Estimating time-variation in measurement error from data revisions; an application to forecasting in dynamic models

George Kapetanios[*]

and

Tony Yates[**]

Working Paper no. 238

[*] Queen Mary and Westfield College 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. We are grateful to Richard Harrison for the insights he brought to our previous joint work from which this paper emerged. We have also benefited from helpful comments from Ian Bond, Alex Brazier, Spencer Dale, Melissa Davey, Graham Elliott, Simon Hayes, Athanasios Orphanides, Jo Paisley, Simon Price, Pierre Siklos, Peter Sinclair, Shaun Vahey and two anonymous referees.

Copies of working papers may be obtained from Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH; telephone 020 7601 4030, fax 020 7601 3298, e-mail mapublications@bankofengland.co.uk

Working papers are also available at www.bankofengland.co.uk/wp/index.html

The Bank of England's working paper series is externally refereed.

ISSN 1368-5562

Contents

Abstract

Over time, economic statistics are refined. This means that newer data are typically less well measured than old data. Time or vintage-variation in measurement error like this influences how forecasts should be made. Measurement error is obviously not directly observable. This paper shows that modelling the behaviour of the statistics agency generates an estimate of this time-variation. This provides an alternative to assuming that the final releases of variables are true. The paper applies the method to UK aggregate expenditure data, and demonstrates the gains in forecasting from exploiting these model-based estimates of measurement error.

Key words: Data uncertainty, measurement error, revisions, real-time data, forecasting.

JEL classification: C32,C53

Summary

Over time many sources of data that are relevant for estimating the current state of the economy are refined. This means that at any one time policymakers will be faced with data that are measured with different amounts of measurement error. Typically, more recently released data will be less well measured than revised data. How measurement error varies across data series and across vintages will affect optimal forecasts: these will, of course, put less weight on data that are less well measured. The problem is that, by construction, the amount of measurement error in any series is never observed. A popular choice of researchers has been to assume that the final or latest release of an observation on some variable is the truth, and then to proxy the variance of the measurement error in earlier releases by the variance of earlier releases around the final release. The drawback of this method is that the final release is itself measured badly.

This paper offers an alternative. The idea is to model the process by which the statistics agency publishes and then revises data. A hypothetical agency is constructed that conducts a series of independent random 'surveys' on a data point. Each time the agency conducts a new survey, the original estimate is revised based on its knowledge of the sampling error in the earlier and latest surveys. Using this assumption, we can describe how the variance of revisions to data between one vintage and any other will be related to the variance of the measurement error in the underlying surveys. We can therefore use something we observe - the variance of revisions in real-time data - to infer what we do not observe - the variance of measurement error. This paper applies the method to a real-time data set for the United Kingdom. We derive estimates of the variance of measurement error in vintages of the quarterly growth rates of private consumption and imports. We find that measurement error in the first release of imports is about six times that in the first release for consumption. We apply our estimates of the variance of measurement error in different vintages in a forecasting environment. We find that forecasts that are adjusted for our model-based estimates of measurement error outperform those that are not.

1 Introduction

The aim of this paper is to produce an estimate of how measurement error varies across data and vintages, by modelling the process that gives rise to real-time data.(1)

Over time, statistics agencies revise and improve estimates of data as they collect more information. It is likely that they measure some data better than others, and that they hone their estimates of data more quickly for some data than for others. If we could quantify the measurement error in different variables and vintages, this information would be useful for estimation and forecasting (and for computing estimates of policy responses that rely on such forecasts). The more noise in data, the less weight we should put on them in forming an estimate of the state of the economy, future inflation, or the right setting for current monetary policy. These points are well known. They have been emphasised recently in papers by Aoki (2003), Swanson (2000), Woodford and Svensson (2003) - who look at optimal monetary policy and measurement error - and, Harrison, Kapetanios and Yates (2004) - who studied how optimal forecasts are affected by measurment errors that vary across vintages. However, since we never observe the true value for any data, we can therefore never observe the measurement error in any data vintage. Some data are based at least in part on surveys. But information about the sampling error in these surveys is typically not available. More importantly, many data are constructed from many different sources other than surveys. Readily quantifiable concepts of the measurement error contained in these other information sources may not exist: judgment is sometimes an important element in reconciling information from conflicting sources, and there is no simple way of quantifying the measurement error associated with that kind of information.

What we do observe are successive vintages of data. One option is to approximate the measurement error in a variable by assuming that the final release of a variable is equal to the true value. Then, the variance of early releases around the final release is an estimate of the variance of those releases about the truth. This approach has been taken by Harrison *et al.* (2004), Coenen, Levin and Weiland (2001) and many others. Rather obviously, the closer the final release is to the true value, the better this is as an approximating method. Ominously, since we never know how close the final release is to the true value for some data, we cannot assess how good an

(1) To stress at the outset, we are concerned with trying to estimate the variance in the difference between observed and true values of variables (what we call ‘measurement error’), as distinct from the variance of shocks that drive true, unobserved variables.

approximation this method is to the ideal. Another option is to use a Kalman Filter. That method uses an assumption about the economic process driving movements in the true variable to get an estimate of the truth. That estimate can then be used to go back and compute how, on average, early releases of data vary around a Kalman-Filter estimate of the truth, based on final or near-final data.

Our paper provides an alternative, exploiting behavioural assumptions about the statistics agency that generates the real-time data. We show how a few assumptions can generate an estimate of the measurement error in different vintages of data from observations on the variance of revisions to data. These assumptions are the following. First, a conjecture about the scheme that the statistics agency uses to weight new information about an observation together with old information. We start out by assuming that the statistics agency weights the surveys optimally as they arrive. We also experiment with another scheme we describe as 'naive': where surveys are weighted equally. Our second assumption is that the arrival of incremental information about a data point can be modelled as a sequence of independent random draws from a population. Crucially, our hypothetical agency does not engage in filtering and forecasting itself in the way that Sargent (1989) described was possible, and Mankiw, Runkle and Shapiro (1984) investigated for releases of US GNP. This assumption is consistent with our assumption about an agency that does optimal weighting. Rational data collection can reasonably be taken to include making use of all future information that is not independent from current information in the current release: and therefore that incremental information is independent. A third assumption relates to the evolution of the quantity of incremental information that arrives about an observation over time. The basic model assumes that the rate at which the flow of new information in surveys changes is fixed. But we relax this assumption and derive a model that allows the data to reveal how the flow of new information evolves.

These assumptions together build a model of a hypothetical statistics agency. The model is not a literal description of the real-world data collection process. Many data are constructed and subsequently revised using information that has many sources. Some of it is literally from surveys. Other information is based on censuses. Still other information is based on economic models that are used to corroborate one information source with another, or judgment, or even filtering and forecasting. Our model will be useful if the flow of information from these many sources can, to a good enough approximation, be described as if it were a sequence of independent random surveys.

This model can be used to compare the predicted variance of revisions to data with the observed variance of revisions. Imagine a statistics agency that conducted ever smaller independent random samples from a population, and weighted them optimally. The revisions to the data that this agency would make would get smaller over time. The reliability of the new information would shrink relative to the ever larger weighted combinations of older information, and so would get less and less weight each time. By being specific about the weighting scheme the statistics agency uses, and about how the sample sizes of these surveys evolve, we can use the variance of the revisions in this way to uncover estimates of measurement error in the published data. As we will show, it turns out that the variance of revisions can be written as a function of the variance of the measurement error. In our simplest model there are only two unknowns. The initial-period measurement error, and the rate at which the measurement error in subsequent surveys grows or declines. We simply choose the combination of these two unknowns that best fits the predicted to the observed variance of revisions, for all possible revisions. Having done that we can construct estimates of the measurement error surrounding any variable, and for any vintage of that variable.

That information can then be used in some optimal forecasting or optimal policy procedure. Armed with our proposed technique, we apply it to some data to illustrate that it can work. This involves first estimating the time-variation in measurement error, and second applying this information to an optimal forecasting problem. We use the UK real-time data set of Castle and Ellis (2002). We compare results for the real growth of consumption and imports expenditure, and find that the first release for imports is about six times worse measured than that for consumption.

