Real-Time Perceptions of Historical GDP Data Uncertainty*

Ana Beatriz Galvão
University of Warwick

James Mitchell
University of Warwick

This version: July 2020

Abstract

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

Key words: data revisions; fan charts; backcasts; density forecast calibration; real-time data

JEL code: C53, E32

*We thank Johnny Runge at NIESR for help in running the expert survey. We also thank two anonymous referees at ESCoE, seminar participants at the Bank of England, Reading, UCL and ESCoE for helpful comments. This research has been funded by the ONS as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). See https://www.escoe.ac.uk/projects/modelling-communicating-data-uncertainty/.

1 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, 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 GDP estimate (using the output approach) around 27 days after the end of the quarter, based on just 44% of the total sample, data revisions should really come as no surprise as GDP estimates 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 five-year moving average and standard deviation of revisions to year-on-year 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 three and four years after the first estimate. Figure 1 shows that these revisions can be substantial. Standard deviation estimates exceed 1% for many of the 5-year windows and there is a tendency for GDP estimates to be revised upwards after first release, given that the moving average is generally positive in Figure 1. We also see that revisions are time-varying and often larger at business cycle turning points, with the standard deviation 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, Eklund, Jeffery, Kapetanios and Labhard (2012), Kishor and Koenig (2012) and Galvão (2017)). In order to understand the underlying "true" data, studies often discriminate between news and

[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

[3] In the summer of 2018 the ONS changed its publication model. The first estimate of quarterly GDP is now available at around 40 days; and it has a higher data content than the first estimate 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 post summer 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 GDP (vintage) dataset downloadable from the ONS website (February 2020 vintage).

[5] An online appendix confirms this visual impression by estimating an econometric model of revisions that allows for time-variation in both the revision mean and its volatility.

noise revisions following the approach of Mankiw and Shapiro (1986), what Aruoba, Diebold, Nalewaik, Schorftheide and Song (2016) call forecast-error and measurement-error approaches. In tandem, national statistical offices and central banks increasingly publish real-time data vintages (e.g. see Croushore and Stark (2001) and Giannone, Henry, Lalik and Modugno (2012)).

But despite growing awareness by statisticians and economists of these and other data uncertainties, national statistical offices continue to communicate GDP point estimates only, certainly in their headline data publications. Manski (2015, 2016) has emphasised that this practice of acknowledging data uncertainties at best qualitatively or verbally, rather than quantitatively, is common across statistical offices; and has called for more transparent communication of GDP data uncertainties. Indeed, at the time of writing, the ONS is qualitatively emphasising data uncertainties in its GDP press releases, due to the challenges faced measuring the economy under Covid-19 induced shutdowns.[6]

In this paper, absent direct quantitative communication by the statistical office, we undertake two empirical exercises to ascertain how well understood data uncertainty really is and how effectively data uncertainty can in fact be quantified in real-time, should one wish to communicate it. We consider perceptions of UK GDP data uncertainty from both the Bank of England, as reported each quarter since 2007 in their *Inflation Report*, and 100 experts as elicited in a specially designed online survey of their probabilistic expectations of historical data uncertainty that we conducted in the first quarter of 2019. This focus is justified as follows. Firstly, experts' perceptions of GDP data uncertainty affect policy and decision making. 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) and Croushore (2011)). Aoki (2003) shows theoretically that as data uncertainty increases, policymakers should attenuate their responses to the data. Clements and Galvão (2017) find that surprises to expected GDP revisions affect financial markets. Secondly, the measurement of GDP data uncertainty is not straightforward. Manski (2015) distinguishes between "transitory" and "permanent" statistical uncertainty. For GDP estimates, permanent uncertainty is mainly caused by surveys' sampling errors. But with a variety of surveys used to measure GDP, including on the income and expenditure side, statistical offices do not in practice publish estimates of these errors.[7] Transitory uncertainty stems from publication of

[6] For example, in its 12 June 2020 GDP data release the ONS writes that its GDP estimates are "subject to more uncertainty than usual": see https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/gdpmonthlyestimateuk/april2020

[7] To quote the ONS: "[t]he estimate of GDP ... is currently constructed from a wide variety of data

early data releases that are revised over time as new information arrives. In the UK, this also involves balancing GDP measures on the output, income and expenditure side.[8] Statistical offices frequently publish analyses of past GDP revisions, and emphasise 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.

Expectations play a key role in macroeconomics. Expert and public surveys eliciting probabilistic replies are increasingly used both to measure subjective expectations of individuals and understand their formation; e.g. see Manski (2004). In the spirit of this work, in this paper we first characterise and then evaluate probabilistic expectations of GDP data uncertainty. We emphasise that these expectations of data uncertainty are formed without any direct quantitative communication of uncertainty by the statistical office. Recent work, assessing how central bank communication affects individuals' expectations of inflation and interest rates (Coibion, Gorodnichenko and Kamdar, 2018; Haldane and McMahon, 2018; Coibion, Gorodnichenko and Weber, 2019; Kryvtsov and Petersen, 2020), has shown how communication can affect expectations.

The remainder of this paper is structured as follows. Section 2 reports features of the Bank of England's probabilistic backcasts for GDP growth. It compares them with model-based alternatives that use historical GDP data revisions data to measure transitory data uncertainty. Specifically, we pick up the suggestion of Fixler, Greenaway-McGrevy and Grimm (2014) and compute GDP data uncertainty intervals from the mean and standard deviation of past revisions, assuming data revisions are normally distributed. This helps us isolate the role of judgement in forming quantitative data uncertainty estimates. While the Bank's fan chart is informed by the data revisions model of Cunningham and Jeffery (2007), ultimately it is subjective. We then provide the first evaluation of the accuracy and calibration of these densities for historical GDP growth. In Section 3 we gauge data uncertainty across a wider set of experts by conducting an online survey. We follow the advice of Manski (2004) and measure uncertainty around the latest GDP point estimate (at the time of running the survey this was the 2018Q3 GDP estimate of 1.5%) by eliciting probabilistic expectations about this GDP estimate in the form of subjective

---

sources, some of which are not based on random samples or do not have published sampling and non-sampling errors available. 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

[8] In some countries, such as the US, income and expenditure estimates of GDP are not balanced (by the statistical office, at least) and remain separate and often divergent; see Aruoba et al. (2016).

histograms. Section 4 concludes. An online appendix contains supplementary econometric results analysing the revisions properties of UK GDP data.

