Density forecast revisions and forecast efficiency

James Mitchell*
National Institute of Economic and Social Research

February 4, 2008

Abstract

We explain that revisions to successive density forecasts of the same outcome, as measured by the Kullback-Leibler Information Criterion, need not be unpredictable, unlike those to conditional mean forecasts, even when the forecaster uses information efficiently. However one can still test the efficiency of fixed-event conditional density forecasts, similarly to conditional mean forecasts, by testing the independence of revisions to an event forecast extracted from the density forecast. In an application we thereby test the efficiency of the fixed-event density forecasts of U.S. inflation and GDP growth supplied by the Survey of Professional Forecasters.

*Address for correspondence: James Mitchell, National Institute of Economic and Social Research, 2 Dean Trench Street, Smith Square, London, SW1P 3HE, U.K. Tel: +44 (0) 207 654 1926. Fax: +44 (0) 207 654 1900. E-Mail: j.mitchell@niesr.ac.uk. Mitchell gratefully acknowledges financial support from the ESRC (Award Reference: RES-000-22-1512). Thanks to Stephen Hall, Martin Weale and seminar participants at Warwick for helpful conversations. Particular thanks to Ken Wallis for written comments.

1 Introduction

Despite the increased use of density forecasts in economics, since they provide a full impression of forecast uncertainty, focus has been on so-called "rolling" density forecasts, i.e. successive ($t = 1, ..., T$) h-step ahead forecasts with evaluation tests proposed by Diebold et al. (1998). These test the *ex post* performance of density forecasts as a 'whole'; in practice this involves using a goodness-of-fit test to establish whether the probability integral transforms of the forecast density with respect to the realisations ($t = 1, ..., T$) of the variable of interest are uniform or via a transformation normal. Scoring rules have also been employed (see Gneiting & Raftery (2007) for a review), although they test *relative* rather than *absolute* performance.

But in practice forecasters are often asked to make successive density forecasts of the same event, i.e. "fixed-event" forecasts, as well as series of forecasts of fixed length h. For example, in the Survey of Professional Forecasters [SPF] forecasters are asked at lags of one to eight quarters what, in probabilistic terms, they expect the outturn for annual U.S. inflation and output growth in a given year to be. There are therefore in total a sequence of eight forecasts of the same event. These indicate how the forecaster has changed their mind. However, little attention has been paid, at least explicitly, to the fixed-event aspect of density forecasts and what (if anything) can be inferred from revisions to successive density forecasts of the same event. This is despite the availability of tests for the rationality of fixed-event point forecasts (Nordhaus (1987) and Clements (1997)), the testable proposition (for weak efficiency) being that, under quadratic loss, forecast revisions should be uncorrelated with past forecast revisions.

In this paper we begin to remedy this shortcoming by showing, in Section 2, that one apparently obvious extension to density forecasts, using the Kullback-Leibler Information Criterion (KLIC) to measure revisions to successive densities, need not convey any information on forecast rationality. This precludes testing the weak efficiency of density forecasts simply, as with point forecasts, by testing the independence of KLIC revisions. But Section 3 notes that one can still test the rationality of fixed-event density forecasts, similarly to fixed-event point forecasts, by reducing them to a probability event forecast. This is based on the fact that if the density forecast, from which a probabilistic event forecast is extracted, is rational then all interval and probabilistic event forecasts must be well-calibrated also. On this basis, in Section 4, we test the efficiency of the fixed-event density forecasts of inflation and GDP growth supplied by the SPF and thereby provide exploratory evidence on predictability in different aspects of the forecast density to the traditional analysis focused on the conditional mean. Section 5 concludes.

2 Fixed-event forecasts and "news"

2.1 Revisions to point forecasts

"News" as traditionally defined with reference to point forecasts, or national accounts data (cf. Faust et al. (2005)), implies that the revision or difference between a forecast made at horizon h and a forecast made at horizon $(h+1)$, so-called "fixed-event" forecasts, reflects information unknown and unpredictable at horizon $(h+1)$. Under quadratic loss the revision between successive fixed-event point forecasts is news and is therefore unpredictable when the forecaster efficiently processes the information available to them at the time they made the forecast (see Nordhaus (1987)).

More formally, let $g(y_t \mid \Psi_{t-h})$ denote the forecast density of y_t $(t = 1, ..., T)$ made h $(h = 1, ..., H)$ periods ahead conditional on the information set Ψ_{t-h} available to the forecaster. Under squared error loss the (optimal) point forecast extracted from $g(y_t \mid \Psi_{t-h})$ will equal the conditional mathematical expectation $\mathsf{E}(y_t \mid \Psi_{t-h})$. Under other loss functions it need not (e.g. see Zellner (1986)). But in any case when $\mathsf{E}(y_t \mid \Psi_{t-h})$ is a rational expectation for the mean of $\{y_t\}$ it should equal the (true) conditional expectation $\mathsf{E}(y_t \mid \Omega_{t-h})$, where $\Omega_{t-h} = \{\Omega_{t-h-1}, \Omega_{t-h-2}, ..\}$ is all available information available at horizon $(t-h)$, where $\Psi_{t-h} \subseteq \Omega_{t-h}$.