We illustrate how the outputs from our procedure can be used, showing how a univariate forecasting model for UK consumption growth can be improved by using the estimated time-variation in the measurement error of UK consumption growth vintages. We estimate a univariate forecasting model for consumption growth, and improve on it, given estimated information about time-varation in data uncertainty coming from our statistics agency model.

Of course, there are many series that are never revised. In the United Kingdom, the retail prices index is one example. The mere fact that series do not get revised is not an indication that those series are measured perfectly. Our method is powerless to uncover anything about measurement error in these cases.

2 A benchmark model of the statistics agency

In this section, we set out a model of a statistics agency, and use it to derive a relationship between quantities that we observe - the variance of revisions - and quantities that we do not observe - the variance of measurement error. We use two variants of our model. The first is a model of a 'rational' agency, one that weights together samples optimally given information about the sample size. The second is a model of a 'naive' agency, one that uses equal weights regardless.

2.1 A model of a rational statistics agency

It is assumed that the statistics agency faces a sequence of problems over time. In the first period, a survey is conducted and an estimate published. In the second period, new information, equivalent to a second survey of the same population, comes in, and the statistics agency has to weight the first and second surveys together to form a combined estimate. The statistics agency never observes the true value of anything (and neither do we as econometricians). But it starts out by trying to minimise the expected variance of its estimates around the true value.

At time t, the agency simply publishes the results of the first survey, on a variable dated t which we will call y. At time $t+1$ the problem is to minimise the expected variance of a weighted average of the first two releases around the true value, $y_{t|T}$:

At time $t+1$ the statistics agency's problem is:

$$\text{Min}E\left[\lambda_{1|1}y_{t|t+1}+(1-\lambda_{1|1})y_{t|t}-y_{t|T}\right]^2. \qquad \textbf{(1)}$$

$\lambda_{1|1}$ should be read as follows: the second subscript indicates which survey release is being weighted; the first subscript indicates when it is being weighted.

At time $t+2$ the agency has to choose weights to minimise the variance of the weighted sum of the first three releases around the true value.

$$\text{Min}E\left[\lambda_{2|0}y_{t|t}+\lambda_{2|1}y_{t|t+1}+\lambda_{2|2}y_{t|t+2}-y_{t|T}\right]^2 \tag{2}$$

The problem for the agency at some time $t+n$ can therefore be written as follows:

$$\text{Min}E\left[\sum_{k=0}^{n}\lambda_{n|k}y_{t|t+k}-y_{t|T}\right]^2, s.t\sum_{k=0}^{n}\lambda_{n|k}=1. \tag{3}$$

This is a stylised assumption to make about the process that generates data revisions. It is illustrative only. We need only to make an assumption about how the statistics agency weights new information. This could be based on solving optimal signal extraction problems like the one set out here. Alternatively, it could be based on assuming the agency follows a rule of thumb. We use the 'optimising' agency as a benchmark, but the general method does not depend on it. Note that, although taken literally our model assumes that all the agency does is conduct surveys, the model's usefulness will depend on whether the more complicated data collection and revision process (involving judgments, corroboration, forecasting, filtering) can be taken as if it behaved like a sequence of independent surveys. The assumption of independence is not overly restrictive as it relates to incremental information which under any concept of rationality is bound to be uncorrelated with (under normality, independent of) past information.

We proceed as follows: first, we solve for the weights that the statistics agency will place on new information over time. Second, we use this to solve for expressions that link the variance of data revisions with the variance of the underlying measurement error, and the rate at which this measurement error decays over time.

It is important to stress that our model is powerless to say anything about measurement error if there are no published revisions to a series. There are many such series that are used for forecasting and monetary policy. The fact that they are not revised is consistent with them being perfectly estimated, but more likely it just indicates that, for whatever reason, no new information is collected that sheds light on previously released data.

We assume that a survey on a data observation is a function of the true value and some measurement error, thus:

$$y_{t|t+k} = y_{t|T} + \upsilon_{t|t+k}. \quad \textbf{(4)}$$

$\upsilon_{t|t+k}$ is the measurement error contained in a survey-based estimate of y_t carried out at some time $t + k$.

We can then write the statistics agency's problem thus:

$$\text{Min} E\left[\sum_{k=0}^{n} \lambda_{n|k}(y_{t|T} + \upsilon_{t|t+k}) - y_{t|T}\right]^2, s.t \sum_{k=0}^{n} \lambda_{n|k} = 1. \quad \textbf{(5)}$$

Since the weights sum to 1, the true ys cancel out. This gives:

$$\text{Min} E\left[\sum_{k=0}^{n} \lambda_{n|k}\upsilon_{t|t+k}\right]^2. \quad \textbf{(6)}$$

We can express this in terms of the variance of the measurement error as follows:

$$\text{Min} \sum_{k=0}^{n} (\lambda_{n|k})^2 \sigma^2_{\upsilon|k}. \quad \textbf{(7)}$$

For the sake of notational simplicity, $\sigma^2_{\upsilon|k} = var(\upsilon_{t|t+k})$. Assuming that each successive survey can be thought of as an independent draw does not imply that the published data are independent. However, it implies that each incremental survey from which the agency forms those published estimates is independent.

The next important assumption we make is that the variance of the measurement error around successive surveys changes at a fixed rate. In terms of our 'survey' metaphor, this means assuming that, each period, the incremental surveys get smaller and smaller, and at a fixed rate. To be more concrete, we assume the following:

$$\sigma^2_{\upsilon|k} = (1+i)^k \sigma^2_{\upsilon|0}. \qquad \textbf{(8)}$$

If i were zero, each incremental survey estimate would have the same variance about the truth, and the statistics agency would weight them equally. Nothing is assumed about i at this stage. It could be positive or negative. Data on the variance of revisions will be used to estimate i. Intuitively, and as it turns out from our (fixed i) estimates, i is some positive, constant number. A positive i leads to revisions getting smaller and smaller with each successive vintage, which is a broad (but not universal) feature of national accounts data.(2) If i is positive, each incremental survey will be a more noisy estimate of the truth than the preceding one. But each published release, which will include more and more surveys, will be a *less* noisy estimate than the preceding one. This is not to say that the statistics agency somehow becomes worse at estimating data over time: it is a way of expressing, in model terms, the fact that the incremental flow of information gradually dries up. If it did not, the size of revisions would never tail off.

Conceptually, it is straightforward to allow for a degree of decay i that varies over k, or to test whether it does or does not. This would have the advantage that it must be more realistic. Certainly we know from the way the national accounts are put together that there are times when revisions are systematically larger (for example, around *Blue Book* publication dates). But allowing for a variable i involves a trade-off, as we shall see. The more i varies over k, the fewer observations there are to estimate it. So although a time-varying i model would be more realistic, estimates of it would be more imprecise. It is also possible to experiment with more complex functional forms that describe the decay of the measurement error than the one here. But for the moment we illustrate the basic approach with a constant i and equation **(8)**.

