

Measuring Data Uncertainty: An Application using the Bank of England’s “Fan Charts” for Historical GDP Growth*

Ana Beatriz Galvao

University of Warwick and ESCoE

James Mitchell

University of Warwick and ESCoE

April 2019

Abstract

Historical economic data are often uncertain due to sampling and non-sampling errors. But data uncertainty is rarely communicated quantitatively. An exception are the “fan charts” for historical GDP growth published at the Bank of England. We propose a generic loss function based approach to extract from these *ex ante* density forecasts a quantitative measure of unforecastable data uncertainty. We find GDP data uncertainty in the UK rose sharply at the onset of the 2008/9 recession; and that data uncertainty is positively correlated with popular estimates of macroeconomic uncertainty.

Key words: data revisions; fan charts; macroeconomic uncertainty; backcasts; ex ante uncertainty; ex post uncertainty; density forecast calibration; real time data

JEL code: C53, E32

*The authors gratefully acknowledge financial support from the Economic Statistics Centre of Excellence (ESCoE). We also thank two anonymous referees at ESCoE, seminar participants at the Bank of England, Reading, UCL and the ESCoE for helpful comments. Corresponding author: Prof. Ana Beatriz Galvao, Warwick Business School, University of Warwick, Coventry, CV4 7AL, U.K.; ana.galvao@wbs.ac.uk

1 Introduction

Macroeconomists are increasingly concerned with the impact that “uncertainty” has on the economy, including its putative negative effects on business investment, consumption and aggregate output. Indeed, Bloom (2014) emphasises how uncertainty is linked to business cycle phases, rising in recessions and falling in booms. Empirical uncertainty measures typically include stock market volatility, macroeconomic forecasting uncertainty, disagreement among professional forecasters and economic policy uncertainty (Bachman, Elstener and Sims, 2013; Jurado, Ludvigson and Ng, 2015; Baker, Bloom and Davis, 2016; Rossi, Sekhposyan and Soupre, 2016; Sekkel and Soojin, 2018). But as these uncertainty measures have been found to differ, both in their historical behaviour and their impact on the macroeconomy (Rossi and Sekhposyan, 2015), attempts to distinguish between them are made. This has involved making the Knightian distinction between risk (*ex ante* uncertainty, “known unknowns”) and uncertainty (“unknown unknowns”) (Rossi et al., 2016), and between macroeconomic and specific, such as financial, uncertainties (Ludvigson, Ma and Ng, 2018). This paper complements these developments by measuring (realised) data uncertainty as the *ex post* mistakes a forecaster makes in her *ex ante* probabilistic assessments of the historical data. Unforecastable uncertainty is defined as the difference between the *ex post* (or realised) and *ex ante* (or expected) values of the loss function chosen by the user of the forecast.

Data uncertainty arises when economic statistics from official surveys are subject to sampling and non-sampling errors.¹ Manski (2015) re-interprets this uncertainty as comprising both “transitory” and “permanent” statistical uncertainty. Transitory statistical uncertainty stems from publication of early data releases that are revised over time as new information arrives. Permanent statistical uncertainty arises due to data incompleteness (e.g. non-response) or the inadequacy of data collection (e.g. sampling uncertainty due to a finite sample) which does not diminish over time.

Official measures of variables like GDP, from national statistical offices, are revised as new information is received and methodological improvements are made.² For example, over the

¹In official statistics, “data uncertainty”, as defined here, is commonly captured by “accuracy and reliability” as in Principle 12 of Eurostat’s European Code of Practice (<https://ec.europa.eu/eurostat/web/quality/principle12>). Accuracy is defined by the *error* that arises when estimates differ from the underlying exact or *true* value - due to sampling and non-sampling errors. Reliability relates to assessment and validation of the source data and statistical processes, with data revisions occurring as a result of improvements to these processes.

²McKenzie (2006) delineates seven reasons for “revisions” including updated sample information, correction of errors, replacement of first estimates derived from incomplete surveys/judgements/statistical techniques, bench-

period of the study in this paper, the Office for National Statistics (ONS) in the UK published its first - so-called “preliminary” - quarterly GDP estimates around 27 days after the end of the quarter. Because this timeliness is achieved by basing their estimate on 44% of the sample, it is (and should be) no surprise to see the ONS subsequently revise these preliminary estimates as more sampling information subsequently becomes available to them.³ As a consequence, many authors have proposed models of data revisions to model and forecast this “transitory” GDP data uncertainty (e.g. see Jacobs and van Norden (2011), Cunningham, Eklund, Jeffery, Kapetanios and Labhard (2012), Kishor and Koenig (2012) and Galvao (2017)).

But while the ONS do discuss data revisions, and emphasise in their press releases and on their website that preliminary estimates of GDP will be revised, their headline estimates of GDP remain *point* estimates - with no accompanying quantitative measures of uncertainty. Manski (2015; 2018) has emphasised how this practice of acknowledging potential data uncertainties at best qualitatively or verbally, rather than quantitatively, is common practice across statistical offices; and has called for more transparent communication of data uncertainties. This paper picks up Manski’s call to measure and communicate data uncertainties, focusing on GDP as a key measure of the macroeconomy.⁴

Data revisions have been found to affect assessments of the current state of the economy and the monetary policy stance (e.g. see Orphanides (2001) and Croushore (2011)). The Deputy Governor for Monetary Policy at the Bank of England (Bean (2007)), quoting former US Secretary of Defence Donald Rumsfeld, has emphasised how monetary policy faces both *data risk* (Rumsfeld’s “known unknowns”) and *data uncertainty* (his “unknown unknowns”).

An important example, and rare illustration, of how (historical) data uncertainty is both communicated and affects policy making is provided by the Monetary Policy Committee (MPC) at the Bank of England. As well as forecasting the future, the MPC provide direct estimates of data risk, of *ex ante* 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

marking, updated seasonal factors, updated base period for constant price estimates and changes in statistical methodology.

³In 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 “preliminary” 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.

⁴We confine attention to measurement of data uncertainties stemming from data revisions, i.e. to “transitory” statistical uncertainty as defined by Manski (2015). Broader uncertainties, including methodological or conceptual uncertainty, are not considered. For a more general classification of uncertainties see van der Bles, van der Linden, Freeman, Mitchell, Galvao, Zaval and Spiegelhalter (2019).

of England’s *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).⁵

Figure 1 provides an example of what these fan charts look like, taken from the May 2017 *Inflation Report*. In Figure 1, “(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 1 shows that the fan becomes progressively narrower as one looks further back into the past (from the perspective of May 2017). 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 1, post May 2017) to these older, more historical estimates that date back to 2013. The ONS’s latest (as of May 2017) estimate of GDP growth is shown in Figure 1 by the solid black line. Cunningham and Jeffery (2007) provide an explanation of the data revisions’ model, used by Bank staff, that along with MPC judgment helps shape the form of these backcast fan charts. Their model exploits historical patterns in ONS revisions and information from qualitative business surveys.

Extracted from successive quarterly issues of the *Inflation Report*, Figure 2 presents, in its top panel, the MPC’s characterisations from 2007Q3 to 2017Q1 of the *ex ante* or expected revision to each *first* (or “preliminary”) estimate of GDP growth (formally defined as $b = 1$, below) that was published by the ONS. And in its bottom panel, Figure 2 presents the MPC’s expectation of the uncertainty (as measured by the standard deviation) associated with this. It should be borne in mind that, due to timings, the MPC has sight of the ONS’s latest “preliminary” estimate for GDP growth when it forms its probabilistic expectations of this first (historical) estimate. Figure 2 shows that the MPC has always expected revisions to be positive - they consistently expected the ONS to subsequently revise upwards their “preliminary” estimate of GDP growth. But the expected size of this revision has varied over time. By contrast, looking at the bottom panel of Figure 2, the MPC has made changes to its expectations of revision 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; as the *Blue Book* typically involves extensive annual revisions to the national

⁵Strictly, 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.

accounts. It is also evident from Figure 2 that the MPC has become more uncertain about GDP revisions over time: expected uncertainty doubled between 2007 and 2017.

To appreciate the size of these *ex ante* uncertainty estimates consider the following back-of-the-envelope calculation. If the latest estimate of GDP growth (year-on-year) were 2% and the expected revision was zero, the 2017 level of *ex ante* uncertainty would suggest that the MPC is expecting the ‘mature’ value of GDP growth to fall, with a 90% probability, between 0% and 4%. Arguably, this is a considerable degree of uncertainty as the MPC are neither ruling out *no* growth nor *strong* (above trend) growth.

Simply eye-balling the width of the fan chart to the left and right of the vertical dashed line in Figure 1 also suggests that data uncertainty matters. If (historical) data uncertainties were small relative to future macroeconomic uncertainties, we should expect the width of the fan chart to widen appreciably, perhaps even sharply, as we move to the right of the vertical dashed line. But inspection of Figure 1 (and we observe a similar picture from other *Inflation Reports* at other forecast origins) reveals only a smooth and modest increase in the width of the fan chart as it moves to the right of the vertical dashed line: data uncertainty appears to matter to the MPC.

The remainder of this paper is structured as follows. Section 2 first establishes the historical characteristics of GDP growth data revisions. As changes over time to data/sampling methods and methodology at the ONS lead us to expect temporal changes to the nature of the observed revisions’ process, and this likely affects policymakers’ ability to forecast revisions, the characteristics of data revisions are summarised over different sub-samples. A parsimonious Unobserved Components-Stochastic Volatility (UC-SV) model, that allows for time-variation both in the revision mean and in the volatilities of the measurement error and revision mean innovations, is also used to provide a flexible representation of how the data revisions’ process has evolved over the last 35 years.

In Section 3 we propose a new measure of uncertainty that captures the unforecastable component to forecasts (in our context, these forecasts are of data revisions). This involves distinguishing between the *ex ante* (expected) and *ex post* (or realised) values of the loss function that are relevant to the economic agent who is seeking to make decisions based on forecasts (in our context, uncertain data). We discuss how this new measure both generalises existing popular uncertainty measures, that assume specific loss functions, and relates to extant density forecast calibration tests. To better understand our uncertainty measure we decompose it into

interpretable components.

Section 4 then uses the MPC’s fan charts to measure *ex ante* uncertainty as this serves as the basis for us producing quantitative estimates of (unforecastable) data uncertainty. In so-doing, we provide the first evaluation of the accuracy and calibration of the MPC’s predictive densities for historical GDP growth values. While there have been numerous studies evaluating the MPC’s forecasts for future GDP (Clements, 2004; Mitchell and Hall, 2005; Groen, Kapetanios and Price, 2009; Galbraith and van Norden, 2012; Independent Evaluation Office, 2015), we are not aware of an evaluation of their probabilistic backcasts. We find that the MPC’s point predictions better anticipate mature ONS growth estimates than the ONS’s own first releases; and that their density estimates are, on average, well-calibrated except for the oldest data - to the extreme left of the fan chart as seen in Figure 1.

In Section 5, we consider the time variation in our measure of data uncertainty. We use a decomposition of our proposed uncertainty measure to understand better what contributes to the observed time-variations in uncertainty. We find that rises in uncertainty are linked to unexpected deteriorations in the reliability of the MPC’s backcasts. We also compare our measure of data uncertainty with popular measures of macroeconomic uncertainty, as computed for the UK by Redl (2017) and Bank of England (2016). We find that data uncertainty is highly correlated with measures of macroeconomic uncertainty, and increases dramatically at the onset of the 2008 recession. Section 6 concludes.

2 Historical Characteristics of GDP Growth Data Revisions

To quote the ONS: “The 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. 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”.⁶

It is, however, possible to provide an indication of “transitory” statistical uncertainties associated with GDP estimates by analysing historical revisions. And national statistical offices and central banks accordingly often now publish real-time data vintages and analyse the implied

⁶See <https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi>

revisions (e.g. see Croushore and Stark (2001)). Other sources of uncertainty, for example due to limitations of the survey methodology, are not represented; and methodological work on measuring non-sampling errors continues (e.g. see Manski (2016)).⁷

Our focus turns first to an examination of these real-time data vintages, before returning to the MPC’s assessment of *ex ante* data uncertainties. In part subjectively formed, these may well seek to capture uncertainties beyond “transitory” statistical ones.

2.1 Real-time Data

We use the real-time dataset of UK real GDP published by the ONS.⁸ From this dataset we extract quarterly vintage estimates, denoted y_t^{t+b} , of year-on-year GDP growth for the reference quarter t , ($t = 1, \dots, T$); with the superscript denoting the vintage date or publication date in quarters ($b = 1, 2, \dots$).⁹ 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 $(t + 1)$. Revisions between the “mature” data, y_t^{t+l} (where $l > b$), and an earlier b -th data release 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.¹⁰ For an analysis of UK data revisions over earlier sample periods see Garratt and Vahey (2006) and references therein.

We identify, and subsequently seek to discriminate between, two measures of “mature” data. This is to acknowledge that there is always debate about what defines a “mature” estimate. Data revisions are an ongoing process, so there is understandably uncertainty about the appropriate value of l ; in turn, there is 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.

⁷In measuring data uncertainty we focus on *epistemic* uncertainty, that is, uncertainty due to lack of knowledge about current and past data. Unlike *aleatory* uncertainty, more relevant for forecasts when there is stochastic uncertainty about future shocks than for backcasts, epistemic uncertainty is expected to reduce. Specifically, in our application, epistemic uncertainty is expected to diminish over time as additional data are collected and both corrections and improvements to these data are made by the statistics office.

⁸See <https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/realtimedatabaseforukgdpabmi>

⁹As 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 first releases, to avoid jumps due to changes in the chain-linking base year.