The law of iterated expectations then implies

$$\mathsf{E}\left[\mathsf{E}(y_t \mid \Psi_{t-h}) \mid \Psi_{t-h-1}\right] = \mathsf{E}(y_t \mid \Psi_{t-h-1}) \Rightarrow \mathsf{E}\left[\mathsf{E}(y_t \mid \Psi_{t-h}) - \mathsf{E}(y_t \mid \Psi_{t-h-1}) \mid \Psi_{t-h-1}\right] = 0 \quad (1)$$

which means revisions to successive conditional mean forecasts are unpredictable using information known to the forecaster when the earlier forecast was made at horizon $(t-h-1)$. When the information set Ψ_{t-h-1} is assumed to consist only of previous or lagged revisions then (1) is classed a test for "weak" efficiency. Indeed, (1) then amounts to saying, using Nordhaus' (p. 673) oft-quoted line: "If I could look at your most recent forecast and accurately say, 'Your next forecast will be 2% lower than today's', then you can surely improve your forecast". Under these conditions forecast revisions are spiky not smooth.

Strong efficiency, or rationality, requires that the information set consists of all available information, including knowledge of the true conditional density. That is, under rationality the conditional density forecast $g\left(y_t \mid \Psi_{t-h}\right) = f(y_t \mid \Omega_{t-h})$, the true conditional density, and the forecasts are fully efficient with respect to the 'true' information set Ω_{t-h}; in other words the KLIC distance between the true, but in general unknown, conditional density and the forecast density equals zero.[1] This is precisely the null hypothesis tested when practitioners apply a goodness-of-fit test to establish whether a sequence of probability integral transforms, computed for a series of rolling density forecasts, are uniform and serially independent at orders greater than or equal to h; see Bao et al. (2007) and Mitchell & Hall (2005). That is, when the KLIC distance equals zero, and $f(y_t \mid \Omega_{t-h}) = g(y_t \mid \Psi_{t-h})$, the probability integral transforms $z_{t|t-h} = \int_{-\infty}^{y_t} g(u)du$ $(t = 1, ..., T)$ are uniformly distributed.

2.2 Revisions to density forecasts

To understand what can be inferred from the movement of and revisions to fixed-event density forecasts requires us to measure the revision between two successive density forecasts of the same event. A natural candidate is the KLIC revision. The KLIC is a measure of *divergence* between two densities. Lahiri & Liu (2006) also used the KLIC to measure the revision between successive density forecasts.

Define $KLIC_{t|t-h}$ as the revision between the $(h + 1)$ and h period ahead density forecasts of y_t:

$$KLIC_{t|t-h} = \int g(y_t \mid \Psi_{t-h-1}) \ln \left\{ \frac{g(y_t \mid \Psi_{t-h-1})}{g(y_t \mid \Psi_{t-h})} \right\} dy_t. \tag{2}$$

Analogously to the case for point forecasts, for the KLIC revision between $g(y_t \mid \Psi_{t-h})$ and $g(y_t \mid \Psi_{t-h-1})$ to constitute "news" it should be unpredictable when the forecaster forms rational forecasts and it should reflect information not known at horizon $(t-h-1)$.

But below we show, via a simple example, that we should not expect the KLIC revisions, $KLIC_{t|t-h}$, to be independent even when the density forecasts are rational. This precludes testing the predictability of KLIC revisions as a means of testing the weak efficiency of fixed-event density forecasts.

[1] Rationality implies a "one-to-one relationship" between the moments of $g\left(y_t \mid \Psi_{t-h}\right)$ and $f(y_t \mid \Omega_{t-h})$; see Pesaran (1987), p. 24.

2.2.1 The predictability of conditional variance forecasts

Define the h-step ahead mean and variance forecasts as "unpredictable" (cf. Clements & Hendry (1998), p.38) with respect to information available to the forecaster at horizon $(h+1)$ when[2]

$$\mathsf{E}\left[\mathsf{E}(y_t \mid \Psi_{t-h}) \mid \Psi_{t-h-1}\right] = \mathsf{E}(y_t \mid \Psi_{t-h-1}) \tag{3}$$

$$\mathsf{E}\left[\mathsf{V}(y_t \mid \Psi_{t-h}) \mid \Psi_{t-h-1})\right] = \mathsf{V}(y_t \mid \Psi_{t-h-1}) \tag{4}$$

For KLIC revisions to be independent we should expect both (3) and (4) to hold when the density forecasts are rational. But this will not usually be the case in economics. This is because the true, but in general unknown, conditional density varies at each horizon, and so, in general, we should not expect the sequence of KLIC revisions to be independent even under forecast rationality. To illustrate consider a simple example.

