Rounding of probability forecasts:

The SPF forecast probabilities of negative output growth

Michael P. Clements*
Department of Economics,
University of Warwick,
UK. CV4 7AL.
m.p.clements@warwick.ac.uk.

July 17, 2009

Abstract

We consider the possibility that respondents to the Survey of Professional Forecasters round their probability forecasts of the event that real output will decline in the future. We make various assumptions about how forecasters round their forecasts, including that individuals adopt constant patterns of rounding of their responses across time, as well as assuming similar rounding practices when they report their probability of decline forecasts and their histogram forecasts. Our primary interest is in whether reporting practices account for the lack of internal consistency of the probability forecasts of a decline in real output and the histograms for annual real output growth.

Journal of Economic Literature classification: C53, E32, E37

Keywords: Rounding, probability forecasts, probability distributions.

*Computations were performed using code written in the Gauss Programming Language.
1 Introduction

There is a long tradition of using forecasts from surveys to test for the rationality of expectations formation. The forecasts of professional forecasters of inflation are often found to pass tests of rationality (see, for example, Carroll (2003)) and to outperform model-based forecasts in terms of accuracy (see, e.g., Ang, Bekaert and Wei (2007)). More recently, a number of papers have documented inconsistencies between the point forecasts and the subjective probability distributions of output growth and inflation made by some professional forecasters (Engelberg, Manski and Williams (2009) and Clements (2009, 2008b)), and also between probability forecasts and the subjective probability distributions. These findings cast doubt on the rationality of forecasters in a general sense, in terms of their ability or willingness to report mutually consistent sets of forecasts. A possible explanation is that forecasters ‘round’ their probability forecasts. In this paper we consider the extent to which reporting practices, in terms of rounding forecasts of probabilities, contributes to these apparent inconsistencies.

Reported forecast probabilities often appear to be ‘rounded’, in the sense that a 33% probability might be reported as 35%, or 30%, or 40%, or 50%, or even 0%, reflecting rounding to the nearest 5%, 10%, 20%, 50%, or to either 0 or 100%. In this paper we consider probability forecasts taken from one of the best known and longest-running surveys of the US economy, the Survey of Professional Forecasters (SPF), and assess whether reporting practices such as rounding account for the mismatch documented by Clements (2009) between respondents’ probability forecasts of a decline in real output and the histograms for annual real output growth. Probability forecasts are often taken at face value, despite the manifest evidence that they are rounded. The question we address is whether ignoring rounding affects the conclusions we reach regarding forecaster behaviour: in our case, the consistency of two different but related types of forecasts made by the SPF respondents. In checking the consistency of the probability forecasts of a decline in real output and the histograms for annual real output growth Clements (2009) implicitly assumes that neither forecasts is rounded. The key finding is that allowing for plausible patterns of rounding behaviour reduces the rate of inconsistent pairs of forecasts from one in three, to one in four, indicating that rounding has a significant effect, yet is not the sole explanation of the inconsistency.

The problem with rounding is that we generally have no idea to what extent the survey participants round their forecasts before reporting them. Individual respondents do not provide in-
formation on their reporting practices. Respondents may round their forecasts for a number of reasons, which include the use of rounding as a means of conveying uncertainty, as well as to simplify reporting. From the patterns of responses by individuals in response to the same survey question across surveys (in our case the probability of a decline) and across different questions for a given survey (in our case, the histograms forecasts), we will investigate the consequences of making some plausible assumptions concerning individuals reporting practices on the inconsistency of the different types of forecast. Our approach is motivated by Manski and Molinari (2008), who show how individual reporting patterns can be used to estimate the degree of rounding inherent in the reported forecasts.

Our interest in the SPF is because it provides histogram forecasts as well as point and probability forecasts over a long historical period, and allows for an assessment of the consistency of the different types of forecasts made by individual respondents: see Engelberg et al. (2009) and Clements (2009, 2008b). These authors consider the relationship between the subjective probability distributions (reported as histograms) and the point forecasts, for real output growth and inflation, and conclude that in a certain number of cases the point forecasts of output growth and inflation are higher and lower, respectively, than measures of central tendency calculated from the histogram forecasts. These inconsistencies are apparent using the bounds approach of Engelberg et al. (2009) to calculate the lower and upper values (the ‘bounds’) on measures of central tendency for the histogram forecasts when we do not wish to make any assumption about the distribution of probability mass within the histogram bins. The point forecast can then be compared to the bounds to see whether the point forecast can be interpreted as that measure of central tendency of the individual’s underlying subjective distribution (which is only partially revealed in the reported histogram). Clements (2009) extends the bounds approach to an analysis of whether the forecast probabilities of output declines are consistent with the assessment implied by the histogram forecasts, where again a bounds approach obviates the need to make an assumption about how the histogram relates to the underlying subjective probability distribution, albeit at the cost of obtaining a lower and upper bound on the imputed probability of a decline (rather than a point estimate). The key finding is that the histograms imply higher probabilities of output declines than the forecast.

---

1 García and Manzanares (2007) report a similar finding for the GDP growth and inflation forecasts of the ECB’s Survey of Professional Forecasters, for the period 1999Q1 to 2006Q4.
probabilities of declines. For approximately one third of all pairs the forecast probability lies below 
the lower bound for this event imputed from the histogram.