A literal reading of our model is that the only discrepancies that arise between surveys, published data and the truth are those that come from the surveys being conducted using samples smaller than the population. There are no additional 'data quality' issues that we model. The real-life data-generating process includes the task of detecting and correcting errors in completing or processing surveys, errors of interpretation, and much more. In many cases, variables measured by statisticians will necessarily only imperfectly correspond to the economic concepts that inform

(2) See Castle and Ellis (2002) and Harrison *et al.* (2004) for a discussion. Notable exceptions are revisions that coincide with *Blue Book* publication dates, which can cause the variance of revisions at some time t to be greater than that at time $t-1$.

policy. To the extent that any of these errors have a component that is random and independent, our model has something to say about them. The less this is the case, the more we must take our model to be abstracting from such errors.[3]

The minimand for the statistics agency can now be written thus:

$$\text{Min} \sum_{k=0}^{n} (\lambda_{n|k})^2 (1+i) \sigma^2_{\upsilon|0}. \tag{9}$$

The first-order conditions (FOCs) for this problem are as follows:

$$\lambda_{n|j} - (1 - \lambda_{n|0} - \lambda_{n|1} - \lambda_{n|2} - ... - \lambda_{n|n-1})(1+i)^n = 0, \quad j = 0, \dots, n-1. \tag{10}$$

From the FOCs, the weights are related by the following equation:

$$\lambda_{n|0} = \lambda_{n|1}(i+1) = ... = \lambda_{n|n-1}(i+1)^{n-1}. \tag{11}$$

An expression for $\lambda_{n|0}$ is therefore as follows:

$$\lambda_{n|0} = \frac{i(1+i)^n}{(1+i)^{n+1} - 1}. \tag{12}$$

From this we can deduce that the general expression for the optimal weight on a survey at some point is thus:

$$\lambda_{n|j} = \frac{i(1+i)^{n-j}}{(1+i)^{n+1} - 1}, \; j = 0, ...n. \tag{13}$$

This expression implies (assuming a positive value of i), that later surveys receive smaller weights. This is intuitive: later surveys, under a positive i, are subject to larger measurement error.

(3) One feature of these other errors that we do not model is that they may be irreducible. Our survey errors are in principle reducible by conducting an infinite number of surveys. Some kinds of error (say, human error in data collection) may not be reducible.

Having solved for the weights the statistics agency places on successive surveys, the next step is to write down an expression for the revision between any two periods, say period n and l. This is given by the following:

$$
\begin{aligned}
Ry_{t|n,l} &= y_{t|P,t+n} - y_{t|P,t+l} && \textbf{(14)} \\
&= (\lambda_{n|0} - \lambda_{l|0})y_{t|t} + (\lambda_{n|1} - \lambda_{l|1})y_{t|t+1} + (\lambda_{n|2} - \lambda_{l|2})y_{t|t+2} + \\
&\quad ... + (\lambda_{n|l} - \lambda_{l|l})y_{t|t+l} + \lambda_{n|l+1}y_{t|t+l+1} + ... + \lambda_{n|n}y_{t|t+n} && \textbf{(15)} \\
&= \sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})y_{t|t+k} + \sum_{k=l+1}^{n}\lambda_{n|k}y_{t|t+k} && \textbf{(16)}
\end{aligned}
$$

We can get an expression for the variance of revisions by substituting in **(4)**, the relation between the observation and the true data. This gives the following:

$$\sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})(y_{t|T} + v_{t|t+k}) + \sum_{k=l+1}^{n}\lambda_{n|k}(y_{t|T} + v_{t|t+k}). \qquad \textbf{(17)}$$

The weights on individual surveys in any vintage release sum to one, which implies that the terms in the true value of y cancel, and that we can therefore write that the revision between any two dates is given by the following equation:

$$Ry_{t|n,l} = \sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})v_{t|t+k} + \sum_{k=l+1}^{n}\lambda_{n|k}v_{t|t+k}. \qquad \textbf{(18)}$$

Setting all covariance terms to zero, we write the variance of a revision between any two dates thus:

$$var(Ry_{t|n,l}) = \sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})^2\sigma^2_{v|k} + \sum_{k=l+1}^{n}(\lambda_{n|k})^2\sigma^2_{v|k}. \qquad \textbf{(19)}$$

Substituting in **(8)**, we get that this revision variance is as follows:

$$var(Ry_{t|n,l}) = \sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})^2(1+i)^k\sigma^2_{v|0} + \sum_{k=l+1}^{n}(\lambda_{n|k})^2(1+i)^k\sigma^2_{v|0}. \qquad \textbf{(20)}$$

We can use **(11)** to get:

$$var(Ry_{t|n,l}) = \sigma_{\upsilon}^{2}\left[\sum_{k=0}^{l}(\lambda_{n|k}-\lambda_{l|k})^{2}(1+i)^{k}+\sum_{k=l+1}^{n}(\lambda_{n|0})^{2}(1+i)^{-k}\right]. \quad \textbf{(21)}$$

Finally, we can expand these summations through some long-winded but basic algebra to get the following expression:

$$\begin{aligned} & var(Ry_{t|n,l}) \quad \textbf{(22)} \\ = \; & \sigma_{\upsilon|0}^{2}\left[\frac{\left((1+i)^{n-l}-1\right)\left(i(1+i)^{l}\right)}{\left((1+i)^{n+1}-1\right)^{2}}\right]\left[\frac{(1+i)^{n-l}-1}{(1+i)^{l+1}-1}+(1+i)^{n-l}\right] \end{aligned}$$

We can now form systems of equations in the variance of revisions (which we observe), the decay parameter i and the variance of the measurement error $\sigma_{\upsilon|0}^{2}$. For example, given values for revisions over two different periods (eg the variance of six and ten-period revisions), we could form two equations in our two unknowns, and solve. In fact, the data may allow us to collect many observations on the variance of revisions at different periods, and this will allow us to construct an estimator for $\sigma_{\upsilon|0}^{2}$ and i. This is simply a standard GMM estimator. Formally, we solve the following problem:

$$Min\sum_{n}\sum_{l}(var(Ry_{t|n,l}))-var(R^{*}y_{t|n,l}))^{2}. \quad \textbf{(23)}$$

The star denotes the variance estimated from the data. Once we have estimates of $\sigma_{\upsilon|0}^{2}, i$, we can get an expression for the variance of the measurement error in any published release. This variance is given by the following expression:

$$var(y_{t|P,t+n}-y_{t|T}) = \sigma_{\upsilon|0}^{2}\sum_{k=0}^{n}(\lambda_{n|0})^{2}(1+i)^{-k}. \quad \textbf{(24)}$$

We can expand the summation term on the right-hand side, to yield the following:

$$var(y_{t|P,t+n} - y_{t|T}) = \sigma^2_{\upsilon|0} \frac{i(1+i)^n}{(1+i)^{n+1}-1}. \tag{25}$$

2.2 A model of a naive or rule-of-thumb statistics agency

So far it is assumed that the statistics agency solves an optimisation problem when it chooses how to weight the incremental surveys together. We turn next to a model where we instead assume that incremental surveys are weighted equally. We do this for three reasons. First, we want to demonstrate that our method does not depend on the 'rational' agency assumption: it just relies on making *some* behavioural assumption. Second, it is plausible that the 'rational' agency model is not the best one to capture the real world. And this might not be because actual agencies are not rational, but because they may solve more complicated problems than the one we have given them here. Third, we want to take our method to the data, and we want some way of exploring how robust the estimates from this method are to choosing alternatives to our basic behavioural model.

Assuming that the statistics agency weights surveys incrementally implies that a revision to a data release between any two periods l and n will be given by:

$$\begin{aligned} R_t y_{t|t+n,t+l} &= y_{t|P,t+n} - y_{t|P,t+l} && (26)\\ &= (\lambda_{t+n} - \lambda_{t+l})(\sum_{k=0}^{n-1} y_{t|t+k}) + \lambda_{t+n}(\sum_{k=l+1}^{n} y_{t|t+k}) && (27)\\ &= \left(\frac{1}{n+1} - \frac{1}{l+1}\right)\left(\sum_{k=0}^{n-1} y_{t|t+k}\right) + \frac{1}{n+1}(\sum_{k=l+1}^{n} y_{t|t+k}). && (28) \end{aligned}$$

We can then show that the variance of revisions is given by:

$$var(R_t y_{t|t+n,t+l}) = (\frac{1}{n+1} - \frac{1}{l+1})^2(\sum_{k=0}^{n-1} \sigma^2_{\upsilon|k}) + (\frac{1}{n+1})^2(\sum_{k=l+1}^{n} \sigma^2_{\upsilon|k}). \quad \textbf{(29)}$$

Substituting in our familiar expression for the relationship between the variance of measurement error around successive surveys, and expanding these summation terms, we can show that the variance of revisions is given by:

$$\begin{aligned} var(R_t y_{t|t+n,t+l}) &= \sigma^2_{\upsilon|0}[(\frac{1}{n+1} - \frac{1}{l+1})^2(\frac{(1+i)^{l+1}-1}{i}) \\ &+(\frac{1}{n+1})^2(\frac{(1+i)^{n+1}-(1+i)^{l+1}}{i})]. \end{aligned} \quad \textbf{(30)}$$

Note that this expression differs from its counterpart in the model of the optimising statistics agency in the previous section, contained in equation **(22)**. Armed with this expression, we go to the data and once again solve the estimation problem **(23)**.