Via the Wold decomposition theorem any zero-mean covariance stationary process $\{y_t\}$ can be written in the form

$$y_t = \mu + \sum\nolimits_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{5}$$

where $\psi_0 = 1$, $\sum\nolimits_{j=0}^{\infty} \psi_j^2 < \infty$ and ε_t is white noise with variance σ^2.

Ignoring parameter uncertainty, Table 1 shows fixed-event conditional forecasts for both the mean and variance of (5) at forecast horizons of $0, 1, ..., \infty$ periods. These are rational density forecasts; the forecaster knows the 'true' underlying model and sets their subjective density forecasts of y_t equal to it such that $\mathsf{E}(y_t \mid \Psi_{t-h}) = \mathsf{E}(y_t \mid \Omega_{t-h})$ and $\mathsf{V}(y_t \mid \Psi_{t-h}) = \mathsf{V}(y_t \mid \Omega_{t-h})$. For convenience we set $\mu = 0$. Although the forecaster's optimal point forecast need not equal the conditional mean, as discussed, if the density forecast is rational then irrespective of the loss function the conditional mean forecast should equal the true mean.[3] Table 1 also shows the revision between successive mean and variance forecasts. The lower panel of Table 1, for the convenience of the reader, extracts the forecasts for an AR(1), a special case of (5) with stationarity imposing $|\phi| < 1$.

From Table 1 we can see that while the mean forecasts are unpredictable when formed

[2] We do not believe the present focus on the first two moments is restrictive. But future work, of course, might think of generalising the analysis to higher moments, if these exist. But our aim here is to demonstrate, for which our focus on the first two moments is sufficient, that revisions to successive density forecasts need not convey any information on forecast rationality.

[3] Indeed as discussed by Engelberg et al. (2006) and Clements (2007) there can be a divergence between the two.

Table 1: The evolution of fixed-event density forecasts for an MA(∞) and AR(1) model

$MA : y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$

Horizon (h)	$\mathsf{E}(y_t \mid \Psi_{t-h})$	$\mathsf{V}(y_t \mid \Psi_{t-h})$	$\mathsf{E}(y_t \mid \Psi_{t-h})-$ $\mathsf{E}(y_t \mid \Psi_{t-h-1})$	$\mathsf{V}(y_t \mid \Psi_{t-h})-$ $\mathsf{V}(y_t \mid \Psi_{t-h-1})$
0	y_t	0		
1	$\sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}$	σ^2	$\psi_1 \varepsilon_{t-1}$	$-\psi_1^2 \sigma^2$
2	$\sum_{i=2}^{\infty} \psi_i \varepsilon_{t-i}$	$(1+\psi_1^2)\sigma^2$	$\psi_2 \varepsilon_{t-2}$	$-\psi_2^2 \sigma^2$
3	$\sum_{i=3}^{\infty} \psi_i \varepsilon_{t-i}$	$(1+\psi_1^2+\psi_2^2)\sigma^2$	$\psi_3 \varepsilon_{t-3}$	$-\psi_3^2 \sigma^2$
4	$\sum_{i=4}^{\infty} \psi_i \varepsilon_{t-i}$	$(1+\psi_1^2+\psi_2^2+\psi_3^2)\sigma^2$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		
∞	0	$(1+\psi_1^2+\psi_2^2+...+\psi_\infty^2)\sigma^2$		

$AR(1) : y_t = \phi y_{t-1} + \varepsilon_t$

Horizon (h)	$\mathsf{E}(y_t \mid \Psi_{t-h})$	$\mathsf{V}(y_t \mid \Psi_{t-h})$	$\mathsf{E}(y_t \mid \Psi_{t-h})-$ $\mathsf{E}(y_t \mid \Psi_{t-h-1})$	$\mathsf{V}(y_t \mid \Psi_{t-h})-$ $\mathsf{V}(y_t \mid \Psi_{t-h-1})$
0	y_t	0		
1	ϕy_{t-1}	σ^2	$\phi \varepsilon_{t-1}$	$-\phi^2 \sigma^2$
2	$\phi^2 y_{t-2}$	$(1+\phi^2)\sigma^2$	$\phi^2 \varepsilon_{t-2}$	$-\phi^4 \sigma^2$
3	$\phi^3 y_{t-3}$	$(1+\phi^2+\phi^4)\sigma^2$	$\phi^3 \varepsilon_{t-3}$	$-\phi^6 \sigma^2$
4	$\phi^4 y_{t-4}$	$(1+\phi^2+\phi^4+\phi^6)\sigma^2$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		
∞	0	$\sigma^2/(1-\phi^2)$		

rationally, since at horizon $(t-h-1)$ we do not know ε_{t-h} because y_{t-h} has not yet been observed, there is a trend to the variance forecasts. This trend is predictable since it depends only on σ^2 and the parameters of the forecasting model, which are assumed known to the forecaster. Indeed the pattern displayed by the fixed-event conditional variance forecasts might be helpful in inferring something about the forecasting model either implicitly or explicitly used. For example, from Table 1 we see that as h decreases the reduction in the variance forecasts is more pronounced the more persistent the time-series, as measured by a higher ϕ, is perceived to be.