Engelberg et al. (2009) consider whether rounding of the histograms may account for the ap-
parent inconsistencies between the histograms and the point forecasts, but find that their results 
are qualitatively unchanged if they allow for rounding. We are interested in whether the appar-
ent inconsistencies between the histograms and forecast probabilities are affected by allowing for 
rounding. It seems more likely that rounding might be an important factor in our comparison, 
as both types of forecast are likely to be subject to rounding, and perhaps especially the forecast 
probabilities of a decline in output. This makes the SPF probability forecasts of particular interest 
from the perspective of analysing reporting behaviour. Because the event of negative real output 
growth is relatively rare, and is perceived as such by the SPF respondents, the forecast probabilities 
are often small, and are sometimes reported as zero. This is especially problematic, as a reported 
forecast probability of zero could mask an actual assessment close to a fifty-fifty chance for ‘coarse 
rounding’, i.e., if actual assessments in the interval [0, 50] are reported as 0, and those [50, 100] 
are reported as 100%. Alternatively a reported probability of 0 could reflect an actual belief the 
event is certain not to occur. The reported histogram forecasts may also reflect rounding of the 
probabilities attached to the different intervals. In our analysis the histograms play a dual role: we 
will use the histogram forecasts in conjunction with the probability of decline forecasts to deduce 
reporting practices by individual respondents; and secondly, rounding of the histogram forecast 
probabilities may directly contribute to the apparent inconsistencies between the probabilities of 
decline and the histograms reported by Clements (2009).

The remainder of the paper is organised as follows. Section 2 describes the SPF data and the 
evidence of rounding of the probability forecasts. We describe some plausible assumptions we will 
make about rounding practices, and the impact these have in terms of replacing the reported num-
bers by intervals. Section 3 reviews the approach of Clements (2009) to determining whether the 
reported probability forecasts of a decline in output are consistent with their probability distribu-
tions, and calculates the impact of allowing for rounding, both of the forecast probabilities, and of 
the histogram forecasts. Section 4 offers some concluding remarks.
2 The Survey of Professional Forecasters (SPF) data and rounding

The SPF quarterly survey began as the NBER-ASA survey in 1968:4 and runs to the present day. The survey questions elicit information from the respondents on their point forecasts for a number of variables; their histograms for output growth and inflation; and the probabilities they attach to declines in real output. The SPF point forecasts and histograms have been widely analysed in separate exercises.\(^2\) The probability forecasts (of the event that output will decline) have received relatively little attention.\(^3\)

We will also make use of the quarterly Real Time Data Sets for Macroeconomists (RTDSM) maintained by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)). This consists of a data set for each quarter that contains only those data that would have been available at a given reference date: subsequent revisions, base-year and other definitional changes that occurred after the reference date are omitted.\(^4\) The RTDSMs tie in with the SPF surveys such that a respondent to, say, the 1995:Q1 survey would have access to the data in the Feb 1995 RTDSM. The datasets contain quarterly observations over the period 1947:Q1 to the quarter before the reference date (1994:4, in our example), and allow us to construct measures based solely on the data vintages available to the SPF respondents at the time they made the forecast.

Our main focus is on the forecasts of a decline in real output (negative output growth), and the extent to which the figures reported by the survey respondents may be rounded. We will also make use of the histograms for annual real output growth, and the point forecasts of quarterly real output. We use data from 1981:3 to 2005:1, as prior to 1981:3 the histograms for output growth referred to nominal output, and point forecasts for real GDP (GNP) were not recorded.\(^5\)

The probability forecasts consist of reported probabilities of a decline in output in the current

\(^2\)For example, the point forecasts have been analysed by Zarnowitz (1985), Keane and Runkle (1990), Davies and Lahiri (1999), and the probability distributions by Diebold, Tay and Wallis (1999), Giordani and Söderlind (2003) and Clements (2006). A detailed academic bibliography of papers that use SPF data is maintained at http://www.phil.frb.org/econ/spf/spfbib.html.

\(^3\)Although the SPF produces an 'anxious index' by averaging the individual respondents' probabilities of declines in real output in the following quarter, and this is shown to be correlated with the NBER business cycle periods of expansion and recession. Recently Clements (2008a) examines the degree of disagreement in these forecasts across individuals, and Lahiri and Wang (2006b, 2006a) consider their accuracy.

\(^4\)Later vintages of data would not serve our purpose, as they typically contain revisions and definitional changes that were largely unpredictable (see Faust, Rogers and Wright (2005)) based on information available at the time, and as such would not have been known to the SPF respondents.

\(^5\)The definition of real output changes from GNP to GDP in 1992:1, and there are also base year changes over the period, but these are not expected to affect our findings.
quarter (the survey date quarter) relative to the previous quarter, for the next quarter relative to the current quarter, and so on up to the same quarter a year ahead relative to three quarters ahead. The probability distributions refer to the annual change from the previous year to the year of the survey, as well as of the survey year to the following year, although we use only the former. The point forecasts are of the level of output in the current year, as well as for the previous quarter, for the current quarter, the next four quarters, and for the current year.

The total number of usable point and probability forecasts across all surveys and respondents is 2462. These forecasts come from the 95 quarterly surveys from 1981:3 to 2005:1, and from 181 different respondents. We restrict the sample to only include regular forecasters - those who have responded to 12 or more surveys. This gives 73 respondents. These regular respondents account for 1969 forecasts, some 80% of the total.