The new expression for the variance of the measurement error in data releases will be different. It is given by the following:

$$var(y_{t|P,t+n} - y_{t|T}) = \sigma^2_{\upsilon|0} \frac{(1+i)^{n+1}-1}{i(n+1)^2}. \quad \textbf{(31)}$$

This, by inspection, is different from our earlier expression under an optimising statistics agency, given in equation (**(25)**).

3 Estimating the term structure of measurement error in a real-time data set

By way of an illustration, we next take the model to some data. We use the real-time data set compiled by Castle and Ellis (2002) based on data published by the United Kingdom's Office for National Statistics. The data set is described in more detail in that article. We will estimate the measurement error contained in the initial release, $\sigma^2_{\upsilon|0}$ and the rate of decay of that measurement error i for real (ie constant-price) growth in private consumption and imports expenditure. The real-time data we use cover releases between 1985 and 2001. For each observation on a series, we have typically around 45 releases. (There are roughly two releases per quarter over this period, although the exact frequency of releases has changed from time to time.) So we can compile a set of variances that records the average variance of revisions between the first and the second release, between the first and the third, between the first and the fourth, and so on. We compute these averages over observations.

Table A shows our estimates for the measurement error and the rates of decay in the two series. The estimates tell us that the variance of the measurement error in the first release of the growth of imports is a little under six times that of consumption growth. And the information flow falls off faster (i is higher) for consumption than for imports. To interpret the magnitude of these measurement errors, suppose that the steady-state consumption growth rate is about 2.5 % a year, or about 0.6 % a quarter (0.006 in the units in the table). That means that the variance of the estimate of the growth rate is is a little over 1/100th of the average growth rate itself; and the standard deviation of the growth rate (about 0.008, or 0.8 %) is therefore roughly of the same order of magnitude as the growth rate itself. Recalling our previous discussion, note that i is estimated to be positive, implying that the predicted variance of revisions tails off from vintage to vintage.

Table A: Initial release measurement errors, and rates of decay

growth in:	$\boldsymbol{\sigma^2_{\upsilon\|0}}$	**i**
imports	4.0E-04	1.1E-06
consumption	7.1E-05	1.1E-02

In the charts that follow, we use the estimated $\sigma^2_{\upsilon|0}$ and i to compute what the model says about the measurement error in different releases of the two series. Chart 1 shows the data for consumption; Chart 2 shows the data for imports.

The measurement error (on the y-axis) shrinks, of course, as we move to later releases (along the x-axis). The charts compare estimates derived by assuming a rational agency with those when we assume a ‘naive’one in the sense set out earlier in the paper. The shape of the curves for

Chart 1: Measurement error in consumption growth

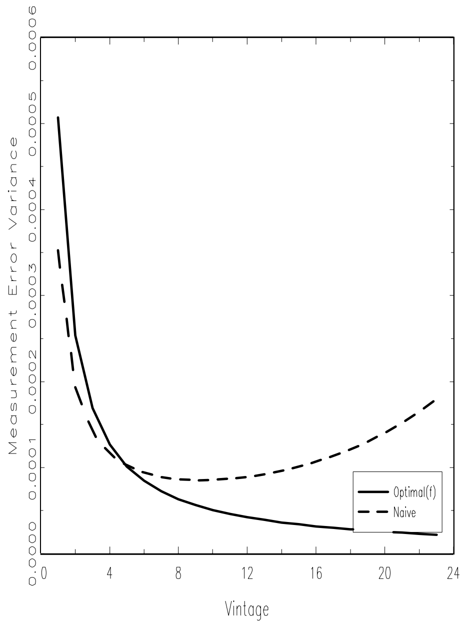
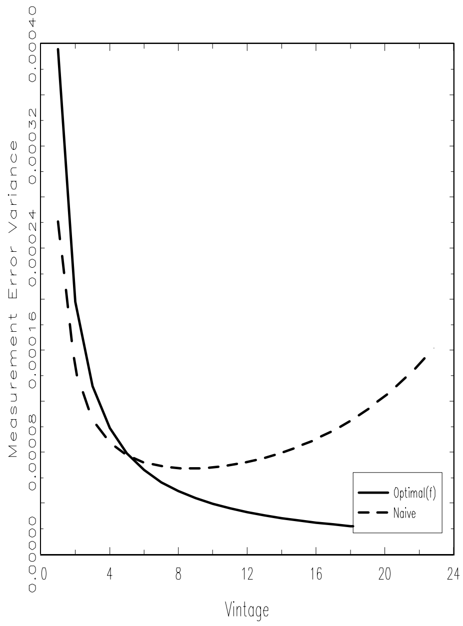


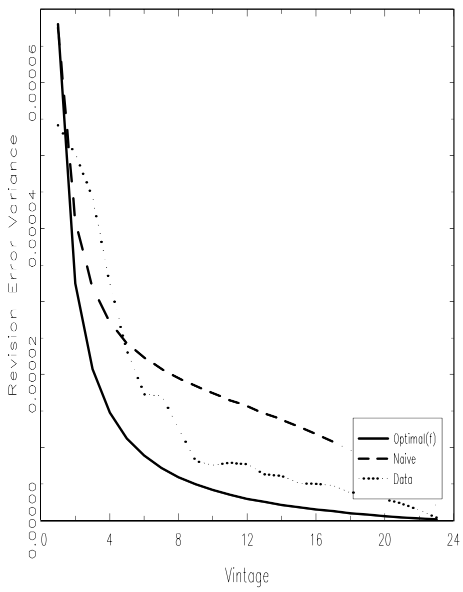
Chart 2: Measurement error in the growth of imports



consumption and imports is the same - that is because we are using the functional forms from the same model in each case. The differences in the two series are apparent from the different y-axis scales. The naive estimates differ in ways that have some intuition. Note that, for the first few releases, the estimates of the measurement error in the two series are pretty close. For that period at least our method is in some measure robust to polar assumptions about the behaviour of the statistics agency. But the naive estimates are lower. This seems counter-intuitive but is not. The model looks at the variance of revisions between two early releases. For the naive agency, it assumes that some of the variance of revisions is due to poor weighting, implying that the underlying sampling error in the early surveys is lower than in the case when it tries to fit an optimal agency through that same observed variance of revisions. Further out, the measurement error for the naive estimates increases exponentially. This is because later surveys, whose measurement error is growing exponentially, are weighted equally. At infinity, the variance of the published estimate for a naive agency will tend to infinity.

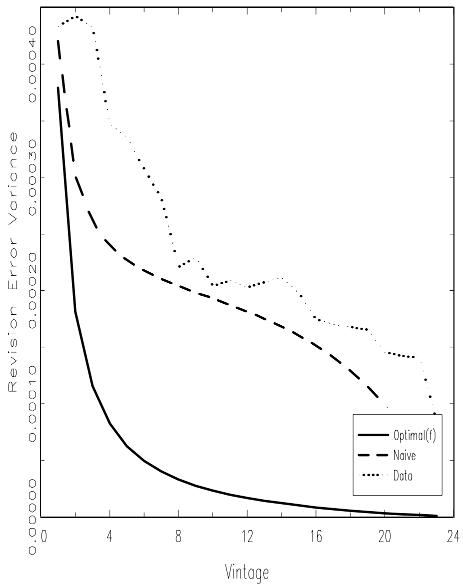
We can get some indication of how well the model fits the data by displaying charts of some model-predicted variance of revisions against actual data. Charts 3 and 4 do this.