¹⁰The ONS made significant changes to its GDP publication model and release calendars in the summer of 2018. Our latest vintage is July 2017, so that our analysis is not affected by these recent changes.

We classify y_t^{t+13} as the 13th quarterly release 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).

Our second candidate measure of “mature” data is data-based, and reflects the time-series properties of historical data revisions. Anticipating the results in Table 1, the mean of ONS revisions to their GDP growth estimates after the third round of annual revisions is still statistically different from zero; but after four years these revisions have a mean that is statistically insignificant from zero. As a consequence we consider y_t^{t+17} , i.e. GDP growth as published by the ONS four years after their preliminary release.

Figure 3 plots the first (preliminary) release, y_t^{t+1} , and the 17th release, y_t^{t+17} , for quarters $t = 1982Q2 - 2013Q1$ and shows that while the first estimate is highly correlated with the later estimate, there are important gaps between the two. These gaps appear bigger, and more persistent, at certain points in time.

2.2 Split-Sample Summary Statistics for Data Revisions

To draw out these apparent temporal changes to the revisions’ process seen in Figure 3 we first analyse data revisions over different sub-samples. To understand the characteristics of the revisions’ process ($y_t^{t+17} - y_t^{t+1}$) over these, we analyse not just the total revision, i.e. ($y_t^{t+17} - y_t^{t+1}$), but that sub-component of the revision, ($y_t^{t+17} - y_t^{t+13}$), that lets us assess our earlier claim that important revisions happen even after three years. This matters as it has a bearing on what measure of “mature” data we choose empirically. Faust, Rogers and Wright (2005) and Aruoba (2008) explain that an elementary condition for a GDP estimate to be *efficient* is that there is no predictable revision bias. As a consequence, we also analyse revisions made between the 17th release and the (currently available) *latest* release, denoted ($y_t^{latest} - y_t^{t+17}$). The latest quarterly data vintage that we use in this paper is y_t^{2017Q2} .

Table 1 presents the mean, standard deviation and mean absolute error of each of these three revision processes. Statistics are computed for three sub-samples starting in 1983Q2, 1993Q2 and 2007Q3, cognisant that the revisions’ process may exhibit temporal instabilities.

Our samples end in $t = 2013Q1$, to reflect the fact we have to wait four years to observe y_t^{t+17} . We note that our last sub-sample, 2007Q3-2013Q1, coincides with the period over which we evaluate the MPC’s backcasts in Section 3 below. And the second sub-sample is determined by an apparent break in the volatility of the UK revision measurement error, as shown by the model we estimate in the next sub-section. Sample mean statistics in bold indicate that the null hypothesis, that the revision mean is zero, is rejected at a 10% significance level. Robust (Newey-West) standard errors are used to compute these t -statistics.

The statistics in Table 1 suggest that revisions between the “mature” GDP estimate, y_t^{t+17} , and the preliminary estimate, y_t^{t+1} , have a statistically significant and positive mean over both the 1983Q2-2013Q1 and 1993Q2-2013Q1 samples. The average revision remains positive, at 0.3 p.p., over the more recent 2007Q3-2013Q1 sample; but due to heightened volatility in the revisions’ process, evidenced by an elevated standard deviation, this is no longer statistically significant from zero. Moreover, over the 1993Q2-2013Q1 period the revision mean between y_t^{t+13} and y_t^{t+17} is statistically different from zero. This result, along with the finding that there is little evidence that revisions made after y_t^{t+17} have a non-zero mean (i.e. that the mean of $(y_t^{latest} - y_t^{t+17})$ is statistically insignificant from zero), serves to motivate empirically our decision to define the “mature” estimate as y_t^{t+17} .

In the online appendix we show analogous statistics for US GDP data. If we measure revision size by the mean absolute error, then UK and US data revisions to GDP growth are similarly sized in the 20-year period between 1993 and 2013.

Additional information on the nature of UK data revisions is provided in Table 2, which presents t -statistics for whether revisions are *news* or *noise*. These tests are based on the forecast efficiency regressions:

$$\left(y_t^{t+l} - y_t^{t+b}\right) = \beta_0^{news} + \beta_1^{news} y_t^{t+b} + \varepsilon_t \quad (1)$$

$$\left(y_t^{t+l} - y_t^{t+b}\right) = \beta_0^{noise} + \beta_1^{noise} y_t^{t+l} + \varepsilon_t. \quad (2)$$

The null hypothesis that data revisions add information (they contain *news*) implies $\beta_1^{news} = 0$. If data revisions remove the measurement error (*noise*) in the initial release then $\beta_1^{noise} = 0$. For additional details on the application of these tests see Clements and Galvão (2010) and references therein. The t -statistics in Table 2 suggests that revisions to the preliminary estimate, y_t^{t+1} , or the 13th release, y_t^{t+13} , are in general a result of the ONS adding new information. In

contrast, revisions made after the 17th release are mainly noise, i.e. data revisions are negatively correlated with the earlier estimates in y_t^{t+17} . This supports the view that earlier rounds of data revisions mainly contain new information. This reinforces our decision to use y_t^{t+17} as our measure of “mature” UK GDP growth.

In summary, ONS preliminary estimates of GDP growth tend to be significantly lower than the values they release four years later. The MPC appear to have been aware of this in real-time too, as seen by the positive expected revision in Figure 2. The statistical evidence that revisions contain news, and of temporal changes to the revisions’ process, indicates that data revisions are likely hard to predict *ex ante*. This, in turn, suggests that unforecastable uncertainties may be important, and that users of data like the MPC perhaps need to be alert to Knightian uncertainty (unknown unknowns).

2.3 A UC-SV Model of Changes in the Data Revisions’ Process

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

$$\begin{aligned} y_t^{t+l} &= y_t^{t+b} + \mu_t + e^{.5(h_0 + \varpi_g \tilde{h}_t)} \zeta_{\varepsilon,t} \\ \mu_t &= \mu_{t-1} + e^{.5(g_0 + \varpi_h \tilde{g}_t)} \zeta_{\eta,t} \\ \tilde{h}_t &= \tilde{h}_{t-1} + \zeta_{h,t} \\ \tilde{g}_t &= \tilde{g}_{t-1} + \zeta_{g,t} \\ \zeta_{\varepsilon,t}, \zeta_{\eta,t}, \zeta_{g,t}, \zeta_{h,t} &\text{ are all } N(0, 1) \end{aligned} \tag{3}$$

implying data revisions, $(y_t^{t+l} - y_t^{t+b}) = rev_t^{(l-b)}$, comprise two unobserved components: (i) a time-varying mean, μ_t ; and (ii) a mean-zero measurement error, $\zeta_{\varepsilon,t}$. Cunningham et al. (2012) also decompose data revisions into bias and measurement error components.¹¹ But our model 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 forecasting error using the earlier release y_t^{t+b} to predict the later release, y_t^{t+l} . By fitting a 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 (2019).

¹¹Other 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.

The model in (3) 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 (2018).

Estimation of the four parameters in (3), i.e. g_0 , ϖ_g , h_0 and ϖ_h , proceeds as in Chan (2018). We use 35 years of full-sample revisions data, $rev_t^{(l-b)}$, from 1983. Further details and statistical evidence to support time variation (i.e. $\varpi_g \neq 0$ and $\varpi_h \neq 0$) are provided in the online appendix; 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*.

2.3.1 Dating Changes in the Revisions' Process: Results from the UC-SV Model

We show here results having estimated the UC-SV model on revisions between the ONS 'preliminary' estimate ($b = 1$) and their 17th release ($l = 17$). Figure 4 plots the smoothed (posterior mean) estimates of the revision mean μ_t , the measurement error and revision mean volatilities and the implied θ_t , alongside 90% credible intervals.

The top panel of Figure 4 shows that the revision mean has fluctuated over time, varying between -3% and +3%. But μ_t is generally positive or statistically equal to zero, except from 1989Q4-1990Q2, 2004Q3-2005Q1, 2006Q4-2007Q1 and 2008Q2-2009Q2. Recall Table 1 showed the revision mean to be positive. But Figure 4 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 4 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. Or that it is easier to produce *good* estimates when growth is relatively stable. The fact that most of the decline in volatility is over by 1993 explains why we present the statistics in Table 1 over a sub-sample beginning in 1993Q2. Table 1 indicates a declining revision mean if earlier observations are ignored.

Revision mean volatility is seen, in Figure 4, 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.

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. Interestingly, as described in Appendix A (Table A1 and Figure A1), we also find evidence of an increase in revision mean volatility in the US during the Great Recession.

2.3.2 The impact of different data maturities

To examine whether revisions' behaviour is similar for data of different maturities, we re-estimate the UC-SV model, (3), for $b = 4, 8$ and 12 . We continue to set $l = 17$. From Figure 5 we see that the revision mean volatility increases during the recessionary period for all values of b . The decline in the trend seen for the measurement error volatility is also observed when $b = 4$, but there are increases later in the sample when $b = 8$ and $b = 12$. An interesting result from Figure 5 is that the revision mean and revision mean volatility have experienced large increases in the 2007-2009 period at $b = 12$: i.e. *late* data revisions are proportionally larger in the 2006-2013 period than in the 1983-2005 period.

3 A Measure of Uncertainty

Having found, based on the summary in-sample analysis above, that there have been marked temporal changes to the ONS data revisions' process, we turn to consider the real-time implications. In so doing, we propose a measure of uncertainty that is based on an underlying full predictive density. We distinguish between *ex ante* (expected) and *ex post* (or realised) values of the loss function (or scoring rule) that are relevant to the economic agent who is seeking to make

decisions under uncertainty. Section 4 then considers empirical measures of data uncertainty, using the probabilistic backcasts provided by the MPC.

3.1 Measuring (unforecastable) Uncertainty

We aim to measure uncertainty by the lack of knowledge about the future realisation. We quantify this lack of knowledge by the difference between the *ex post* (or realised) and the *ex ante* (or expected) values of the loss function (or scoring rule) chosen by the user of the forecast.¹² In other words, extending the logic of Jurado et al. (2015) from quadratic loss and point forecasts to a generic loss function and to density forecasts, we propose that when measuring uncertainty what matters is what is unforecastable, as stressed by Jurado et al. (2015), but also *relevant*. Relevance is defined by the user's loss function.

Let $f_{t|t-h}$ denote the *density forecast* formed in period $(t-h)$ for the realisation for period t . If, as in the case of backcasts, predictions for the revised values for period t , realised at time $(t+l)$ where $l > b$, are made at time $(t+b)$ then $h = -b$, and the density is $f_{t|t+b}$, as the user is uncertain about the past rather than the future.

In generic terms, we propose to measure uncertainty, $Unc_{t|t-h}$, as:

$$Unc_{t|t-h} = L(f_{t|t-h}, y_t) - E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)] \quad (4)$$

($t = 1, \dots, T$), where $L(f_{t|t-h}, y_t)$ is the realised, *ex post* value of the chosen loss function or scoring rule (defined to have negative orientation, so that the lower the score the better) whose arguments are the density forecast $f_{t|t-h}$ (with associated cumulative distribution function, $F_{t|t-h}$) of the random observation Y_t with subsequent realisation y_t .

The (*ex ante*) unconditional expectation, $E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)]$, with respect to the forecast, $f_{t|t-h}$, is formed by the user at time $t-h$, so is made before the realisation, y_t , is subsequently observed. There is always a question about how economic agents form their expectations. We assume - in the context of measuring unforecastable uncertainty - that they are computed *optimally* so as to minimise loss; i.e. $E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)] = \arg \min_{f_{t|t-h}(\cdot)} E[L(f_{t|t-h}, Y_t)]$; this implies

$$E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)] = \int f_{t|t-h}(y) L(f_{t|t-h}, y) dy \leq \int f_{t|t-h}(y) L(f_{t|t-h}^*, y) dy = E_{f_{t|t-h}} [L(f_{t|t-h}^*, Y_t)] \quad (5)$$

¹²This user need not be the producer of the forecast.

for any alternative density forecast, $f_{t|t-h}^* \neq f_{t|t-h}$. These *ex ante* forecasts are in effect computed by (honest and loss-minimising) agents who assume (*ex ante*) that their forecast is as good as it can be. But there is no presumption that this forecast is well-calibrated *ex post* (see section 3.2 below). For (5) to hold, we restrict $L(f_{t|t-h}, Y_t)$ to the class of “proper” scoring rules (see Gneiting and Raftery (2007)).¹³

The expected loss, $E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)]$, can also be interpreted as the “risk” of the forecast; e.g. see Elliott and Timmermann (2016). Gneiting and Raftery (2007) call this expected loss (or score) the entropy function. See equation (9) below for further discussion of entropy; below we also see that for Gaussian predictive densities and some specific loss functions $E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)]$ relates directly to the *ex ante* predicted variance. Loss is minimised when $f_{t|t-h} = g_t$, where $Y_t \sim g_t$.

3.2 Testing calibration for a given loss function

A scoring rule or loss function summarises the calibration and sharpness of the predictive density $f_{t|t-h}$; see Gneiting and Raftery (2007). Here we explain how calibration can be tested for a specific scoring rule; and relate to calibration tests that are robust to the chosen scoring rule.

Given the loss function $L(f_{t|t-h}, y_t)$, correct unconditional average calibration of the forecast $f_{t|t-h}$ with respect to the realisation y_t , is defined as when

$$H_0^U : E(Unc_{t|t-h}) = 0; \quad (6)$$

in other words, when the values of the *ex ante* and *ex post* loss functions are equal in expectation.