2 The Bank of England's MPC "fan charts" for historical GDP growth

An important example, and rare illustration, of how historical GDP data uncertainty is both communicated in real-time and affects policy making is provided by the Monetary Policy Committee (MPC) at the Bank of England.[9] 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, the Bank of England provide in real-time 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 their *Inflation Report*. 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." (Bank of England (2007), p.39). This apparent focus on measuring the "mature" estimate of GDP need not imply that the MPC are focused exclusively on the measurement of transitory uncertainties, due to data revisions. They may take the view that "true" GDP is never observed (as in Jacobs and van Norden (2011), Cunningham et al. (2012) and Aruoba et al. (2016)) and their expected uncertainty estimates may reflect permanent data uncertainties too.

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; Independent Evaluation Office, 2015), previous research has not characterised 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.

2.1 Measures of mature GDP

To evaluate accuracy, we will discriminate between two observed measures of "mature" GDP. This is to acknowledge that there is always debate about what defines the "mature" and what

[9] 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.

might be considered the "true" value of GDP. As Figure 1 indicates, the properties of data revisions depend on what later estimate the first GDP estimate is compared against.

Let y_t^{t+b} denote the ONS's first 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 $(b = 1, 2, ..., B)$. 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, $rev_t^{(l-b)}$, between more mature data, y_t^{t+l} (where $l > b$), and the earlier b-th estimate are then defined as $rev_t^{(l-b)} = (y_t^{t+l} - y_t^{t+b})$. 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.

Our first candidate measure of mature data is y_t^{t+13} $(l = 13)$, i.e. the GDP growth estimate for quarter t published by the ONS three 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 13^{th} 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. Aruoba (2008) and Clements and Galvão (2012) adopt similar approaches 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 (see their "Analysis of GDP Revisions in Blue Books: 2019")[10]. Informal conversations with Bank of England staff suggest that the 13^{th} estimate is also their preferred "mature" estimate.

Our second candidate measure of mature data is data-based. It reflects the fact that there appear to be some major revisions published even after 3 years, in particular for reference quarters after 2008, as seen in Figure 1 which plots both $(y_t^{t+17} - y_t^{t+1})$ and $(y_t^{t+13} - y_t^{t+1})$. This impression is confirmed by higher average revision or bias estimates at 4 years relative to 3 years (cf. Tables A1 and A2 in the online appendix). As a consequence, we also consider y_t^{t+17}, i.e., the GDP growth estimate for quarter t published by the ONS four years after their first

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

estimate.

2.2 Benchmark Probabilistic Backcasts: Data-Based Historical Fan Charts

To help assess the judgemental contribution to the MPC's fan charts, we will compare their features with an "unconditional" model-based benchmark. Following the suggestion of Fixler et al. (2014), we construct these benchmark density estimates of transitory data uncertainty assuming a normal distribution with the means and standard deviations estimated from historical data revisions alone (with additional variables not used to try and help predict data revisions). 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 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 may involve trying to capture permanent as well as transitory data uncertainties, a comparison of their backcasts with the mechanically produced univariate benchmark helps us evaluate this subjective aspect to the MPC's densities.

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

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

where the moments of this density are recursively estimated from ONS revisions, $rev_\tau^{(l-b)}$, between the l^{th} and the b^{th} estimates:

$$\begin{aligned}
\hat{y}_t^{t+b,r} &= y_t^{t+b} + \hat{\mu}_t^{t+b,r} && (2)\\
\hat{\mu}_t^{t+b,r} &= \frac{1}{t-l}\sum_{\tau=1983Q2}^{\tau=t-l+1} rev_\tau^{(l-b)} && (3)\\
\hat{\sigma}_t^{t+b,r} &= \sqrt{\frac{1}{t-l}\sum_{\tau=1983Q2}^{\tau=t-l+1}\left(rev_\tau^{(l-b)} - \hat{\mu}_t^{t+b,r}\right)^2} && (4)\\
rev_\tau^{(l-b)} &= y_\tau^{\tau+l} - y_\tau^{\tau+b}. && (5)
\end{aligned}$$

Importantly, in computing (1), we only use data revisions data that would have actually been available at each point in real-time. For example, $\hat{y}_{t-b}^{t,r}$ and $\hat{\sigma}_{t-b}^{t,r}$ are the mean and standard deviation computed in quarter t using revisions data, $(y_t^{t+l} - y_t^{t+b})$, for reference quarters up to $(t-(l-1))$; i.e., these backcasts are conditional on y_{t-b}^{t}, but the time-series of past revisions employed to compute the moments is only available up to l quarters ago, i.e. up to $\left(y_{t-l}^{t} - y_{t-l}^{t-l+b}\right)$.

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 the online appendix). 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 the online appendix; 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.

2.3 Features of the MPC's Historical Fan Charts

Figure 2 illustrates what a typical MPC fan chart looks like, taken from the February 2018 *Inflation Report*. In Figure 2, "(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 revisions' process is complete." (November 2007; *Inflation Report*, p. 39). 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, post 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.

From May 2018 the fan chart's format in the *Inflation Report* was modified slightly. While the MPC continued to communicate data uncertainty, via the fan chart, they no longer included their backcast for expected revisions. This means that while data uncertainty was still acknowledged, the MPC no longer communicated an expected bias: so from May 2018 the ONS's estimates fall exactly in the middle of each fan chart. Also, from July 2018, because of later publication of the ONS's new first GDP estimate, the MPC no longer have sight of the ONS's first release estimate when their fan charts are published. Our empirical analysis is unaffected by these changes, given its focus on the earlier sample period (given we have to wait 3 or 4 years after publication of the fan chart to observe the mature GDP data).

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 shape the form 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 ONS estimates asymptote to the truth as they mature, and in this sense captures transitory but not permanent data uncertainties. 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 standard deviation (which we call, expected data uncertainty).