Table 1 reports the proportion of reported probabilities which are multiples of 5, i.e., reported as 0, 5, 10 etc. A striking feature of the table is that over a half of all forecasts are reported as multiples of 10, and that around 90% are reported as multiples of either 5 or 10 for each of the five forecast horizons. It is clear that these values are being reported far more often than chance would dictate. Also evident is that a decline in output is perceived to be a low probability event. 30% of forecasts assign a zero probability to output declining in the current quarter, and even at the four quarter horizon nearly 40% of the responses are forecasts of either 0, 5 or 10.

Clearly, not all the forecasts reported as multiples of 5 correspond to underlying subjective probability forecasts which are exact multiples of 5. Manski and Molinari (2008) argue that whilst the common practice of ignoring the possibility that responses may be rounded leads to precise point data it is not credible. Suppose instead that probability forecasts reported as multiples of 5 i.e., 0, 5, 10, . . . , 100 reflect actual probability assessments in the intervals [0,2.5], [2.5,7.5], . . . , [97.5,100]. We refer to this relatively precise degree of rounding as ‘M5’. Of course, respondents may round to the nearest 10. In which case all we know is that reported multiples of 10 (i.e., 0,10,
Table 1: Reported probabilities of decline of SPF respondents

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Current quarter</th>
<th>1-quarter</th>
<th>2-quarter</th>
<th>3-quarter</th>
<th>4-quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.300</td>
<td>.134</td>
<td>.076</td>
<td>.071</td>
<td>.074</td>
</tr>
<tr>
<td>5</td>
<td>.132</td>
<td>.157</td>
<td>.145</td>
<td>.117</td>
<td>.119</td>
</tr>
<tr>
<td>10</td>
<td>.111</td>
<td>.174</td>
<td>.204</td>
<td>.202</td>
<td>.191</td>
</tr>
<tr>
<td>15</td>
<td>.034</td>
<td>.068</td>
<td>.115</td>
<td>.114</td>
<td>.098</td>
</tr>
<tr>
<td>20</td>
<td>.070</td>
<td>.103</td>
<td>.130</td>
<td>.153</td>
<td>.155</td>
</tr>
<tr>
<td>25</td>
<td>.031</td>
<td>.060</td>
<td>.062</td>
<td>.082</td>
<td>.082</td>
</tr>
<tr>
<td>30</td>
<td>.044</td>
<td>.060</td>
<td>.069</td>
<td>.068</td>
<td>.077</td>
</tr>
<tr>
<td>35</td>
<td>.017</td>
<td>.016</td>
<td>.022</td>
<td>.028</td>
<td>.023</td>
</tr>
<tr>
<td>40</td>
<td>.034</td>
<td>.039</td>
<td>.034</td>
<td>.028</td>
<td>.037</td>
</tr>
<tr>
<td>45</td>
<td>.010</td>
<td>.008</td>
<td>.003</td>
<td>.006</td>
<td>.007</td>
</tr>
<tr>
<td>50</td>
<td>.029</td>
<td>.034</td>
<td>.023</td>
<td>.020</td>
<td>.031</td>
</tr>
<tr>
<td>55</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>60</td>
<td>.009</td>
<td>.007</td>
<td>.009</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td>65</td>
<td>.003</td>
<td>.004</td>
<td>.000</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>70</td>
<td>.006</td>
<td>.004</td>
<td>.002</td>
<td>.003</td>
<td>.002</td>
</tr>
<tr>
<td>75</td>
<td>.006</td>
<td>.007</td>
<td>.005</td>
<td>.003</td>
<td>.003</td>
</tr>
<tr>
<td>80</td>
<td>.008</td>
<td>.005</td>
<td>.003</td>
<td>.001</td>
<td>.002</td>
</tr>
<tr>
<td>85</td>
<td>.001</td>
<td>.001</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>90</td>
<td>.011</td>
<td>.002</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td>95</td>
<td>.004</td>
<td>.001</td>
<td>.000</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>100</td>
<td>.016</td>
<td>.002</td>
<td>.000</td>
<td>.000</td>
<td>.001</td>
</tr>
</tbody>
</table>

Proportion reported as:

<table>
<thead>
<tr>
<th>Proportion</th>
<th>10 or 5</th>
<th>10</th>
<th>5 (excl. multiples 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 or 5</td>
<td>0.872</td>
<td>0.884</td>
<td>0.901</td>
</tr>
<tr>
<td>10</td>
<td>0.636</td>
<td>0.563</td>
<td>0.550</td>
</tr>
<tr>
<td>5 (excl. multiples 10)</td>
<td>0.236</td>
<td>0.321</td>
<td>0.351</td>
</tr>
</tbody>
</table>


... 100 correspond to underlying probability assessments that fall in the intervals [0,5], [5,15], ..., [95,100]. If we couple this assumption with the assumption that forecasts reported as multiples of 5 (but excluding those reported as multiples of 10) are treated as in M5, then we have strategy M10. The assumption that a grosser form of rounding occurs would be that reported multiples of 50 (i.e., 0,50,100) would actually reflect intervals of [0,25], [25,75] and [75,100], and this assumption with M10 gives M50. The grossest from of rounding behaviour would indicate that reported forecasts of 0 and 100 are associated with the intervals [0,50] and [50,100]. This is referred to as M100.

In table 2 we report the distribution of the widths of the interval forecast generated by assuming rounding of the form M5, M10, M50 and M100 for each of the five forecast horizons. Note that zero widths result when the reported forecasts are not a multiple of 5, where formally we assume that the upper and lower values of the interval are equal to the reported value. Under M5, the maximum width is 5 percentage points, corresponding to a reported point forecast which is a multiple of 5 other than 0 or 100 - boundary forecasts of 0 or 100 give rise to intervals of 2.5 under M5. Allowing for rounding to the nearest 10, M10, results in nearly half the forecasts being replaced with intervals of width of 10 at the longer horizons. We regard M5 and M10 as perhaps the most plausible rounding schemes for professional forecasters, while M50 and M100 are included to illustrate the impact of allowing the grosser forms of rounding. The intervals naturally become relatively uninformative.

Rather than this mechanical form of rounding, Manski and Molinari (2008) suggest using response patterns by individual respondents to infer the degree of rounding undertaken by each individual. They consider individual patterns of responses to different questions in a given survey. We begin by looking at individual patterns of responses to the same question (the assessment of the probability of a decline in output, at a given horizon) across different surveys. As we will show, the use of the information provided by these patterns is valuable in markedly narrowing the intervals that replace the reported probability forecasts. In section 3.3 we consider response patterns across questions.

Assume that each individual applies the same rounding rule when they respond to the SPF question on the probability of decline, but that the rounding rule may differ by individual. Adapting the algorithm proposed by Manski and Molinari (2008), we proceed as follows. If all the responses by individual \( j \) are either 0 or 100, we assume that individual \( j \) engages in gross rounding, so that
Table 2: Widths of intervals assuming the reported probabilities reflect different degrees of rounding

<table>
<thead>
<tr>
<th>‘Rounding’</th>
<th>Percentage of intervals with a width of percentage points:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0  2.5   5   10   25  50</td>
</tr>
<tr>
<td>Horizon: current</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>12.8 31.6 55.6 0.0 0.0 0.0</td>
</tr>
<tr>
<td>M10</td>
<td>12.8 0.0 55.2 32.0 0.0 0.0</td>
</tr>
<tr>
<td>M50</td>
<td>12.8 0.0 23.6 29.2 31.6 2.9</td>
</tr>
<tr>
<td>M100</td>
<td>12.8 0.0 23.6 29.2 0.0 34.5</td>
</tr>
<tr>
<td>‘Individual’</td>
<td>76.7 6.4 16.5 0.3 0.0 0.0</td>
</tr>
<tr>
<td>Horizon: 1-quarter ahead</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>11.6 13.6 74.9 0.0 0.0 0.0</td>
</tr>
<tr>
<td>M10</td>
<td>11.6 0.0 45.7 42.8 0.0 0.0</td>
</tr>
<tr>
<td>M50</td>
<td>11.6 0.0 32.1 39.4 13.6 3.4</td>
</tr>
<tr>
<td>M100</td>
<td>11.6 0.0 32.1 39.4 0.0 16.9</td>
</tr>
<tr>
<td>‘Individual’</td>
<td>57.7 5.8 35.9 0.6 0.0 0.0</td>
</tr>
<tr>
<td>Horizon: 2-quarter ahead</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>9.9 7.6 82.5 0.0 0.0 0.0</td>
</tr>
<tr>
<td>M10</td>
<td>9.9 0.0 42.8 47.3 0.0 0.0</td>
</tr>
<tr>
<td>M50</td>
<td>9.9 0.0 35.1 45.0 7.6 2.3</td>
</tr>
<tr>
<td>M100</td>
<td>9.9 0.0 35.1 45.0 0.0 10.0</td>
</tr>
<tr>
<td>‘Individual’</td>
<td>52.7 1.7 45.1 0.4 0.0 0.0</td>
</tr>
<tr>
<td>Horizon: 3-quarter ahead</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>9.9 7.1 83.1 0.0 0.0 0.0</td>
</tr>
<tr>
<td>M10</td>
<td>9.9 0.0 42.0 48.1 0.0 0.0</td>
</tr>
<tr>
<td>M50</td>
<td>9.9 0.0 34.9 46.1 7.1 2.0</td>
</tr>
<tr>
<td>M100</td>
<td>9.9 0.0 34.9 46.1 0.0 9.1</td>
</tr>
<tr>
<td>‘Individual’</td>
<td>52.1 2.8 45.1 0.0 0.0 0.0</td>
</tr>
<tr>
<td>Horizon: 4-quarter ahead</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>9.1 7.5 83.4 0.0 0.0 0.0</td>
</tr>
<tr>
<td>M10</td>
<td>9.1 0.0 40.7 50.2 0.0 0.0</td>
</tr>
<tr>
<td>M50</td>
<td>9.1 0.0 33.2 47.0 7.5 3.1</td>
</tr>
<tr>
<td>M100</td>
<td>9.1 0.0 33.2 47.0 0.0 10.7</td>
</tr>
<tr>
<td>‘Individual’</td>
<td>54.6 2.7 42.7 0.0 0.0 0.0</td>
</tr>
</tbody>
</table>