Following Giacomini and White (2006), a conditional calibration test can also be defined with respect to a k -vector of “test functions”, w_{t-h} , representing the information set assumed known at the time the forecast, $f_{t|t-h}$, was made:

$$H_0^C : E(Unc_{t|t-h} | w_{t-h}) = 0. \quad (7)$$

Noting that $w_{t-h} = 1$ in (6), Wald-type test statistics

$$GW_T = T \left(T^{-1} \sum_{t=1}^T w_{t-h} Unc_{t|t-h} \right)' \widehat{\Sigma}^{-1} \left(T^{-1} \sum_{t=1}^T w_{t-h} Unc_{t|t-h} \right) \quad (8)$$

¹³Our focus on strictly proper scoring rules means that ‘truth telling’ (namely quoting the true density as the forecast density) is the optimal strategy in expectation; see Gneiting and Raftery (2007).

can then be used to test H_0^C and H_0^U , where $\hat{\Sigma}$ is a consistent estimator for the asymptotic variance of $(w_{t-h}Unc_{t|t-h})$ and $GW_T \xrightarrow{d} \chi_k^2$.

As we elaborate on in the following sub-section, non-zero values for $Unc_{t|t-h}$ - indicating that the forecaster is found *ex post* to have been unable to characterise the *ex ante* probability distribution correctly - can be interpreted as measures of (realised) Knightian uncertainty.¹⁴ If $Unc_{t|t-h} \neq 0$ the forecaster has made forecasting errors or ‘mistakes’, albeit perhaps probabilistic ones. These errors indicate that unknown (or unexpected) unknowns - Knightian uncertainty - have impacted y_t beyond both any *ex ante* uncertainty, or risk, that was anticipated at time $(t-h)$ and any inability or reluctance of the agent to quantify probabilistic forecasts for certain states. Movements in $Unc_{t|t-h}$ may be caused by shocks to confidence or risk, the definition of ambiguity in Ilut and Schneider (2014).

We emphasise the importance of our assumption that the *ex ante* forecasts are computed *optimally* - so as to minimise *ex ante* loss, (5). One can, of course, always imagine another *inferior* forecaster, who perhaps uses a misspecified forecasting model, who makes more forecasting errors than $f_{t|t-h}$: for this forecaster, $Unc_{t|t-h}$ confounds Knightian uncertainty (unknown unknowns) with mistakes that were in principle avoidable *ex ante* (known unknowns); or at least they were known by a better forecaster - see Example 1 below. In other words, one forecaster’s *unknown* unknowns are always a superior forecaster’s *known* unknowns. In trying to provide quantitative measures of unforecastable uncertainty it is therefore important empirically to select a *good* forecaster.

3.3 Understanding Uncertainty: A Decomposition

To understand better what $Unc_{t|t-h}$ measures note that, in principle (as computation can be difficult), *ex post* one can decompose any loss function into interpretable components. With g_t continuing to denote the ideal density, we can decompose $L(f_{t|t-h}, y_t)$ into *reliability* and *entropy* components:

$$L(f_{t|t-h}, y_t) = \underbrace{L(f_{t|t-h}, y_t) - L(g_t, y_t)}_{\text{reliability}} + \underbrace{L(g_t, y_t)}_{\text{entropy}} \quad (9)$$

where the entropy, $L(g_t, y_t)$, is the minimal achievable *ex post* loss under $L(\cdot)$. Entropy is a feature of the data, and the chosen loss function, not of the forecasts themselves. Given $L(g_t, y_t)$

¹⁴One can, of course, never measure unrealised (over the sample, T) *unknown* unknowns using our approach.

is the minimum *ex post* loss, $reliability \geq 0$.

In turn, (9) implies that our uncertainty measure comprises both an (*ex post*) *reliability* component and, what we call, expected (*ex ante*) reliability, $reliability^e$:

$$Unc_{t|t-h} = \underbrace{[L(f_{t|t-h}, y_t) - L(g_t, y_t)]}_{reliability_t} - \underbrace{[E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)] - L(g_t, y_t)]}_{reliability_t^e}. \quad (10)$$

Specific values for $reliability_t$ and some of the decomposition components depend on the loss function.¹⁵ Some analytical expressions are available for the Briers score (Yates, 1982), Mean Squared Error (MSE) loss (Murphy and Winkler, 1987), the Kullback-Leibler divergence (Weijs, van Nooijen and van de Giesen, 2010), the quantile score (Bentzien and Friederichs, 2014) and Continuous Ranked Probability Score (CRPS) loss (Rossi et al., 2016).

We can draw out properties of these generic decompositions for well-calibrated densities, as defined under H_0^U . If $reliability = 0$, given that $f_{t|t-h}$ is formed *optimally*, then $reliability^e = 0$ and $Unc_{t|t-h} = 0$ implying no (realised) Knightian uncertainty. When $reliability = 0$ no alternative forecast can achieve lower *ex post* loss.

However, when $reliability \geq 0$ it is still possible that $Unc_{t|t-h} = 0$ if $reliability = reliability^e$, i.e. if forecasters offset *ex post* errors, that explain positive values for *reliability*, by *ex ante* expecting a greater loss than if they did know the ideal density, g_t .¹⁶ In this situation it is arguably a matter of semantics whether there is Knightian uncertainty or not. On one hand, $Unc_{t|t-h} = 0$ tells us that $f_{t|t-h}$ is well-calibrated which indicates that unknown unknowns have not led to forecast errors. But on the other hand, since another forecast g_t exists which would deliver a lower *ex post* loss, there are unquantified unknown unknowns. And if these shocks were anticipated *ex ante*, i.e. if g_t were used instead of $f_{t|t-h}$, $reliability = 0$ and the user would have made lower loss decisions. We use an example to illustrate this point.

¹⁵Further intuition might be gained by decomposing entropy around a benchmark unconditional density forecast, b_t (often called the climatological forecast), whose moments match those of the data, y_t , *ex post*:

$$Unc_{t|t-h} = \underbrace{[L(f_{t|t-h}, y_t) - L(g_t, y_t)]}_{reliability} + \{ \underbrace{L(b_t, y_t)}_{realised\ variance} - \underbrace{[L(b_t, y_t) - L(g_t, y_t)]}_{resolution} - \underbrace{E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)]}_{expected\ loss} \} \quad (11)$$

so that uncertainty now comprises *reliability*, $L(b_t, y_t)$ which can be interpreted as a measure of realised variance, *resolution* and expected loss. Resolution or sharpness measures the informational content of the unconditional forecast and equals zero when $b_t = g_t$.

¹⁶When the conditional and unconditional data densities differ, such that $resolution > 0$, $E_{f_{t|t-h}} [L(f_{t|t-h}, Y_t)] < L(b_t, y_t)$, and the expected loss of forecasting using $f_{t|t-h}$ is lower (better) than if b_t were used. For a well-calibrated density, expected loss should equal the realised variance only if no resolution is required to fit the true density.

Example 1 Following Gneiting, Balabdaoui and Raftery (2007), consider the true density $g_t = N(\mu_t, 1)$ where $\mu_t = N(0, 1)$; and let $L(f_{t|t-h}, y_t)$ be the logarithmic scoring rule.¹⁷ The unconditional or so-called climatological forecaster $b_t = N(0, 2)$ is such that $Unc_{t|t-h} = 0$. This is because $reliability_t^e > 0$ as this forecaster in a sense deliberately but optimally overstates the variance relative to the conditional truth, to compensate for missing μ_t . Ignoring constants, this is seen as follows:

$$\begin{aligned}
Unc_{t|t-h} &= \underbrace{[L(f_{t|t-h}, y_t) - L(g_t, y_t)]}_{\text{reliability}} - \underbrace{\{[L(b_t, y_t) - L(g_t, y_t)] - \underbrace{L(b_t, y_t)}_{\text{realised variance}}\}}_{\text{resolution}} + \underbrace{E_{f_{t|t-h}}[L(f_{t|t-h}, Y_t)] - \}_{}}_{\text{expected loss}} \\
0 &= \underbrace{[\ln(\sqrt{2}) - \ln(1)]}_{\text{reliability}} - \underbrace{[\ln(\sqrt{2}) - \ln(1)] - \underbrace{\ln(\sqrt{2})}_{\text{realised variance}} + \underbrace{\ln(\sqrt{2})}_{\text{expected loss}}}_{\text{reliability}^e}
\end{aligned} \tag{12}$$

3.4 Uncertainty for Four Loss Functions

By defining uncertainty in terms of a generic loss function we both nest special cases from the existing literature, such as Jurado et al. (2015) considered further below, and also directly link our measures of uncertainty to tests in the tradition of Dawid (1984) and Diebold, Gunther and Tay (1998) for the absolute calibration of $f_{t|t-h}$.

Here we consider four choices of $L(f_{t|t-h}, y_t)$. Conveniently, as it aids interpretation, we show below that for Gaussian predictive densities with predicted mean, $\hat{y}_{t|t-h}$, and predicted variance, $\hat{\sigma}_{t|t-h}^2$, analytical expressions for $Unc_{t|t-h}$ can be defined for each of these four loss functions. Let Φ denote the CDF of a standardised Gaussian distribution, ϕ denote the associated PDF and $z_t = \frac{(y_t - \hat{y}_{t|t-h})}{\hat{\sigma}_{t|t-h}}$ denote the standardised point forecasting error.

The first loss function is MSE loss, defined as $L^{MSE}(f_{t|t-h}, y_t) = (y_t - \hat{y}_{t|t-h})^2$. Under L^{MSE} , uncertainty is:

$$Unc_{t|t-h}^{MSE} = (y_t - \hat{y}_{t|t-h})^2 - \hat{\sigma}_{t|t-h}^2 \tag{13}$$

since $E_{f_{t|t-h}}(y_t - \hat{y}_{t|t-h})^2 = \hat{\sigma}_{t|t-h}^2$ for the *optimal* forecaster. Jurado et al. (2015) emphasise the importance of removing the forecastable component (in their case for the conditional mean) when measuring uncertainty (about the mean). Their uncertainty measure is the *ex ante* expectation of the MSE loss function, i.e. $E_{f_{t|t-h}}(y_t - \hat{y}_{t|t-h})^2 = \hat{\sigma}_{t|t-h}^2$. What Jurado et al. (2015) do not address is how ‘good’ a forecast this was of *ex post* uncertainty, measured under MSE loss as $(y_t - \hat{y}_{t|t-h})^2$. By contrast, our measure of uncertainty, $Unc_{t|t-h}^{MSE}$, under MSE loss as

¹⁷See (16) for a definition of the logarithmic score.

assumed by Jurado et al. (2015), isolates this ‘error’. As it is this ‘error’, deviation between the *ex ante* and *ex post* values of the loss function, that captures (realised) Knightian uncertainty for the *optimal* forecaster. Under the null $E(Unc_{t|t-h}^{MSE}) = 0$, our test using GW_T is similar to the test proposed in Clements (2014) to compare *ex ante* and *ex post* forecasting uncertainty measures.

The second loss function scores the coverage for $(1 - \alpha)$, $0 < \alpha < 1$ prediction intervals. The lower and upper limits of these intervals are the predicted quantiles $\hat{y}_{t|t-h}^{(\alpha/2)}$ and $\hat{y}_{t|t-h}^{1-(\alpha/2)}$; and the loss function is:

$$L^{Int}(f_{t|t-h}, y_t) = 1 - (I(y_t < \hat{y}_{t|t-h}^{(\alpha/2)}) + I(y_t > \hat{y}_{t|t-h}^{1-(\alpha/2)})), \quad (14)$$

where $I(\cdot)$ is an indicator function. Under L^{Int} , uncertainty is:

$$Unc_{t|t-h}^{Int} = \alpha - (I(y_t < \hat{y}_{t|t-h}^{(\alpha/2)}) + I(y_t > \hat{y}_{t|t-h}^{1-(\alpha/2)})) \quad (15)$$

since $E_{f_t|t-h}(L^{Int}(f_{t|t-h}, Y_t)) = 1 - \alpha$ for the *optimal* forecaster. $Unc_{t|t-h}^{Int}$ will be positive when the observed coverage of the prediction interval is larger than the nominal rate, and negative when there is under-coverage. We could also compute uncertainty for value-at-risk forecasts, since they also normally require the computation of a quantile forecast $\hat{y}_{t|t-h}^{(\alpha)}$ implying that $Unc_{t|t-h}^{Int} = I(y_t < \hat{y}_{t|t-h}^{(\alpha)}) - \alpha$.

The third loss function is the logarithmic score loss function:

$$L^{\log S}(f_{t|t-h}, y_t) = \log(\hat{\sigma}_{t|t-h}) + 0.5 \log(2\pi) + 0.5 \left(\frac{(y_t - \hat{y}_{t|t-h})}{\hat{\sigma}_{t|t-h}} \right)^2, \quad (16)$$

implying that:

$$Unc_{t|t-h}^{\log S} = 0.5 z_t^2 - 0.5 \quad (17)$$

since

$$E_{f_t|t-h} \left[L^{\log S}(f_{t|t-h}, Y_t) \right] = 0.5(1 + \log 2\pi) + \log(\hat{\sigma}_{t|t-h}) \quad (18)$$

as the *optimal* forecaster sets $E_{f_t|t-h} (y_t - \hat{y}_{t|t-h})^2 = \hat{\sigma}_t^2$.