Revisions to the mean forecasts are therefore white noise but, consistent with the "fan" shape of density forecasts published by the Bank of England and others, the variance declines as we get closer to the event of interest. Therefore unlike the point forecasts, revisions to fixed-event variance forecasts should not be white noise even under rationality. Importantly they decline in a manner which is predictable. This means our best guess of $\mathsf{V}(y_t \mid \Psi_{t-h})$ at horizon $(t-h-1)$ is not $\mathsf{V}(y_t \mid \Psi_{t-h-1})$. If parameter uncertainty is accommodated, the conditional variances need no longer be monotonically increasing in h; see Chong & Hendry (1986). However they remain predictable.

Therefore when the variance is forecastable, which as we have discussed it often is for economic forecasts, KLIC revisions are not news even if the forecaster were rational. This implies $\mathsf{E}(KLIC_{t|t-h} \mid KLIC_{t|t-h-1}) \neq 0$. We can see this analytically when $g(y_t \mid \Psi_{t-h})$ and $g(y_t \mid \Psi_{t-h-1})$ are assumed normal, so that the KLIC revision is given as:

$$KLIC_{t|t-h} = -0.5 - 0.5 \ln \frac{\mathsf{V}(y_t \mid \Psi_{t-h-1})}{\mathsf{V}(y_t \mid \Psi_{t-h})} + \frac{(\mathsf{E}(y_t \mid \Psi_{t-h}) - \mathsf{E}(y_t \mid \Psi_{t-h-1}))^2}{2\mathsf{V}(y_t \mid \Psi_{t-h})} + 0.5 \frac{\mathsf{V}(y_t \mid \Psi_{t-h-1})}{\mathsf{V}(y_t \mid \Psi_{t-h})}. \tag{6}$$

Therefore assuming, as in Table 1, that (i) $\mathsf{E}\left[\mathsf{E}(y_t \mid \Psi_{t-h}) \mid \Psi_{t-h-1}\right] = \mathsf{E}(y_t \mid \Psi_{t-h-1})$ and (ii) $\mathsf{E}\left[V(y_t \mid \Psi_{t-h}) \mid \Psi_{t-h-1})\right] = V(y_t \mid \Psi_{t-h})$ we see that the expected KLIC difference with respect to information available at $(t-h-1)$ is

$$\mathsf{E}\left(KLIC_{t|t-h} \mid \Psi_{t-h-1}\right) = -0.5 \ln \frac{V(y_t \mid \Psi_{t-h-1})}{V(y_t \mid \Psi_{t-h})} + 0.5 \frac{V(y_t \mid \Psi_{t-h-1})}{V(y_t \mid \Psi_{t-h})} \tag{7}$$

This implies

$$\mathsf{E}(KLIC_{t|t-h} \mid KLIC_{t|t-h-1}) \neq 0. \tag{8}$$

Therefore, the KLIC revision is not unpredictable even though the forecaster is assumed rational. $\mathsf{E}(KLIC_{t|t-h} \mid KLIC_{t|t-h-1}) = 0$ if and only if both (3) and (4) hold. We might expect these conditions to hold under rationality only when the true conditional density $f(y_t \mid \Omega_{t-h}) = f(y_t)$ for all h. This requires $\{y_t\}$ to be independent, i.e. unpredictable with respect to the information set Ω_{t-h}; cf. Clements & Hendry (1998), p.35. Only then does $KLIC_{t|t-h}$ measure only the effect of 'news' arriving between periods $t-h-1$ and $t-h$.

3 Revisions to event forecasts

Density forecasts can always be reduced to forecast probabilities of particular events. Consider $p_{t|t-h}$ to be the probability forecast made h-periods ahead of an event (such as a breach of the inflation target, a) happening at time t; $p_{t|t-h} = P(y_t \geq a \mid \Psi_{t-h})$. Let I_t

denote a binary variable equal to unity when the event occurs at time t (i.e. when actual inflation turns out greater than a), 0 otherwise. For a given a, $p_{t|t-h}$ can simply be read off the density forecast.

Similarly to Christoffersen's (1998) test for the conditional efficiency of interval forecasts, a probability event forecast $p_{t|t-h}$ is conditionally efficient with respect to the information set Ψ_{t-h} when:

$$\mathsf{E}\left[I_t \mid \Psi_{t-h}\right] = p_{t|t-h}. \tag{9}$$

We should expect (9) to hold for all (irrespective of a) probability event forecasts $p_{t|t-h}$ extracted from rational density forecasts. This follows from the fact that if the density forecast, from which the probabilistic forecast is extracted, is rational then all interval and probabilistic event forecasts (extracted from the density forecast rather than declared, as optimal, by the forecaster) must be correctly calibrated also. That is, if $g(y_t \mid \Psi_{t-h}) = f(y_t \mid \Omega_{t-h})$ then $\mathsf{E}\left[I_t \mid \Psi_{t-h}\right] = \mathsf{E}\left[I_t \mid \Omega_{t-h}\right] = p_{t|t-h}$, irrespective of the user's loss function. However, similarly to how under a general loss function a user's optimal point forecast need not equal the conditional mean, a user's optimal probabilistic event forecast (as perhaps elicited by a survey in a separate question to the question which asks for their density forecast) need not equal the conditional expectation $\mathsf{E}\left[I_t \mid \Psi_{t-h}\right]$. Nevertheless, it is of interest to note that both the quadratic probability score, an analogue of the widely used root mean squared error criterion, and the log probability score, two popular but specific loss functions used to evaluate probability event forecasts, are maximised when $p_{t|t-h} = \mathsf{E}(I_t \mid \Psi_{t-h})$.

For probability event forecasts which are conditionally efficient we again know from the law of iterated expectations that:

$$\mathsf{E}\left\{\mathsf{E}(I_t \mid \Psi_{t-h}) \mid \Psi_{t-h-1}\right\} = \mathsf{E}(I_t \mid \Psi_{t-h-1}). \tag{10}$$

This implies

$$\mathsf{E}(p_{t|t-h} - p_{t|t-h-1} \mid \Psi_{t-h-1}) = 0, \tag{11}$$

which says that the revision to the probability event forecast is orthogonal to information available at $(t-h-1)$, including lagged revisions to the probability event forecasts. Thus a testable proposition for (weakly) efficient fixed-event density forecasts is that revisions to probability forecasts, extracted from the density forecast, are independent. When there is a clear objective, such as a central bank keeping inflation less than 2%, it is obvious what a to consider. But in any case, for the density forecast to be well calibrated overall, i.e. rational, (11) needs to hold for all possible a's. Since an infinity of event forecasts can be extracted from the density forecast, in an application it is important, as we do below, to evaluate both over a large number of arbitrary events and over events of specific interest.

4 SPF density forecasts for inflation and GDP growth

Started in 1969 the Survey of Professional Forecasters (SPF) is the oldest quarterly survey of forecasters in the United States. Initially administered by the ASA-NBER it has been run by the Philadelphia Fed since 1990. Over its history the survey has asked various questions, but since 1981 it has recorded quarterly subjective density forecasts of both inflation and output growth at both the individual and aggregate (mean across individuals) levels. The panel of forecasters are asked for their density forecasts of both inflation and output growth in both the current and next calendar years. Thus there are eight successive forecasts of the same event. Prior to 1981q3 respondents were asked only about the current year and not the next year and we therefore confine attention to the post 1981q3 era. From 1981q3-1992q1 output growth refers to real GNP, thereafter to real GDP. Following Diebold et al. (1999) we focus on the aggregate SPF density forecasts.

Since respondents' replies are in the form of tabular histograms these marginal densities are in fact known only in discrete form.[4] Prior to analysis we therefore fit normal distributions, following Giordani & Söderlind (2003). This approach was also followed by Lahiri & Liu (2006). We ignore the uncertainty associated with these estimates (see also Engelberg et al. (2006)) and henceforth assume the mean and variance forecasts are known with certainty.

4.1 Forecast efficiency

We have seen that revisions to conditional mean forecasts (given by the mathematical expectation) should be unpredictable if the forecaster uses information efficiently. While revisions to density forecasts need not be even when formed rationally, revisions to event forecasts should be. Therefore we test the weak efficiency of fixed-event density forecasts

[4]Major changes in the probability distribution questions occurred in the 1981q3 and 1992q1 surveys. From 1981q3 respondents were asked to attach a probability to six bins. From 1992q1 this was increased to 10 bins.

by reducing them to probability event forecasts. Since an infinity of event forecasts can be extracted from the density forecast we evaluate both over a large number of arbitrary events and over events of specific interest. We define the events of interest as the probability that inflation falls in its comfort zone of 1-2% and the probability that GDP growth is 'around trend', namely 2-4%.

4.1.1 Testing forecast efficiency

To test whether lagged revisions, $r_{t|t-h-1}$, whether to the conditional mean forecast (as in previous studies) or the event forecast, explain the current revision, $r_{t|t-h}$, separately for each year ($t = 1, ..., T$) traditionally the following regression is considered:

$$r_{t|t-h} = \alpha_1^t r_{t|t-h-1} + u_{t|t-h} \ (h = 1, ..., H-2), \tag{12}$$

where $r_{t|t-h} = p_{t|t-h} - p_{t|t-h-1}$, and practitioners test the null hypothesis of weak efficiency $H_0 : \alpha_1^t = 0$. Squared and higher power lagged values of $r_{t|t-h-1}$ might also be included as regressors in (12) in an attempt to pick up nonlinear dependence. Here we focus on linear dependence. This means our results provide a lower bound on the true degree of inefficiency.