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

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

denote the MPC's density estimate for mature GDP growth for reference quarter $t-b$ made b quarters later $(b = 1, ..., (l-1))$ in quarter t (the *backcasting origin*). Note this means that the effective backcast horizon, h, declines with b, and is given as $h = (l-b)$. $f_{t-b|t}^{mpc}$ 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 July 2018), in principle, 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 $y_{t-2}^{t}, ..., y_{t-B}^{t}$, before publishing their own historical estimates, $\hat{y}_{t-1}^{t}, ..., \hat{y}_{t-B}^{t}$ and $\hat{\sigma}_{t-1}^{t}, ..., \hat{\sigma}_{t-B}^{t}$, in the quarter t *Inflation Report*. In practice, the MPC tend to look $B = 16$ quarters back into the past in 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 standard deviation 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 note that, assuming Gaussianity and that the expected revision is zero, this implies the MPC is expecting the '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 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 extracted from the 2007Q3 to 2018Q1 *Inflation Reports* of the 'expected revision' and expected data uncertainty. The top panel presents for $b = 1, 4, 8, 12, 16$ their expected revisions, computed as the difference between the MPC estimate of GDP growth, $\hat{y}_{t-b}^{t}$, and the ONS estimate, y_{t-b}^{t}. The bottom panel plots the MPC's estimates of expected data uncertainty, $\hat{\sigma}_{t-b}^{t}$.

The top 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 expected revisions to change underlying GDP values even for heavily revised data. The expected revision only becomes zero for the sixteenth estimate ($b = 16$) and even then from 2012Q3 onwards. Thirdly, the expected size of revisions has varied over time, consistent with Figure 1 and the time-varying properties of UK GDP data revisions analysed econometrically in the online Appendix. 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. Reorganising the fan charts by reference quarter (to analyse $\hat{y}_{t}^{t+b}$), in Figure A2 in the online appendix 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 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 publicised double-dip recession.

Turning to the bottom panel of Figure 3, we firstly see that the MPC has 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' standard deviation reported 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 2.4.2.

2.3.1 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 standard deviation bands) for "mature" GDP growth extracted from their backcast density $f^{mpc}_{t|t+b}$ 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 $f^{mpc}_{t|t+b} = N(\hat{y}^{t+b}_t, \hat{\sigma}^{2,t+b}_t)$ to be compared with the "mature" outturns, y^{t+l}_t.

The figure also includes 68% confidence intervals from our unconditional benchmark density ($f^{r}_{t|t+b}$) defined in (1). We superimpose on Figure 4 the "mature" estimate at $l = 13$, y^{t+13}_t. We order the plots in Figure 4 from the shortest to the longest backcast horizons. So we might might expect the intervals to widen, as they in fact do, as we look down from Figure 4A to Figure 4C and are, in effect, 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. The MPC perceived more data uncertainty than suggested by the history of data revisions. 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. Anticipating 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 appear to perceive 'too much' data uncertainty. While the mature GDP estimate (the *outturn*) does fall within the 68% interval on 68% of occasions when only a first estimate of GDP growth is available ($b = 1$), as b increases the mature GDP outturns increasingly fall within the 68% interval (they fall inside on 74% of occasions when $b = 6$ and on 97% of occasions when $b = 12$). This tendency of the MPC to perceive too much uncertainty is confirmed by comparison with the model-based perceptions of data uncertainty from $f^r_{t|t+b}$. The unconditional intervals, that use only information on past revisions to assess the degree of data uncertainty, are still too wide but narrower. Their ex post coverage rates are 61% ($b = 1$), 55% ($b = 6$) and 87% ($b = 12$).

2.4 Evaluating the Historical Fan Charts

To evaluate the accuracy of the probabilistic backcasts from the MPC we both test their coverage ex post and compare them against the unconditional benchmark density 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 uncertainty. The evaluation sample is relatively small, as the MPC have produced their forecasts quarterly for only twenty years; and this should be borne in mind when interpreting our evaluation results. Acknowledging the small sample sizes, we emphasise empirical coverage rates of the confidence intervals rather than statistical tests of coverage, although we report both. This preference for exploratory data analysis also explains our focus on coverage of the confidence intervals implied by the density estimates rather than formal statistic tests for whether the density 'as a whole' is calibrated according to tests deployed on the probability integral transforms; e.g. see Mitchell and Wallis (2011).

2.4.1 Evaluating Interval Estimates of Historical Data Uncertainty

We consider the empirical coverage of the real-time 50%, 75% and 90% central intervals implied by the MPC and unconditional benchmark densities. We consider backcasting horizons $b = 1, ..., B = l - 1$; and report and test the coverage of their ex ante intervals against the mature estimates of GDP, y_t^{t+l}, published 3 years ($l = 13$) and 4 years ($l = 17$) after the ONS first estimate.

When the intervals are well-calibrated, and correctly assess transitory data uncertainty i.e. the spread of the underlying density generating the mature data, we should expect their empirical coverage (ex post) to match the nominal size of the (ex ante) interval. As a consequence, an evaluation of empirical coverage helps us understand whether the MPC and the benchmark unconditional model can correctly assess in real-time the data uncertainty caused by future data revisions. Tables 1 and 2 report their empirical coverage rates and p-values of the Christoffersen (1998) test for correct (unconditional) coverage.[11]

The results in Tables 1 and 2 confirm the visual impression from Figure 4 that the MPC tend to perceive, in real-time, more data uncertainty than suggested by past data revisions as represented by the unconditional density. The coverage rates of the MPC's interval estimates in Tables 1 and 2 are equal to or higher than for the unconditional density in all cases but one, when $l = 13$, $b = 1$ and we focus on the 90% interval. Policymakers perceived more uncertainty than the history of data revisions, as represented by the unconditional density, suggests.