The fourth loss function is CRPS loss:

$$\begin{aligned} L^{CRPS}(f_{t|t-h}, y_t) &= \hat{\sigma}_{t|t-h} [z_t(2\Phi(z_t) - 1) + 2\phi(z_t) - (1/\sqrt{\pi})] \\ &= ((y_t - \hat{y}_{t|t-h})) (2\Phi(z_t) - 1) + 2\hat{\sigma}_{t|t-h}\phi(z_t) - (\hat{\sigma}_{t|t-h}/\sqrt{\pi}), \end{aligned} \quad (19)$$

such that:

$$Unc_{t|t-h}^{CRPS} = ((y_t - \hat{y}_{t|t-h})) (2\Phi(z_t) - 1) + 2\hat{\sigma}_{t|t-h}\phi(z_t) - 2(\hat{\sigma}_{t|t-h}/\sqrt{\pi}), \quad (20)$$

since $E_{f_{t|t-h}}(L^{CRPS}(f_{t|t-h}, Y_t)) = (\hat{\sigma}_{t|t-h}/\sqrt{\pi})$ for the *optimal* forecaster.

Similarly to (20), Rossi et al. (2016) also measure uncertainty using $L^{CRPS}(f_{t|t-h}, y_t)$. Our point of departure is to strip out *ex ante* uncertainty by subtracting $E_{f_t}(L^{CRPS}(f_{t|t-h}, y_t))$ from $L^{CRPS}(f_{t|t-h}, y_t)$. As well as facilitating a comparison with measures of uncertainty for other loss functions, this lets us directly relate Unc_t^{CRPS} to tests for the absolute calibration of the density forecast.

For $Unc_{t|t-h}^{\log S}$ and $Unc_{t|t-h}^{CRPS}$ we can relate the Wald test to tests for absolute calibration as commonly deployed on density forecasts. These test whether the probability integral transforms (pits), $F_{t|t-h}(y_t)$, are uniformly distributed, or equivalently (see Berkowitz (2001)), whether z_t follows a standard normal distribution. When the predictive densities are well calibrated and $z_t \sim N(0, 1)$, $E(Unc_{t|t-h}^{\log S}) = 0$ since $z_t^2 \sim \chi_1^2$ with a mean of 1. Similarly, $E(Unc_{t|t-h}^{CRPS}) = 0$ when $z_t \sim N(0, 1)$. Moreover, when $z_t \sim N(0, 1)$, $E(Unc_{t|t-h}) = 0$ for all strictly proper loss functions, $L(f_{t|t-h}, Y_t)$. This is analogous to the result in Diebold et al. (1998) that when a forecast coincides with the true data generating process, it will be preferred by all forecast users, regardless of their loss function.

4 Empirical Measures of Data Uncertainty using the MPC's Fan Chart

Operationally, our measure of uncertainty in (4) and the accompanying test for calibration, as set out in (6), require *ex ante* (density) probabilistic forecasts, $f_{t|t-h}$ - from the *optimal* forecaster. The ONS do emphasise data revisions in their communications. But, as discussed in the Introduction, it is the MPC at the Bank of England that, to our knowledge uniquely in an international context, provide direct quantitative estimates of the likely *ex ante* uncertainty

around past GDP values. We describe these backcasts in the next section; and while hesitant to classify the MPC as the *optimal* forecaster, we do believe their forecasts provide a competitive and an appropriate benchmark to measure unforecastable data uncertainty. This is because of international evidence that judgment-based forecasts, from institutional and professional forecasters like the MPC, typically outperform model-based forecasts, especially at short horizons (e.g. see Sekkel and Soojin (2018) and references therein). We therefore use the MPC’s *ex ante* density forecasts to measure unforecastable data uncertainty for each of the four loss functions considered in section 3.4. And we provide the first evaluation of the accuracy of the MPC’s predictive densities for historical GDP growth. We do so via both absolute and relative tests for calibration. Sample sizes are small, as the MPC have produced their forecasts quarterly for only twenty years; and this should be borne in mind when interpreting our statistical tests for calibration. Nevertheless, we do believe there is now a sufficient track-record to render these statistical tests meaningful.

4.1 The MPC’s Probabilistic Expectations of Future Data Revisions

The MPC have published probabilistic backcasts for “mature” GDP growth on a quarterly basis since November 2007.¹⁸ These backcasts take the form of Gaussian densities that represent the MPC’s view, formed in quarter t , of revised growth ending in quarter $b = t - 1, \dots, t - B$. In practice the MPC have looked $B \geq 16$ quarters back into the past.

The MPC’s probabilistic backcasts are represented as symmetric Gaussian densities. They are typically published near the beginning of the second month of quarter t . This means that the MPC were, in principle, able to observe the ONS’s “preliminary” GDP estimate for the previous quarter y_{t-1}^t , along with the ONS’s (perhaps revised) estimates for historical growth $y_{t-2}^t, \dots, y_{t-B}^t$, before they published their own estimates, $\hat{y}_{t-1}^t, \dots, \hat{y}_{t-B}^t$ and $\hat{\sigma}_{t-1}^t, \dots, \hat{\sigma}_{t-B}^t$, in the quarter t *Inflation Report*. This means, to give a concrete example, that when during quarter $(t + 4)$ the MPC publish their (fourth) density backcast for mature growth for reference quarter t , given as $N(\hat{y}_t^{t+4}, \hat{\sigma}_t^{2,t+4})$, they do so conditioning on the ONS’s (fourth) estimate, y_t^{t+4} .

The MPC’s sequence of probabilistic backcasts for future revisions to GDP growth for reference quarter t are therefore characterised as

$$f_{t|t+b} = N(\hat{y}_t^{t+b}, \hat{\sigma}_t^{2,t+b}), \quad (21)$$

¹⁸The spreadsheet that contains the parameters used are downloadable at <https://www.bankofengland.co.uk/inflation-report/2017/november-2017>

where $t + b$ ($b = 1, \dots, B$) denotes the quarter when the backcast is computed. As discussed in Section 2, we assume the “mature” estimate, y_t^{t+l} , is observed at $l = 17$. This implies the “forecast horizon” is $(l - b)$.

To illustrate, Figure 6 plots the 90% prediction intervals for “mature” GDP growth implied by the MPC’s estimates of \hat{y}_t^{t+1} and $\hat{\sigma}_t^{t+1}$, as seen in Figure 2 previously, for $t = 2007Q3, \dots, 2017Q1$. In anticipation both of measuring unforecastable data uncertainty, $Unc_{t|t+b}$, and the calibration tests, we superimpose on Figure 6 the ONS’s “mature” estimate, y_t^{t+17} . Figure 6 indicates that the MPC’s 90% intervals have widened appreciably since 2011; and that the actual ONS values fell outside these predictive intervals during the recessionary period. Over the sample as a whole, 17% of the ONS outturns fell outside the 90% bands. Under correct calibration, we should of course expect 10%. Figure 6 also reminds us why our evaluation tests and our measures of data uncertainty (reported below) end in 2013Q1: the *ex post* realisations arrive 17 quarters later.

To draw out further features of the MPC’s probabilistic backcasts, the top panel of Figure 7 presents, for $b = 1, \dots, 16$, their *ex ante* revision uncertainty (standard deviation) averaged over the $T = 23$ observations in the $t = 2007Q3, \dots, 2013Q1$ sample: $\bar{\sigma}^{t+b} = \frac{1}{T} \sum_t \hat{\sigma}_t^{t+b}$. This shows that average *ex ante* revision uncertainty tends to decrease as b increases - in other words as the forecast horizon decreases. The MPC became less uncertain about mature GDP as intermediate ONS estimates became available to them. Expected uncertainty, however, does still appear to be large - at 0.75 p.p. - even at $b = 16$.

4.2 Measuring Uncertainty: Comparing *ex ante* and *ex post* Loss

In the absence of direct knowledge of the MPC’s or indeed others users’ loss functions, that they use to interpret the fan chart, we provide empirical measures of uncertainty:

$$Unc_{t|t+b} = L(f_{t|t+b}, y_t^{t+17}) - E_{f_{t|t+b}} [L(f_{t|t+b}, Y_t^{t+17})], \quad (22)$$

for each of the four loss functions seen in Section 3.4.

The bottom panel of Figure 7 presents uncertainty estimates averaged across time, $\overline{Unc}_{t+b} = \frac{1}{T} \sum_t Unc_{t|t+b}$ ($T = 23$), for the MSE, *Int*, *logS* and CRPS loss functions for $b = 1, \dots, 16$. For the interval loss function we set $\alpha = 0.10$, as in Figure 6, and multiply values by -1 . This means that values $\overline{Unc}_{t+b} > 0$ (< 0) imply the MPC under (over) estimated uncertainty, as the

expected values for the loss function were lower (higher) than the realised values.

Looking at the bottom panel of Figure 7 we see that uncertainty, albeit with some ups and downs, declines steadily as the forecasting horizon $(17 - b)$ shortens for both the MSE and logS loss function. The more discrete, step declines observed after $b = 4$, and then again after $b = 12$, appear to time with ONS’s publication of the *Blue Book*. But these declines are flatter for the interval and the CRPS loss functions. For example, MSE data uncertainty at $b = 8$ is half that at $b = 1$. In contrast, for the interval loss function, \overline{Unc}_{t+b} is highest at $b = 10$ and 11 instead of at $b = 1$. These differences might be explained by how extreme outturns (or surprises) are less heavily penalised by the interval and CRPS than MSE and logS loss functions. Nevertheless, all four uncertainty measures agree that, conditional on reasonably mature ONS data ($b = 15, 16$), the MPC *ex ante* tended to over-estimate loss. They anticipated a wider variety of GDP revision outcomes than actually happened. As a result, average data uncertainty is negative at the shortest forecast horizons.

4.3 Testing Calibration of the MPC’s Probabilistic Backcasts

Under $H_0^U : E(Unc_{t|t+b}) = 0$ the MPC’s predictive densities $f_{t|t+b}$ are well calibrated. Table 3 presents GW_T type test statistics of H_0^U for each of the four loss functions. The test statistics reported are robust t -statistics, except for the interval-based loss function, where p -values for the asymptotically equivalent unconditional coverage test of Christoffersen (1998) are reported instead at $(1 - \alpha) = 0.90, 0.75, 0.50$. Test statistics or p -values in bold indicate rejection of H_0^U at the 10% level.

Table 3 shows that according to both the MSE and CRPS loss functions the MPC’s predictive densities appear well-calibrated, except at $b = 16$. At $b = 16$ we again see evidence that, *ex ante*, the MPC over-estimated the likely mean square or CRPS ‘errors’. Under the logS loss function, there is evidence of mis-calibration not just at $b = 16$, but also $b = 8, 9, 10, 11$. The interval loss function results echo this, with evidence against calibration at $b = 16$ (for $(1 - \alpha) = 0.75, 0.50$) and for $b = 8, \dots, 12$ for $(1 - \alpha) = 0.90$. But overall, of 96 tests (undertaken independently) calibration is rejected, at 10%, only on 15 occasions.

As discussed in the Introduction, the Bank of England have stated that they consult, *inter alia*, timely data from qualitative business surveys to predict data revisions better. Given this we also undertake conditional calibration tests, of H_0^C , using the average value over the reference quarter of the PMI Services survey published by Markit; these data are used as conditioning

information, w_{t+b} , to test (7). We use the PMI Services survey, as services comprise around 80% of UK GDP. Table 4 presents analogous results to Table 3 for this conditional test for the MSE, CRPS and logS loss functions. The evidence for mis-calibration is in fact even weaker than in Table 3. H_0^C is rejected only at $b = 16$ and twice at $b = 6$.

Overall, our tests for absolute calibration of the Bank’s probabilistic backcasts suggest that, on average since their introduction in 1997, the Bank has produced well-calibrated probabilistic backcasts relative to the ONS’s mature GDP growth estimates. This is so except at the shortest horizons, when the Bank appears *ex ante* to over-estimate their likely losses consistently.

4.4 Relative Performance of the MPC’s backcasts

Having tested absolute calibration of the MPC’s density backcasts, for four loss functions, we now test their relative performance. This is important in lending credibility to our claim that the MPC’s forecasts are *optimal*, as defined in (5). We compare the MPC against both an unconditional predictive benchmark and, for the point forecasts only (as ONS do not publish probabilistic estimates of GDP growth), against the ONS’s own early estimates.

4.4.1 Benchmark Probabilistic Backcasts

To evaluate the relative performance of the MPC’s backcasts we use an “unconditional” benchmark. Clements (2018) has similarly argued for the use of unconditional benchmark densities. We chose to use an unconditional model rather than the UC-SV model considered above, in (3). This is because although SV models have been found to be useful to measure, in real-time, the volatility around ‘reasonably’ unbiased forecasts, as seen in Clark et al. (2019), in our application experimentation (see Table A3 in the online appendix) revealed that when estimated recursively the UC-SV model generated excessively volatile *ex ante* predictions for y_t^{t+l} . This is as at time $t - b$ the UC-SV model’s long horizon predictions ($l - b$ steps ahead) set the revision mean, at all horizons, equal to the last observed value. Comparing Table A3 with Table 5 (below), we see strong evidence that the UC-SV model’s forecasts are outperformed by the “unconditional” benchmark model that we now introduce.