Chart 3: The variance of consumption growth revisions, predicted and actual



These are not the only revisions to which the model is fitted. The model is fitted solving the minimisation problem set out above. That involves looking at revisions between all possible pairs of releases. The chart plots the variance of revisions between some release and the final release on

Chart 4: The variance of revisions to imports, predicted and actual



the x-axis. But these fits tell us something. Crudely, the optimal agency model seems to do better, though only just, at capturing the slope of revisions to consumption growth; but the naive agency does better for imports.

4 A model with a variable rate of decay

4.1 Theory

Up to this point, we have assumed a fixed rate of decay i, the quantity that determines the size of the sample of next period's incremental survey relative to this period's sample. We now want to relax this assumption and allow the data to determine how i evolves over time. The flow of information may not decline at a fixed rate over time in reality. This is a straightforward extension of the model, though the algebra becomes a little involved. An appendix presents the derivation in full, but here we state the key points of departure only.

The relationship between the measurement error of one survey relative to another, the counterpart to equation **(8)** in the fixed-decay model, is now given by:

$$\sigma^2_{\upsilon|k} = \prod_{j=0}^{k}(1+i_j)\sigma^2_{\upsilon|0}, k = 0. \quad \textbf{(32)}$$

To simplify notation, henceforth we write $\prod_{j=l}^{k}(1+i_j) \equiv \eta_{k,l}$.We can now write out the statistics agency's minimisation problem, given by **(7)** for the fixed-decay model, in terms of the first period's survey measurement error, thus:

$$\text{Min} \sum_{k=0}^{n}(\lambda_{n|k})^2 \eta_{k,1}\sigma^2_{\upsilon|0}. \quad \textbf{(33)}$$

After some straightforward algebra, we get the following expression for the relationship between our observables (the variance of revisions) and our unknowns (the i_js and $\sigma^2_{\upsilon|0}$):

$$var(Ry_{t|n,l}) = \sigma^2_{\upsilon|0} \frac{\eta^2_{n,i}\left(\sum_{k=l+1}^{n}\eta^{-1}_{k,1}\right)\left(1+\sum_{i=0}^{l-1}\eta_{l,l-i}\right)^2 + \left(1+\sum_{i=0}^{n-l-2}\eta_{n,n-i}\right)^2 \eta_{l,1}\sum_{k=0}^{l}\eta_{l,k+1}}{\left(1+\sum_{i=0}^{n-1}\eta_{n,n-i}\right)^2\left(1+\sum_{i=0}^{l-1}\eta_{l,l-i}\right)^2}. \quad \textbf{(34)}$$

And, in the same way as before, we note that this gives us a system of equations in our observables and unknowns, and find the optimal choice of the i_js and $\sigma^2_{\upsilon|0}$ that minimises an objective function involving the observed and predicted variance of revisions. We turn next to apply this model to the data.

4.2 Time-varying decay: an application

Allowing for time-varying is makes the model more realistic, but there is a cost: we will have more parameters to estimate, and will inevitably uncover estimates that are correspondingly less precise. The modelling framework allows us to fix some of the is to be equal if we want to. To speed up the numerical maximisation, we decided to fix the is beyond the 15th release.

Charts 5 and 6 show, for consumption growth, the implied model-based estimates of the variance of measurement error around different releases, and a plot of the fit of the model-based revisions to data on actual revisions.

The time-varying model based estimate of the variance of measurement error in the first release of consumption growth (σ_{v}^{2}) is $7.5 * 10^{-5}$, not much different from the time-invariant case ($7.1 * 10^{-5}$). But the is, the implied rates of decay of the incremental sample sizes are very different, both from the time-invariant case, and from release to release. Starting from the decay between release 1 and 2, and moving on, the first few of these is are given by $\{3.3, 4.8 * 10^{-4}, 4.8 * 10^{-3}, 7.5 * 10^{-5}, 1.4 * 10^{-3}....\}$, which compares with the fixed i estimate of $1.1 * 10^{-2}$. These figures generate substantially different estimates of the variance of measurement error for consumption growth, differences that get larger as we move through to later releases.

The charts plotting how the models look against (some of) the data on the variance of revisions indicate that the time-varying i model does a little better at describing the data.

5 Data uncertainty and optimal forecasting

We move on now to provide an illustration of how the estimates based on models like ours could be used for practical purposes. We choose a forecasting example that derives from some earlier work of ours. In Section 4 of Harrison *et al.* (2004), we presented an example of how to compute optimal forecasts in a dynamic model subject to the constraint that the forecast only uses data as old as the longest lag in the forecasting equation. We will describe that procedure briefly here. Readers interested in a fuller description should go back to the original paper. Note that the outputs of the procedure described above would have many more general uses: as an input into Kalman-Filter-based forecasts, or models of optimal monetary policy under uncertainty. We present a particular application merely to illustrate that our estimates of the variation in data uncertainty across vintages can make a material difference in a real setting.

The basic set-up is in the context of a univariate model for some true data, where we denote true data y_t^* using asterisks. The univariate model for the true variable is given by:

Chart 5: Model estimates of measurement error in consumption growth

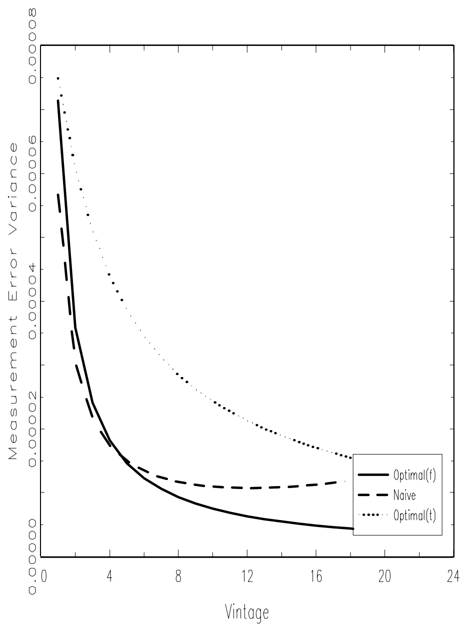
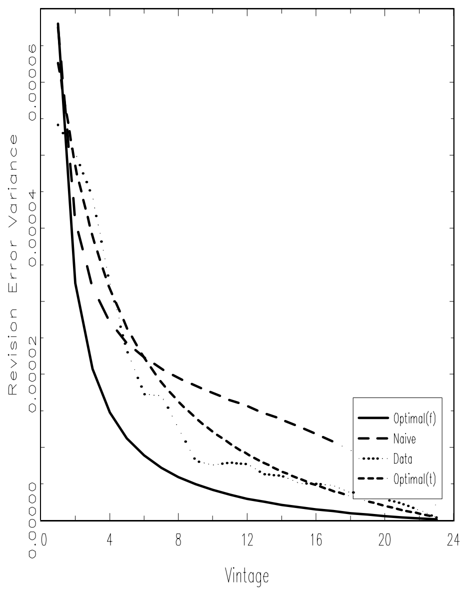


Chart 6: Variance of consumption revisions, actual and alternative predicted values



$$y_t^* = \sum_{i=1}^{p} a_i y_{t-i}^* + e_t. \qquad \textbf{(35)}$$

e_t is a white noise shock and the a_is are coefficients. We write the measurement model as:

$$\mathbf{y}_t = \mathbf{y}_t^* + \mathbf{v}_t, (\mathbf{y}_t = \{y_t, y_{t-1}, ...y_{t-p+1}\}). \qquad \textbf{(36)}$$

As before, υ_t is the white noise measurement error. The problem is to minimise the one-step-ahead forecast error:

$$\mathbf{y}_{t+1}^* - \widehat{\mathbf{y}}_{t+1} = \mathbf{A}_1\mathbf{y}_t^* + \boldsymbol{\epsilon}_t - \tilde{\mathbf{A}}_1\mathbf{y}_t^* + \tilde{\mathbf{A}}_1\mathbf{v}_t = (\mathbf{A}_1 - \tilde{\mathbf{A}}_1)\mathbf{y}_t^* + \tilde{\mathbf{A}}_1\mathbf{v}_t + \boldsymbol{\epsilon}_{t+1}. \qquad \textbf{(37)}$$

The hat superscript denotes the forecast. Here, the $\mathbf{A}_1$ is a vector of the a_is used to compute the one-step-ahead forecast (hence the subscript '1'). The choice variable in this minimisation problem is the matrix of coefficents $\tilde{\mathbf{A}}_1$. As we showed in Harrison *et al.* (2004), and state briefly here, the optimal choice for $\tilde{\mathbf{A}}_1$ involves weighting a variable according to its signal, and according to the measurement error (hence it takes the estimates of this from our model as an input).