There are 1969 forecasts. M5 to M100 indicate increasingly grosser degrees of rounding, as described in the text. ‘Individual’ indicates that the degree of rounding is inferred from individual-level behaviour.
responses of 0 are replaced by the interval $[0,50]$, and responses of 100 by in the interval $[50,100]$. If all responses are 0, 50, 100, we assume that the respondent is rounding to the nearest 50, and these point forecasts are replaced by the following intervals, $[0,25]$, $[25,75]$, $[75,100]$. Similarly if all responses are multiples of 10, and if all responses are multiples of 5. This can be codified formally as follows, where the reported point forecasts made by $j$ to survey $t_j$, $p_{jt_j}$, is replaced by the interval $[p_{jt_j,L}, p_{jt_j,U}]$ according to the following scheme:

* If $p_{jt_j} = (0,100)$, for all $t_j$, then $[p_{jt_j,L}, p_{jt_j,U}] = [\max (0, p_{jt_j} - 50), \min (p_{jt_j} + 50, 100)]$.
* If $p_{jt_j} = (0,50,100)$, for all $t_j$, then $[p_{jt_j,L}, p_{jt_j,U}] = [\max (0, p_{jt_j} - 25), \min (p_{jt_j} + 25, 100)]$.
* If $p_{jt_j} = (0,10,20,\ldots,100)$, for all $t_j$, then $[p_{jt_j,L}, p_{jt_j,U}] = [\max (0, p_{jt_j} - 5), \min (p_{jt_j} + 5, 100)]$.
* If $p_{jt_j} = (0,5,10,15,\ldots,100)$, for all $t_j$, then $[p_{jt_j,L}, p_{jt_j,U}] = [\max (0, p_{jt_j} - 2.5), \min (p_{jt_j} + 2.5, 100)]$.
* Otherwise $[p_{jt_j,L}, p_{jt_j,U}] = [p_{jt_j}, p_{jt_j}]$.

Note that $t_j$ will one of the quarterly surveys from 1981Q3 to 2005Q1. None of the individuals responded to all the surveys, so that for each individual $j$, $t_j$ is an element of a subset of the quarterly forecasts. The study of Manski and Molinari (2008) replaces non-responses with uninformative intervals $[0,100]$. We ignore the non-responses as it seems reasonable to assume that failure to file a return to a particular survey provides no information on an individual’s rounding behaviour.

The distribution of the widths of the resulting intervals are recorded in table 2 in the rows ‘Individual’. The narrowing of the intervals when individual response pattern information is used is marked. For example, for the current quarter forecasts over three quarters (76.7%) of the intervals are in fact points, and virtually no intervals are of greater width than 5 at any of the horizons. Use of this schema effectively rules out rounding to anything other than a multiple of 5. Although there are many instances of forecasts reported as multiples of 10, 50 and 100, as evident from table 1, virtually all respondents produce at least one forecast which is a multiple of five or has a finer value, ruling out the grosser forms of rounding under the maintained assumption that individuals operate with the same rounding rule for all their forecasts.

Comparing the width distributions for the intervals calculated using individual response patterns for the current quarter and longer horizon forecasts it is apparent that the degree of rounding increases with the forecast horizon. This is evident from the increase in the width of the intervals as the horizon increases. Specifically, going from the current quarter to the 1-quarter ahead forecasts
we observe a reduction in the percentage of 0-width intervals from 77% to 58%, and a doubling in the percentage of intervals of width 10 (17 to 36%). We interpret this positive relationship between the degree of rounding and the forecast horizon as suggesting that SPF respondents round their forecasts to convey ‘ambiguity’ rather than to ‘simplify communication’ (in the parlance of Manski and Molinari (2008)). If they were rounding to simplify communication there would be no reason to expect the degree of rounding to increase with the forecast horizon. So it appears that rounding is used to convey uncertainty in the underlying subjective probability. Notice also that the degree of rounding changes little as the horizon increases beyond 1-quarter ahead. This suggests that the current-quarter forecasts are relatively less certain than the longer-horizon forecasts, but that the longer-horizon forecasts (1 to 4-quarters ahead) are all approximately equally uncertain.9

3 The consistency of the probability forecasts and the probability distributions

In this section we investigate the potential impact of rounding on the assessment of the consistency of the probability forecasts and the histogram forecasts. We first explain the approach taken by Clements (2009), who takes the reported probability forecasts at face value, and assesses the consistency of these forecasts and the respondents’ probability distributions by adapting the bounds approach of Engelberg et al. (2009). We then allow for rounding of the forecast probabilities, and finally consider the possible impact of the respondents also rounding their histogram forecasts.

3.1 Review of the bounds approach to determining the consistency of the forecast probabilities of decline and the histogram forecasts.

Engelberg et al. (2009) use a bounds approach to assess the consistency of the point forecasts and probability distributions. Because the probability distribution is reported as a histogram, the survey return provides an incomplete picture of a respondent’s underlying subjective probability distribution, and it is only possible to calculate point measures of, say, the mean (in Engelberg et al.’s case) if we are prepared to make assumptions about the distribution of the probability mass

9This ties in with the finding of Lahiri and Wang (2006a) that the SPF probability forecasts beyond two quarters ahead have little ‘skill’ or ‘value’ in a sense which they explain.
within the histogram intervals. If we do not wish to make such assumptions we can still bound the values that the mean (or any of the other measures of central tendency) can take - instead of a point estimate we obtain a lower and upper value for the particular measure of central tendency. Engelberg et al. (2009) compare point forecasts of annual output growth and inflation from SPF to bounds to see whether the point forecasts and histograms are consistent (see also Clements (2009, 2008b)).

