecasting models for UK consumption: standard and uncertainty corrected

	AR(1)	AR(2)	AR(3)	AR(4)
Standard	$-0.112_{(0.082)}$	-	-	-
	$-0.101_{(0.083)}$	$0.101_{(0.083)}$	-	-
	$-0.110_{(0.083)}$	$0.114_{(0.083)}$	$0.114_{(0.084)}$	-
	$-0.104_{(0.083)}$	$0.123_{(0.083)}$	$0.104_{(0.084)}$	$-0.081_{(0.084)}$
optimal	$-0.084_{(0.062)}$	-	-	-
	$-0.073_{(0.058)}$	$0.079_{(0.062)}$	-	-
	$-0.070_{(0.052)}$	$0.081_{(0.058)}$	$0.079_{(0.062)}$	-
	$-0.062_{(0.048)}$	$0.078_{(0.052)}$	$0.067_{(0.058)}$	$-0.063_{(0.062)}$

use the most recent data. This conclusion is not affected even if we allow for adjustments in the parameters of the dynamic model to optimise the forecast in terms of mean square error. Finally, we have provided a general method of determining the optimal forecasting model in the presence of data measurement and revision errors with known covariance structure. An empirical illustration on forecasting consumption was also considered.

References

Bates, J. M., and C. J. W. Granger (1969): "The combination of forecasts," *Operational Research Quarterly*, 20, 133–167.

Busetti, F. (2001): "The use of preliminary data in econometric forecasting: an application with the Bank of Italy Quarterly Model," Bank of Italy Discussion Paper.

Castle, J., and C. Ellis (2002): "Building a real-time database for GDP(E)," *Bank of England Quarterly Bulletin*, Spring, 42–49.

Christoffersen, P., and F. Diebold (1998): "Co-integration and Long-Horizon Forecasting," *Journal of Business and Economic Statistics*, 16, 450–458.

Croushore, D., and T. Stark (2001): "A Real Time Dataset for Macroeconomists," *Journal of Econometrics*, 105, 111–130.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Eggington, D., A. Pick, and S. Vahey (2002): "Keep it real!: A real-time data set for macroeconomists: does the vintage matter?," *Economics Letters*, forthcoming.

Faust, J., J. Rogers, and J. Wright (2001): "News and noise in G-7 GDP announcements," Federal Reserve Board of Governors Working Paper.

Harvey, A., C. R. Mckenzie, and D. P. C. B. M. J. Desai (1983): "Irregular Data Revisions," .

Hendry, D., and M. P. Clements (2001): "Economic forecasting: some lessons from recent research," Mimeo.

Hoffman, D. L., and R. H. Rasche (1996): “Assessing forecast performance in a cointegrated system,” *Journal of Applied Econometrics*, 11, 495–517.

Howrey, E. P. (1978): “The use of preliminary data in econometric forecasting,” *Review of Economic Statistics*, 60, 193–200.

Koenig, E. F., S. Dolmas, and J. Piger (2003): “The Use and Abuse of 'Real-Time' Data in Economic Forecasting,” *Review of Economics and Statistics*, 85, 618–628.

Mankiw, N. G., D. E. Runkle, and M. D. Shapiro (1984): “Are Preliminary Estimates of the Money Stock Rational Forecasts?,” *Journal of Monetary Economics*, 14, 14–27.

Orphanides, A. (2000): “The quest for prosperity without inflation,” Sveriges Riksbank working paper number 93.

Orphanides, A., and S. Van-Norden (2001): “The reliability of inflation forecasts based on output gaps in real time,” mimeo.

Patterson, K. D., and S. M. Hervai (1991): “Data revisions and expenditure components of GDP,” *Economic Journal*, 101, 887–901.

Sargent, T. J. (1989): “Two Models of Measurement and the Investment Accelarator,” *Journal of Political Economy*, 97, 251–287.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm