OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12542

Real-Time Perceptions of Historical GDP Data Uncertainty*

ANA BEATRIZ GALVÃO† and JAMES MITCHELL‡

†*Warwick Business School, University of Warwick, Coventry, UK (e-mail: ana.galvao@wbs.ac.uk)*
‡*Federal Reserve Bank of Cleveland, Cleveland, OH, USA*

Abstract

GDP is measured with error. But data uncertainty is rarely communicated quantitatively in real-time. An exception are the fan charts for historical real GDP growth published by the Bank of England. To assess how well data uncertainty is understood, we first evaluate the accuracy of the historical fan charts. We find that data uncertainties can be accurately quantified, even without judgement, using past revisions data. Secondly, we conduct an online survey to gauge perceptions of GDP data uncertainty across a wider set of experts. Our results call for greater communication of data uncertainties to anchor experts' dispersed expectations.

I. Introduction

Economic history is continuously rewritten as data are revised.[1] As a result, the path of the UK's economic recovery since the global financial crisis looks quite different today than it did in its immediate aftermath. As an example, that attracted media attention at the time, Gross Domestic Product (GDP) data revisions in 2013 revised away the UK's 'double-dip' recession, previously believed to have occurred in early 2012.[2] Given that the Office for National Statistics (ONS) published its first quarterly UK real GDP estimate (using the output approach) around 27 days after the end of the quarter, but based on just 44% of the total sample, data revisions should really come as no surprise. GDP estimates

JEL Classification numbers: C53, E32.
*We thank Johnny Runge at NIESR for help in running the expert survey. We also thank two anonymous referees, Simon van Norden, seminar participants at the Bank of England, Reading, UCL and ESCoE for helpful comments. A previous version of this paper was circulated as 'Measuring data uncertainty: an application using the Bank of England's ''fan charts'' for historical GDP growth'. This research has been funded by the ONS as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

[1]McKenzie (2006) delineates seven reasons for 'revisions', including updated sample information, correction of errors, replacement of first estimates derived from incomplete surveys/judgements/statistical techniques, benchmarking, updated seasonal factors, updated base period for constant price estimates and changes in statistical methodology.
[2]See https://www.bbc.co.uk/news/business-23079082.

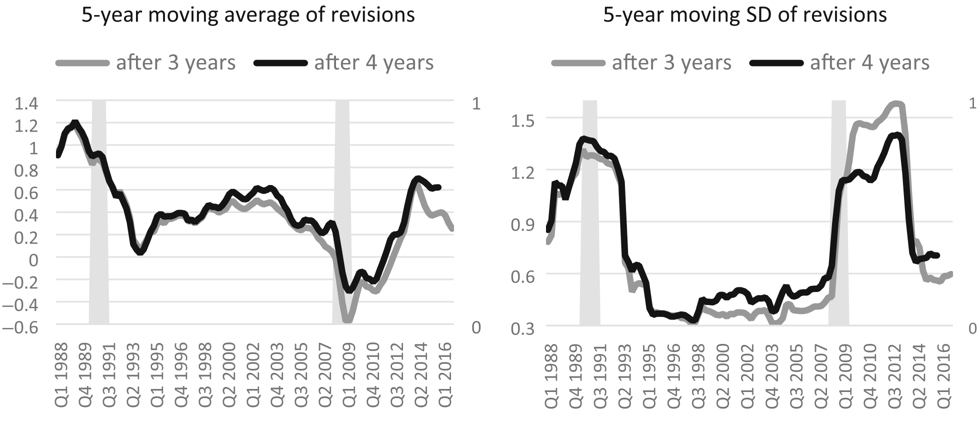

Figure 1. Data revisions between the first and the 'mature' estimate of year-on-year GDP growth, with 'mature' GDP growth measured 3 or 4 years after the first estimate (in %)
Notes: Underlying revisions data dates from 1983Q2. Dates on the *x*-axes refer to the final quarter in the 5-year moving average. Shaded areas indicate recession quarters, as identified via the Bry−Boschan dating algorithm applied to the levels GDP data; see Galvão and Kara (2020)

are updated (and balanced) with the arrival of more sampling information (including on the income- and expenditure-side of GDP).[3]

Figure 1 provides historical perspective, plotting the 5-year moving average and SD of revisions to year-on-year (real) GDP growth estimates in the UK using data back to 1983.[4] Revisions are measured as the difference between the ONS's first estimate and more mature estimates published 3 and 4 years after the first estimate. Figure 1 shows that these revisions can be substantial. SD estimates exceed 1 percentage point for many of the 5-year windows. There is a tendency for GDP estimates to be revised upwards after first release, given that the moving average in Figure 1 is generally positive. We also see that revisions are time-varying and often larger at business cycle turning points, with the SD rising around recessions.[5]

Accordingly, aware that data revisions matter (and not just for UK GDP), a now large 'real-time' literature has developed to analyse and model data revisions across variables and countries (e.g., see Faust, Rogers, and Wright 2005; Jacobs and van Norden 2011; Cunningham *et al.* 2012 Kishor and Koenig 2012; Galvão 2017). In order to understand the underlying 'true' data, following Mankiw and Shapiro (1986) studies often discriminate between news and noise revisions. In tandem, national statistical offices and central banks increasingly publish real-time data vintages (e.g. see Croushore and Stark, 2001; Giannone *et al.*, 2012).

[3]In the summer of 2018 the ONS changed its publication model. From then on, the first estimate of quarterly GDP became available at around 40 days; and has a higher data content than the first estimates considered for the period analysed in this paper. In due course the modelling exercise in this paper can be repeated using these new data, as data accumulate from 2018.
[4]We focus our analysis on data revisions from 1983, since earlier data vintages were based on a release calendar that differs from the subsequent one. Data revisions are constructed from the real-time real GDP (vintage) dataset downloadable from the ONS website.
[5]Appendix S1 confirms this visual impression by estimating an econometric model of revisions that allows for time-variation in both the revision mean and its volatility.

But despite growing awareness by statisticians and economists of these and other data uncertainties, national statistical offices continue to emphasize GDP *point* estimates only, certainly in their headline data publications. In supporting documentation, and increasingly via linked real-time databases, they do acknowledge data uncertainties, but they do not directly quantify these for GDP growth in their data releases.[6] Manski (2015, 2016) has emphasized that the practice of acknowledging data uncertainties at best qualitatively or verbally, rather than quantitatively, is common across statistical offices. He has called for more transparent communication of GDP data uncertainties.

In this paper, absent direct quantitative communication by the statistical office, we undertake two empirical exercises to ascertain both how well understood data uncertainty is and how effectively data uncertainty could be quantified in real-time, should one wish to communicate it. We first consider probabilistic perceptions of UK GDP data uncertainty from the Bank of England, as reported quarterly since 2007 in their *Inflation Report*. Secondly, we elicit and then characterize the probabilistic perceptions of GDP data uncertainty from 100 experts by undertaking a specially designed online survey.

Our focus on data uncertainty is justified as follows. Firstly, the use of early GDP estimates, due to their limited data content and ensuing revisions, has been found to lead to misleading real-time views about the state of economy and the monetary policy stance (e.g. see Orphanides, 2001; Croushore, 2011). In turn, Clements and Galvão (2017) find that surprises to expected GDP revisions affect financial markets. Understanding experts' probabilistic perceptions of these early GDP estimates may therefore enable a better understanding of both the historical and desired ('optimal') relationship between policy decisions, financial markets and these early data releases. Certainly, some theoretical models, such as that of Aoki (2003), show that as data uncertainty increases policymakers should attenuate their responses to the data. It is also of note that the extensive recent literature on macroeconomic uncertainty and its impact on GDP (e.g. see Jurado, Ludvigson, and Ng, 2015) defines macroeconomic uncertainty to be high when it is harder to forecast the macroeconomic future. The fact that the Bank of England's 'fan chart' for GDP growth is almost as wide one quarter in the past as it is one quarter into the future therefore suggests that data uncertainty may be an important component of macroeconomic uncertainty.

Secondly, measurement of GDP data uncertainty is not straightforward. This is understood by following Manski (2015) and decomposing data uncertainty into 'permanent' and 'transitory' data uncertainty.[7] Permanent data uncertainty arises due to data incompleteness (e.g., survey non-response) or the inadequacy of data collection (e.g. sampling uncertainty due to a finite sample) and does not diminish over time. With a variety of surveys used to measure GDP, statistical offices do not in practice publish estimates of these sampling errors.[8] Transitory data uncertainty arises because data collection takes

[6]For example, the ONS subject their GDP estimates to periodic revisions analysis (see https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/articles/analysisofgdprevisionsinbluebooks2019/2020-03-13) and provide real-time databases at https://cy.ons.gov.uk/economy/grossdomesticproductgdp/datasets/realtimedatabaseforukgdpcomponentsfortheexpenditureapproachtothemeasureofgdp.

[7]Manski (2015) also distinguishes conceptual uncertainty which arises from a lack of understanding about what the statistics measure.

[8]To quote the ONS: '[t]he estimate of GDP ... is currently constructed from a wide variety of data sources, some of which are not based on random samples or do not have published sampling and non-sampling errors available.

time. Early GDP data are revised over time as new information arrives. In the UK, this also involves balancing contrasting estimates on the output, income and expenditure-side of GDP.[9] Transitory data uncertainty should decline as data accumulate. Statistical offices frequently publish analyses of past GDP revisions, and they emphasize revisions in their press releases and on their websites. However, their headline GDP estimates remain *point* estimates – with no accompanying quantitative measures of transitory and/or permanent uncertainty.

The remainder of this paper is structured as follows. Section II reports features of the Bank of England's probabilistic backcasts for GDP growth and discusses their interpretation. It compares the bank's estimates with model-based alternatives, that use historical GDP data revisions data to measure transitory data uncertainty. We then provide the first evaluation of the accuracy and calibration of these densities for historical GDP growth. In section III we gauge data uncertainty across a wider set of experts by conducting an online survey. We measure uncertainty around the latest GDP point estimate (at the time of running the survey) by eliciting probabilistic expectations about this GDP estimate in the form of subjective histograms. Section IV concludes. An Appendix S1 contains supplementary econometric results analysing the revisions properties of UK GDP data and undertakes robustness checks.

II. The Bank of England's 'fan charts' for historical GDP growth

An important example, and rare illustration, of how historical (real) GDP data uncertainty is both communicated in real-time and affects policymaking is provided by the Monetary Policy Committee (MPC) at the Bank of England.[10] Indeed, we are only aware of one other instance of regular public real-time communication of historical GDP data uncertainties, by the Riksbank in Sweden. As well as forecasting the future, in real-time the Bank of England provide direct estimates of uncertainty for past values of GDP growth via their well-known fan charts. These fan charts have been published each quarter since November 2007 in the Bank of England's *Inflation Report*.

While the literature has provided numerous analyses of the MPC's fan chart *fore*cast (Clements, 2004; Mitchell and Hall, 2005; Groen, Kapetanios, and Price, 2009; Galbraith and van Norden, 2012), previous research has neither characterized and drawn out features of their probabilistic forecasts of the past (their '*back* casts') nor evaluated their accuracy *ex post*. Doing so is a necessary first step both in understanding expert perceptions of data uncertainty and in assessing how accurate (useful) they are.

As such it is very difficult to measure both error aspects and their impact on GDP. While development work continues in this area, like all other G7 national statistical institutes, we don't publish a measure of the sampling error or non-sampling error associated with GDP'; see https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi.

[9]Developing the seminal work of Stone, Champernowne, and Meade (1942) that reconciled these contrasting estimates of 'true' GDP, a recent literature has explored how to reconcile GDP estimates measured on the income- and expenditure-sides acknowledging data uncertainties due to data revisions; e.g. see Jacobs *et al.* (2022) and Koop *et al.* (2022). Combining these contrasting estimates should reduce both transitory and permanent data uncertainties. Quantifying these is an interesting topic for future research.

[10]Strictly, the fan charts in the *Inflation Report* reflect the (collective) view of the (nine members of the) MPC not necessarily the views of the Bank of England.

Measuring mature GDP

Let y_t^{t+b} denote the ONS's estimate of year-on-year GDP growth for the *reference* quarter t published during the (*backcasting origin*) quarter $(t+b)$. The superscript denotes the vintage date or publication date in quarters $(b = 1, 2, \ldots, B)$. Therefore, for example, y_t^{t+1} denotes the ONS's ‘ preliminary’ (or first) estimate of year-on-year GDP growth for the reference quarter t published during quarter $(t+1)$. Over the period of study in this paper, ONS published this preliminary estimate towards the end of the first month of quarter $(t+1)$. Revisions, $\text{rev}_t^{(l-b)}$, between more mature data, y_t^{t+l} (where $l > b$), and the earlier bth estimate are then defined as $\text{rev}_t^{(l-b)} = (y_t^{t+l} - y_t^{t+b})$.

As data revisions are an ongoing process, there is understandably uncertainty about the appropriate value of l. In turn, this reflects uncertainty about what types of revision (cf. McKenzie, 2006) should be modelled and quantified. We focus in the main paper on y_t^{t+13} $(l = 13)$, as our measure of mature GDP. This is the GDP growth estimate for quarter t published by the ONS 3 years after the preliminary release. By this time, GDP growth estimates in the UK have gone through at least three annual (*Blue Book*) revisions at the ONS. We refer to y_t^{t+13} as the 13th quarterly estimate of UK growth, even though revised values may have not have been incorporated into all intermediate quarterly data releases, i.e. there may have been fewer than 13 revisions. Clements and Galvão (2012) adopt a similar approach when studying US GDP growth data revisions, making the assumption that after three annual revisions, revisions to growth are mainly benchmark revisions. Benchmark revisions, in general, are not modelled in data revision models based on the view that they are unpredictable; see Croushore (2011). Recent evaluations of UK data revisions performed by the ONS also consider revisions up to 3 years.[11]

In Appendix S1 we test the sensitivity of this assumption, by reporting results (summarised below) where mature GDP is defined as the estimate published by ONS after 4 years: $l = 17$. This reflects the fact, seen in Figure 1, that there are some major revisions even after 3 years, in particular since 2008. This impression is confirmed by higher average revision or bias estimates at 4 years relative to 3 years (cf. Tables S1 and S2).

Benchmark backcasts: Data-based historical fan charts

To help assess the judgemental contribution to the MPC's fan charts, we compare their features with an ‘unconditional’ model-based benchmark. Following the suggestion of Fixler, Greenaway-McGrevy, and Grimm (2014), we construct these benchmark density estimates of transitory data uncertainty assuming a normal distribution with the means and SDs estimated from historical data revisions alone (with revisions again considered up to 3 years, $l = 13$, so that the impact of benchmark revisions is mitigated). Clements (2018) has similarly argued for the use of unconditional benchmark densities.

An important characteristic of this benchmark is that it uses the statistical properties of past revisions to predict the likely path of future revisions. We do not include information

[11] See https://www.ons.gov.uk/releases/nationalaccountsarticlesanalysisofrevisionsinbluebooksandpinkbooks2019.

from quantitative predictors and/or expert judgement about the likely path; and the benchmark is not designed to capture permanent statistical uncertainties. As the MPC's backcasts are ultimately judgement-based, which as discussed in section II.3 below may involve trying to capture permanent as well as transitory data uncertainties, a comparison of their backcasts with the mechanically produced benchmark helps us evaluate this subjective aspect to the MPC's densities.

Benchmark (r) unconditional probabilistic backcasts for growth in *reference* quarter t, made in *backcasting origin* quarter $(t+b)$, are produced using historical data revisions data as follows:

$$f^{r}_{t|t+b} = N\left(\hat{y}^{t+b,r}_{t}, \hat{\sigma}^{2,t+b,r}_{t}\right), \tag{1}$$

where the moments of this Gaussian (N) density are recursively estimated from ONS revisions, $rev^{(l-b)}_{\tau}$, between the lth and the bth estimates:

$$\hat{y}^{t+b,r}_{t} = y^{t+b}_{t} + \hat{\mu}^{t+b,r}_{t}, \tag{2}$$

$$\hat{\mu}^{t+b,r}_{t} = \frac{1}{t-l}\sum_{\tau=1983Q2}^{\tau=t-l+1} \mathrm{rev}^{(l-b)}_{\tau}, \tag{3}$$

$$\hat{\sigma}^{t+b,r}_{t} = \sqrt{\frac{1}{t-l}\sum_{\tau=1983Q2}^{\tau=t-l+1}\left(\mathrm{rev}^{(l-b)}_{\tau} - \hat{\mu}^{t+b,r}_{t}\right)^{2}}, \tag{4}$$

$$\mathrm{rev}^{(l-b)}_{\tau} = y^{\tau+l}_{\tau} - y^{\tau+b}_{\tau}. \tag{5}$$

Importantly, in computing (1), we only use data revisions data that would have actually been available at each point in real-time. That is, this benchmark unconditional density is formed *ex ante*, rather than after having observed the latest revision(s). For example, $\hat{y}^{t,r}_{t-b}$ and $\hat{\sigma}^{t,r}_{t-b}$ are the mean and standard deviation computed in quarter t using revisions data, $(y^{t+l}_{t} - y^{t+b}_{t})$, for reference quarters up to $(t-(l-1))$; i.e. these backcasts are conditional on y^{t}_{t-b}, but the time-series of past revisions employed to compute the moments is only available up to l quarters ago, i.e. up to $\left(y^{t}_{t-l} - y^{t-l+b}_{t-l}\right)$.

We use revisions data back to 1983 to estimate (1). We did experiment with rolling windows of 5 years, as in Figure 1, to accommodate possible changes in the revision process, as discussed in the Introduction (and expanded upon in Appendix S1). But we did not find their use improved the accuracy of the unconditional backcasts; so here we focus on use of expanding windows of revisions data back to 1983. We also experimented with use of the econometric model used to model data revisions in Appendix S1; while its flexibility improves fit in-sample, its real-time accuracy was clearly inferior to the simpler benchmark, (1), that again we accordingly focus on here.

Features and interpretation of the MPC's historical fan charts

Figure 2 shows what a typical MPC fan chart looks like. It is taken from the February 2018 *Inflation Report*. Figure 2 shows that the fan becomes progressively narrower as one looks further back into the past, from the perspective of February 2018. This is to be expected, as the data revisions' process is more complete and fewer revisions are expected to be made in the future (in Figure 2, after February 2018) to these older more historical estimates that date back to 2013. The ONS's latest (as of February 2018) estimate of GDP growth is shown in Figure 2 by the solid black line. The fact that this line does not lie precisely in the middle of the fan chart reflects the MPC's perception that expected revisions (to the ONS's estimates) are non-zero. In Figure 2, the MPC expected GDP to be revised upwards, to the degree that the ONS estimate lies beneath the mean of the MPC's fan chart.

Looked at through the lens of the Manski (2015) classification, it is unclear whether the MPC are communicating 'transitory' and/or 'permanent' data uncertainty. The notes to Figure 2, in the MPC's words, explain that '[t]he fan chart depicts the probability of various outcomes for GDP growth ... To the left of the vertical dashed line, the distribution reflects the likelihood of revisions to the data over the past'. At first sight, this seems to suggest that the MPC are quantifying data revisions (i.e. transitory) data uncertainty alone. Under this interpretation, presumably if the MPC plotted their fan charts further back into the past than the 3–4 years shown in Figure 2, then any remaining transitory data uncertainty would disappear and the fan chart would collapse at a point mass on the bold line.

But other MPC documentation sheds some doubt on this interpretation. As one example, to quote from the November 2007 *Inflation Report* (p. 39): '(t)o the left of the first vertical dashed line, the centre of the darkest band of the fan chart gives the Committee's best collective judgement of the most likely path for GDP growth once the

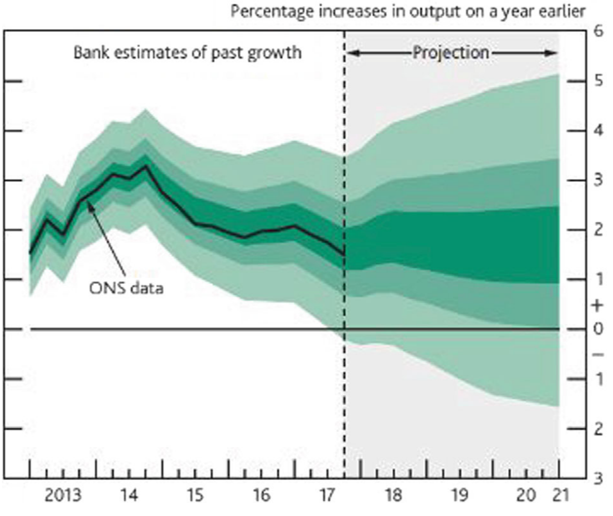

Figure 2. Illustrative fan chart for GDP growth (*Inflation Report*, February 2018)
Notes: Bank of England's fan chart for GDP growth (from the February 2018 'Inflation Report'). In their notes to this chart the Bank write: ''The fan chart depicts the probability of various outcomes for GDP growth ... To the left of the vertical dashed line, the distribution reflects the likelihood of revisions to the data over the past; to the right, it reflects uncertainty over the evolution of GDP growth in the future ... The fan chart is constructed so that outturns are also expected to lie within each pair of the lighter green areas on 30 occasions. In any particular quarter of the forecast period, GDP growth is therefore expected to lie somewhere within the fan on 90 out of 100 occasions''

revisions' process is complete.' As another example, also taken from the November 2007 *Inflation Report* (p. 39): the charts should be interpreted as 'the MPC's best collective judgement of the most likely path for the mature estimate of GDP growth, and the uncertainty around it, both over the past and into the future'.

These two quotes might be interpreted as suggesting that the backcast part of the fan chart quantifies the MPC's impression of what 'mature' GDP growth and its uncertainty will be – once the revisions' process is complete and 'transitory' uncertainty vanishes. Uncertainty around the 'mature' estimate of GDP then comprises 'permanent' data uncertainty. If the MPC thought that the mature data had no (permanent) uncertainty, then presumably the fan chart would collapse on a point mass. So the fact that the fan chart in Figure 2 has a positive variance, certainly 3–4 years into the past, suggests that the MPC is quantifying more than 'transitory' data uncertainty.[12]

Even though the MPC's density forecasts are two-piece normal (see Wallis, 2014), thereby allowing for asymmetries, their backcasts take the form of Gaussian densities. So we can characterise their features fully via examination of their mean and SD. We call the latter 'expected data uncertainty'.

Organising the density backcasts by *backcasting origin* to look at historical growth estimates made in quarter t, let:

$$f_{t-b|t}^{\mathrm{mpc}} = N\left(\hat{y}_{t-b}^{t}, \hat{\sigma}_{t-b}^{2,t}\right), \tag{6}$$

denote the MPC's density estimate for mature GDP growth for reference quarter $t-b$ made b quarters later ($b = 1, \ldots, (l-1)$) in quarter t (the *backcasting origin*). Note this means that the effective backcast horizon, h, declines with b, and is given as $h = (l-b)$. $f_{t-b|t}^{\mathrm{mpc}}$ are typically published near the beginning of the second month of quarter t. This means, to give an example when $b = 1$, that the MPC were (prior to the ONS changing its publication model for GDP in the summer of 2018) able to observe the ONS's latest 'preliminary' GDP estimate for the previous quarter y_{t-1}^{t}, along with their (perhaps revised) estimates for historical growth $y_{t-2}^{t}, \ldots, y_{t-B}^{t}$, before publishing their own historical estimates, $\hat{y}_{t-1}^{t}, \ldots, \hat{y}_{t-B}^{t}$ and $\hat{\sigma}_{t-1}^{t}, \ldots, \hat{\sigma}_{t-B}^{t}$, in the quarter t *Inflation Report*. In practice, the MPC tend to look $B = 16$ quarters back into the past when publishing their fan charts.

We can infer from Figure 2 that the MPC expects a considerable degree of uncertainty around (at the time) the ONS's latest estimate (of 1.5%) of GDP growth in 2017Q4. The SD of the GDP growth estimate in 2017Q4, as reported in the spreadsheets underlying this published fan chart, is 1.1%. To appreciate the size of this estimate note that, assuming Gaussianity and that the expected revision is zero, it implies the MPC is expecting the

[12]Cunningham and Jeffery (2007) and Cunningham *et al.* (2012) provide an explanation of the data revisions' model, used by bank staff, that along with MPC judgement helps inform the shape of these backcast fan charts. Their model exploits historical patterns in ONS revisions and information from qualitative business surveys to deliver backcasts of ' true' GDP growth. The model assumes that 'true' GDP is an unobserved variable and that estimates of GDP converge (as a point mass) on this true value as the number of revisions tends to infinity. This assumption implies that the model's 'true' GDP is not capturing the sort of permanent data uncertainties that Manski (2015) suggests do exist, even after the data revisions' process is complete. Rather we might interpret Cunningham *et al.* (2012) model as measuring the (latent) value 'true' GDP would attain absent both transitory and permanent data uncertainty, i.e. if there were no sampling errors and/or measurement errors.

'mature' value of GDP growth to fall, with a 95% probability, somewhere between −0.7% and 3.7%. So in early 2018 the MPC was actually uncertain whether the economy was growing or contracting 4−7 months ago (relative to one year prior to this).

To provide historical perspective, Figure 3 presents the MPC's characterisations of the 'expected revision' and expected data uncertainty, as extracted from the 2007Q3 to 2018Q1 *Inflation Reports*. The left panel presents for $b = 1, 4, 8, 12, 16$ their expected revisions, computed as the difference between the MPC estimate of GDP growth, $\hat{y}^t_{t-b}$, and the ONS estimate, y^t_{t-b}. The right panel plots the MPC's estimates of expected data uncertainty, $\hat{\sigma}^{\,t}_{t-b}$.

The left panel of Figure 3 reveals some interesting features about the MPC's expected revision for different data maturities. First, the MPC generally expected revisions to be positive – they consistently expected the ONS to revise upwards their estimates of GDP growth. The expected revision for a first GDP estimate ($b = 1$) is always positive; i.e. the MPC expected revisions to raise the initial ONS estimate of GDP growth. Secondly, they continue to expect non-zero revisions even for more mature data, implying that the MPC thought data revisions would change underlying GDP values even for heavily revised data. The expected revision becomes zero only for the 16th estimate ($b = 16$), and even then only from 2012Q3. Thirdly, the expected size of revisions has varied over time. This is consistent with Figure 1 and econometric evidence in Appendix S1 that the distribution of UK data revisions is time-varying. From 2012 we also see a decline in the absolute value of expected revisions. For the fan charts published in 2017, the expected revision values are all less than 0.3%. This is also in line with the decline in average revisions seen in Figure 1 for reference quarters from 2014 onwards. Re-organising the fan charts

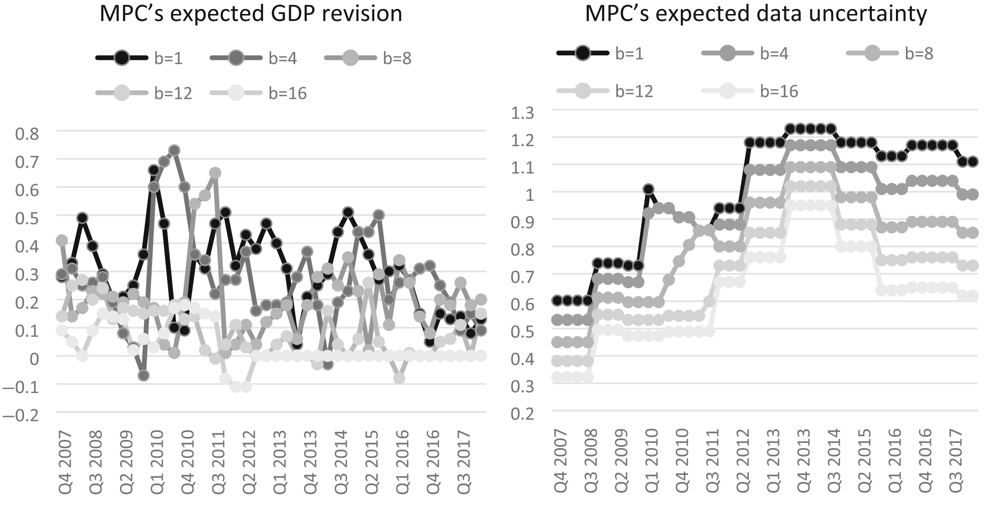

Figure 3. The expected revision and revision uncertainty (SD) for GDP growth – extracted from the Bank of England MPC's fan charts

Notes: The dates in the horizontal axis refer to the quarter where the MPC's fan charts were published and denote by t in the explanation that follows. The expected revision is $\hat{y}^t_{t-b} - y^t_{t-b}$; y^t_{t-b} is the ONS's estimate of year-on-year growth for quarter $t-b$ made in quarter t; and $\hat{y}^t_{t-b}$ is the MPC's prediction of 'mature' GDP growth for quarter $t-b$ published later in quarter t. Expected uncertainty refers to the MPC's *ex ante* prediction made in quarter t of the uncertainty (SD) of the 'mature' value of GDP growth for reference quarter $t-b$

by reference quarter (to analyse $\hat{y}_t^{t+b}$), in Figure S2 we see the direct implications of the revisions by plotting the Bank's evolving expectations of mature GDP growth in quarter t as estimated b ($b = 1, 4, 8, 12$) quarters later. We see that the MPC's view of the onset of the recession in 2008 has changed. We also see an upward revision in their growth rate estimate for 2012, at the time of the much-publicized but vanishing double-dip recession referred to in the Introduction.

Turning to the right panel of Figure 3, we firstly see that the MPC made changes to its expectations of data uncertainty in a more discrete manner. Changes tend to occur for the Q3 value of GDP growth (as published by the Bank of England in November) following publication of the *Blue Book* by the ONS. The *Blue Book* publication typically involves extensive annual revisions to the national accounts. Secondly, consistent with the transitory characteristics of GDP data uncertainty explained by data revisions, Figure 3 shows that the MPC expect data uncertainty to decrease with the maturity of the data; i.e. uncertainty decreases with b. Finally, it is evident from Figure 3 that the MPC has become more uncertain over time. Expected data uncertainty, for a given b, tended to double between 2007 and 2018. This is consistent with the rise in the data revisions' SD seen in Figure 1. Note, however, that the decline in expected data uncertainty for all maturities seen in Figure 3 from 2015 onwards is not as substantial as the decline indicated by the historical analysis in Figure 1. We explore further the differences between these MPC's perceptions of data uncertainty and uncertainty estimates formed from past data revisions data alone in section II below.

Confidence intervals for 'mature' GDP growth

To further understand perceptions of historical data uncertainty from the Bank of England's MPC, Figure 4 plots 68% confidence intervals (equivalent under Gaussianity to one SD bands) for 'mature' GDP growth extracted from their published backcast density $f_{t|t+b}^{\text{mpc}}$ at $b = 1, 6, 12$. Note that we are now re-organising (6) by reference quarter, as in (1), to enable the interval forecasts implied by $f_{t|t+b}^{\text{mpc}} = N(\hat{y}_t^{t+b}, \hat{\sigma}_t^{2,t+b})$ to be compared with the 'mature' out-turns, y_t^{t+l}.

The figure also includes 68% confidence intervals from our unconditional benchmark density ($f_{t|t+b}^{r}$) defined in (1). We superimpose on Figure 4 the 'mature' estimate at $l = 13$, y_t^{t+13}. We order the plots in Figure 4 from the shortest to the longest backcast horizons. So we might expect the intervals to widen, as they do, as we look down from the first to the second panel in Figure 4, as then, in effect, we are inspecting longer horizon backcasts, about which there is more uncertainty.

Figure 4 indicates that the MPC's intervals are consistently wider than the benchmark density, particularly since 2012. Looking furthest back into the past (to $b = 12$), we see that the MPC's intervals are in fact always wider. This is consistent with the MPC either over-estimating transitory data uncertainty, relative to the benchmark model, or seeking to capture some aspect of permanent, as well as transitory, statistical uncertainty in their historical fan charts. In advance of our more formal evaluation of the accuracy of these probabilistic backcasts in the next section, we note that the MPC's intervals in general appear 'too wide' as b increases: they perceive 'too much' data uncertainty. While the mature GDP estimate (the *outturn*) does fall within the 68% interval on 68% of

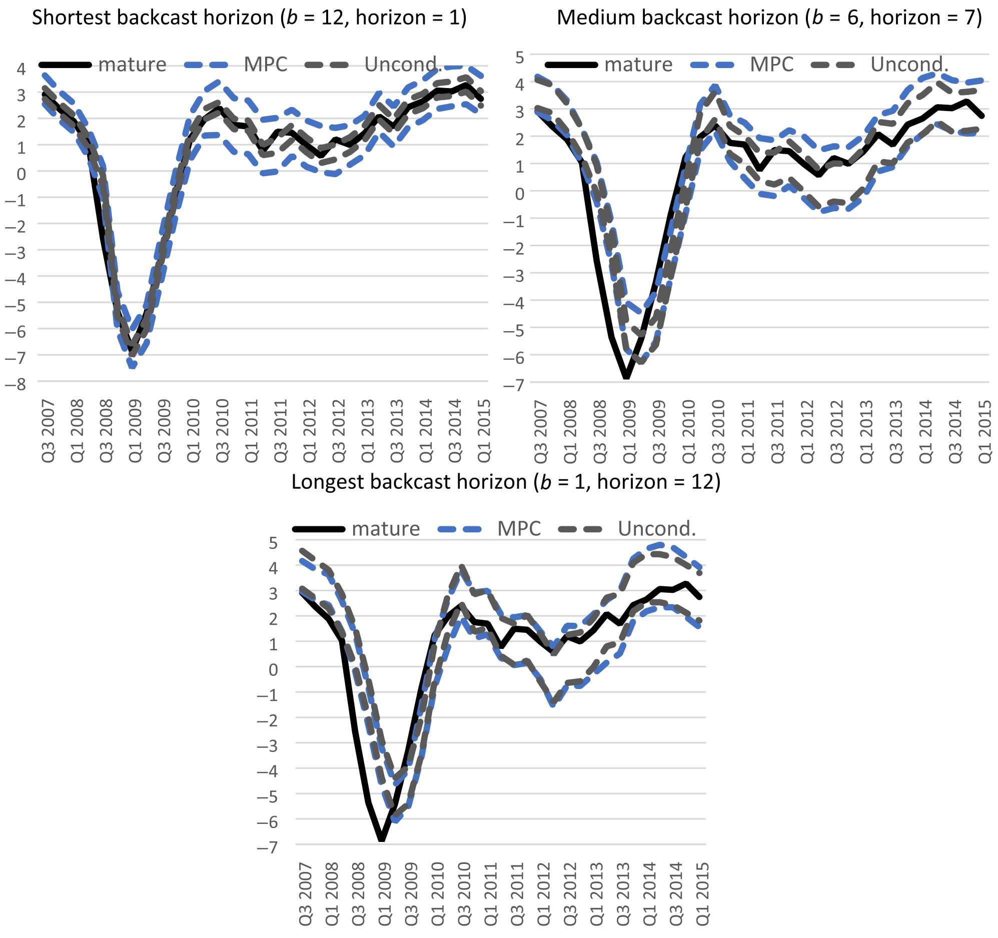

Figure 4. Bank of England (MPC) and unconditional 68% ex ante confidence intervals for three backcast horizons, alongside the ex post 'mature' ONS values

Notes: The confidence intervals are computed using data up to $t + b$ with $b = 12$, $b = 6$ (both upper panel) and $b = 1$ (lower panel) for the mature values of each reference quarter in the horizontal axis. The 'mature' values are the GDP growth estimate published by the ONS 3 years ($l = 13$) after their first estimate. The backcast horizon is then $13 - 12 = 1$ and $13 - 6 = 7$ in the middle panel and $13 - 1 = 12$ in the lower panel

occasions when only a first estimate of GDP growth is available ($b = 1$), as b increases the mature GDP out-turns increasingly fall within the 68% interval. They fall inside on 74% of occasions when $b = 6$ and on 97% of occasions when $b = 12$. The unconditional intervals, that use information on past revisions only to assess the degree of data uncertainty, are still too wide. But they are narrower. Their *ex post* coverage rates are 61% ($b = 1$), 55% ($b = 6$) and 87% ($b = 12$).

Evaluating the historical fan charts

To evaluate, respectively, the absolute and relative accuracy of the probabilistic backcasts from the MPC, we evaluate their *ex post* calibration and compare them against the *ex ante* unconditional benchmark density seen in (1). We emphasise that we can only evaluate real-time perceptions of transitory uncertainty (due to data revisions), given that statistical offices do not publish sampling errors for GDP estimates, an important source of permanent data uncertainty. Our evaluation sample is relatively small, as the MPC have produced their forecasts on a quarterly basis for only 20 years. This should be borne in mind when interpreting our evaluation results.

Assessing the calibration of the backcast densities

We use probability integral transforms (PITs) to assess the absolute calibration of the MPC and unconditional benchmark backcast densities relative to the mature estimates of GDP. Under the null hypothesis of correct calibration, such that the backcast densities accurately assess transitory data uncertainty and characterise the underlying density generating the mature data, the empirical cumulated PITs (calculated by integrating, over time, the backcast density up to the realised mature estimate) should fall on the 'theoretical' 45-degree lines shown in Figure 5. As a consequence, visual inspection of the cumulated PITs helps us understand if the MPC and the benchmark unconditional model did, in real-time, correctly quantify data uncertainty due to future data revisions. Deviations of the empirical Cumulative Density Function (CDF) from the 45-degree line can, in turn, help identify reasons for calibration failure.

We complement visual inspection of the cumulated PITs with more formal tests for the null hypothesis of correct calibration. We do so using the critical values derived by Rossi and Sekhposyan (2019) for the Kolmogorov–Smirnov test that the empirical PITs do all fall on the 45-degree line. These critical values are presented in Figure 5 as 95% intervals. The width of these intervals is inversely proportional to the sample size. The bands are therefore quite wide in our application. An attractive feature of the Rossi and Sekhposyan (2019) test is that it can accommodate serial dependence in the PITs. For backcasting horizons where $l - b > 1$ we should expect serial dependence in the PITs, even for correctly calibrated densities, given the overlapping sample implied by, what are in our application, multi-step-ahead backcasts.[13] Accordingly, in Figure 5 we use the block bootstrap recommended by Rossi and Sekhposyan (2019) to compute the critical value bands, except when $l - b = 1$ as then standard critical values are valid.

Figure 5 plots the cumulated PITs at backcasting horizons $b = 1, 6, 12$ for the MPC and unconditional backcasting densities when compared against mature estimates of GDP, y_t^{t+13}. Figure 5 shows that both MPC and unconditional backcasting densities are pretty well calibrated for more recent data (when $b = 1$ and $b = 6$), in the sense that the empirical cumulated PITs do not fall outside the critical value bands. Their calibration is also relatively similar. But the S-shape of the cumulated empirical PITs at $b = 12$ indicates that when thinking about older data, the MPC either overstate transitory data uncertainty or attempt to quantify permanent as well as transitory uncertainties. This evidence of under-confidence confirms the visual impression from Figure 4 showing that the MPC tend to perceive, in real-time, 'too much' data uncertainty, especially for older data (i.e. as b increases).

[13] A well-calibrated ('optimal') one-step-ahead point (or density) backcast/forecast should still deliver serially independent back/forecast errors (or PITs). It is at backcast/forecast horizons greater than one that we expect dependence, even under the null of correct calibration.

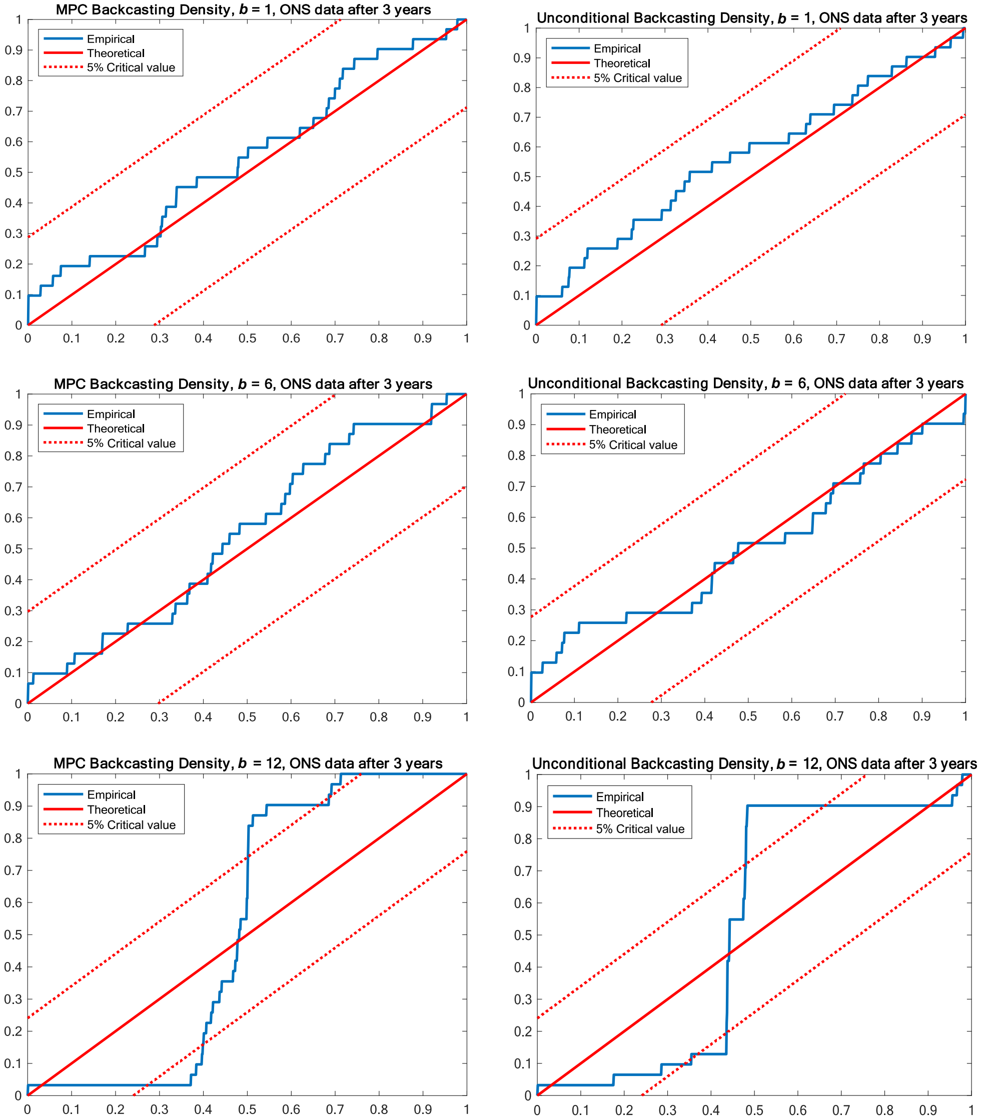

Figure 5. Evaluation of MPC and unconditional benchmark density backcasts against ONS data published 3 years after the first release (y_t^{t+13}): Empirical PITs CDF, with critical value bands from Rossi and Sekhposyan (2019)

Notes: If the empirical PITs (in blue) are outside the 5% bands, then the null of correct calibration (specification) of the predictive density is rejected. Critical values are computed via bootstrap (accounting for the serial correlation of multi-step-ahead predictions)

Figure S5 shows that the MPC's perceptions of data uncertainty are better calibrated for older data (i.e. when $b = 12$) if we assess calibration against those GDP values published by the ONS 4 years after the first release. The cumulated empirical PITs are then closer to the 45-degree line. Recall that $b = 12$ represents a $h = 1$ prediction in Figure 5, but is a $h = 5$ prediction in Figure S5. Figure S5 also reveals greater differences between the MPC and unconditional densities. The MPC's judgement appears to hinder at

$b = 1$. But it helps at $b = 6$ and $b = 12$, in the sense that the MPC's PITs are then closer to the 45-degree line than those from the unconditional density. We saw in Figure 4 that the MPC perceive more data uncertainty than the unconditional density, especially as b increases.

Relative performance

To compare the MPC's real-time perceptions of data uncertainty directly against the unconditional benchmark, we evaluate relative performance across three loss functions. These loss functions are designed to evaluate different aspects of probabilistic performance. We use the root mean squared forecast error (RMSE) to evaluate the accuracy of the mean estimates from the MPC and benchmark densities; we also report the RMSE of the ONS's own earlier estimates, y_t^{t+b}, against y_t^{t+l} (for each $b < l$).

Then we measure the accuracy of the density estimates, $g_{t|t+b} = f_{t|t+b}^{\text{mpc}}$ or $g_{t|t+b} = f_{t|t+b}^{r}$, over all possible events on the support of the density using both the logarithmic score (logscore) and the continuous ranked probability score (CRPS):

$$\log \text{score}_{t|t+b}^{t+l} = -\log g_{t|t+b}\left(y_t^{t+l}\right), \tag{7}$$

$$\text{CRPS}_{t|t+b}^{t+l} = \int_{-\infty}^{+\infty} \left[G_{t|t+b}(y) - I(y_t^{t+l} \leq y)\right]^2 dy, \tag{8}$$

where $G_{t|t+b}(.)$ is the CDF associated with the density forecast $g_{t|t+b}(.)$ and $I(y_t^{t+l} \leq y)$ denotes an indicator function equal to one if $y_t^{t+l} \leq y$ and zero otherwise. Diebold and Mariano (1995)-type t-statistics for equal forecast accuracy of two competing forecasts are computed for each of the three loss functions using Newey and West (1987) SEs.[14]

While, for easy-reading, in the table that follows we place rejections of the null of equal forecast accuracy in bold type face, when significant at the 10% level, we emphasise that these tests should be taken merely as a general guide. They likely have poor (joint) size and power properties. This is because: (i) we are undertaking multiple t-tests individually; (ii) of the overlapping nature of the data (when $b > 1$); and (iii) we cannot be confident of the size of the Newey-West correction in our small sample. Nevertheless, as we will explain, we believe it is reassuring that, despite this qualification, the assessment that follows is consistent with that in sections II.4.1 and II.4.3 (below).

Table 1 first reports the RMSE statistics of the ONS's own earlier estimates alongside those of the MPC and the unconditional density. Accuracy is again measured against y_t^{t+13} with supplementary tables for y_t^{t+17} in Appendix S1 (see Table S4). From Table 1 we see, as expected, that for the ONS these RMSE estimates decrease as b increases. Comparison with Table S4 shows that accuracy tends to be better against y_t^{t+17} than y_t^{t+13}. But looking at the RMSE ratios in Table 1, we see that the mean estimates from the MPC provide more accurate point estimates of mature ONS data than the ONS's own earlier estimates. The

[14]The Newey–West SEs are computed using a truncation lag of 2, for all b, as suggested by what Lazarus *et al.* (2018) call the 'NW textbook' rule of thumb, i.e. $0.75T^{1/3}$. Our sample size precludes consideration of much larger lag lengths.

TABLE 1

Point and density accuracy and tests for equal accuracy: Evaluation of the relative performance of ONS, Bank of England (MPC) and unconditional backcasts against 'mature' GDP data observed 3 years after the first ONS data release

	Point					*Density*					
	ONS	*MPC*		*Uncond.*		*MPC*	*Uncond.*		*MPC*	*Uncond.*	
b	*RMSE*	*Ratio*	*t-stat*	*Ratio*	*t-stat*	*logsc*	*Diff.*	*t-stat*	*CRPS*	*Ratio*	*t-stat*
1	1.272	0.979	−0.305	1.056	0.551	2.18	0.14	0.96	0.68	1.06	1.41
2	1.183	0.952	−0.645	1.032	0.369	1.94	0.29	1.07	0.62	1.07	1.24
3	1.195	0.923	−1.263	1.023	0.318	1.96	0.28	1.08	0.60	1.08	1.25
4	1.174	0.940	−0.940	1.011	0.223	1.94	0.92	1.29	0.60	1.08	1.01
5	1.126	0.956	−0.619	1.013	0.299	1.78	0.89	1.27	0.58	1.07	0.91
6	1.087	0.966	−0.479	1.018	0.413	1.72	1.00	1.29	0.56	1.08	0.98
7	0.974	0.970	−0.382	1.019	0.440	1.57	1.13	1.23	0.50	1.05	0.59
8	0.839	1.001	0.014	1.030	0.722	1.40	0.72	0.96	0.45	0.99	−0.22
9	0.739	1.061	0.820	1.040	0.961	1.34	0.59	0.66	0.41	0.89	**−1.93**
10	0.722	1.061	0.834	1.028	0.967	1.31	0.68	0.67	0.39	0.87	**−2.17**
11	0.654	1.020	0.537	1.017	0.862	1.18	0.81	0.64	0.34	0.82	**−2.28**
12	0.418	1.011	1.172	1.011	0.811	0.90	0.60	0.39	0.25	0.60	**−2.91**

Notes: Point accuracy is measured by the RMSE. The RMSE estimates for the mean estimates from the Bank of England MPC and unconditional densities are reported as ratios relative to the ONS RMSE. Ratios <1 indicate improved point forecast accuracy relative to the ONS's own first estimate. *t*-stat is the *t*-statistic of the null hypothesis of equal accuracy with the ONS under the null (or the no change forecast). t-statistics in bold if statistically significant at a 10% level. Density accuracy is measured using the log-score (logsc), equation (7), and CRPS, equation (8). Statistics for the unconditional densities are reported relative to the MPC. For the log-score this is a difference (Diff.), so that positive values indicate the unconditional benchmark has worse performance than the MPC. For the CRPS, ratios are reported, so that values >1 indicate superior performance of the MPC. *t*-stat is the *t*-statistic of the null hypothesis of equal accuracy, for the chosen loss function, of the MPC and unconditional benchmark densities, such that negative statistics indicate superior performance for the unconditional benchmark. *t*-statistics are in bold are statistically significant at a 10% level. Evaluation period: 2007Q3–2015Q1.

MPC has been correct historically, as shown in Figure 3, to expect ONS first-release GDP data to be revised upwards over time, as seen from the historical time-series of revisions (cf. Figure 1 and Table S2). This holds for the shorter horizon backcasts (up to $b = 8$). (It holds for all values of b when the outcome is y_t^{t+17}). The RMSE gains for the MPC are typically between 3% and 8%, but are statistically insignificant (they are significant on just three occasions for y_t^{t+17}). The mean estimates from the benchmark statistical model are not as competitive as those from the MPC or ONS. However, they do outperform the ONS's own early estimates for y_t^{t+17}, with RMSE ratios less than one 14 out of 16 times, although these gains are never statistically significant.

Table 1 also reports the logscore and CRPS statistics, for a given b, averaged over the evaluation sample. Smaller values of both statistics indicate more accurate probabilistic backcasts. The logscore and CRPS statistics for the benchmark unconditional density are reported relative to those of the MPC.

Table 1 shows that accuracy increases further back into the past; i.e. the logscore and CRPS values decrease as b increases (and the horizon correspondingly decreases). This is as we might expect, as the MPC is having to predict fewer future data revisions as b increases. Comparing against the unconditional benchmark, we see that the MPC is always more accurate according to the logscore. But these gains are never statistically significant. Using the CRPS to compare the MPC and unconditional densities, we again see that the MPC is more accurate for smaller b. But for older data, $b > 8$, we find that the unconditional density is more accurate. These gains are statistically significant for $b = 9, \ldots, 12$. This is consistent with us finding that the MPC tend to overestimate data uncertainty for older data (i.e. for higher values of b). But, emphasising the importance of how we define 'mature' data, when evaluated against more revised data, $l = 17$, the MPC densities become competitive again against the unconditional density (see Table S4). The MPC's density backcasts are better at anticipating the ONS GDP estimate published 4 years after the first release, than the one published one year earlier ($l = 13$).

Overall, we find that the MPC provide more accurate, although usually not significantly so, point estimates of mature GDP values than the comparably timed estimates published by the ONS. These gains are observed more frequently if the 'mature' estimate is defined as the ONS GDP estimate published 4 years after the first estimate. Relative to the unconditional benchmark density, and looking across the logscore and CRPS loss functions, the MPC provide a similar level of accuracy, except about the older historical data. This suggests that, except for these older data, MPC judgement on data uncertainty is informed by models of data revisions. The fact that the MPC overstate data uncertainty for older data is consistent with them aiming to quantify permanent as well as transitory data uncertainties.

This evaluation serves to illustrate that historical data-based and more judgemental methods, as used by the MPC, can be used to reliably measure GDP transitory data uncertainties. But deviations of the empirical PITs from the 45-degree line, especially for older data, are a reminder that measurement of uncertainty remains a challenge. Given our earlier discussion of how the statistical properties of GDP data revisions have changed over time, we now examine whether there are temporal variations in the accuracy of the MPC and unconditional benchmark densities.

Time-variation in relative accuracy

Figure 6 breaks down the average CRPS statistics, reported in Table 1, by plotting their evolution over time, by reference quarter, t. Thereby, we assess whether the accuracy of the MPC density has changed over time relative to the unconditional benchmark density.

The left panel of Figure 6 plots the CRPS estimates for the earlier density estimates ($b = 1, 4$); while the right panel considers the later estimates ($b = 8, 12$). Both panels of Figure 6 clearly show that the accuracy of both the $f^{\text{mpc}}_{t|t+b}$ and $f^{r}_{t|t+b}$ densities deteriorates substantially during the recessionary period, 2008-09. We also see that it is during 2008−09 that we observe more differences between the two densities, with $f^{\text{mpc}}_{t|t+b}$ delivering gains.

This serves as additional evidence that the data uncertainty information communicated in the MPC's fan chart captures more than past data revisions, at least as captured by the unconditional benchmark. Figure 6 suggests that this supplementary information is particularly helpful during the 2008−09 turbulent period.

To complement this analysis, and assess the extent to which this time-variation in realised data uncertainty, as captured by the CRPS, represents 'skill' on the behalf of the MPC, we next study 'scaled resolution'.

Scaled resolution is measured, following Hersbach (2000) and Rossi, Sekhposyan, and Soupre (2016), by decomposing the CRPS into the sum of three components:

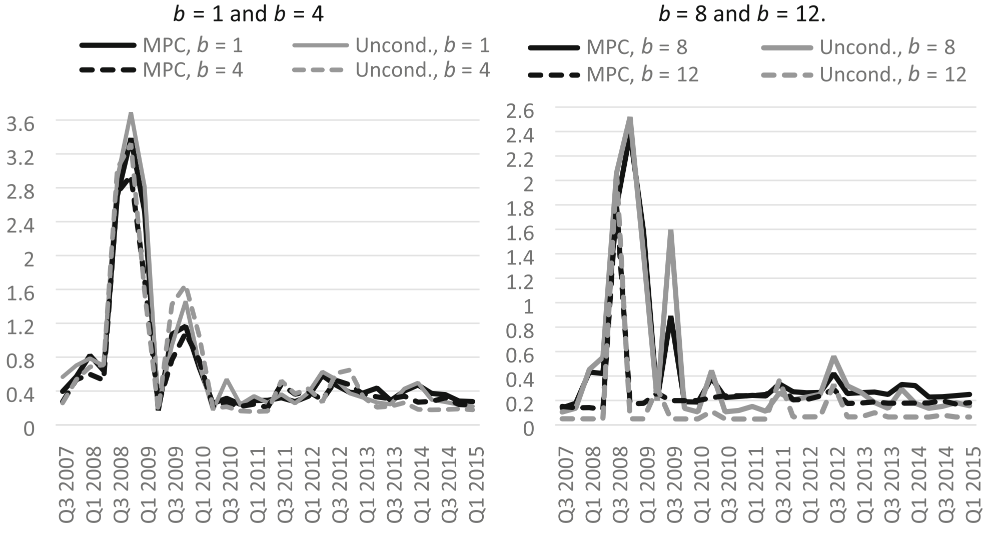

Figure 6. CRPS statistics for the Bank of England (MPC) and unconditional backcast densities against the ONS data published 3 years after the first release (y_t^{t+13})
Notes: Backcasts are computed using GDP releases up to $t+b$. Dates in the horizontal axis are reference quarters as in Figure 4, instead of the *Inflation Report* release dates from Figure 3)

reliability (the negative of) resolution and entropy (or the uncertainty inherent in the outcome variable). Scaled resolution equals resolution divided by entropy. As proposed by Galbraith and van Norden (2012), scaled resolution achieves its maximum at unity, when the probability assessments accurately identify and discriminate between low and high probability outcomes. This three-component decomposition of the CRPS, and scaled resolution itself, are understood by noting, as shown in Hersbach (2000) and Rossi *et al.* (2016), that the CRPS is the integral of the well-known Briers score over all possible probability-event thresholds, the y seen in equation (8). The Briers score, used to measure the accuracy of forecasts of binary outcomes, is commonly decomposed into these three components (albeit they are sometimes named differently) to aid interpretation of probability forecasts; see Murphy (1973). As Galbraith and van Norden (2012) explain, 'reliability' measures calibration, namely the consistency between the probability assessments and the outcomes: our PITs analysis in section II.4.1 provided one test of density calibration. 'Resolution', in turn, measures the ability of the MPC and benchmark unconditional probability backcasts, that are both formed *ex ante*, to capture time-variation in the true data density, relative to the *ex post* unconditional density. Computationally we follow Rossi *et al.* (2016), gratefully acknowledging use of their Matlab code, and in Figure 7 plot scaled resolution computed over rolling windows of four quarters.

Figure 7 shows that the scaled resolution of the MPC backcasts falls during the 2008−09 recession and again during the aforementioned vanishing double-dip recession of 2012. But the MPC's scaled resolution is higher than the unconditional benchmark during 2008-09. So the greater accuracy of the MPC over this period, as shown by the lower CRPS values in Figure 6, is associated with more 'skill' on their behalf. On average (looking across the whole sample period), resolution is higher for the MPC than the unconditional benchmark for newer data ($b=1,4$), but it is lower for more mature

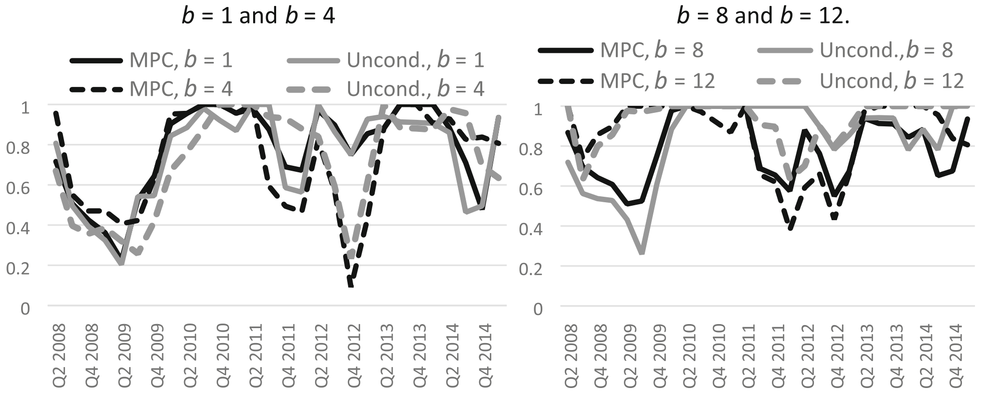

Figure 7. Scaled resolution for the Bank of England (MPC) and unconditional backcast densities. *Notes*: 'Mature' data is defined as (y_t^{t+13}). Scaled resolution = (resolution/entropy), as suggested by Galbraith and Van Norden (2012). Scaled resolution equals unity at the maximum resolution (resolution = entropy). Resolution measures how well the predictive density captures time variation in the underlying data density. Computed over rolling windows of four observations, as in Rossi *et al.* (2016). Dates refer to the last observation in the window

data ($b = 8, 12$).[15] Interestingly, in their analysis of the scaled resolution of the MPC's forecasts, Galbraith and van Norden (2012) find more evidence for resolution when the MPC's forecasts are evaluated against data realisations that have been through fewer rounds of revision. The fact that we find MPC resolution averages more at ($b = 8, 12$) than at ($b = 1, 4$) indicates that the MPC's backcasts also have better resolution when evaluated against outturns published sooner rather than later after the backcast.

III. Experts' perceptions of data uncertainty: a case study

In this section, we present the results of a survey we undertook to gauge perceptions of GDP data uncertainty across a wider set of experts than the MPC. We asked a specially commissioned sample of survey respondents to provide their probabilistic assessments of real GDP growth, having first reminded them of the ONS's latest quarterly growth point estimate for 2018Q3 (the latest estimate at the time of running the survey). Based on their reported individual histograms, we compute expectations for both the mature GDP growth estimate and data uncertainty.

Similarly to the MPC's perceptions of data uncertainty, these experts' probabilistic assessments of data uncertainty are formed without any direct quantitative communication of data uncertainty by the statistical office. All the respondents were told was the ONS's latest GDP point estimate, alongside the accompanying ONS press release. Some experts may have read the latest Bank of England *Inflation Report*. At the time of running the survey, the latest fan chart (from the February 2019 *Inflation Report*) indicated that the expected mature (real) GDP value was equal to the current ONS estimates, implying that the expected revision was zero. Data uncertainty about the 2018Q3 GDP estimate

[15]Averaged over the sample shown in Figure 7, the scaled resolution of the MPC for $b = 1, 4, 8$ and 12 is 0.78, 0.73, 0.79 and 0.85, respectively. Equivalent values for the unconditional benchmark are 0.75, 0.72, 0.83 and 0.93.

of 1.5% was 1.1%. Our case-study survey therefore lets us assess whether other experts' perceptions of data uncertainty are in line with MPC views, or whether they take ONS point estimates of GDP at face-value and do not perceive any data uncertainty. As mentioned above, the ONS changed its publication model for GDP in the summer of 2018. This structure change may contribute to additional uncertainties when the experts quantified their impressions of data uncertainty. As when the experts replied to our survey, there was insufficient data since the structure change for anyone to form a meaningful data-based assessment of its implications for GDP data uncertainty.

Survey details

We conducted a targeted online survey of more than 100 experts. These experts are professionals (many of whom are economists), working mainly in government institutions, industry and academia. The survey was aimed at maximising the number of respondents across a range of expert user groups (industry, government institutions and academia), rather than ensuring representativeness.

The survey asked about data uncertainty perceptions for the ONS's latest GDP point estimate; in effect, this means the experts are asked about GDP growth when $b = 1$. At the time of running the survey, in early 2019, this concerned the GDP estimate for 2018Q3 published by ONS on the 9 November 2018. The online expert survey was disseminated through the ESCoE (Economic Statistics Centre of Excellence) email list, social media particularly Twitter and emailing personal contacts and asking them to forward to colleagues. The recruitment period lasted for 4 weeks, between 18 February and 17 March 2019. The survey received 104 completed responses.

After collecting a range of background data, the online survey informed respondents that:

`On 9th November 2018, the ONS published its latest GDP first quarterly estimate: 'UK gross domestic product (GDP) in volume terms is estimated to have increased by 0.6% between Quarter 2 (Apr to June) and Quarter 3 (July to Sept) 2018. Compared with the same quarter a year ago, the UK economy has grown by 1.5%'.`

The experts were then asked to '`provide (best-guess) estimates of the percentage probabilities you would attach to various outcomes for GDP growth. The probabilities should sum to 100%`'. The outcomes were GDP less than 0%, 0% to 0.5%, 0.5% to 1%, 1% to 1.5%, 1.5% to 2%, 2% to 2.5%, 2.5% to 3% and more than 3%.

This sort of probabilistic/histogram question, as suggested by Manski (2004) and popular in the Surveys of Professional Forecasters run by the Philadelphia Fed in the USA and the European Central Bank in Europe, facilitates interpersonal comparisons of uncertainty. This contrasts questions that elicit qualitative uncertainty statements.

From the background questions we learn that most experts are regular users of GDP statistics. 74% used GDP and national account statistics during the past 12 months. Most experts use GDP statistics either quarterly (23%), monthly (25%) or weekly (18%). The expert survey covers all age brackets from 18, but only 29% of the sample identified as female. The most represented employment sectors are academia and research (32%), ONS and Bank of England (17%), Government departments (15%) and private business (10%). We do not find any evidence that perceptions of data uncertainty vary by these characteristics, although our survey is too small and not designed to explore what explains experts' perceptions of data uncertainty. Rather it is designed to measure experts' quantitative perceptions of UK GDP growth data uncertainty.

Perceptions of the expected revision and of data uncertainty

Given that each expert characterises data uncertainty via a histogram, which raises questions about how to quantify the uncertainty, we construct three measures of uncertainty from our sample of histograms using both non-parametric and parametric methods. We then: (i) relate these measures of data uncertainty to experts' point expectations; (ii) assess the heterogeneity across experts; and (iii), by comparing results across the different estimation methods, examine whether experts' perceptions of data uncertainty are symmetric.

Specifically, the non-parametric method lets us estimate the mean and SD of each individual's reported histogram without making specific parametric assumptions about any underlying continuous density that the experts may subjectively have. But, as the first and last intervals are open-ended, an assumption is still required about the range over which the individual histograms are defined. Following Abel *et al.* (2016) and others, we assume that the first and last intervals have a length double that of the central intervals. Results are not especially sensitive to this assumption. Following Zarnowitz and Lambros (1987), we assume that the probability mass is uniformly distributed within each interval rather than concentrated at the midpoint of each interval, although results are again robust to this.

The mean, μ_i, and SD, σ_i, of individual i's histogram are estimated as:

$$\mu_i = \sum_{j=1}^{8} \left(\frac{(u_j - l_j)}{2} \right) p_{i,j}, \tag{9}$$

$$\sigma_i = \sqrt{\left[\sum_{j=1}^{8} \left(\frac{\left(u_j^3 - l_j^3\right)}{3\left(u_j - l_j\right)} \right) p_{i,j} - \left[\sum_{j=1}^{8} \left(\frac{\left(u_j^2 - l_j^2\right)}{2\left(u_j - l_j\right)} \right) p_{i,j} \right]^2 - \frac{w^2}{12} \right]}, \tag{10}$$

where u_j and l_j the upper and lower limits of the jth interval, w is the width of the central intervals (0.5 percentage points in our case) and $p_{i,j}$ is the probability that forecaster i assigns to the jth interval ($j = 1, \ldots, 8$). The last term in the formula for σ_i is the commonly applied Sheppard correction for the variance.

We contrast these nonparametric estimates with estimates assuming that each expert's underlying density takes a specific form. Firstly, following Giordani and Soderlind (2003), we assume that experts' underlying probabilistic beliefs are Gaussian. For each expert, i, we estimate the mean and variance of their Gaussian distribution by minimising:

$$\min_{\mu_i,\sigma_i^2} \sum_{j=1}^{8} [F_n(t_j;\, \mu_i, \sigma_i^2) - F_i(t_j)]^2,$$

where $F_n(t_j;\, \mu_i, \sigma_i^2)$ is the CDF of the Gaussian and $F_i(t_j)$ are the observed cumulated histogram data, for each expert, at each of the $t_1, \ldots, t_8$ right endpoints of the eight intervals used in the histogram question: $F_i(t_J) = \sum_{j=1}^{J} p_{i,j}$ We follow Engelberg, Manski, and Williams (2009) and for those six experts (just under 6% of our sample of experts) that reply to the histogram question by assigning their non-zero probabilities to just two (adjacent in our case) intervals, we fit triangular distributions that provide symmetric characterizations of the underlying distributions. No expert in our survey saw data uncertainty as confined to a single bin.

Secondly, following Engelberg *et al.* (2009), we assume that experts' subjective densities are a member of the generalised Beta family and we again use the histogram data to fit the parameters. The generalised Beta is a Beta distribution, defined by two parameters (a and b), scaled to have support (l, r), where l and r are two additional parameters defining the left and right bounds. The two shape parameters, a and b, allow for considerable flexibility in characterising expert's perceptions of data uncertainty. While we enforce unimodality, via the restriction that $a > 1$ and $b > 1$ (an assumption supported by our histogram sample, given that there is no evidence of multimodality), unlike the Gaussian density the Beta allows for possible asymmetries in experts' perceptions of data uncertainty. When an expert attaches non-zero probabilities to interior intervals only, l and r are set equal to the left and right endpoints of the intervals with positive probability. But when there is mass in either or both outer intervals, as in Engelberg *et al.* (2009), we treat l and/or r as free parameters to be estimated.

Before analysing these three contrasting estimates of data uncertainty, and looking for evidence for asymmetry, we note that of our 104 experts, 17% reply to the histogram question in even or odd multiples of 10. A further 29% reply in even or odd multiples of 5. This suggests that nearly 40% of experts are not sufficiently confident to give precise numerical values for their subjective probabilities of data uncertainty. Rounding is a common feature of surveys that elicit probabilistic expectations (e.g. see Manski and Molinari, 2010). It can be interpreted as one manifestation of uncertain uncertainty.

Turning to the experts' perceptions of data uncertainty, the top three panels of Figure 8 plot, for each expert, their mean and SD estimates – as estimated from the reported histograms using the three different estimation methods. Figure 8 shows that these experts, like the Bank of England's MPC, do think probabilistically. They too expect GDP data uncertainty. But perhaps understandably, given experts were given no explicit guidance from ONS or us (in the survey) about uncertainty, there is considerable heterogeneity across experts as to the degree of perceived uncertainty - with standard deviation estimates in the range $[0.1\%, 1.4\%]$, with a mean SD of 0.6%. A mean SD estimate of 0.6% compares with the higher 1.1% SD reported by the MPC for this same 2018Q3 GDP estimate (when $b = 1$).

The experts' mean estimates, as reported in the top panel of Figure 8, are pretty well anchored around the GDP growth point estimate of 1.5%, of course communicated to the experts in the survey. But the mass of these estimates, across experts, is somewhat

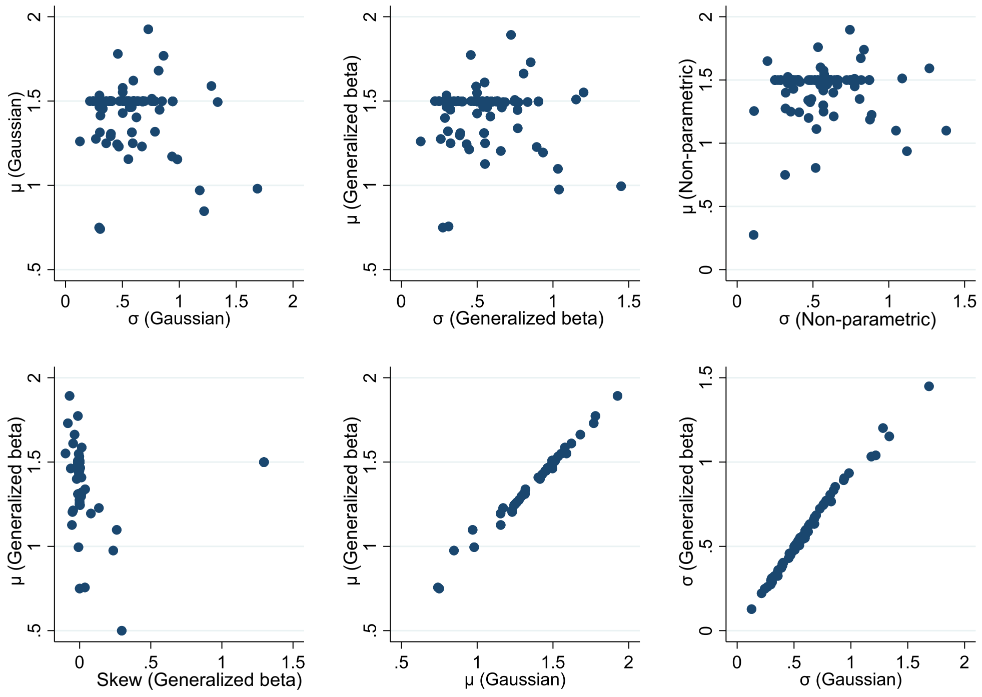

Figure 8. Mean vs. SD/skew of 104 experts' reported subjective density estimates of GDP growth (year-on-year, in %) data uncertainty (b=1): estimates computed non-parametrically and having fitted a Gaussian or generalised Beta distribution to the underlying histogram data

below 1.5%, providing some evidence that, on average, unlike the MPC (cf. Figure 3), experts expect data revisions to lower, not raise, GDP growth. While data revisions are an ongoing process, so this may change, it is interesting that the latest (as of the time of writing) GDP data vintage suggests that data revisions have raised growth for 2018Q3 from this first estimate of 1.5% to 1.8%.

The bottom three panels of Figure 8 further explain these results. We see that inference about the mean and standard deviation of the histogram-based perceptions of data uncertainty is, in general, robust to which of the three estimation methods is used. But while the bottom-left panel of Figure 8 confirms this assessment, in evidencing little statistical evidence for skew when the generalised Beta is fitted, we do see tentative evidence that some experts, especially those with lower means, did expect modest positive skew, i.e. upside risks to data uncertainty. Skew is measured here as the mean of the generalised Beta, for each expert, minus the median. The results though, in general, provide empirical support for the symmetry assumption, of course a maintained assumption when estimating the Gaussian density, given that the mean and SD estimates from the normal and generalised Beta densities are so similar.