Since H is small in our application (the SPF provides only eight forecasts of the same event, $H = 8$) results based on estimation of (12) separately for each event, t, are at best suggestive. We therefore follow Clements et al. (2007) and pool across t in order to increase the power of the tests. Related approaches have been considered by Davies & Lahiri (1995) and Clements (1997).

Since revisions to forecasts of different targets (t) made at the same time ($t - h$) are likely correlated, since when a forecaster changes her mind about her forecast for tomorrow she will probably also change her forecast for the day after, it is important to model the likely correlation structure. This involves decomposing the forecasting error as follows

$$I_t - p_{t|t-h} = \alpha + \eta_{t|t-h} \tag{13}$$

where $\eta_{t|t-h} = \sum_{j=1}^{h} u_{t|t-j}$ is the sum of the shocks occurring between the time the forecast was made and the realisation, where assuming homoscedastic shocks $\mathsf{E}(u_{t|t-h}^2) = \sigma_u^2$. Assuming rational forecasts we should expect $\alpha_1 = 0$ in (12), since $\mathsf{E}(r_{t|t-h} r_{t|t-h-1}) = \mathsf{E}(u_{t|t-h-1} u_{t|t-h-2}) = 0$. The alternative but more commonly used model of Davies and

Lahiri assumes $\eta_{t|t-h} = \sum_{j=1}^{h} u_{t|t-j} + u^*_{t|t-h}$, where $\sum_{j=1}^{h} u_{t|t-j}$ is interpreted as the common shock and $u^*_{t|t-h}$ as the idiosyncratic shock. This is inconsistent with rational forecasts, since $\mathsf{E}(r_{t|t-h} r_{t|t-h-1}) = \mathsf{E}(u_{t|t-h-1} + u^*_{t|t-h-1} - u^*_{t|t-h}, u_{t|t-h-2} + u^*_{t|t-h-2} - u^*_{t|t-h-1}) \neq 0$, as long as $\mathsf{E}(u^*_{t|t-h} u^*_{t|t-h}) \neq 0$. In other words, to be consistent with the (weak) rationality of fixed-event forecasts, as implied by the law of iterated expectations, it is important to assume that the forecaster has access to 'private information' (see Davies & Lahiri (1995), footnote 13, and also Clements et al. (2007)).

In practice, first we follow Clements et al. (2007) and estimate the following set of regressions across $t = 1, ..., T$:

$$r_{t|t-1} = \alpha_1 r_{t|t-2} + e_{t|t-1} \quad (14)$$

$$r_{t|t-2} = \alpha_2 r_{t|t-3} + e_{t|t-2} \quad (15)$$

$$\vdots \quad (16)$$

$$r_{t|t-6} = \alpha_6 r_{t|t-7} + e_{t|t-6} \quad (17)$$

and test whether $\alpha_i = 0$ $(i = 1, ..., 6)$.

Secondly, we test for dependence using a pooled version of (14)-(17); see Clements et al. (2007):

$$r^* = \alpha^* r^*_{-1} + e^* \quad (18)$$

where $r_{.t-h}$ denotes the T-dimensional vector comprising $r_{t|t-h}$ stacked across t ($t = 1, ..., T$), $r^* = (r_{.t-1}, r_{.t-2,...,} r_{.t-6})$ and $r^*_{-1} = (r_{.t-2}, r_{.t-3,...,} r_{.t-7})$. The covariance structure of the $6T$- error vector e^* follows from our variant of the Davies-Lahiri model whereby under rationality $\mathsf{E}(r_{t|t-h} r_{t+k|t-h}) = \mathsf{E}(u_{t|t-h} u_{t+k|t-h}) = \sigma^2_u$ $(k = 1, 2, ...)$. σ^2_u is estimable from the OLS residuals of (18).

We also consider variants of these tests which accommodate possible aggregation bias. Efficiency tests with aggregated rather than individual-level forecasts, as in our application, can result in aggregation bias if, as seems to be the case with the SPF, the consensus forecast made at $(t - h - 1)$ only becomes available to individual forecasters at horizon $(t - h)$. A consistent test for $H_0 : \alpha^t_1 = 0$ then requires the first lag in (12) to be replaced with the second lag, $r_{t|t-h-2}$; see Isiklar (2005).

4.1.2 Results for forecast efficiency

Table 2 presents t-tests for the significance of the lagged revisions for both inflation and GDP growth using (i) equations (14)-(17) and (ii) the pooled equation, equation (18). Results, for comparative purposes as these are the traditional focus, are also reported for the conditional mean forecasts extracted from the density forecasts. Table 3 then presents results replacing the first with second lag.

In Table 2 the evidence against weak efficiency appears to be greater for the fixed-event mean forecasts of inflation than GDP growth. The pooled test rejects at a 95% significance level the efficiency of the inflation, but not the GDP, conditional mean forecasts. Interestingly, Davies & Lahiri (1999) also rejected the efficiency of the inflation (point not conditional mean) forecasts from the SPF.[5] Table 3 shows that the inflation forecasts remain inefficient when the first lag is replaced by the second.