The “unconditional” (nonparametric) benchmark model is motivated as follows. Because time-series models of stationary data converge to the unconditional mean and variance at long horizons, we recursively estimate (as if in real-time) the unconditional mean and standard deviation of historical revisions. The probabilistic backcasts for quarter t , using information up

to $t + b$, exploiting ONS revisions between the l^{th} data release and the b^{th} release up to $t - l + 1$, are then defined as $f_{t|t+b}^{unc} = N(\hat{y}_t^{t+b,un}, \hat{\sigma}_t^{2,t+b,un})$, where

$$\hat{y}_t^{t+b,un} = y_t^{t+b} + \hat{\mu}_t^{t+b,un} \text{ for } t = 2007Q3, \dots, 2013Q1 \quad (23)$$

$$\hat{\mu}_t^{t+b,un} = \frac{1}{t-l} \sum_{\tau=1983Q2}^{\tau=t-l+1} rev_{\tau}^{(l-b)} \quad (24)$$

$$\hat{\sigma}_t^{t+b,un} = \sqrt{\frac{1}{t-l} \sum_{\tau=1983Q2}^{\tau=t-l+1} \left(rev_{\tau}^{(l-b)} - \hat{\mu}_t^{t+b,un} \right)^2} \quad (25)$$

$$rev_{\tau}^{(l-b)} = y_{\tau}^{\tau+l} - y_{\tau}^{\tau+b}. \quad (26)$$

That is, at each prediction origin $t + b$, this benchmark model uses a sample of revisions between the ONS's mature estimate and the ONS's b -th estimate, as available in real-time for observations up to $t - l + 1$. This sample is then used to compute a data-based predictive density for the b -quarter backcast. Recall, as b increases we are in effect shortening the backcasting horizon. Note that our sample of revisions dates back to 1983Q2, and that the evaluation sample increases with b .

4.4.2 MPC vs. Unconditional Benchmark

Table 5 evaluates the relative accuracy of the MPC and benchmark model probabilistic backcasts using both point and density forecast evaluation statistics. We use two measures of point forecast accuracy. The first is the bias: the sample average of $(y_t^{t+l} - \hat{y}_t^{t+b})$. If the MPC or the “unconditional” model *improves* over ONS, we should expect their bias to be smaller than the revision mean. This revision mean is the average value of $(y_t^{t+l} - y_t^{t+b})$ and is denoted “ONS” in Table 5. We also look at whether the root mean squared error (RMSE) is lower than the standard deviation of the ONS revision $(y_t^{t+l} - y_t^{t+b})$, again denoted “ONS” in the RMSE column. We also compute the ratio of the unconditional or MPC RMSE to the ONS RMSE. If this ratio is less than 1, then the MPC or unconditional benchmark produces more accurate estimates of ONS “mature” growth than the ONS's own estimates.

The statistics in Table 5 suggest that, for all values of b , the MPC provide more accurate point estimates of mature ONS data than the ONS's own earlier estimates. The RMSE gains are between 3% and 7%. The average bias of the MPC is small for all b (between -0.07 and 0.09). This suggests that the MPC are good at anticipating ONS revisions. The bias of the “unconditional” model is larger at -0.2 when $b = 1$.

Comparing the probabilistic performance of the MPC and benchmark forecasts using the interval loss function we see from Table 5 that the MPC provide better nominal coverage in the majority of cases. But at $b = 16$ the MPC is beaten by the unconditional model. At $b = 16$, the MPC’s confidence intervals are too wide such that, for example, 91% of outturns fall inside their 50% prediction interval.

Looking next in Table 5 at the logarithmic score results, we see the relative performance of the “unconditional” model also worsens with b . But in terms of the CRPS, the benchmark model actually improves over the MPC at $b = 12, 16$. This is as for these older data the point backcasts perform similarly. In this case, the CRPS prefers predictions with smaller predicted variance which is the case for the unconditional model that embeds the assumption that mature data are released by the ONS at $t + l$.

Overall, we find that the MPC provides more accurate point estimates of revised GDP values than the comparably timed estimates published by the ONS themselves. The MPC’s probabilistic estimates are also more accurate, except at $b = 16$, than those from a benchmark model that, in effect, bias-corrects the latest ONS estimates.

5 Historical Characteristics of Data Uncertainty

Having found that - except at the shortest horizon ($b = 16$) - the Bank’s probabilistic backcasts appear, on average, both well calibrated (importantly across a range of loss functions), such that $E(Unc_{t|t+b}) = 0$, and that they are accurate relative to a benchmark, we consider the evolution of $Unc_{t|t+b}$ over time.

We focus here on measuring $Unc_{t|t+b}$ using the CRPS loss function. Equation (4), of course, provides a measure of uncertainty for a generic loss function, $L(f_{t|t-h}, y_t)$. But in the absence of knowledge of a given user’s specific loss function, we choose here to follow Rossi et al. (2016) and focus exclusively on the CRPS. An attraction of the CRPS is arguably its intuitive appeal: the CRPS cumulates the forecast errors for each possible predicted event in the forecast density:

$$L^{CRPS}(f_{t|t-h}, y_t) = \int_{-\infty}^{\infty} (u_{t|t-h}(r))^2 dr \quad (27)$$

where $u_{t|t-h}(r) = (F_{t|t-h}(r) - I(y_t \leq r))^2$ is the mean squared forecast error or Briers score, i.e. the squared difference between the probability forecast of the binary event ($y_t \leq r$) at the threshold $r \in \mathbb{R}$, the cumulative predictive density $F_{t|t-h}(r)$, and the binary outcome, $I(y_t \leq r)$.

In the spirit of Rossi et al. (2016), we have argued that we can interpret $Unc_{t|t+b}$ as (realised) mistakes the MPC made when measuring risk (*ex ante* uncertainty). $Unc_{t|t+b}$ thereby provides, for the MPC, a measure of unforecastable (or Knightian) uncertainty.

Figure 8 plots $Unc_{t|t+b}^{CRPS}$ for $b = 1, \dots, 7$ in the top panel (Figure 8A) and for $b = 8, \dots, 16$ in the bottom panel (Figure 8B). We indicate the line of correct calibration, $Unc_{t|t+b}^{CRPS} = 0$, via a thick black line. Arguably, data uncertainty at $b = 1, \dots, 7$ has more impact on policy making than uncertainty about older observations.

Looking at Figure 8A first, we observe a large positive spike in $Unc_{t|t+b}^{CRPS}$ at the onset of the 2008/2009 recession. Recall, in section 2 we also found that the data revision mean switched from positive to negative, and that revision mean volatility increased, with the onset of this recession. These instabilities in the revisions' process support the view that this spike in data uncertainty is explained by unexpected or unknown unknowns (Knightian uncertainty) surprising the MPC. That is, from 2008 to 2009 in real-time, *ex ante*, the MPC were unable to assess correctly probabilities for all revised GDP growth states of nature. So while the onset of the global financial crisis was no doubt making the future economic outlook more uncertain, Figure 8A shows how the MPC were also battling elevated uncertainty about where the economy was currently and indeed where it had been in the recent past. But from late 2009 $Unc_{t|t+b}^{CRPS}$ falls and thereafter fluctuates around zero. This is consistent with how, on average, the calibration tests in Table 3 find these deviations from 0 to be statistically insignificant.

Turning to the older historical data ($b = 8, \dots, 16$) in Figure 8B, we see data uncertainty, $Unc_{t|t+b}^{CRPS}$, as we might hope for older data, is smaller. $Unc_{t|t+b}^{CRPS}$ is mainly positive from 2007 to 2010 and negative from 2010. So since the recession the MPC have, in general, over-estimated the degree of data uncertainty for these older ($b > 8$) data. This may reflect difficulties in predicting benchmark revisions (recall in section 2 we found that the size of these later revisions changed) and a reluctance of the MPC to downsize *ex ante* data uncertainty as the recession ended.

5.1 Decomposing Data Uncertainty

Figure 9 decomposes $Unc_{t|t+b}^{CRPS}$ into $reliability_t$ and $reliability_t^e$ components using (10). These components are plotted for $b = 1, 4, 8$. In practice, computation requires pooled estimation over a sample of data. We use four quarter rolling windows; the dates in Figure 9 refer to the last observation in these rolling windows.

Figure 9 reveals the rapid increase in unforecastable data uncertainty, $Unc_{t|t+b}^{CRPS}$, at the onset of the 2008/2009 recession to have been associated with both an increase in *ex post* reliability (i.e. an increase in *unreliability*, as the MPC's *ex ante* predictions proved less accurate) and falling levels of expected reliability. The fact that $reliability_t^e$ became increasingly negative in the initial stages of the recession contributed to, rather than detracted from, overall uncertainty as measured by $Unc_{t|t+b}^{CRPS}$. But from 2009, as also reflected in Figure 2, the MPC did begin to anticipate, *ex ante*, more data uncertainty. Accordingly we see $reliability_t^e$ begin to rise, and then stabilise above zero at around 0.5, for all three values of b ($b = 1, 4, 8$). This suggests, as in Example 1, that the MPC were wisely compensating for expected forecasting errors by raising the *ex ante* variance of their density backcasts; this action helped deliver better calibrated densities, as seen by $Unc_{t|t+b}^{CRPS}$ falling from its peaks in 2008 to values closer to zero. The relatively small, but negative, values of $Unc_{t|t+b}^{CRPS}$ observed for all b in the immediate aftermath of the recession reflect the fact that the MPC began to over-estimate data uncertainty ($reliability_t^e > reliability_t$). Nevertheless, since 2010 $reliability_t$ and $reliability_t^e$ do track each other fairly closely. But as they do so at positive values there is scope for the MPC to improve their forecasts further.

5.2 Data Uncertainty and Macroeconomic Uncertainty

It is helpful to contrast our estimates of (historical) data uncertainty with established existing estimates of uncertainty, notwithstanding that methodologically these macroeconomic measures, for a specific loss function, tend to focus either on *ex ante* uncertainty, as in Jurado et al. (2015), or *ex post* uncertainty, as in Rossi et al. (2016), rather than the difference between the two - as captured by $Unc_{t|t-h}$.

We compare our data uncertainty estimates with two alternative estimates of macroeconomic uncertainty. The first, seen in Redl (2017), estimates uncertainty in the UK using the approach of Jurado et al. (2015). More specifically, Redl (2017) uses 33 monthly series to compute an aggregate *ex ante* uncertainty measure under MSE loss (cf. (13)). The second measure is from Bank of England (2016). It is computed by extracting a principal component from a set of uncertainty measures. These measures include the standard deviation of GDP growth consensus forecasts, the number of newspaper reports citing *uncertainty*, a survey measure of household confidence, and the implied volatilities of options' contracts on the FTSE 100 and the sterling effective exchange rate. Both the Redl (2017) and Bank of England estimates of uncertainty

are available monthly, but for comparative purposes are presented in Figure 10 averaged over the quarter.¹⁹

Figure 10 indicates that these two measures of macroeconomic uncertainty are positively correlated with our estimates of data uncertainty, $Unc_{t|t+b}^{CRPS}$, at both $b = 1, 4$. To investigate this apparent correlation further, we look at their dynamic correlations in Table 6. Specifically we look at the correlations between macroeconomic uncertainty in quarter t and data uncertainty at lags $(t - 1, t - 2)$ and leads $(t + 1, t + 2)$. We report these correlation coefficients against the Redl (2017) macroeconomic uncertainty measure (in Table 6A) and the Bank of England (2016) uncertainty measure (in Table 6B) for $b = 1, 4, 8, 12$. Coefficients in bold indicate that the correlation is statistically different from zero (using asymptotic bounds).

As anticipated, data uncertainty for more mature ($b = 8, 12$) GDP growth values is not so strongly correlated with macroeconomic uncertainty. But data uncertainty for earlier GDP estimates ($b = 1, 4$) has a high correlation with macroeconomic uncertainty. The strongest correlation, of between 76% and 82%, is at lag 1. On the face of it this suggests that data uncertainty leads macroeconomic uncertainty by one quarter. Note, however, that our data uncertainty measure is organised by reference quarter date, and the “preliminary” ONS GDP estimate and the initial backcasts from the MPC are available with a one quarter delay. So this implies that data uncertainty and macroeconomic uncertainty are coincident indicators.

6 Conclusion

This paper provides and emphasises the importance of quantitative estimates of unforecastable data uncertainty for UK GDP growth. We find that data uncertainty does, as suggested by Bean (2007), contribute to a more uncertain economic environment for policymakers. We show that data uncertainty rises at the onset of recessions. It is also positively correlated with popular measures of macroeconomic uncertainty.

Methodologically we define unforecastable *uncertainty* as a lack of knowledge about the future or the past; we quantify this lack of knowledge by the difference between the *ex post* (or realised) and the *ex ante* (or expected) values of the loss function (or scoring rule) relevant to the user of the underlying density forecast. We show how our uncertainty measures can be used as the basis for the construction of tests for calibration of probabilistic forecasts. In so-doing, we

¹⁹We are grateful to Chris Redl for his help and providing these measures of UK macroeconomic uncertainty.

link scoring rules commonly used to evaluate relative density forecast performance to absolute density forecast evaluation tests.

We commend the MPC at the Bank of England for quantifying *ex ante* GDP data uncertainty in their quarterly fan charts; and, echoing Manski (2015; 2018), 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.²⁰ We do believe, however, that the MPC 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, which precludes *ex post* evaluation without making some assumption.

We hope that our approach to measuring data uncertainty, that relies on *ex ante* probabilistic expectations for historical GDP data, can be applied in other countries. While the MPC at the Bank of England is, we believe, unique in providing direct judgement-based *ex ante* estimates of data uncertainty, other countries could and, we believe, should follow. Our preliminary analysis of US GDP growth data revisions (in the online appendix) suggests that data uncertainty also matters in the US; and that the challenge, for forecasters of data revisions, will be accommodating temporal changes to the mean and variance of data revisions especially at business cycle turning points. But direct, quantitative estimates of data uncertainty will help policy and other decision makers better understand the macroeconomic uncertainties they face when making decisions.

²⁰Note 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.