This perception by the MPC of higher data uncertainty does appear overstated (as an estimate of transitory data uncertainty) when we look, in Table 1, at the ex post coverage rates against ONS estimates published 3 years after the first estimate ($l = 13$). The coverage rates of the MPC's 50% intervals exceed the ex ante levels across all backcasting horizons, b.[12] Nevertheless, this over-estimate of data uncertainty is only rejected statistically for the 50% intervals at longer backcasting horizons. And the wider ex ante 75% and 90% intervals achieve coverage rates closer to the desired levels, albeit they are again too wide (at $b = 12$) when we look furthest back into the past.

The MPC's perceptions of data uncertainty achieve better coverage rates, both in absolute terms and relative to the unconditional density, when we assess coverage against the values published by the ONS 4 year after the first release (y_t^{t+17}). Although the MPC's assessment of uncertainty at the longest backcasting horizon ($b = 16$) remains too wide, with coverage rates higher than the nominal levels at 50%, 75% and 90%. At the shorter horizons, up to $b = 9$, the MPC's ex ante perceptions of data uncertainty are never rejected statistically according to the coverage test. In contrast, the unconditional density is far more frequently rejected across horizons, b. This contrasts its better performance when evaluating against the y_t^{t+13} outturns in Table 1.

Summarising, the MPC do a better job at matching the ex ante and ex post coverage rates

[11] The p-values are computed using the chi-squared LR test of unconditional coverage in Christoffersen (1998). Tables A1 and A2 in the online appendix provide supplementary information on the calibration of their densities by testing whether the mean of the MPC and benchmark unconditional density offers an unbiased estimate of the mature ONS estimate. In general, we find little evidence for bias of these point estimates, especially at $l = 17$. This suggests that calibration of the interval forecasts depends primarily on expected data uncertainty and the maintained assumption of Gaussianity rather than the accuracy of the predictions for the expected revision.

[12] Under Gaussianity, over-coverage of the (central) interval is possible only if the forecast variance is too high. It cannot be explained away by bias of the mean forecast. In any case, as seen in Tables A1 and A2 (in the online appendix), the evidence for bias of the mean estimates from the MPC and the unconditional densities is weak especially at $l = 17$ when there are no rejections of the null hypothesis of unbiasedness (Table A2).

and, in general, are well-calibrated against ONS GDP estimates published 4 years after the first release. But their uncertainty bands, especially for older data (higher b), tend to be too wide when evaluated against estimates published 3 years after the first release suggesting either over-estimation of transitory data uncertainty or an attempt to capture permanent as well as transitory uncertainties. In turn, the unconditional data-based densities are more accurate against the outcomes observed 3 years after the initial release, but tend to underpredict the uncertainty associated with data published later (after 4 years).

2.4.2 Relative Performance

To compare directly the MPC's real-time perceptions of data uncertainty 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, $g_{t|t+b} = f_{t|t+b}^{mpc}$ 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):

$$\log score_{t|t+b}^{t+l} = -\log g_{t|t+b}(y_t^{t+l}) \tag{7}$$

$$CRPS_{t|t+b}^{t+l} = \int_{-\infty}^{+\infty} \Big[G_{t|t+b}(y) - I(y_t^{t+l} \leq y)\Big]^2 dy \tag{8}$$

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-Mariano type t-statistics for equal forecast accuracy of two competing forecasts are computed for each of the three loss functions using Newey-West standard errors.

Table 3 reports the RMSE statistics of the ONS's own earlier estimates alongside those of the MPC and the unconditional density. Accuracy is measured against both y_t^{t+13} and y_t^{t+17} as measures of mature GDP growth. From Table 3 we see, as expected, that for the ONS these RMSE estimates decrease as b increases; and accuracy tends to be better against y_t^{t+13} than y_t^{t+17}. But looking at the RMSE ratios in Table 3, 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 MPC is correct, as shown in Figure 3, to expect ONS first release GDP data to be revised upwards over time, which is also consistent with the historical time-series of revisions (cf. Figure 1 and Table A2 in the online appendix). This is the case for all values of b when the outcome is y_t^{t+17}, and for the shorter horizon backcasts (up to $b = 8$) when $l = 13$. The RMSE gains for the MPC are typically between 3% and 8%, but are statistically significant on just 3 occasions (for y_t^{t+17}). The mean estimates from the benchmark statistical model are not as competitive as those from the MPC, but again outperform the ONS's own early estimates for y_t^{t+17}, with RMSE ratios less than one 14 out of 16 times.

Table 4 then reports the logscore and CRPS statistics, for a given b, averaged over the evaluation sample for the MPC densities. 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 4 confirms that 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. Accuracy also increases, as expected, further back into the past; i.e. the logscore and CRPS values decrease as b increases. Comparing against the unconditional benchmark, we see the MPC is always more accurate according to the logscore, except at $b = 1$ when $l = 17$, although the gains are never statistically significant. Using the CRPS to compare the MPC and unconditional densities, we again see that the MPC is more accurate except at the longer horizons. The gains at longer horizons associated with the unconditional density are stronger against the less mature data ($l = 13$) and are statistically significant when $b = 9, ..., 12$. The CRPS prefers backcasts with less uncertainty, which holds for the longer horizon unconditional benchmark. The MPC overestimate (transitory) uncertainty at the longest horizons ($b = 12$, when $l = 13$, or $b = 16$ when $l = 17$).

Overall, we find that the MPC provide more accurate point estimates of mature GDP values than the comparably timed estimates published by the ONS themselves, especially if the "mature" estimate is taken to be the ONS GDP estimate published 4 years after their first estimate. The MPC's probabilistic backcasts also tend to be more accurate than those from the benchmark model that exploits patterns in past data revisions, except at longer horizons (further back in the past) when the MPC overstates transitory data uncertainty. Model-based alternatives tend to suggest less data uncertainty than the MPC and are better calibrated against less mature data. While the MPC is disposed to overstate transitory uncertainty, this diminishes when later vintage data are used as the "target" and is rarely sufficient to indicate calibration failure that is statistically significant. Their density estimates are, on average, well-calibrated except for the oldest data.

This evaluation serves to illustrate that model-based and more judgemental methods can be used to help measure GDP transitory data uncertainties. But it is also a reminder that measurement of uncertainty can be hard. We now consider whether this is explained by temporal variations, and perhaps heightened uncertainties, at business cycle turning points.

2.4.3 Time-Variation in Relative Accuracy

Figure 5 breaks down the average CRPS statistics, reported in Table 4, 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. We focus on evaluation of the densities against the $l = 13$ mature estimate.

Figure 5A plots the CRPS estimates for the earlier density estimates ($b = 1, 4$); while Figure 5B considers the later estimates ($b = 8, 12$). Both panels of Figure 5 show clearly that the accuracy of both the $f^{mpc}_{t|t+b}$ and $f^{r}_{t|t+b}$ densities deteriorates substantially during the recessionary period, 2008-10. We also see that it is during 2008-10 that we observe more differences between the two densities, with $f^{mpc}_{t|t+b}$ delivering gains.

This serves as additional evidence that the data uncertainty information communicated in the fan chart captures more than the information contained in the history of past data revisions. Figure 5 suggests that this supplementary information is particularly helpful around business cycle turning points.

3 Experts' Perceptions of Data Uncertainty: A Case Study

In this section, we present the results of a survey designed to gauge perceptions of GDP data uncertainty across a wider set of experts than the MPC. We asked survey respondents to provide their probabilistic assessments of 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 is the ONS's latest GDP point estimate, alongside the accompanying ONS press release. But given that our survey is of experts, some may have read the latest Bank of England *Inflation Report*, which, as analysed above, does provide a quantitative assessment of data uncertainty. At the time of running the survey, the latest fan chart (from the February 2019 *Inflation Report*) indicated that the expected mature GDP value was equal to the current ONS estimates, implying that the expected revision was equal to zero. Data uncertainty about the 2018Q3 GDP estimate of 1.5% was 1.1%. Our case-study survey lets us assess whether other experts' perceptions of data uncertainty are in line with MPC views or whether they take ONS point estimates at face-value and do not perceive any data uncertainty.

3.1 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 being 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 9th November 2018. The online expert survey was disseminated through the ESCoE (Economic Statistics Centre of Excellence) emailing list, social media particularly Twitter and emailing personal contacts and asking them to forward to colleagues. The recruitment period lasted for four weeks, between 18 February and 17 March 2019. The survey received 104 completed responses. The survey included additional questions on qualitative measures of uncertainty that we do not evaluate in this paper, as we are mainly interested in quantitative measures,. Additional details on survey design are available in Galvão, Mitchell and Runge (2019).

After collecting a range of background data, respondents were informed 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%".

After being asked, for some qualitative impressions of data uncertainty e.g. How accurate do you think the annual estimate of GDP growth of 1.5% is likely to be? (possible replies on a four-point scale: from very accurate through to very inaccurate), the experts were then asked: Please provide (best-guess) estimates of the percentage probabilities you would attach to various outcomes for GDP growth. The probabilities should sum to 100% as indicated:

Probability of GDP growth (for the year ending in 2018Q3) being in the following ranges:

	Year ending in 2018Q3
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%	
More than 3%	
TOTAL	100%

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 US and the European Central Bank in Europe (e.g. see Abel, Rich, Song and Tracy (2016)), facilitates interpersonal comparisons of uncertainty. This contrasts questions that elicit qualitative uncertainty statements. It is therefore our focus.

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 with only 29% of the sample identifying 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 uncertainty vary by these characteristics, although our survey is too small and not designed to explore what explains experts' perceptions of data uncertainty, as interesting as this would be. It is simply designed, as a first step in this direction, to measure experts' quantitative perceptions of UK GDP growth data uncertainty.

3.2 Perceptions of the Expected Revision and of Data Uncertainty

We estimate the mean and standard deviation 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. And 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 standard deviation, σ_i, of individual i's histogram are estimated as:

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

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

where u_j and l_j the upper and lower limits of the jth interval, w is the width of the central intervals and $p_{i,j}$ is the probability that forecaster i assigns to the jth interval. The last term in the formula for σ_i is the commonly applied Sheppard correction for the variance.

Figure 6 plots, for each expert, their mean and standard deviation estimates as estimated from the reported histograms. Figure 6 shows that these experts, like the Bank of England's MPC, do expect 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 their degrees of perceived uncertainty - with standard deviation estimates in the range $[0.1\%, 1.4\%]$ with a mean standard deviation of 0.6%. This compares with the higher 1.1% standard deviation reported by the MPC for this same 2018Q3 GDP estimate (also when $b = 1$). In contrast, the experts' mean estimates in Figure 6 are better anchored around the GDP growth point estimate of 1.5%, of course communicated to the experts in the survey. The reported mean, across experts, falls slightly below 1.5% at 1.4%, providing modest evidence that, on average, unlike the MPC (cf. Figure 3), experts expect data revisions to lower, not raise, GDP growth.

Recent research finds that effective communication by central banks can improve the accuracy of individuals' expectations (Kryvtsov and Petersen, 2020). 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 about GDP data uncertainty is large. We conjecture that this dispersion is, at least in part, attributable to the fact that statistical offices do not communicate interval estimates, or other uncertainty measures, for their GDP estimates. Future research will consider how alternative data uncertainty communication strategies may affect individual expectations about GDP data uncertainty. In fact, the ONS opened a consultation (in April 2020) to consider the communication of uncertainty intervals around GDP estimates.[13]

4 Conclusion

The ONS do emphasise data revisions in their communications. But it is the MPC at the Bank of England that, rarely in an international context, provide direct quantitative estimates of the likely uncertainty around historical GDP values. This paper provides the first direct examination and evaluation of the accuracy of these predictive densities for historical GDP growth.

The MPC's perceptions of data uncertainty often imply uncertainty about whether, even three to four years in the past, the economy was growing or contracting. This is so even when GDP growth, on the face of it, looked to have been fairly strong. Ex post evaluation of the MPC's data uncertainty estimates, with respect to mature GDP outturns, indicates that the MPC is disposed to overstate transitory data uncertainty, although this diminishes when later vintage data are used as the "target" and is rarely sufficient to reject the calibration of their interval estimates, except for the oldest data. The MPC's fan chart for historical data would have been improved, as an estimate of transitory data uncertainty, if the intervals for older data were narrower. Model-based alternatives, that exploit the information in historical data

[13] See https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/communicatinggrossdomesticproduct/2020-04-16

revisions, tend to suggest less data uncertainty and are better calibrated against earlier vintage data.