Clements (2009) shows that the SPF histograms and probability forecasts can be compared for the Q4 surveys using a bounds approach. For the Q4 surveys, current-quarter probabilities of decline are calculated from the probability distributions, which can be compared to the directly-reported current-quarter forecast probabilities of decline. Given the realized values of output in the seven quarters up to and including the third quarter of the year (taken from the RTDSM for the fourth quarter of the year), we can infer the year-on-year rate of growth that equates the Q4 level of output with that of the preceding quarter. The implied current-quarter probability of decline from the histogram is then the probability that year-on-year output growth will not exceed this rate. As in the case of calculating means from the histograms, the required calculation could be performed by assuming uniformity within intervals, or by approximating the histograms by normal densities, amongst other methods. But without making any such assumptions about the relationship between the histogram and a respondent’s actual beliefs, we can calculate a lower and upper bound on the probability.

We provide a concrete example that illustrates the bounds approach and serves to make explicit the assumptions we make. The following example uses the actual forecasts of a respondent to a specific survey, the fourth quarter of 1992. This respondent’s histogram is shown in table 3 - non-zero probabilities are attached to the annual output growth rate in 1992 being between 1 and 1.9 (0.65) and being between 2 and 2.9 (0.35). At the time of the survey, the latest estimates of the level of real output in the quarters 1991:1 to 1992:3 inclusive were: 4796.7, 4817.1, 4831.8, 4838.5, 4873.7, 4892.4, 4924.5, respectively. Hence real output will decline in the fourth quarter of the year if the level of real output is less than 4924.5, the third quarter figure. The respondent provides a direct forecast of the probability of this event: it is \( p_0 = 0.01 \) (the subscript denotes a ‘current quarter’ forecast). We can also infer this probability from the histogram. Given the annual level of output in 1991 of 4821.025 (from averaging the relevant quarters), we can calculate the annual
Table 3: Example of a histogram return to 1992:4 survey, and histogram allowing for rounding

<table>
<thead>
<tr>
<th>Interval</th>
<th>Reported probabilities</th>
<th>Rounded I</th>
<th>Rounded II</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3 to -2.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>-2 to -1.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>-1 to -0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0 to 0.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.05</td>
</tr>
<tr>
<td>1 to 1.9</td>
<td>0.65</td>
<td>0.60</td>
<td>65.0</td>
</tr>
<tr>
<td>2 to 2.9</td>
<td>0.35</td>
<td>0.35</td>
<td>0.30</td>
</tr>
<tr>
<td>3 to 3.9</td>
<td>0.0</td>
<td>0.05</td>
<td>0.0</td>
</tr>
<tr>
<td>4 to 4.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>5 to 5.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>6 to 6.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The highest interval (bottom row) was open-ended in the survey questionnaire (‘6+’), as was the lowest (‘ < −2’). Rounded I and Rounded II are the two cases when the reported bin probabilities are assumed to have been rounded, and bound the histogram under the assumption of rounding up/down to a multiple of 0.05.

The highest interval (bottom row) was open-ended in the survey questionnaire (‘6+’), as was the lowest (‘ < −2’). Rounded I and Rounded II are the two cases when the reported bin probabilities are assumed to have been rounded, and bound the histogram under the assumption of rounding up/down to a multiple of 0.05.

growth rate for 1992 which equates the 1992:4 level of output to the previous quarter level: this is 1.7. From the histogram, we can then infer the probability of annual output growth being less than 1.7, which is the histogram probability that output will decline in the fourth quarter relative to the third quarter. This probability could be calculated by fitting a normal distribution to the histogram, as in Giordani and Söderlind (2003), which gives an estimated probability of 0.299, or by linear interpolation of the histogram, resulting in an estimated probability of 0.506.¹⁰ Instead we can calculate a bound on this probability by assuming that all the mass in the 1 to 1.9 interval lies below 1.7, which creates an upper bound on the probability of 0.65, and then that all the mass in this interval lies above 1.7, creating a lower bound on the probability of 0.¹¹ Bounds calculated in this way satisfy $u, l \in [0, 1]$, and $u - l \in [0, 1]$, where $u$ and $l$ are the upper and lower bounds on the probability, where e.g., $l = u = 1$ would result if 1.7 exceeds the upper limit of the right-most histogram interval containing non-zero mass.

The first row of table 4 shows that over a third of the probability forecasts lie below the

³⁰That is, $(1.7 - 1)/(1.9 - 1)$ of the probability (0.65) attached to the interval 1 to 1.9 is less than or equal to 1.7, so linear interpolation gives $(0.7/0.9) \times 0.65 = 0.506$.