References

- Aruoba, S. (2008). Data revisions are not well-behaved, *Journal of Money, Credit and Banking* **40**: 319–340.
- Bachman, R., Elstener, S. and Sims, E. R. (2013). Uncertainty and economic activity: evidence from business survey data, *American Economic Journal: Macroeconomics* **5**: 217–49.
- Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring economic policy uncertainty, *Quarterly Journal of Economics* **131**: 1593–1636.
- Bank of England (2007). *Inflation Report, November 2007*.
- Bank of England (2016). *Inflation Report, August 2016*.
- Bean, C. (2007). Risk, uncertainty and monetary policy, *Bank of England Quarterly Bulletin* **47**(4): 600–606.
- Bentzien, S. and Friederichs, P. (2014). Decomposition and graphical portrayal of the quantile score, *Quarterly Journal of the Royal Meteorological Society* **140**(683): 1924–1934.
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics* **19**(4): 465–474.
- Bloom, N. (2014). Fluctuations in uncertainty, *Journal of Economic Perspectives* **28**: 153–76.
- 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. (2019). Modeling time-varying uncertainty of multiple-horizon forecast errors, *Review of Economics and Statistics* . Forthcoming.
- Clements, M. P. (2004). Evaluating the Bank of England density forecasts of inflation, *Economic Journal* **114**: 855–877.

- Clements, M. P. (2014). Forecasting uncertainty - ex ante and ex post: U.S. inflation and output growth, *Journal of Business and Economic Statistics* **32**: 206–16.
- Clements, M. P. (2018). Are macroeconomic density forecasts informative?, *International Journal of Forecasting* **34**: 181–198.
- Clements, M. P. and Galvão, A. B. (2010). First announcements and real economic activity, *European Economic Review* **54**: 803–817.
- 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.
- 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 .
- Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach, *Journal of the Royal Statistical Society, Series A* **147**(2): 278–292.
- Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**(4): 863–83.
- Elliott, G. and Timmermann, A. (2016). Forecasting in economics and finance, *Annual Review of Economics* **8**(1): 81–110.
- 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.

- 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.
- Galvao, A. B. (2017). DSGE models and data revisions, *Journal of Econometrics* **196**: 215–232.
- Garratt, A. and Vahey, S. P. (2006). UK real-time macro data characteristics, *Economic Journal* **116**: F119–F135.
- Giacomini, R. and White, H. (2006). Tests of conditional predictive ability, *Econometrica* **74**: 1545 – 1578.
- Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society Series B* **69**(2): 243–268.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**: 359–378.
- 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.
- Ilut, C. L. and Schneider, M. (2014). Ambiguous business cycles, *American Economic Review* **104**(8): 2368–99.
- 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.
- Jurado, K., Ludvigson, S. C. and Ng, S. (2015). Measuring uncertainty, *American Economic Review* **105**: 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* **30**: 181–190.
- Ludvigson, S., Ma, S. and Ng, S. (2018). Uncertainty and business cycles: exogenous impulse or endogenous responses?, *New York University, mimeo* .

- 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.
- Manski, C. F. (2018). Communicating uncertainty in policy analysis, *Proceedings of the National Academy of Sciences* .
- 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. (2018). R2 bounds for predictive models: What univariate properties tell us about multivariate predictability, *Journal of Business and Economic Statistics* **Forthcoming**.
- Murphy, A. H. and Winkler, R. L. (1987). A general framework for forecast verification, *Monthly Weather Review* **115**(7): 1330–1338.
- Office for National Statistics (2016). International comparison of GDP revisions, *ONS Article Series* .
- 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.
- Redl, C. (2017). The impact of uncertainty shocks in the United Kingdom, *Bank of England Staff Working Paper no. 695* .
- Rossi, B. and Sekhposyan, T. (2015). Macroeconomic uncertainty indices based on nowcast and forecast error distributions, *American Economic Review: Papers and Proceedings* **105**(5): 650–55.

- Rossi, B., Sekhposyan, T. and Soupre, M. (2016). Understanding the sources of macroeconomic uncertainty, *CEPR Discussion Papers* **n. 11415**.
- Sekkel, R. and Soojin, J. (2018). Macroeconomic uncertainty through the lens of professional forecasters, *Journal of Business and Economic Statistics* **Forthcoming**.
- Stock, J. H. and Watson, M. W. (2007). Why has U.S. Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking* **Supplement to Vol. 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.
- Weijs, S. V., van Nooijen, R. and van de Giesen, N. (2010). Kullback-leibler divergence as a forecast skill score with classic reliability, *Monthly Weather Review* **138**(9): 3387–3399.
- Yates, J. (1982). External correspondence: Decompositions of the mean probability score, *Organizational Behavior and Human Performance* **30**: 132–156.

Table 1: Summary Statistics for GDP Growth Data Revisions (source: ONS, real-time database)

Obs Period:	$y^{17} - y^1$			$y^{17} - y^{13}$			$y^{latest} - y^{17}$		
	mean	std	MAE	mean	std	MAE	mean	std	MAE
1983Q2-2013Q1	0.462	0.906	0.759	0.065	0.451	0.294	0.231	0.709	0.609
1993Q2-2013Q1	0.358	0.807	0.674	0.112	0.483	0.307	0.066	0.646	0.526
2007Q3-2013Q1	0.306	1.328	1.131	0.229	0.836	0.696	0.001	0.600	0.420

Notes: Values in bold indicate that the data revision mean is statistically different from zero at a 10% significance level using robust (Newey-West) standard errors. MAE is the mean absolute error, or mean absolute revision. The MPC started releasing their backcasts in 2007Q3. Because we have to wait 4 years to observe our measure of “mature” data, the last observation for which we observe a 17th release is 2013Q1. We use the July 2017 data vintage to measure y^{latest} .

Table 2: t-statistics for tests for *News* and *Noise*

Obs Period:	$y^{17} - y^1$		$y^{17} - y^{13}$		$y^{latest} - y^{17}$	
	H0: news	H0: noise	H0: news	H0: noise	H0: news	H0: noise
1983Q2-2013Q1	0.59	3.52	0.45	1.67	-1.84	0.55
1993Q2-2013Q1	0.63	2.41	0.48	1.30	-2.81	-0.22
2007Q3-2013Q1	1.66	2.90	1.09	1.78	-3.05	2.58

Notes: t-statistics employ Newey-West standard errors. Values in bold indicate that the null hypothesis is rejected at the 10% significance level.

Table 3: GW t-statistics testing H_0^U for the MSE, CRPS and Logarithmic score loss functions and p-values for the Interval loss function at $(1-\alpha)=0.9, 0.75$ and 0.50

	MSE	CRPS	LogS	Int: 90%	Int: 75%	Int: 50%
b=1	1.34	1.38	1.61	0.28	0.30	0.53
b=2	1.22	1.36	1.47	0.28	0.14	0.53
b=3	1.29	1.45	1.53	0.10	0.14	0.14
b=4	1.19	1.31	1.40	0.64	0.14	0.30
b=5	0.92	0.98	1.20	0.64	0.30	0.83
b=6	0.75	0.82	1.11	0.64	0.71	0.83
b=7	0.86	1.12	1.36	0.28	0.90	0.30
b=8	1.19	1.57	1.90	0.03	0.30	0.14
b=9	1.20	1.57	1.89	0.03	0.30	0.14
b=10	1.23	1.60	2.00	0.01	0.30	0.30
b=11	1.08	1.38	1.80	0.01	0.56	0.30
b=12	0.74	1.04	1.43	0.03	0.90	0.30
b=13	0.03	0.41	0.95	0.10	0.71	0.83
b=14	-0.19	0.09	0.79	0.10	0.71	0.30
b=15	-1.20	-1.02	0.07	0.64	0.38	0.02
b=16	-5.49	-4.82	-2.90	0.31	0.04	0.00

Notes: Computed using the 23 observations in the 2007Q3-2013Q1 period. Values in bold imply rejection of H_0^U at the 10% level. The t-statistics are computed using robust standard errors. The interval coverage test used is the Christoffersen (1998) unconditional LR test.

Table 4: GW conditional Wald-statistics for H_0^C using the PMI data to measure w_{t+b}

	MSE	CRPS	LogS
b=1	3.58	4.05	3.25
b=2	2.95	3.21	2.65
b=3	2.60	2.88	2.54
b=4	3.43	3.31	2.58
b=5	4.16	4.46	3.59
b=6	4.85	4.80	3.07
b=7	4.54	3.65	3.06
b=8	2.03	2.86	3.82
b=9	1.85	2.72	3.72
b=10	1.90	2.80	4.16
b=11	1.43	2.06	3.39
b=12	0.57	1.15	2.38
b=13	2.37	1.88	2.84
b=14	2.65	2.52	3.29
b=15	3.24	3.34	2.29
b=16	43.41	36.45	29.10

Notes: Computed using the 23 observations in the 2007Q3-2013Q1 period. Values in bold imply rejection of H_0^C at the 10% level. Chi squared statistics computed using robust standard errors, $q=2$. Thanks to Raffaella Giacomini for code to implement the GW test.

Table 5: Performance of the MPC's backcasts and the unconditional benchmark model: point and density forecast evaluation statistics at $b=1, 4, 8, 12, 16$. Evaluation statistics reported using ONS data after 4 years ($l = 17$) as the outturn (prediction horizon $= l-b$)

		Bias	RMSE	RMSE Ratio to ONS	LogS	CRPS	Int: 90%	Int: 75%	Int: 50%
b=1	ONS	0.306	1.327	-	-	-	-	-	-
	MPC	-0.046	1.243	0.937	2.070	0.702	83%	65%	43%
	Uncond.	-0.223	1.335	1.006	2.118	0.735	74%	61%	43%
b=4	ONS	0.266	1.184	-	-	-	-	-	-
	MPC	-0.024	1.138	0.961	1.898	0.655	87%	61%	39%
	Uncond.	-0.034	1.173	0.990	2.631	0.687	62%	38%	23%
b=8	ONS	0.145	0.984	-	-	-	-	-	-
	MPC	-0.070	0.953	0.968	1.513	0.559	74%	65%	35%
	Uncond.	-0.067	0.971	0.987	2.055	0.595	67%	37%	20%
b=12	ONS	0.186	0.810	-	-	-	-	-	-
	MPC	0.092	0.770	0.951	1.215	0.520	74%	74%	39%
	Uncond.	0.139	0.806	0.995	2.533	0.483	62%	44%	29%
b=16	ONS	0.011	0.286	-	-	-	-	-	-
	MPC	-0.025	0.283	0.989	0.571	0.196	96%	91%	91%
	Uncond.	-0.005	0.283	0.989	1.166	0.110	89%	89%	84%

Notes: out-of-sample evaluation samples: b=1 backcasts: 2007Q3-2013Q1; b=4 backcasts 2006Q4-2013Q1; b=8 backcasts 2005Q4-2013Q1; b=12 backcasts 2004Q4-2013Q1; b=16 backcasts 2003Q4-2013Q1. Unconditional forecasts using data from 1983Q2. Entries for intervals are the empirical coverages at each nominal coverage rate (90%, 75% and 50%).

Table 6: Dynamic correlations of $\text{Unc}_{t|t+b}^{\text{CRPS}}$ with the macroeconomic uncertainty estimates of RedI (2017) and the Bank of England (*Inflation Report*, Aug 2016)

Table 6A: Correlations with Macroeconomic Uncertainty as computed by RedI (2017)

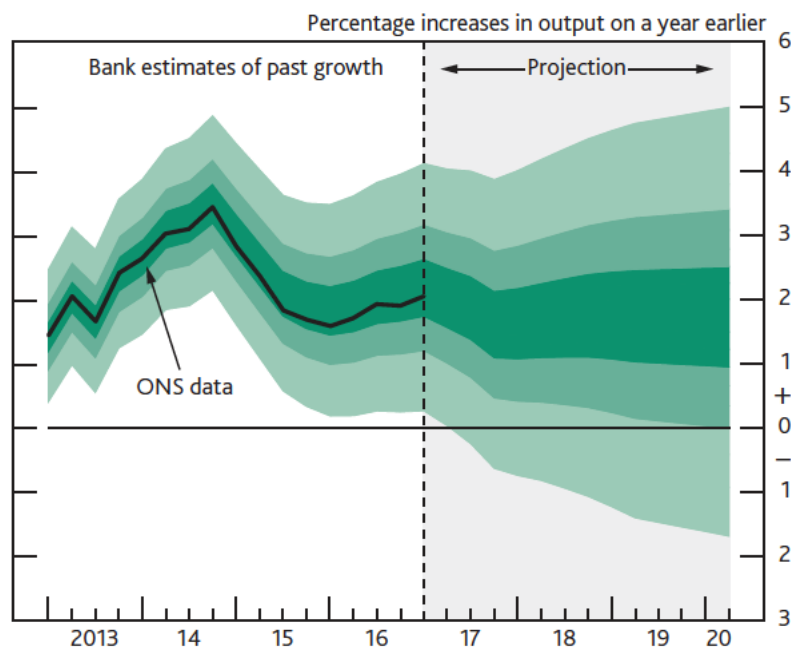
	t-2	t-1	t	t+1	t+2
b=1	0.633	0.815	0.691	0.375	0.085
b=4	0.664	0.761	0.530	0.272	0.048
b=8	0.594	0.728	0.453	0.225	0.162
b=12	0.543	0.400	0.050	0.014	0.270

Table 6B: Correlations with Macroeconomic Uncertainty as published by the Bank of England in the *Inflation Report*, Aug 2016.

	t-2	t-1	t	t+1	t+2
b=1	0.565	0.769	0.596	0.176	-0.163
b=4	0.600	0.756	0.426	0.039	-0.202
b=8	0.471	0.508	0.228	-0.053	-0.088
b=12	0.238	0.183	-0.160	-0.204	0.073