We commend the MPC at the Bank of England for regularly quantifying GDP data uncertainty in their quarterly fan charts. Echoing Manski (2015, 2016), we encourage statistical offices to do more to measure and communicate uncertainties associated with their earlier GDP estimates by providing quantitative measures of the "accuracy and reliability" of these estimates.[14] We especially welcome recent ONS work to communicate transitory data uncertainty, via presentation of confidence intervals around GDP estimates based on data revisions.[15] Other sources of uncertainty, for example due to limitations of the survey methodology, are not represented; and methodological work on measuring these permanent data uncertainties and the uncertainties associated with the increased use of administrative data (such as firm-level tax data) when measuring GDP continues (e.g. see Manski (2016) and Hand (2018), respectively).

The evidence in this paper that uncertainty estimates formed in real-time from the history of GDP data revisions can provide reasonably accurate impressions of transitory data uncertainty provides further encouragement. It is possible to measure transitory data uncertainty accurately by modelling past revisions data. But there are challenges, especially at business cycle turning points. We also believe that the Bank of England would itself improve communication further, if they stated explicitly what data vintage they seek to forecast rather than leaving it as the latent "mature" GDP estimate. This precludes *ex post* evaluation without making some assumption as in this study.

A further advantage of direct communication of data uncertainty by statistical offices would be that this may help anchor users' perceptions of data uncertainty. The illustrative survey evidence that we gathered, by undertaking a targeted ad hoc survey of 100 expert users, indicates that, like the Bank of England, these experts do not take initial point estimates of GDP at face-value: they also expect data uncertainty. But these experts have heterogeneous expectations about the degree of data uncertainty, and on average perceive less uncertainty than the Bank of England.

An important open question, the subject of ongoing research (see van der Bles, van der Linden, Freeman, Mitchell, Galvao, Zaval and Spiegelhalter (2019) and Galvão et al. (2019)), is if

[14] Note that here "accuracy and reliability" are defined as in Eurostat's European Code of Practice, Principle 12. They refer to sampling and non-sampling errors and the measurement of data revisions.

[15] See https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/communicatinggrossdomesticproduct/2020-04-16

and how different methods, formats and visualisations of communicating GDP data uncertainty affect perceptions of historical GDP values, users' understanding of data uncertainty and their trust in the data and data producer.

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* **31**(3): 533–550.

Aoki, K. (2003). On the optimal monetary policy response to noisy indicators, *Journal of Monetary Economics* **50**(3): 501–523.

Aruoba, S. (2008). Data revisions are not well-behaved, *Journal of Money, Credit and Banking* **40**: 319–340.

Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorftheide, F. and Song, D. (2016). Improving GDP measurement: a measurement-error perspective, *Journal of Econometrics* **191**: 384–397.

Bank of England (2007). *Inflation Report, November 2007*.

Chan, J. C. C. (2018). Specification tests for time-varying parameter models with stochastic volatility, *Econometric Reviews* (8): 807–823.

Chan, J. and Hsiao, C. (2014). Estimation of stochastic volatility models with heavy tails and serial dependence, *in* I. Jelizakov and X. Yang (eds), *Bayesian inference in the social sciences*, John Wiley & Sons, pp. 159–180.

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

Clark, T. E., McCracken, M. W. and Mertens, E. (2020). Modeling time-varying uncertainty of multiple-horizon forecast errors, *Review of Economics and Statistics* **102**: 17–33.

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

Clements, M. P. (2018). Are macroeconomic density forecasts informative?, *International Journal of Forecasting* **34**: 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* **30**: 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, forthcoming* **35**: 591–604.

Coibion, O., Gorodnichenko, Y. and Kamdar, R. (2018). The formation of expectations, inflation and the phillips curve, *Journal of Economic Literature* **56**: 1447–1491.

Coibion, O., Gorodnichenko, Y. and Weber, M. (2019). Monetary policy communications and their effects on household inflation expectations, *NBER Working Paper Series, n. 25482* .

Croushore, D. (2011). Frontiers of real-time data analysis, *Journal of Economic Literature* **49**(1): 72–100.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists, *Journal of Econometrics* **105**(1): 111–130.

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* **30**: 173–80.

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

Faust, J., Rogers, J. H. and Wright, J. H. (2005). News and noise in G-7 GDP announcements, *Journal of Money, Credit and Banking* **37 (3)**: 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* **94**: 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* **175**: 713–727.

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

Galvão, A. B. and Kara, A. (2020). The impact of GDP data revisions on identifying and predicting UK recessions, *ESCoE Discussion Paper* .

Galvão, A. B., Mitchell, J. and Runge, J. (2019). Communicating Data Uncertainty: Experimental Evidence for U.K. GDP, *Economic Statistics Centre of Excellence (ESCoE) Discussion Papers ESCoE DP-2019-20*, Economic Statistics Centre of Excellence (ESCoE).

Garratt, A. and Vahey, S. P. (2006). UK real-time macro data characteristics, *Economic Journal* **116**: F119–F135.

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* **94**(4): 1000–1013.

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* **25**: 74–80.

Haldane, A. and McMahon, M. (2018). Central bank communications and the general public, *American Economic Review* **1**: 1–6.

Hand, D. J. (2018). Statistical challenges of administrative and transaction data, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **181**(3): 555–605.

Independent Evaluation Office (2015). Evaluating forecast performance, *Technical report*, Bank of England.
URL: *https://www.bankofengland.co.uk/-/media/boe/files/independent-evaluation-office/2015/evaluating-forecast-performance-november-2015*

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

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

Kryvtsov, O. and Petersen, L. (2020). Central bank communication that works: Lessons from lab experiments, *Journal of Monetary Economics* **forthcoming**.

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

Manski, C. F. (2004). Measuring expectations, *Econometrica* **72**(5): 1329–1376.

Manski, C. F. (2015). Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern, *Journal of Economic Literature* **53**(3): 631–653.