We end with a remark. Recent research finds that effective communication by central banks can stabilise the public's macroeconomic expectations (Coibion, Gorodnichenko, and Weber, 2019; Kryvtsov and Petersen, 2021). By attempting to anchor expectations, central banks also aim to reduce the dispersion of expectations across individuals. Our results indicate that the dispersion of experts' expectations of GDP data uncertainty is large. We conjecture that this dispersion, at least in part, is attributable to the fact that statistical offices do not communicate interval estimates, or other uncertainty measures, for their GDP estimates. Future research should consider how alternative data uncertainty communication strategies may affect individuals' expectations of GDP data uncertainty.

IV. Conclusion

Statistical offices emphasise data revisions in their communications. But it is the MPC at the Bank of England that, rarely in an international context, provides direct quantitative estimates of the likely uncertainty around historical GDP values. They have now done so for more than 20 years. The ONS changed its publication model for GDP in the summer of 2018. In this paper we therefore use pre-2018 GDP data vintages to assess the MPC's predictive densities for historical real GDP growth. With the qualification that this is a relatively short sample that covers only one business cycle, this paper provides the first direct examination and *ex post* evaluation of the accuracy of the MPC's densities of historical GDP data uncertainty.

The MPC's perceptions of data uncertainty often imply uncertainty about whether the economy was growing or contracting, even 3 to 4 years in the past. To gauge data uncertainty across a wider set of experts, we conduct an online survey. This reveals that these experts also do not take GDP point estimates, as published by the statistical office, at face-value. They too expect data uncertainty, albeit not as much as the MPC. But while their estimates of data uncertainty are centred around the ONS's point estimate of GDP, there is considerable heterogeneity in their variances. Like the MPC, these experts tend to perceive data uncertainties as symmetrically distributed around the ONS's point estimate.

Our *ex post* evaluation of the MPC's data uncertainty estimates indicates that the MPC's judgement-based probabilistic assessments of data uncertainty accurately measure 'transitory' uncertainty of the sort explained by data revisions, except for the oldest data. The MPC do overstate the uncertainty associated with the oldest data, perhaps as they aim to quantify permanent as well as transitory data uncertainties. The extent of this overstatement depends on when 'mature' data are assumed to be observed. If observed 3 years after the first release, the over-estimation is more sizeable than if mature data are observed after 4 years given that large revisions can be made to the GDP estimates even after 3 years. The fact that the MPC's estimates of data uncertainty are comparable, in terms of their accuracy, to those from a benchmark statistical model is consistent with the view that MPC assessments of data uncertainty are informed by statistical models.

We hope that our results will encourage data producers to do more to measure and communicate the uncertainties associated with their estimates, by providing quantitative measures of their uncertainty. Direct communication of data uncertainty may help anchor users' perceptions of data uncertainty that, as our paper has shown, otherwise tend to be very heterogeneous.

Final Manuscript Received: September 2020

References

Abel, J., Rich, R., Song, J. and Tracy, J. (2016). ‘The measurement and behavior of uncertainty: evidence from the ECB survey of professional forecasters’, *Journal of Applied Econometrics*, Vol. 31, pp. 533–550.

Aoki, K. (2003). ‘On the optimal monetary policy response to noisy indicators’, *Journal of Monetary Economics*, Vol. 50, pp. 501–523.

Clements, M. P. (2004). ‘Evaluating the Bank of England density forecasts of inflation’, *Economic Journal*, Vol. 114, pp. 855–877.

Clements, M. P. (2018). ‘Are macroeconomic density forecasts informative?’, *International Journal of Forecasting*, Vol. 34, pp. 181–198.

Clements, M. P. and Galvão, A. B. (2012). ‘Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models’, *Journal of Business and Economic Statistics*, Vol. 30, pp. 554–562.

Clements, M. P. and Galvão, A. B. (2017). ‘Predicting early data revisions to US GDP and the effects of releases on equity markets’, *Journal of Business and Economic Statistics*, Vol. 35, pp. 591–604.

Coibion, O., Gorodnichenko, Y. and Weber, M. (2019). *Monetary Policy Communications and Their Effects on Household Inflation Expectations*. NBER Working Paper Series No. 25482.

Croushore, D. (2011). ‘Frontiers of real-time data analysis’, *Journal of Economic Literature*, Vol. 49, pp. 72–100.

Croushore, D. and Stark, T. (2001). ‘A real-time data set for macroeconomists’, *Journal of Econometrics*, Vol. 105, pp. 111–130.

Cunningham, A. and Jeffery, C. (2007). ‘Extracting a better signal from uncertain data’, *Bank of England Quarterly Bulletin*, Vol. 2007Q3.

Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G. and Labhard, V. (2012). ‘A state space approach to extracting the signal from uncertain data’, *Journal of Business & Economic Statistics*, Vol. 30, pp. 173–180.

Diebold, F. X. and Mariano, R. S. (1995). ‘Comparing predictive accuracy’, *Journal of Business and Economic Statistics*, Vol. 13, pp. 253–263.

Engelberg, J., Manski, C. F. and Williams, J. (2009). ‘Comparing the point predictions and subjective probability distributions of professional forecasters’, *Journal of Business & Economic Statistics*, Vol. 27, pp. 30–41.

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

Fixler, D. J., Greenaway-McGrevy, R. and Grimm, B. T. (2014). ‘The revisions to GDP, GDI, and their major components’, *Survey of Current Business*, Vol. 94, pp. 1–23.

Galbraith, J. W. and van Norden, S. (2012). ‘Assessing gross domestic product and inflation probability forecasts derived from Bank of England fan charts’, *Journal of the Royal Statistical Society: A (Statistics in Society)*, Vol. 175, pp. 713–727.

Galvão, A. B. (2017). ‘DSGE models and data revisions’, *Journal of Econometrics*, Vol. 196, pp. 215–232.

Galvão, A. B. and Kara. A. (2020). ’The impact of GDP Data Revisions on Identifying and Predicting UK Recessions’, ESCOE Working Paper n. 2020-12.

Giannone, D., Henry, J., Lalik, M. and Modugno, M. (2012). ‘An area-wide real-time database for the Euro area’, *The Review of Economics and Statistics*, Vol. 94, pp. 1000–1013.

Giordani, P. and Soderlind, P. (2003). ‘Inflation forecast uncertainty’, *European Economic Review*, Vol. 47, pp. 1037–1059.

Groen, J. J., Kapetanios, G. and Price, S. (2009). ‘A real time evaluation of Bank of England forecasts of inflation and growth’, *International Journal of Forecasting*, Vol. 25, pp. 74–80.

Hersbach, H. (2000). ‘Decomposition of the continuous ranked probability score for ensemble prediction systems’, *Weather and Forecasting*, Vol. 15, pp. 559–570.

Jacobs, J. P. A. M. and van Norden, S. (2011). ‘Modeling data revisions: measurement error and dynamics of ‘true’ values’, *Journal of Econometrics*, Vol. 161, pp. 101–109.

Jacobs, J. P. A. M., Sarferaz, S., Sturm, J.-E. and van Norden, S. (2022). ‘Can GDP measurement be further improved? data revision and reconciliation’, *Journal of Business and Economic Statistics*, Vol. 40, pp. 423–431.

Jurado, K., Ludvigson, S. C. and Ng, S. (2015). 'Measuring uncertainty', *American Economic Review*, Vol. 105, pp. 1177–1216.

Kishor, N. K. and Koenig, E. F. (2012). 'VAR estimation and forecasting when data are subject to revision', *Journal of Business and Economic Statistics*, Vol. 30, pp. 181–190.

Koop, G., McIntyre, S., Mitchell, J. and Poon, A. (2022). 'Reconciled estimates of monthly GDP in the United States', *Journal of Business and Economic Statistics*. forthcoming.

Kryvtsov, O. and Petersen, L. (2021). 'Central bank communication that works: Lessons from lab experiments', *Journal of Monetary Economics*, Vol. 117, pp. 760–780.

Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W. (2018). 'HAR inference: recommendations for practice', *Journal of Business and Economic Statistics*, Vol. 36, pp. 541–574.

Mankiw, N. G. and Shapiro, M. D. (1986). 'News or noise: an analysis of GNP revisions', *Survey of Current Business.*, Vol. 66, pp. 20–25.

Manski, C. F. (2004). 'Measuring expectations', *Econometrica*, Vol. 72, pp. 1329–1376.

Manski, C. F. (2015). 'Communicating uncertainty in official economic statistics: an appraisal fifty years after Morgenstern', *Journal of Economic Literature*, Vol. 53, pp. 631–653.

Manski, C. F. (2016). 'Credible interval estimates for official statistics with survey nonresponse', *Journal of Econometrics*, Vol. 191, pp. 293–301.

Manski, C. F. and Molinari, F. (2010). 'Rounding probabilistic expectations in surveys', *Journal of Business & Economic Statistics*, Vol. 28, pp. 219–231.

McKenzie, R. (2006). *Undertaking Revisions and Real-Time Data Analysis Using the OECD Main Economic Indicators Original Release Data and Revisions Database*. OECD Statistics Working Papers No. 2006/02.

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

Murphy, A. H. (1973). 'A new vector partition of the probability score', *Journal of Applied Meteorology and Climatology*, Vol. 12, pp. 595–600.

Newey, W. K. and West, K. D. (1987). 'A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix', *Econometrica*, Vol. 55, pp. 703–708.

Orphanides, A. (2001). 'Monetary policy rules based on real-time data', *American Economic Review*, Vol. 91, pp. 964–985.

Rossi, B. and Sekhposyan, T. (2019). 'Alternative tests for correct specification of conditional forecast densities', *Journal of Econometrics*, Vol. 208, pp. 638–657.

Rossi, B., Sekhposyan, T. and Soupre, M. (2016). *Understanding the Sources of Macroeconomic Uncertainty*. CEPR Discussion Papers No. 11415.

Stone, R., Champernowne, D. G. and Meade, J. E. (1942). 'The precision of national income estimates', *Review of Economic Studies*, Vol. 9, pp. 111–125.

Wallis, K. F. (2014). 'The two-piece normal, binormal, or double Gaussian distribution: Its origin and rediscoveries', *Statistical Science*, Vol. 29, pp. 103–112.

Zarnowitz, V. and Lambros, L. A. (1987). 'Consensus and uncertainty in economic prediction', *Journal of Political Economy*, Vol. 95, pp. 591–621.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Supporting information