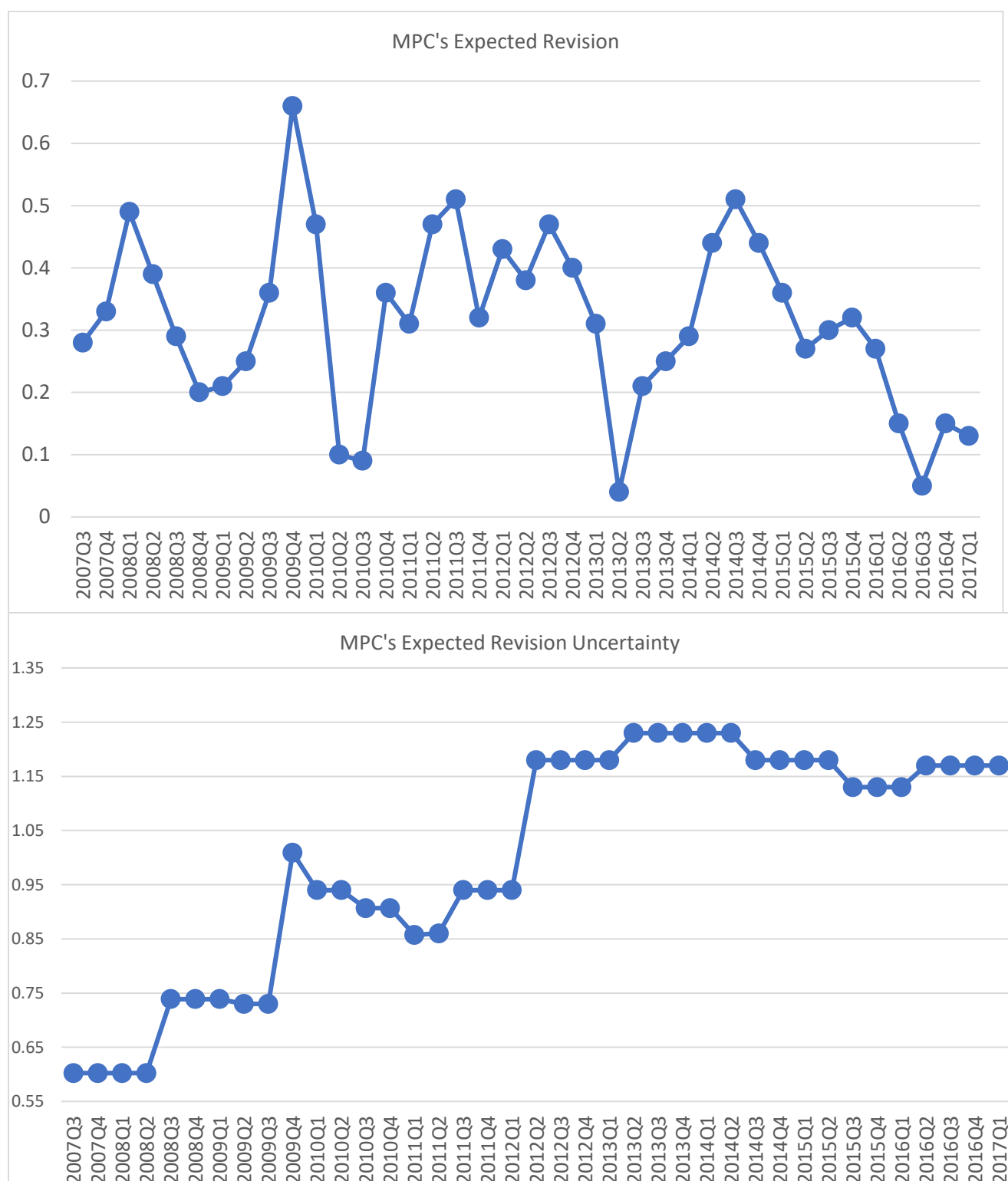
Notes: values in bold indicate statistically significant correlations using the asymptotic 2-standard deviation interval ($2/\sqrt{T}$).

Figure 1: The MPC's Fan Chart for GDP Growth (*Inflation Report*, May 2017)



Notes: Bank of England's fan chart for GDP growth (from the May 2017 "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."

Figure 2: The expected revision and revision uncertainty for GDP growth one quarter in the past (b=1) – using the MPC's backcasts



Notes: The expected revision, using notation formally defined below, is $\hat{y}_t^{t+1} - y_t^{t+1}$; y_t^{t+1} is the ONS's 'preliminary' or first estimate of year-on-year growth for quarter t made in quarter $(t+1)$; and \hat{y}_t^{t+1} is the MPC's prediction of 'mature' GDP growth for quarter t made later in quarter $(t+1)$. Expected uncertainty refers to the MPC's *ex ante* prediction made in quarter $(t+1)$ of the uncertainty (standard deviation) of $(y_t^{t+l} - \hat{y}_t^{t+1})$, where y_t^{t+l} denotes the future ($l > 1$) "mature" value of GDP growth.

Figure 3: The ONS's "Preliminary" (First Release) and "Mature" GDP Growth Estimates

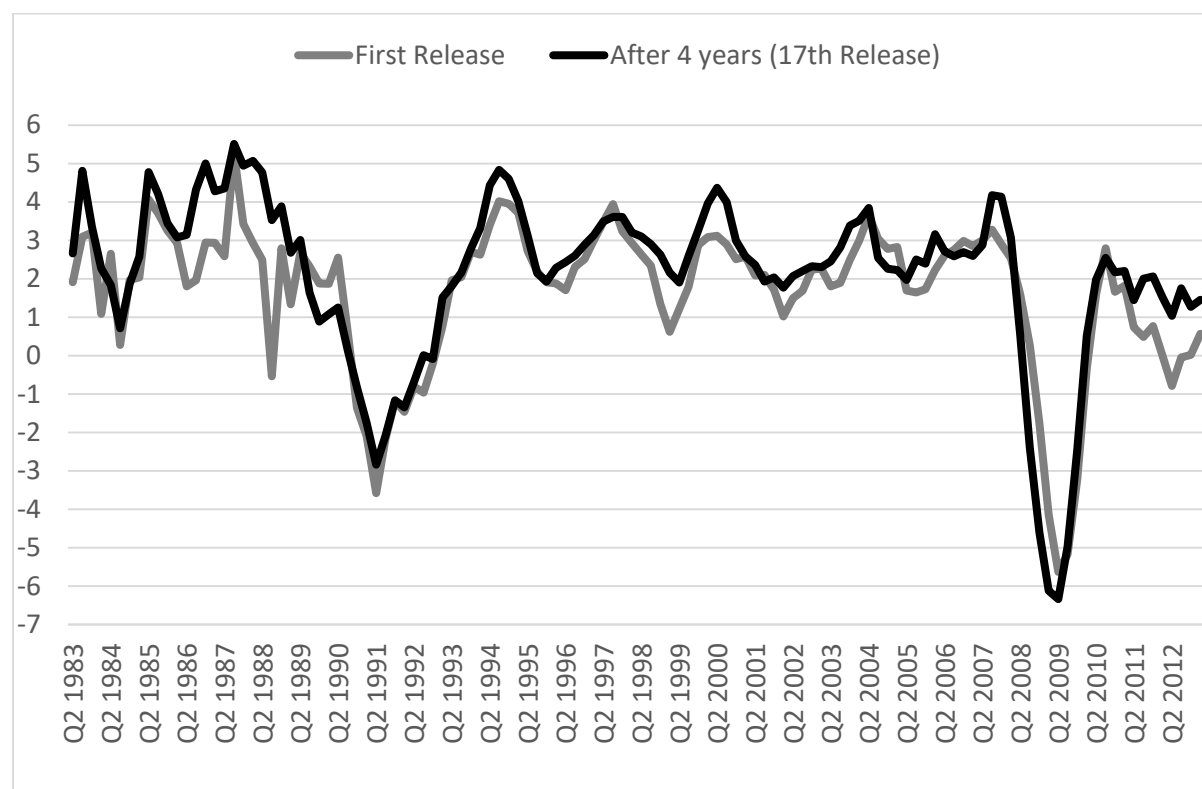


Figure 4: UC-SV Results: Estimated Components for GDP Growth Data Revisions, $rev^{16} = (y^{17} - y^1)$

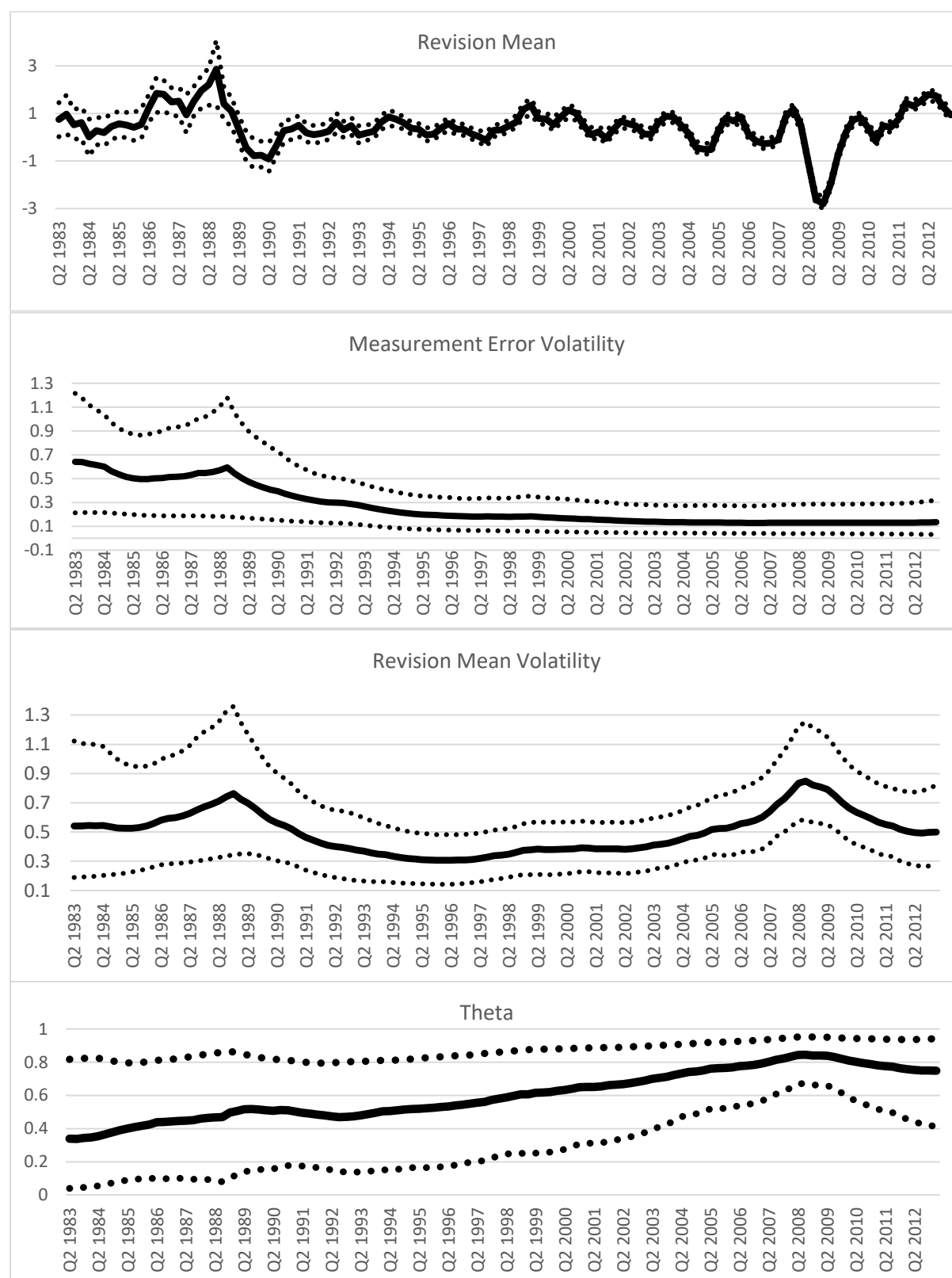


Figure 5: UC-SV results for data of different maturities: estimated components for GDP growth data revisions for $b=1$ ($rev^{16} = (y^{17} - y^1)$), $b=4$ ($rev^{13} = (y^{17} - y^4)$), $b=8$ ($rev^9 = (y^{17} - y^8)$) and $b=12$ ($rev^5 = (y^{17} - y^{12})$)