Turning to the probability event forecasts of inflation being in its comfort zone and the forecasts of GDP growth being around trend, across Tables 2-3 again we see stronger evidence against the efficiency of the inflation forecasts. But in Table 3 there is also increased evidence, as judged by equations (14)-(17), against the efficiency of the GDP growth forecasts with efficiency rejected at 95% in two of the five cases.

Table 4 then presents more detailed results which rather than focusing on events of interest, test for the weak efficiency of the fixed-event probability event forecasts over an arbitrary 700 events, with a lower range (in increments of 0.2%) between -2% to 5% and width 1%-5% (in increments of 0.2%). The table reports the proportion of times weak efficiency is rejected at a 95% level of confidence by equations (14)-(17) and the pooled equation (18). It indicates that a higher proportion of all possible inflation forecasts are inefficient than those for GDP growth. The pooled test indicates that over 46% of all possible probability event forecasts of inflation are inefficient when the first lag is included, down to around 27% when it is excluded. In contrast only about 33% of GDP forecasts are judged inefficient, falling to 19% when the first lag is replaced by the second. This casts considerable doubt on the efficiency, in particular, of fixed-event density forecasts for inflation.

Table 2, and to a lesser degree Table 3, also show that the revisions, whether to the conditional mean or the probability event forecasts, tend to be positively autocorrelated. As Nordhaus (1987) explains this is consistent with forecast smoothing. This is seen by

[5]Our results are also in line with those of Batchelor & Dua (1991) who in an application to Blue Chip forecasts find increased evidence for rationality of real GNP growth than inflation (point) forecasts.

re-writing (12) in a form similar to adaptive expectations, whereby the forecast $p_{t|t-h}$ is revised linearly in response to the previous revision

$$p_{t|t-h} = p_{t|t-h-1} + \alpha_1^t (p_{t|t-h-1} - p_{t|t-h-2}) + u_{t|t-h} \tag{19}$$

revealing that α_1^t can be interpreted as a smoothing or gain parameter. In an analysis of FED Greenbook forecasts Clements et al. (2007) also find evidence of forecast smoothing. They suggest this is motivated by forecasters' desire both to maintain credibility and minimise embarrassment, generated when they are seen to 'change their mind'. Tables 2 and 3 show that a similar story might also be told about the SPF forecasts, and not just the conditional mean forecast but event forecasts extracted from their density forecasts.

5 Conclusion

Economists are making increasing use of density forecasts, or more popularly "fan" charts. This reflects the fact that point forecasts, namely the "central tendency" of the forecast, are better seen as the central points of ranges of uncertainty. In contrast density forecasts provide a full impression of forecast uncertainty. In this paper we seek to understand what if anything can be inferred from the movement of and revisions to successive density forecasts of the same event, so-called "fixed-event" density forecasts. This involves considering how to test whether these forecasts contain all information available at the time the forecast was made.

We explain that revisions to density forecasts as measured by the KLIC, a measure of *divergence* between two densities, need not be unpredictable, unlike those to conditional mean forecasts, even when the forecaster uses information efficiently. Revisions to density forecasts do not comprise entirely of "news"; there is a predictable component. This follows from the fact that forecast rationality does not rule out predictability of the conditional variance forecasts. Only if there is a common target at all horizons, or in other words if the objective is to forecast the unconditional rather than conditional true (objective) density, will KLIC revisions be independent when information is used efficiently. In general, certainly in economics, we should expect the conditional variance to be lower the closer the forecaster to the terminal event.

However measuring the revision to a density forecast by reducing it to a revision to an event forecast we explain that one can then evaluate the rationality of conditional density forecasts similarly to point forecasts. We show that unpredictability remains a feature of rational forecasts. However, in contrast to point forecasts (unless they equal the conditional expectation which they need not under asymmetric loss) revisions to probability event forecasts, extracted from the density rather than declared separately by the forecaster, should be independent whatever the loss function of the forecaster. This follows because under rationality the forecast is assumed to be "correct"; more specifically it is assumed to be conditionally efficient with respect to information available at the time the forecast was made and therefore equal to the true conditional probability of the event.

In an application to the SPF we find that fixed-event density forecasts for inflation do not appear efficient. But there is less evidence that those for GDP growth are inefficient. This is consistent with the view that the SPF smooth their inflation but less so their GDP growth density forecasts.