Manski, C. F. (2016). Credible interval estimates for official statistics with survey nonresponse, *Journal of Econometrics* **191**(2): 293–301.

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* **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 Bullettin of Economics and Statistics* **67**: 995–1033.

Mitchell, J., Robertson, D. and Wright, S. (2019). R2 bounds for predictive models: What univariate properties tell us about multivariate predictability, *Journal of Business and Economic Statistics* **37**(4): 681–695.

Mitchell, J. and Wallis, K. F. (2011). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness, *Journal of Applied Econometrics* **26**: 1023–1040.

Office for National Statistics (2017). Improvements to early estimates of GDP through the 2008 to 2009 downturn, *Economic Review* **April**: 18–25.

Orphanides, A. (2001). Monetary policy rules based on real-time data, *American Economic Review* **91(4)**: 964–985.

Stock, J. H. and Watson, M. W. (2007). Why has U.S. Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking* **39**: 3–33.

van der Bles, A. M., van der Linden, S., Freeman, A., Mitchell, J., Galvao, A. B., Zaval, L. and Spiegelhalter, D. (2019). Communicating uncertainty about facts, numbers and science, *Royal Society Open Science* **6**: 18170.

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

Zarnowitz, V. and Lambros, L. A. (1987). Consensus and uncertainty in economic prediction, *Journal of Political Economy* **95**(3): 591–621.

A Online Appendix: Modelling Data Revisions

A.1 UK GDP Data Revisions

We use the real-time dataset of UK real GDP published by the ONS.[16] From this dataset we extract quarterly vintage estimates, denoted y_t^{t+b}, of year-on-year GDP growth for the reference quarter t, $(t = 1, ..., T)$; with as in the main paper the superscript denoting the vintage date or publication date in quarters $(b = 1, 2, ...)$.[17] So, for example, y_t^{t+1} denotes the ONS's so-called "preliminary" (or first) estimate of year-on-year GDP growth for the reference quarter t published during quarter $(t + 1)$. In general, $\{y_t^{t+b}\}_{t=1}^{t=T}$ is the time series of the b^{th} estimate of GDP growth. Revisions between the "mature" data, y_t^{t+l} (where $l > b$), and an earlier b-th estimate are then given as $(y_t^{t+l} - y_t^{t+b})$. 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.[18] For an analysis of UK data revisions over earlier sample periods see Garratt and Vahey (2006) and references therein.

Figure A1 plots alongside the first (preliminary) estimate of GDP our two candidate measures of "mature" GDP: y_t^{t+13}, the 13^{th} estimate, and the 17^{th} estimate, y_t^{t+17}, for reference quarters from $t = 1982Q2$. The Figure shows that while the first estimate is highly correlated with later estimates, there are gaps between them. These gaps appear bigger, and more persistent, at certain points in time. Tables A1 and A2 summarise across b the average size of these revisions, $(y_t^{t+l} - y_t^{t+b})$, - the bias of the ONS estimates - against both measures of "mature" GDP over the post 2007 evaluation period considered in the main paper. The positive bias estimates reported confirm the impression that, on average, ONS first release data tend to be revised upwards, especially against the more mature data observed after 4 years. But across time Tables A1 and A2 indicate that there is, at best, weak evidence that this bias is statistically significant with t-statistics always less than the 95% critical value.

[16] See https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/realtimedatabaseforukgdpabmi

[17] As with the fan chart, the y_t are computed as $100((Y_t/Y_{t-4}) - 1)$ where Y_t is the level of real GDP in quarter t. We compute growth rates within each vintage, before obtaining the time series of each estimate, to avoid jumps due to changes in the chain-linking base year.

[18] The ONS made changes to its GDP release calendars in the summer of 2018. The delay on the publication of the first estimate of GDP increased from 25 to 40 days. The vintages of real GDP in the ONS real-time dataset are monthly, and these are converted into quarterly vintages using the vintage that includes a first release of quarterly GDP. In this paper, we consider revisions for observations up to 2016Q4, so the ONS new release calendar has no impact in our results as it does not affect the timings of the publication of annual revisions.

A.1.1 A UC-SV Model of Changes in the Data Revisions' Process

To describe possible time variation in the data revision process of UK GDP growth, we use a model to decompose data revisions into a revision mean and a measurement error component.

We let the "mature" data, y_t^{t+l} (where $l > b$), relate to the earlier b-th estimate as follows:

$$
\begin{aligned}
y_t^{t+l} &= y_t^{t+b} + \mu_t^{(l-b)} + e^{.5(h_0+\varpi_h \tilde{h}_t)}\zeta_{\varepsilon,t} \\
\mu_t^{(l-b)} &= \mu_{t-1}^{(l-b)} + e^{.5(g_0+\varpi_g \tilde{g}_t)}\zeta_{\eta,t} \\
\tilde{h}_t &= \tilde{h}_{t-1} + \zeta_{h,t} \\
\tilde{g}_t &= \tilde{g}_{t-1} + \zeta_{g,t} \\
\zeta_{\varepsilon,t}, \zeta_{\eta,t}, \zeta_{g,t}, \zeta_{h,t} &\text{ are all } N(0,1)
\end{aligned} \tag{11}
$$

implying data revisions to the b^{th} estimate, $(y_t^{t+l} - y_t^{t+b}) = rev_t^{(l-b)}$, comprise two unobserved components (UC): (i) a time-varying mean, $\mu_t^{(l-b)}$; and (ii) a mean-zero measurement error, $\zeta_{\varepsilon,t}$. Cunningham et al. (2012) also decompose data revisions into bias and measurement error components.[19] But our modelling differs by using a model that allows for changes both in the mean and volatility of these components, of the sort popularised by Stock and Watson (2007) when modelling US inflation. An alternative interpretation is that $rev_t^{(l-b)}$ is the $(l-b)$-step-ahead backcasting error using the earlier estimate y_t^{t+b} to predict the later estimate, y_t^{t+l}. By fitting a stochastic volatility (SV) model to data revisions, and allowing for possible shifts in data uncertainty over time, we mimic how SV models are used to improve measures of forecast uncertainty; e.g. see Clark, McCracken and Mertens (2020).