¹¹More generally, the lower bound sums the probabilities of all the histogram intervals whose upper limits are less than 1.7, while the upper bound includes in this sum the probability attached to the histogram interval containing the 1.7. value.
Table 4: Bounds on histogram probabilities of decline and directly-reported probabilities from Q4 surveys (497 forecasts)

<table>
<thead>
<tr>
<th>Rounding?</th>
<th>% below bounds</th>
<th>% above bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>None - taken at face value</td>
<td>35.2</td>
<td>1.4</td>
</tr>
<tr>
<td>M5</td>
<td>32.2</td>
<td>1.2</td>
</tr>
<tr>
<td>M10</td>
<td>30.6</td>
<td>1.2</td>
</tr>
<tr>
<td>M50</td>
<td>22.5</td>
<td>1.0</td>
</tr>
<tr>
<td>M100</td>
<td>20.9</td>
<td>1.0</td>
</tr>
<tr>
<td>‘Individual’</td>
<td>34.4</td>
<td>1.4</td>
</tr>
<tr>
<td>M5, M5&lt;sub&gt;h&lt;/sub&gt;</td>
<td>26.4</td>
<td>1.0</td>
</tr>
<tr>
<td>M10, M10&lt;sub&gt;h&lt;/sub&gt;</td>
<td>22.1</td>
<td>1.0</td>
</tr>
<tr>
<td>M5&lt;sub&gt;U&lt;/sub&gt;M5&lt;sub&gt;h&lt;/sub&gt;</td>
<td>28.6</td>
<td>1.2</td>
</tr>
<tr>
<td>M10&lt;sub&gt;U&lt;/sub&gt;M10&lt;sub&gt;h&lt;/sub&gt;</td>
<td>25.2</td>
<td>1.2</td>
</tr>
</tbody>
</table>

M5 to M100 indicate increasingly grosser degrees of rounding of the forecast probabilities of a decline, as described in the text. ‘Individual’ indicates that the degree of rounding is inferred from individual-level responses to the probability forecast of decline question. M5<sub>h</sub> and M10<sub>h</sub> indicate assumed rounding of the histogram bin probabilities.

histogram bounds on the probability of a decline, when the directly reported probabilities and the histograms are taken at face value. The illustrative example above would constitute an example of a consistent pair of histograms and probability forecasts, as 0.01 ∈ [0, 0.65], but note that for just over one out of every three of the 497 pairs of forecasts we find \( p_0 < l \).

3.2 Rounding of probability forecasts

Of interest is whether the tendency to round probability forecasts described in section 2 contributes to the finding that one out of three pairs of forecasts are inconsistent. The rows labelled M5 to M100 in table 4 report the percentage of probability of decline interval forecasts which are inconsistent with the histogram bounds for the rounding schemes described in section 2.\(^{12}\) We find that for the more realistic rounding schemes (M5 and M10) there is only a relatively small reduction in the number of inconsistent pairs. For grosser forms of rounding such as M50 the number of probability of decline intervals that lie below the lower bound on the imputed histogram probability of a decline

\(^{12}\)That is, the percentage of pairs of forecasts for which the upper bound on the probability of decline (\( p_U \)) calculated allowing for rounding is below the lower bound on the histogram probability, \( l \).
falls to around 22%. But recall that this form of rounding supposes that forecast probabilities of 0, say, reflect underlying probability assessments in the range [0, 25]: assuming such a lack of precision is perhaps unduly conservative for professional forecasters.

If we assume that individuals use the same rounding rule for all their current-quarter decline probability forecasts and adopt the scheme described in section 2, then it is apparent from table 4 that replacing the point forecasts by interval forecasts has virtually no effect on the degree of inconsistency. This is inkeeping with the relatively narrow interval forecasts that result under this scheme, as shown in table 2. Note that in calculating the intervals we have used all the current quarter forecasts to infer individuals rounding behaviour, not just the current quarter forecasts made in response to Q4 surveys. This is because the Q4-survey current forecasts resemble the responses to the surveys held in the other quarters of the year in terms of reported forecasts which are multiples of 5 and 10, etc. (not shown to save space). Hence the sample for which we can match current quarter probability forecasts and implied decline probabilities from the histograms appears to be typical of all the current quarter probability forecasts, and we draw on all those forecasts.

### 3.3 Rounding of probability forecasts and the histogram forecasts

So far we have allowed the probability of decline forecasts to reflect rounding but have taken the histogram forecasts at face value. But as noted in the introduction, the probabilities attached to the different histogram intervals may also be rounded. In order to assess the impact of this form of rounding on the relationship between the histogram forecasts and the point forecasts Engelberg et al. (2009, Appendix, pp. 40-1) propose the following scheme. If in each bin of a given histogram forecast all the probability mass is a multiple of 0.05 (as in Table 3) then two sets of calculations are performed. In the first 0.05 is subtracted from each bin used by the forecaster and added to the next higher bin. This amounts to subtracting 0.05 from the lowest bin, and adding it to the bin adjacent to the highest bin used by the forecaster: see column denoted Rounded I in table 3. Because it is unclear whether forecasters round up or down, a second set of calculations subtracts 0.05 from the highest bin used by the forecaster and adds this to the bin immediately below the lowest bin used by the forecaster (as illustrated by the column headed Rounded II in table 3).