The mean squared error is written as:

$$(\mathbf{A}_1 - \tilde{\mathbf{A}}_1)\mathbf{o}(\mathbf{A}_1 - \tilde{\mathbf{A}}_1)' + \tilde{\mathbf{A}}_1\mathbf{6}_\upsilon^T\tilde{\mathbf{A}}_1' + \sigma_\epsilon^2. \qquad \textbf{(38)}$$

Where $\mathbf{o} = E(\mathbf{y}_t^*\mathbf{y}_t^{*'})$, and the elements of this we can draw from $\sigma_\epsilon^2[\mathbf{I}_{p^2} - \mathbf{A} \otimes \mathbf{A}]^{-1}$.

Importantly, we assume that the covariances of the signal ϵ and the noise υ are assumed to be zero.

Differentiating the expression for the mean squared forecast error with respect to $\tilde{\mathbf{A}}_1$, and setting equal to zero, we get the following:

$$\tilde{\mathbf{A}}_1^{opt\,\prime} = (\mathbf{o} + \mathbf{6}_{\upsilon}^{T})^{-1}\mathbf{o}\mathbf{A}_1'. \qquad \textbf{(39)}$$

Note that the greater the measurement error surrounding a particular vintage, the lower the implied corresponding element in $\tilde{\mathbf{A}}_1^{opt\,\prime}$. Or, in short, the more noise in a variable, the less weight it has in an optimal forecast.

We apply this procedure to a univariate model for the quarterly growth rate of UK private consumption. The two tables below set out the results.

Table B: Whole-period coefficients in AR forecasting models for UK consumption growth: standard and uncertainty corrected

	AR(1)	AR(2)	AR(3)	AR(4)
standard	$-0.063_{(0.123)}$			
	$-0.057_{(0.122)}$	$0.156_{(0.122)}$)		
	$-0.118_{(0.118)}$	$0.169_{(0.117)}$	$0.291_{(0.120)}$	
	$-0.073_{(0.124)}$	$0.196_{(0.119)}$	$0.272_{(0.120)}$	$-0.158_{(0.125)}$
optimal	$-0.059_{(0.114)}$			
	$-0.053_{(0.110)}$	$0.146_{(0.113)}$		
	$-0.100_{(0.101)}$	$0.153_{(0.106)}$	$0.267_{(0.111)}$	
	$-0.054_{(0.089)}$	$0.166_{(0.101)}$	$0.241_{(0.108)}$	$-0.150_{(0.113)}$

Table B shows the effect on the forecasting equations of carrying out the procedure. These are estimates on data from 1980-98. Most coefficients fall. The model is a model of de-meaned consumption growth, so this implies putting more weight on the mean. There is some slight tendency to put more weight on older data relative to newer data. For example, take the AR(3). The ratio of the AR(1) to the AR(3) coefficients in the standard model is about 1:2.5. In the uncertainty-adjusted case that ratio is 1:2.7.

Table C shows recursive out-of-sample Diebold-Mariano[4] forecast evaluation tests on our two forecasting models. The whole period refers to out-of-sample tests for 1988-98: the two

(4) See, for details, Diebold and Mariano (1995).

Table C: MSE ratios and Diebold-Mariano tests

	Whole period		First subperiod		Second subperiod	
Model	MSE Ratio	D-M Test	MSE Ratio	D-M Test	MSE Ratio	D-M Test
AR(1)	0.987	2.46*	0.987	1.73*	0.987	2.93*
AR(2)	0.974	2.48*	0.968	2.33*	0.989	1.57*
AR(3)	0.977	1.55	0.975	1.26	0.983	1.38*
AR(4)	0.965	1.73*	0.959	1.52	0.980	1.08*

* denotes significance at the 5% level.

subperiods divide that sample into two equal halves. The Diebold-Mariano test compares the adjusted and the unadjusted root mean squared errors to see whether the forecasts can be said to be statistically significantly different from one another. The test results show that many of them are. The ratios of the mean squared forecast errors are all less than 1, for all the models we considered, implying that the measurement error corrected forecasts are better. The Diebold-Mariano test statistics have a critical value of 1.96 at the 5% level. A majority of these in Table C are greater than that value.

However good our model of the statistics agency is, using the estimates from it has some pay-off here in forecasting performance.

6 Conclusions

Knowing how well one series is measured relative to another, or how much more reliable older, revised data are than more recent data is useful in many situations: estimation, forecasting and monetary policy making. We have presented a method for extracting estimates of measurement error from observations on the variance of revisions in a data series. This method involves a conjecture about how the reliability of the incremental information that a statistics agency obtains declines over time, and about how the agency weights the information together to form a new estimate of a data point. We chose to illustrate our method using an assumption that the measurement error in incremental surveys grows exponentially, to capture the idea that each period less and less new information arrives; and that this information is weighted optimally to form new estimates of the data.

But our method does not depend on this precise assumption. We showed that by deriving results for a variable rate of decay; and by assuming a 'naive' statistics agency that gives as much weight to later, less well measured surveys as to earlier, better ones. Applying our method to real-time data on quarterly growth rates of UK private consumption and imports, we get estimates that suggest that the measurement error in the growth of imports is almost six times larger than that for consumption growth. Finally, we used our estimates of measurement error in a simple forecasting excercise described in Harrison *et al.* (2004). Using AR models for the quarterly growth in private consumption, we showed how the out-of-sample forecasting performance of model-based, measurement-error-corrected forecasts significantly outperform forecasts from unadjusted OLS equations.

Appendix A: A rule-of-thumb statistics agency, measurement error, and data revisions

In this appendix, we derive the relationship between observed revisions and the unobserved parameters of the measurement error function, under the assumption that data observations are weighted equally, rather than optimally.

Suppose that measured and true variables are related thus:

$$y_{t|t+n} = y_{t|T} + \upsilon_{t|t+n}. \qquad \textbf{(A-1)}$$

The variance of the measurement error around successive surveys is given by:

$$\sigma^2_{\upsilon|k} = (1+i)^k \sigma^2_{\upsilon|0}. \qquad \textbf{(A-2)}$$

If the statistics agency weights observations equally, then successive releases will be given by:

$$t : y_{t|P,t} = y_{t|t} \qquad \textbf{(A-3)}$$

$$t+1 : y_{t|P,t+1} = \lambda_{t+1} y_{t|t} + \lambda_{t+1} y_{t|t+1}, \lambda_{t+1} = 1/2 \qquad \textbf{(A-4)}$$

$$t+2 : y_{t|P,t+2} = \lambda_{t+2} y_{t|t} + \lambda_{t+2} y_{t|t+1} \lambda_{t+2} y_{t|t+2}, \lambda_{t+2} = 1/3 \qquad \textbf{(A-5)}$$

$$t+n : y_{t|P,t+n} = \sum_{k=0}^{n} \lambda_{t+n} y_{t|t+k}, \lambda_{t+n} = \frac{1}{n+1}. \qquad \textbf{(A-6)}$$

One-period revisions to published data will then be given by:

$$
\begin{aligned}
R_t y_{t|t+1} &= y_{t|P,t+1} - y_{t|P,t} = (\lambda_{t+1} - 1) y_{t|t} + \lambda_{t+1} y_{t|t+1} \\
&= 1/2 y_{t|t} + 1/2 y_{t|t+1}. \qquad \textbf{(A-7)}
\end{aligned}
$$