Table 2: Testing the weak efficiency of fixed-event forecasts: t-tests for the null hypothesis of efficiency

	Conditional Mean		Probability Event	
	Inflation	GDP Growth	Inflation 1-2%	GDP Growth 2-4%
Pooled	2.07	1.82	3.10	1.01
CJS_{12}	1.70	-0.83	3.67	1.62
CJS_{23}	0.89	0.33	1.32	0.33
CJS_{34}	1.73	1.75	1.49	0.37
CJS_{45}	0.75	0.36	-0.20	-0.14
CJS_{56}	0.31	1.61	1.34	0.53
CJS_{67}	0.52	0.69	0.14	0.24

Notes: T-statistics for the null hypothesis of weak efficiency: $\alpha = 0$. Pooled refers to the test that $\alpha^* = 0$ from (18) and CJS_{jk} refer to the tests of $\alpha_i = 0$ $(i = 1, ..., 6)$ from equations (14)-(17), where j and k, $j, k = 1, ..., 6$ denote $r_{t|t-j}$ and $r_{t|t-k}$. Results are based on estimation from 1988 to 2006

Table 3: Testing the weak efficiency of fixed-event forecasts with the second lag replacing the first: t-tests for the null hypothesis of efficiency

	Conditional Mean		Probability Event	
	Inflation	GDP Growth	Inflation 1-2%	GDP Growth 2-4%
Panel	-2.13	-0.33	2.71	-0.53
CJS_{13}	-1.11	0.49	1.33	2.72
CJS_{24}	1.51	0.68	2.50	0.49
CJS_{35}	-0.56	-2.00	0.56	-2.26
CJS_{46}	-0.65	-0.81	-0.03	-1.74
CJS_{57}	-2.78	2.34	0.45	-0.37

Notes: See notes to Table 2

Table 4: Proportion of Weakly Inefficient Probability Event Forecasts (out of 700) at a 95% significance level

	Inflation	GDP growth
CJS_{12},...,CJS_{67}	0.197	0.185
Pooled	0.455	0.330
Replacing the 1st lag with the 2nd		
CJS_{13},...,CJS_{57}	0.211	0.159
Pooled	0.270	0.191

References

Bao, Y., Lee, T.-H. & Saltoglu, B. (2007), 'Comparing density forecast models', *Journal of Forecasting* **26**, 203–225.

Batchelor, R. & Dua, P. (1991), 'Blue Chip rationality tests', *Journal of Money, Credit and Banking* **23**(4), 692–705.

Chong, Y. Y. & Hendry, D. F. (1986), 'Econometric evaluation of linear macro-economic models', *Review of Economic Studies* **53**, 671–690.

Christoffersen, P. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**, 841–862.

Clements, M. P. (1997), 'Evaluating the rationality of fixed-event forecasts', *Journal of Forecasting* **16**, 225–239.

Clements, M. P. (2007), Internal consistency of survey respondents forecasts: Evidence based on the Survey of Professional Forecasters. Department of Economics, Warwick University.

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

Clements, M. P., Joutz, F. & Stekler, H. O. (2007), 'An evaluation of the forecasts of the Federal Reserve: a pooled approach', *Journal of Applied Econometrics* **22**(1), 121–136.

Davies, A. & Lahiri, K. (1995), 'A new framework for analyzing survey forecasts using three-dimensional panel data', *Journal of Econometrics* **68**, 205–227.

Davies, A. & Lahiri, K. (1999), Re-examining the rational expectations hypothesis using panel data on multi-period forecasts, *in* Hsiao, C. and Pesaran, M H and Lahiri, K and Lee, L F, ed., 'Analysis of panels and limited dependent variable models', Cambridge University Press, Cambridge.

Diebold, F. X., Gunther, A. & Tay, K. (1998), 'Evaluating density forecasts with application to financial risk management', *International Economic Review* **39**, 863–883.

Diebold, F. X., Tay, A. S. & Wallis, K. F. (1999), Evaluating density forecasts of inflation: the Survey of Professional Forecasters, *in* R. Engle & H. White, eds, 'Cointegration, causality and forecasting: a festschrift in honour of Clive W. J. Granger', Oxford University Press.

Engelberg, J., Manski, C. F. & Williams, J. (2006), Comparing the point predictions and subjective probability distributions of professional forecasters. Department of Economics, Northwestern University.

Faust, J., Rogers, J. & Wright, J. (2005), 'News and noise in G-7 GDP announcements', *Journal of Money, Credit, and Banking* **37**, 403–417.

Giordani, P. & Söderlind, P. (2003), 'Inflation forecast uncertainty', *European Economic Review* **47**, 1037–1059.

Gneiting, T. & Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association* **102**, 359–378.

Isiklar, G. (2005), 'On aggregation bias in fixed-event forecast efficiency tests', *Economics Letters* **89**(3), 312–316.

Lahiri, K. & Liu, F. (2006), 'Modelling multi-period inflation uncertainty using a panel of density forecasts', *Journal of Applied Econometrics* **21**, 1199–1219.

Mitchell, J. & Hall, S. G. (2005), 'Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR "fan" charts of inflation', *Oxford Bulletin of Economics and Statistics* **67**, 995–1033.

Nordhaus, W. D. (1987), 'Forecast efficiency: concepts and applications', *Review of Economics and Statistics* **69**, 667–674.

Pesaran, M. H. (1987), *The Limits to Rational Expectations*, Basil Blackwell., Oxford.

Zellner, A. (1986), 'Biased predictors, rationality and the evaluation of forecasts', *Economics Letters* **21**, 45–48.