The effect on the bounds on the histogram estimate of the probability of a decline in output are simple to calculate. Denoting by \([l, u]\) the bounds when no allowance is made for rounding,
the new bound becomes \([\max (0, l - 0.05), \min (u + 0.05, 1)]\), so that the bounds necessarily widen. When the probability masses are not all multiples of 0.05, the bounds are unchanged. We refer to this assumption about rounding behaviour as M5h. We allow for rounding to 0.10: when all the histogram bin probability masses are multiples of 0.10, we proceed as above, and calculate new bounds as \([\max (0, l - 0.10), \min (u + 0.10, 1)]\): this we denote by M10h. M10h include 5% rounding for histograms which have all bins multiples of 0.05 other than multiples of 0.10.

Table 4 reports the proportion of inconsistent pairs of histogram and probability of decline forecasts when the two types of forecasts are rounded according to M5h and M5, respectively, and according to M10h and M10 respectively. These rows are denoted in table 4 as ‘M5, M5h’ and ‘M10, M10h’. Assuming ‘rounding to 10’, the number of inconsistent pairs falls to 22%. This is a marked reduction from the 35% when both types of forecast are taken at face value. This degree of rounding would appear to be relatively mild.

These results do not require that an individual respondent adopts a common rounding practice for the two questions. For example, consider ‘M5, M5h’. For a particular pair, the histogram bounds may reflect rounding (because the bins for that histogram are multiples of 0.05) but the forecast probability is treated ‘as is’ \((p_L = p_U)\) because it is not a multiple of 0.05. If we believe it reasonable to assume that a respondent will adopt a common rounding practice in reporting their forecasts to a particular survey, this may serve to restrict the widths of the intervals. In the above example, because the forecast probability is not rounded, we would also assume that the histogram bin probabilities of 0.05 do not reflect rounding. Assuming a common rounding practice in response to the two questions in a particular survey is denoted in the table by ‘M5\(\cap\)M5h’ and ‘M10\(\cap\)M10h’. Even under the assumption of common rounding across the two types of forecasts, with ‘M10\(\cap\)M10h’ downplaying the likely impact of rounding relative to ‘M10,M10h’, we still find that allowing for rounding reduces the proportion of inconsistent pairs of forecasts to around one in four.

4 Conclusions

Forecasts of economic variables from surveys of the macroeconomic outlook or micro-level prospects sometimes take the form of probabilities, as is the case with the SPF forecasts of a decline in output.
which we consider in this paper. Reported probability forecasts are often taken at face value, despite the evidence that respondents round their assessments before reporting them.

We provide evidence that the SPF probability of decline forecasts are often rounded. Whether this matters or not depends on the use to which the forecasters are to be put. In previous work the reported forecasts have been compared to imputed probabilities of a decline calculated from the respondents’ histograms, and an apparent mismatch between the two types of forecasts in over a third of all cases casts doubt on the ability of professional forecasters to generate consistent forecasts of related phenomena. Of interest is the extent to which rounding affects the conclusions we reach regarding forecaster behaviour. Our key finding is that allowing for plausible patterns of rounding behaviour reduces the rate of inconsistent pairs of forecasts from one in three, to one in four, indicating that rounding has a significant effect, yet is not the sole explanation of the inconsistency.

We consider a number of rounding schemes, two of the most plausible being that a reported probability forecast \( p \) which is a multiple of 0.05 corresponds to an assessment of the underlying probability being in the interval \([p - 0.025, p + 0.025]\) (for \( p \neq 0, 1 \)), and that a reported probability forecast \( p \) which is a multiple of 0.1 corresponds to an assessment of the underlying probability being in the interval \([p - 0.05, p + 0.05]\) (again, for \( p \neq 0, 1 \)). If we allow for rounding of the probability forecasts of this type, and also suppose that histogram forecasts are rounded, we find that the percentage of inconsistent pairs of forecasts falls by around a third, from 35% to 22%. However, it seems plausible that individual respondents apply similar rounding schemes to both their probability forecasts and histograms. Hence when a histogram has not been rounded, we infer that the forecast probability has not been rounded either, irrespective of whether it is a multitude of 0.05, or 0.1. Supposing that individuals act in this way when responding to a particular survey limits the uncertainty about the true values of the probability forecasts and histograms, and therefore also the extent to which rounding can account for the inconsistencies that have been documented. Nevertheless, the number of inconsistencies still falls to around one in four under this assumption about rounding practice.

We also allowed the possibility that an individual applies the same rounding practice on each occasion they respond to the probability forecast question across surveys. This results in narrow intervals about the probability forecasts, but is perhaps questionable. Even an individual who
responded to consecutive surveys would need to follow the same rounding practice over a three year period to deliver the minimum of twelve responses. Our assumption of a common rounding practice across the two questions of a single survey seems more plausible, but naturally results in wider intervals.

Given that we do not observe the true assessments and the degree of rounding has to be inferred, the plausibility of the assumptions we make about rounding patterns takes centre stage. There are likely to be differences of opinion over what is plausible, but even so, testing the robustness of findings to the possibility that probability forecasts are rounded would appear preferable to simply taking the probability forecasts at face value. Allowing for what we regard as a plausible rounding scheme reduces the fraction of inconsistent forecasts from over a third to a quarter.

As to the reasons why forecasters round their probability assessments, we find a positive relationship between the degree of rounding and the forecast horizon (at least for the current and 1-quarter ahead forecasts), which leads us to tentatively conclude that SPF respondents primarily round their forecasts to convey uncertainty.
References