More generally, an l-period revision will be given by:

$$
\begin{aligned}
R_t y_{t|t+n,t+l} &= y_{t|P,t+n} - y_{t|P,t+l} \\
&= (\lambda_{t+n} - \lambda_{t+l})(\sum_{k=0}^{n-1} y_{t|t+k}) + \lambda_{t+n}(\sum_{k=l+1}^{n} y_{t|t+k}) \\
&= (\frac{1}{n+1} - \frac{1}{l+1})(\sum_{k=0}^{n-1} y_{t|t+k}) + \frac{1}{n+1}(\sum_{k=l+1}^{n} y_{t|t+k}). \qquad \textbf{(A-8)}
\end{aligned}
$$

Substituting in the fact that the published data equal the true data plus the measured data, it turns out that the terms in the true value of y cancel out to give an expression solely in terms of the measurement error. This is shown below:

$$
\begin{aligned}
R_t y_{t|t+n,t+l} &= (\frac{1}{n+1} - \frac{1}{l+1})(l+1) y_{t|T} + \sum_{k=0}^{n-1} \upsilon_{t|t+k}) \\
&\quad + \frac{1}{n+1}(n-l) y_{t|T} + (\sum_{k=l+1}^{n} \upsilon_{t|t+k}) \\
&= (\frac{1}{n+1} - \frac{1}{l+1})(\sum_{k=0}^{n-1} \upsilon_{t|t+k}) + \frac{1}{n+1}(\sum_{k=l+1}^{n} \upsilon_{t|t+k}). \qquad \textbf{(A-9)}
\end{aligned}
$$

The variance of an l-period revision can be written as:

$$var(R_t y_{t|t+n,t+l}) = (\frac{1}{n+1} - \frac{1}{l+1})^2 (\sum_{k=0}^{n-1} \sigma^2_{v^k}) + (\frac{1}{n+1})^2 (\sum_{k=l+1}^{n} \sigma^2_{v^k}). \quad \textbf{(A-10)}$$

Substituting in our assumption about how measurement error in different surveys is related ($\sigma^2_{v|k} = (1+i)^k \sigma^2_{v|0}$) we find the following:

$$\begin{aligned} var(R_t y_{t|t+n,t+l}) &= (\frac{1}{n+1} - \frac{1}{l+1})^2 (\sum_{k=0}^{n-1} (1+i)^k \sigma^2_{v|0}) \\ &+ (\frac{1}{n+1})^2 (\sum_{k=l+1}^{n} (1+i)^k \sigma^2_{v|0}). \end{aligned} \quad \textbf{(A-11)}$$

This implies that:

$$var(R_t y_{t|t+n,t+l}) = \sigma^2_{v|0} [(\frac{1}{n+1} - \frac{1}{l+1})^2 (\sum_{k=0}^{n-1} (1+i)^k) + (\frac{1}{n+1})^2 (\sum_{k=l+1}^{n} (1+i)^k)]. \quad \textbf{(A-12)}$$

We can expand the summation terms as follows:

$$\sum_{k=0}^{n-1} (1+i)^k = \frac{(1+i)^{l+1} - 1}{i}, \sum_{k=l+1}^{n} (1+i)^k = \frac{(1+i)^{n+1} - (1+i)^{l+1}}{i}. \quad \textbf{(A-13)}$$

Doing so gives:

$$
\begin{aligned}
var(R_t y_{t|t+n,t+l}) &= \sigma^2_{\upsilon|0}[(\frac{1}{n+1} - \frac{1}{l+1})^2(\frac{(1+i)^{l+1}-1}{i}) \\
&\quad +(\frac{1}{n+1})^2(\frac{(1+i)^{n+1}-(1+i)^{l+1}}{i})]. \qquad \textbf{(A-14)}
\end{aligned}
$$

We can also get the variance of the published data around the true data, for some vintage. The expression is as follows:

$$
var(y_{t|P,t+n} - y_{t|T}) = \sigma^2_{\upsilon|0}\frac{(1+i)^{n+1}-1}{i(n+1)^2}. \qquad \textbf{(A-15)}
$$

Appendix B: A variable rate of decay model

Here we derive our results for the main paper under the assumption that the rate of decay of the size of the sample of incremental surveys changes over time. Recall that in the main paper this quantity, i, was assumed to be fixed. In particular, we will assume that the measurement error surrounding incremental surveys is given by the following expression:

$$\sigma^2_{\upsilon|k} = \prod_{j=0}^{k}(1+i_j)\sigma^2_{\upsilon|0}, k = 0 \tag{B-1}$$

To simplify notation we introduce $\prod_{j=l}^{k}(1+i_j) \equiv \eta_{k,l}$.We can now write out the statistics agency's minimisation problem in terms of the first period's survey measurement error, thus:

$$\text{Min} \sum_{k=0}^{n}(\lambda_{n|k})^2\eta_{k,1}\sigma^2_{\upsilon|0}. \tag{B-2}$$

We write the problem out as:

$$\begin{gathered}\lambda^2_{n|0}\sigma^2_{\upsilon|0} + \lambda^2_{n|1}\eta_{1,1}\sigma^2_{\upsilon|0} + \lambda^2_{n|2}\eta_{2,1}\sigma^2_{\upsilon|0} + \lambda^2_{n|3}\eta_{3,1}\sigma^2_{\upsilon|0} + ...+ \\ \lambda^2_{n|n-1}\eta_{n-1,1}\sigma^2_{\upsilon|0} + (1-\lambda_{n|0}-\lambda_{n|1}-\lambda_{n|2}-...-\lambda_{n|n-1})\eta_{n,1}\sigma^2_{\upsilon|0}.\end{gathered} \tag{B-3}$$

We drop $\sigma^2_{\upsilon|0}$ and so the minimand can be rewritten as:

$$\begin{aligned}&\lambda^2_{n|0} + \lambda^2_{n|1}\eta_{1,1} + \lambda^2_{n|2}\eta_{2,1} + \lambda^2_{n|3}\eta_{3,1} + ... \\ &+\lambda^2_{n|n-1}\eta_{n-1,1} + (1-\lambda_{n|0}-\lambda_{n|1}-\lambda_{n|2}-...-\lambda_{n|n-1})^2\eta_{n,1}.\end{aligned} \tag{B-4}$$

The FOCs for this problem are as follows:

$$\lambda_{n|j} - (1 - \lambda_{n|0} - \lambda_{n|1} - \lambda_{n|2} - ... - \lambda_{n|n-1})\eta_{n,1} = 0,\ j = 0, \ldots, n-1. \quad \textbf{(B-5)}$$

We note that the weights are related thus:

$$\lambda_{n|0} = \lambda_{n|1}\eta_{1,1} = ... = \lambda_{n|n-1}\eta_{n-1,1}. \quad \textbf{(B-6)}$$

So using (**(B-5)**) for $j = 0$ we arrive at:

$$\lambda_{n|0} - \left(1 - \lambda_{n|0} - \lambda_{n|0}/\eta_{1,1} - \lambda_{n|0}/\eta_{2,1} - ... - \lambda_{n|0}/\eta_{n-1,1}\right)\eta_{n,1} = 0. \quad \textbf{(B-7)}$$

This can be written thus:

$$\lambda_{n|0} - \left(\eta_{n,1} - \lambda_{n|0}\eta_{n,1} - \lambda_{n|0}\eta_{n,2} - \lambda_{n|0}\eta_{n,3} - ... - \lambda_{n|0}\eta_{n,n}\right) = 0. \quad \textbf{(B-8)}$$

Grouping the $\lambda_{n|0}$ gives:

$$\lambda_{n|0} = \frac{\eta_{n,1}}{1 + \eta_{n,n} + \eta_{n,n-1} + ... + \eta_{n,1}}. \quad \textbf{(B-9)}$$

We can therefore write the expression for the weight placed on any survey as:

$$\lambda_{n|j} = \frac{\eta_{n,j}}{1 + \eta_{n,n} + \eta_{n,n-1} + ... + \eta_{n,1}},\ j = 1, \ldots n. \quad \textbf{(B-10)}$$

Recall from the main body of the paper that the variance of revisions is given by:

$$\upsilon ar(Ry_{t|n,l}) = \sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})^2\sigma^2_{\upsilon|k} + \sum_{k=l+1}^{n}(\lambda_{n|k})^2\sigma^2_{\upsilon|k}. \tag{B-11}$$