The model in (11) allows for time-varying effects of permanent and transitory shocks to the revisions' process. It implies that $\Delta rev_t^{(l-b)}$ has a first-order time-varying Moving Average, MA(1), representation; with the size of the MA coefficient θ_t increasing as the variance of the transitory shocks increases relative to the variance of the permanent shocks. Higher values of θ_t imply less predictability for the true but unobserved process generating $rev_t^{(l-b)}$, relative to this univariate model; see Mitchell, Robertson and Wright (2019).

Estimation of the four parameters in (11), i.e. g_0, ϖ_g, h_0 and ϖ_h, proceeds as in Chan (2018). We consider revisions for GDP growth reference quarters from 1983Q3. Note that if

[19] Other data revisions models, such as Jacobs and van Norden (2011) and Cunningham et al. (2012), model successive rounds of data revisions. By focusing on a specific revision, $rev_t^{(l-b)}$, we model changes to the revisions process more parsimoniously.

$l = 17$, then $rev_t^{(l-b)}$ for $t = 2015Q4$ will be observed 4 years later, that is, in the GDP vintage published in 2020Q1. Further details and statistical evidence to support time variation (i.e. $\varpi_g \neq 0$ and $\varpi_h \neq 0$) are provided in section A.2 below; this includes robustness checks showing that our results are little changed if we allow either for serial correlation in the transitory component or for t-distributed innovations to accommodate *outliers*.

A.1.2 Dating Changes in the Revisions' Process: Results from the UC-SV Model

We consider estimates of the UC-SV model in eq. (11) with $b = 1$, that is, we model revisions from the ONS 'preliminary' estimate up to the "mature" estimate observed at $t+l$. As discussed previously, we consider both $l = 13$ and $l = 17$ as alternative measures of "mature" data. Figure A3 plots the smoothed (posterior mean) estimates of the revision mean $\mu_t^{(l-b)}$, the measurement error and revision mean volatilities and the implied θ_t, alongside 90% credible intervals, for $l = 17$.

The top panel of Figure A3 shows that the revision mean has fluctuated over time, varying between -3% and $+3\%$. But $\mu_t^{(16)}$ is generally positive or statistically equal to zero for the 16 year period between 1991 and 2007. This is consistent with the visual impression from Figure 1, that ONS first release data tended to be revised upwards, and also the evidence in Tables A1 and A2 that the average (over time) revision is positive, especially relative to mature data published after 4 years. But Figure A3 suggests that underlying this are sizeable swings. Interestingly, at the time (late 2007) that the MPC started publishing its backcasts in the *Inflation Report*, the revision mean switched from positive to negative. This period, of course, also coincides with the onset of the Great Recession.

Figure A3 also plots both volatility estimates. Measurement error volatility declines significantly over time, from 0.7 in 1983 to 0.1 in 2012. We might interpret this as evidence that the ONS have *improved* their estimates of early GDP releases over time. Revision mean volatility is seen, in Figure A3, to increase during the 2008/2009 recession and also at the onset of the recession in 1990. Data revisions increase in recessions. When combining both volatilities to backout θ_t, we see that θ_t increases over time - as transitory shocks increase in size relative to permanent shocks. The posterior mean estimates for θ_t peak at the onset of the 2008 recession, suggesting a very low degree of revision predictability during the recession.

Figure A4 repeats Figure A3 but for $l = 13$, that is, we consider "mature" data is observed after 3 years. The (posterior mean) estimates of the revision mean $\mu_t^{(12)}$ exhibit less variability than when $l = 17$. For example, $\mu_t^{(12)}$ is negative for reference quarters between 2006Q3 and 2009Q1. But 3 quarters of positive mean revisions are identified during this period for $\mu_t^{(16)}$. The increase in the revision mean volatility for $rev^{(12)}$ at the onset of the 2008 recession is sharper, and the decline of the predictability over time observed in θ_t is steeper.

In summary, we find evidence of temporal changes to the GDP growth data revisions' process. The revision mean has increased in size, and changed sign, with the onset of the 2008/2009 recession. Measurement error volatility declined rapidly between 1983 and 1993; but revision mean volatility increased during the turbulent 2007/2009 period. A possible reason for these changes, as suggested by Office for National Statistics (2017), is that the statistical office's models do not perform as well around turning points.

A.2 UC-SV Model for Data Revisions: Additional Results

Here we present additional details about estimation of the UC-SV model in Section A.1.1.

We estimate the UC-SV model by Gibbs sampling using the algorithm in Chan (2018). In order to test whether we need both stochastic volatility factors, we consider the Bayes' factor as in Chan (2018). Table A3 shows the Bayes' factors (with standard deviations over 5 chains of 100,000 kept draws in parentheses) testing whether each volatility process is time varying. Positive Bayes' factors suggest that there is indeed time-variation to the volatility. Bayes' factors are presented for each volatility process and also imposing the joint restriction that $\varpi_g = \varpi_h = 0$. The results are presented assuming that the 17^{th} release is the final revised value ($l = 17$); and making different assumptions about the earlier release: $b = 1, 4, 8, 12$. The statistics in Table A3 suggest, for all revision processes considered, that there is time variation in both the measurement error and local revision mean volatility.

We also considered a specification that allows for serial correlation in the transitory component, by writing the first equation of (11) as:

$$rev_t^{(l-b)} = \mu_t + \rho(rev_{t-1}^{(l-b)} - \mu_{t-1}) + e^{.5(h_0 + \varpi_h \widetilde{h}_t)}\zeta_{\varepsilon,t}. \tag{12}$$

After estimating this modified model for $l = 17$ and $b = 1$, we found that ρ is estimated to be about 0.05 so, qualitatively, the results in Figure A3 are unchanged.

We then allowed the innovations to have a t-distribution, so that the density for data revisions could exhibit fat tails. We use the methods described in Chan and Hsiao (2014) to estimate both stochastic volatility processes with a t-distribution. We find estimates of the t-distribution degrees of freedom around 25, implying that, qualitatively, the results are again similar to those shown in Figure A3 using the normal distribution.