Given the relationship between the $\sigma^2_{\upsilon|k}$'s, this is now, in the variable rate of decay case:

$$\upsilon ar(Ry_{t|n,l}) = \sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})^2\left(\eta_{k,1}\sigma^2_{\upsilon|0}\right) + \sum_{k=l+1}^{n}(\lambda_{n|k})^2\left(\eta_{k,1}\sigma^2_{\upsilon|0}\right). \tag{B-12}$$

Note that the weights are related as follows:

$$\lambda_{n|0} = \lambda_{n|1}\eta_{1,1} = ... = \lambda_{n|n-1}\eta_{n-1,1}. \tag{B-13}$$

We can therefore write:

$$Var(Ry_{t|n,l}) = \sigma^2_{\upsilon|0}\left[\sum_{k=0}^{l}(\lambda_{n|k} - \lambda_{l|k})^2\eta_{k,1} + \sum_{k=l+1}^{n}(\lambda_{n|0})^2\eta^{-1}_{k,1}\right]. \tag{B-14}$$

Expanding the summations in these expressions is laborious, but leads us to the following equation:

$$\upsilon ar(Ry_{t|n,l}) = \sigma^2_{\upsilon|0}\frac{\begin{array}{c}\eta^2_{n,i}\left(\sum_{k=l+1}^{n}\eta^{-1}_{k,1}\right)\left(1+\sum_{i=0}^{l-1}\eta_{l,l-i}\right)^2 \\ +\left(1+\sum_{i=0}^{n-l-2}\eta_{n,n-i}\right)^2\eta_{l,1}\sum_{k=0}^{l}\eta_{l,k+1}\end{array}}{\left(1+\sum_{i=0}^{n-1}\eta_{n,n-i}\right)^2\left(1+\sum_{i=0}^{l-1}\eta_{l,l-i}\right)^2} \tag{B-15}$$

This gives us our system of equations linking the observed variance of revisions to the unknown variable rates of decay i_j and the initial period's measurement error $\sigma^2_{\upsilon|0}$.

References

Aoki, K (2003), 'On the optimal monetary policy response to noisy indicators', *Journal of Monetary Economics*, pages 501–23.

Castle, J and Ellis, C (2002), 'Building a real-time database for GDP(E)', *Bank of England Quarterly Bulletin*, Vol. 42, No. 1, pages 42–49.

Coenen, G , Levin, A and Weiland, W (2001), 'Data uncertainty and the role of money as an information variable in monetary policy', *ECB Working Paper no 84*.

Diebold, F X and Mariano, R S (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics,* Vol. 13, No. 3, pages 253–62.

Harrison, R , Kapetanios, G and Yates, T (2004), 'Forecasting with measurement errors in dynamic models', *Bank of England Working Paper*.

Mankiw, N G, Runkle, D and Shapiro, M D (1984), 'Are preliminary announcements of the money stock rational forecasts?', *Journal of Monetary Economics*, Vol. 14, pages 15–27.

Sargent, T (1989), 'Two models of measurement and the investment accelerator', *Journal of Political Economy*, pages 251–87.

Swanson, E (2000), 'On signal extraction and non-certainty equivalence in optimal monetary policy rules', *Econometric Society World Congress*.

Woodford, M and Svensson, L (2003), 'Indicator variables for monetary policy', *Journal of Monetary Economics*, pages 691–20.

Bank of England working paper series

A complete list of working papers is available from the Bank's web site at www.bankofengland.co.uk/wp/index.html, where abstracts of all papers may be found. Papers published since January 1997 are available in full, in PDF.

Publication date in italics

204	The dynamics of consumers' expenditure: the UK consumption ECM redux *(October 2003)*	Emilio Fernandez-Corugedo Simon Price Andrew Blake
205	Empirical determinants of emerging market economies' sovereign bond spreads *(October 2003)*	Gianluigi Ferrucci
206	The rise in US household debt: assessing its causes and sustainability *(October 2003)*	Sebastian Barnes Garry Young
207	A quantitative framework for commercial property and its relationship to the analysis of the financial stability of the corporate sector *(October 2003)*	John Whitley Richard Windram
208	A matching model of non-employment and wage pressure *(November 2003)*	Andrew Brigden Jonathan Thomas
209	Settlement bank behaviour and throughput rules in an RTGS payment system with collateralised intraday credit *(November 2003)*	Simon Buckle Erin Campbell
210	Company accounts based modelling of business failures and the implications for financial stability *(December 2003)*	Philip Bunn Victoria Redwood
211	An empirical analysis of the dynamic relationship between investment-grade bonds and credit default swaps *(December 2003)*	Roberto Blanco Simon Brennan Ian W Marsh
212	Crisis spillovers in emerging market economies: interlinkages, vulnerabilities and investor behaviour *(January 2004)*	Michael Chui Simon Hall Ashley Taylor
213	Investment-specific technological change and growth accounting *(February 2004)*	Nicholas Oulton
214	An empirical model of household arrears *(March 2004)*	John Whitley Richard Windram Prudence Cox
215	How can the IMF catalyse private capital flows? A model *(April 2004)*	Adrian Penalver
216	IMF lending and creditor moral hazard *(April 2004)*	Andrew G Haldane Jörg Scheibe
217	International financial rescues and debtor-country moral hazard *(April 2004)*	Prasanna Gai Ashley Taylor
218	Health, disability insurance and labour force participation *(April 2004)*	Brian Bell James Smith
219	Bail out or work out? Theoretical considerations *(April 2004)*	Andrew G Haldane Gregor Irwin Victoria Saporta
220	Does job insecurity affect household consumption? *(May 2004)*	Andrew Benito

221	Female labour force participation in the United Kingdom: evolving characteristics or changing behaviour? *(May 2004)*	Maria Gutiérrez-Domènech Brian Bell
222	The roles of expected profitability, Tobin's Q and cash flow in econometric models of company investment *(June 2004)*	Stephen Bond Alexander Klemm Rain Newton-Smith Murtaza Syed Gertjan Vlieghe
223	Real exchange rates and the relative prices of non-traded and traded goods: an empirical analysis *(June 2004)*	Jan J J Groen Clare Lombardelli
224	The informational content of empirical measures of real interest rate and output gaps for the United Kingdom *(July 2004)*	Jens D J Larsen Jack McKeown
225	Exploring the relationship between credit spreads and default probabilities *(July 2004)*	Mark J Manning
226	Corporate capital structure in the United Kingdom: determinants and adjustment *(July 2004)*	Philip Bunn Garry Young
227	The Phillips curve under state-dependent pricing *(July 2004)*	Hasan Bakhshi Hashmat Khan Barbara Rudolf
228	The UK labour force participation rate: business cycle and trend influences *(August 2004)*	Mark Schweitzer David Tinsley
229	On the resolution of banking crises: theory and evidence *(August 2004)*	Glenn Hoggarth Jack Reidhill Peter Sinclair
230	Financial interlinkages in the United Kingdom's interbank market and the risk of contagion *(September 2004)*	Simon Wells
231	Real exchange rate persistence and systematic monetary policy behaviour *(September 2004)*	Jan J J Groen Akito Matsumoto
232	Evolving post-World War II UK economic performance *(September 2004)*	Luca Benati
233	The efficient resolution of capital account crises: how to avoid moral hazard *(September 2004)*	Gregor Irwin David Vines
234	Intertemporal substitution and household production in labour supply *(September 2004)*	Guillermo Felices David Tinsley
235	Rule-based monetary policy under central bank learning *(September 2004)*	Kosuke Aoki Kalin Nikolov
236	The effects of stock market movements on consumption and investment: does the shock matter? *(October 2004)*	Stephen Millard John Power
237	Forecasting with measurement errors in dynamic models *(October 2004)*	Richard Harrison George Kapetanios Tony Yates
238	Estimating time-variation in measurement error from data revisions; an application to forecasting in dynamic models *(October 2004)*	George Kapetanios Tony Yates