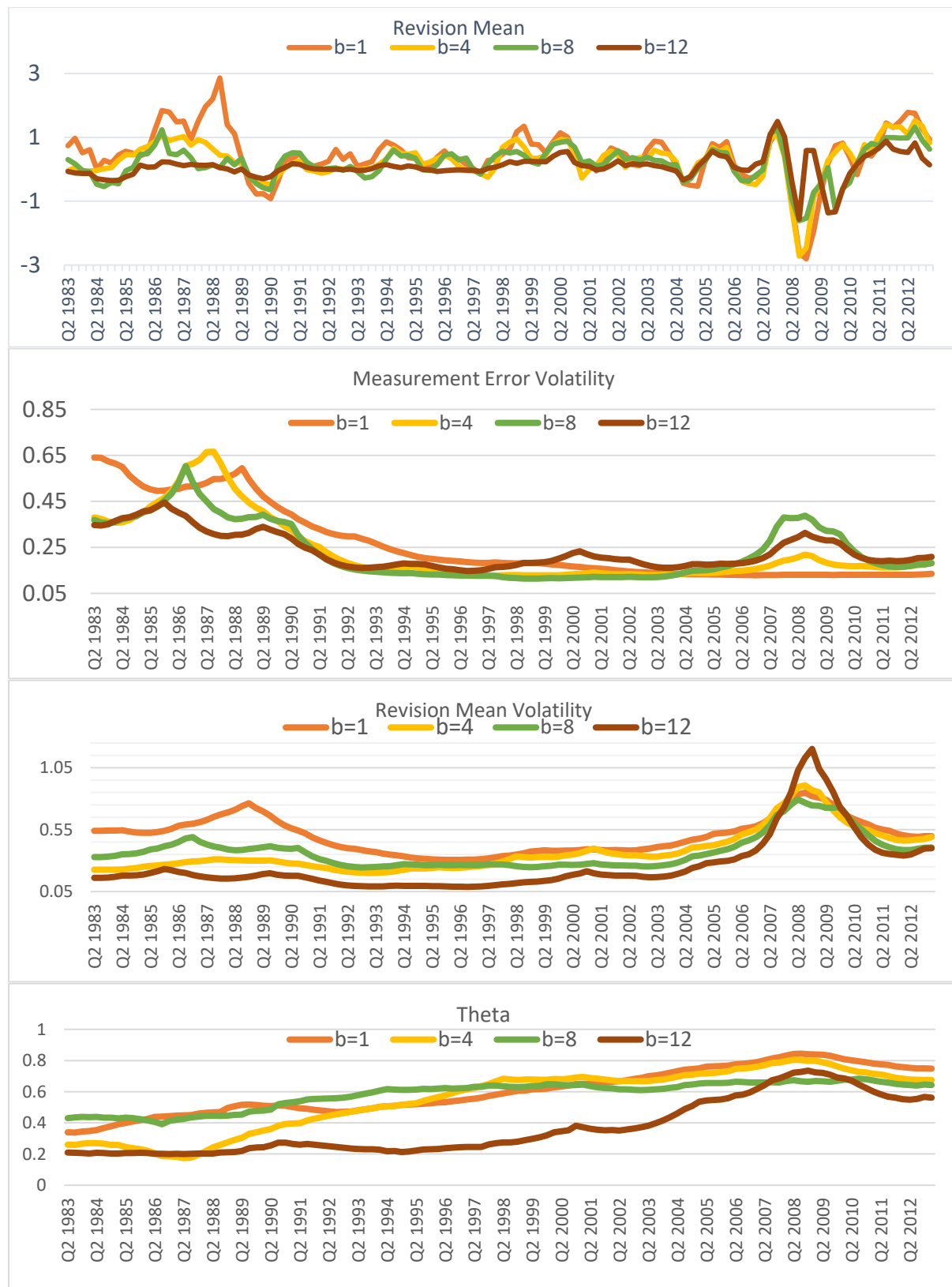


Figure 6: MPC backcasts for GDP growth one quarter in the past ($b=1$): 90% prediction intervals alongside subsequent ONS 'mature' outturns

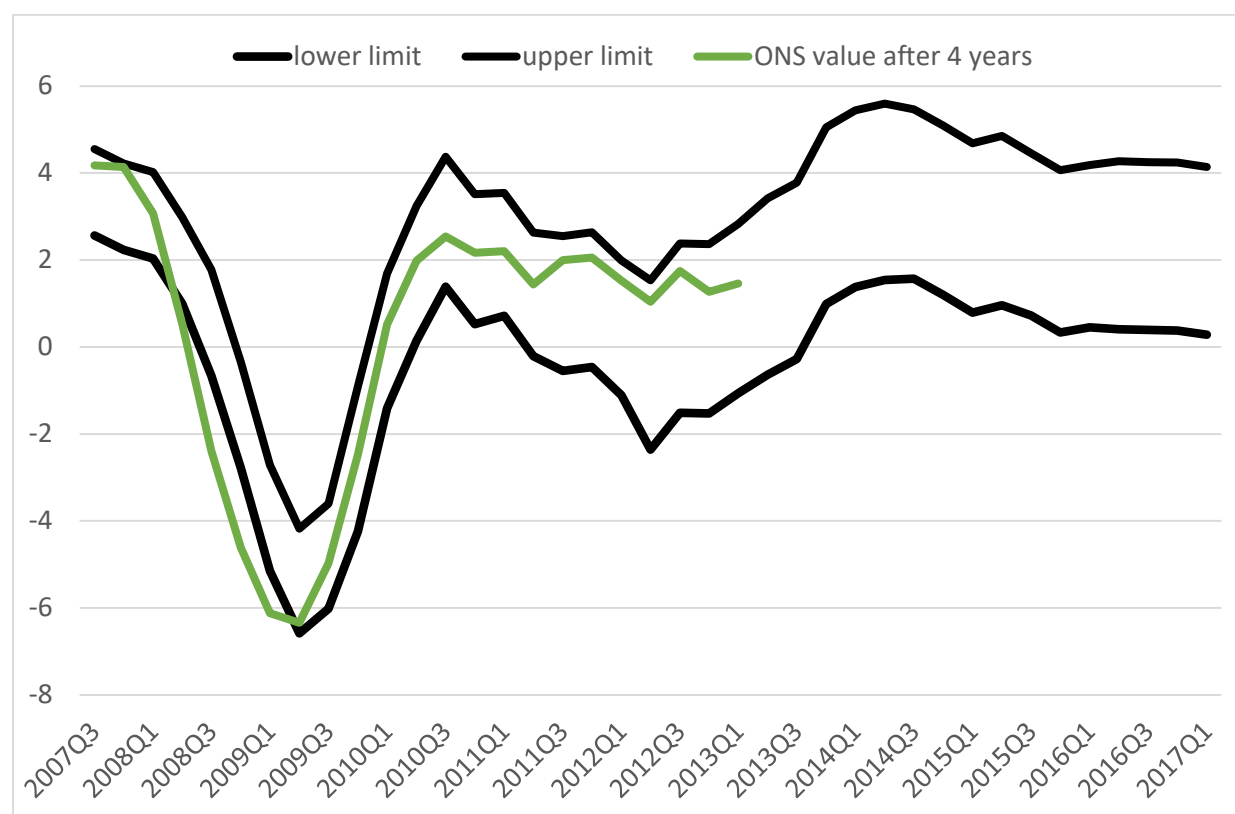
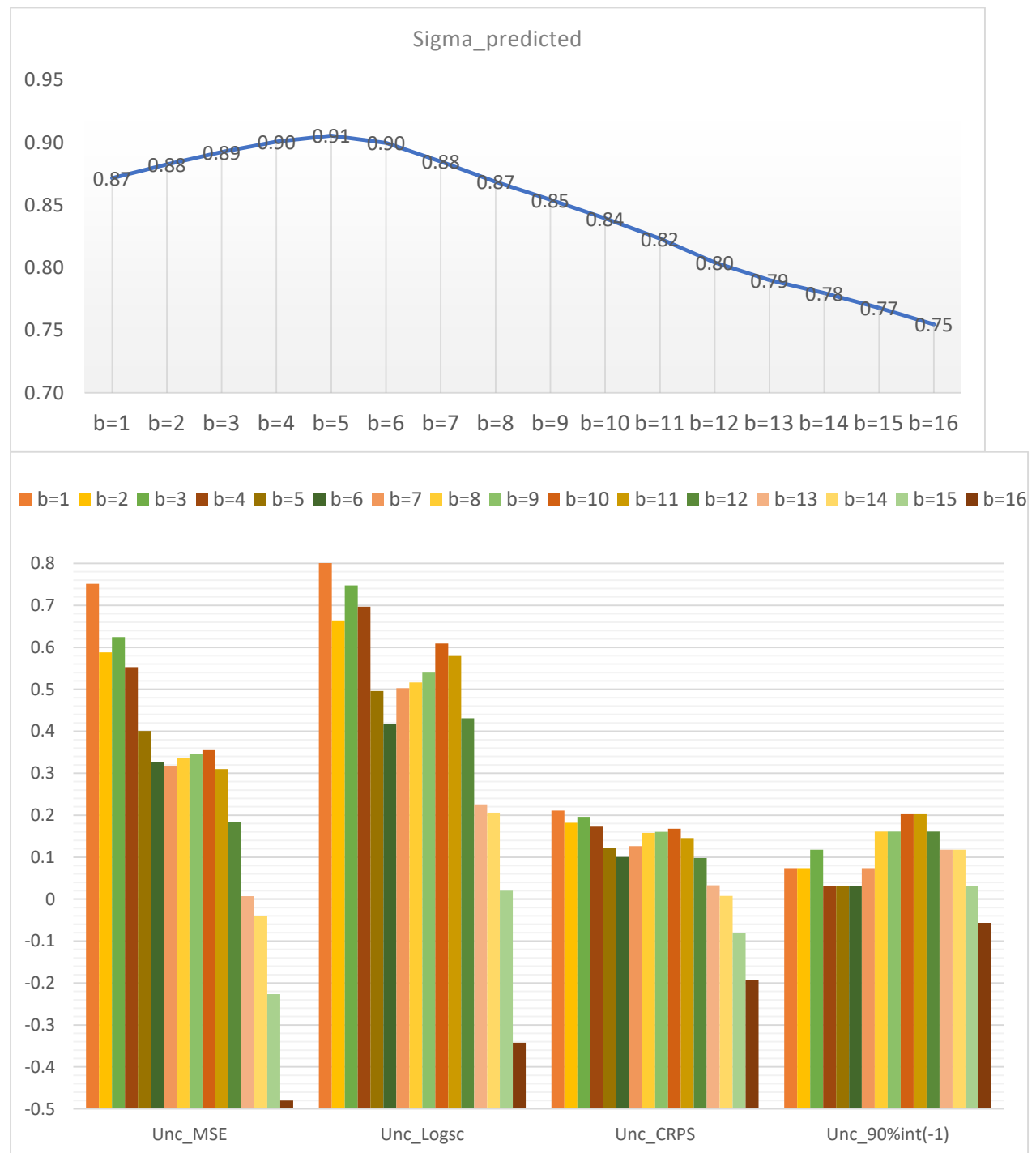


Figure 7: MPC uncertainty: averaged (over the sample) *ex ante* uncertainty, $\hat{\sigma}_t^{t+b}$, and averaged unforecastable uncertainty, $Unc_{t|t+b}^{MSE}$, $Unc_{t|t+b}^{LogS}$, $Unc_{t|t+b}^{CRPS}$, $Unc_{t|t+b}^{90\%Int}$, at $b=1, \dots, 16$ (sample includes forecasts made in 2007Q3 through to 2013Q1)



Notes: $Unc_{t|t+b}^{90\%Int}$ is multiplied by -1 so that positive (negative) deviations indicate under (over)-coverage. Predictions horizons decline with b as the forecast horizon is $(17-b)$.

Figure 8: CRPS-based Data Uncertainty estimates, $Unc_{t|t+b}^{CRPS}$

Figure 8A: $b=1, \dots, 7$

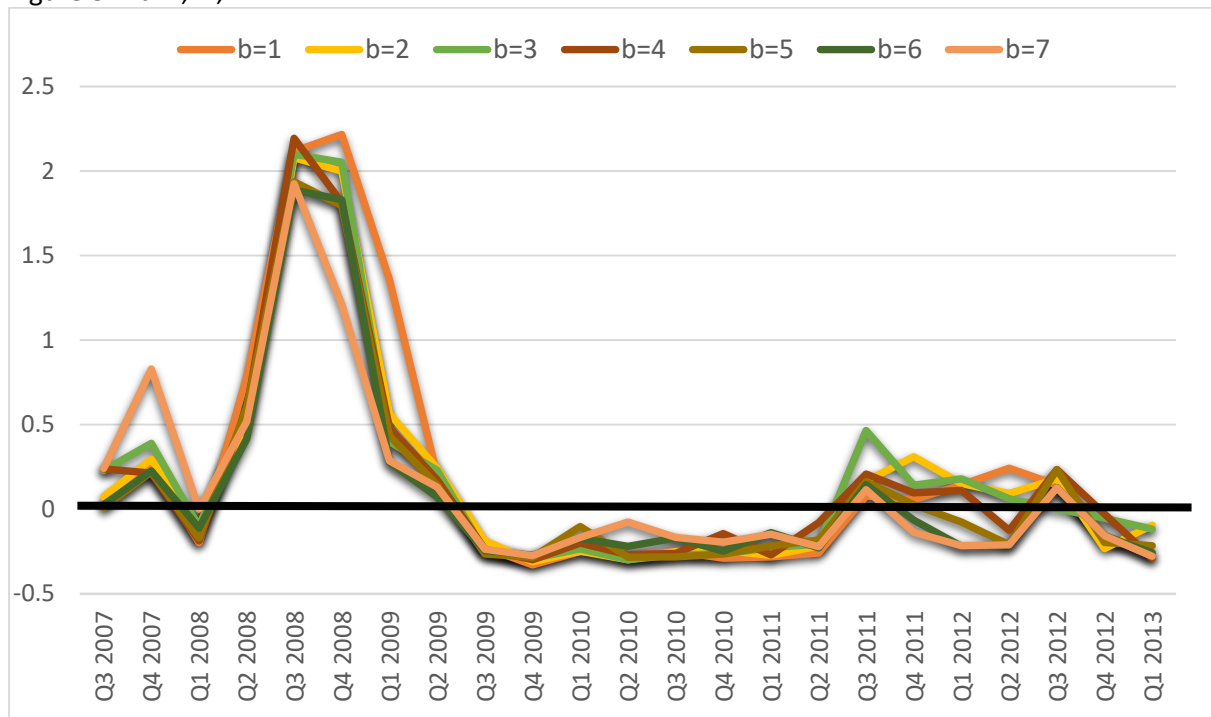


Figure 8B: $b=8, \dots, 16$

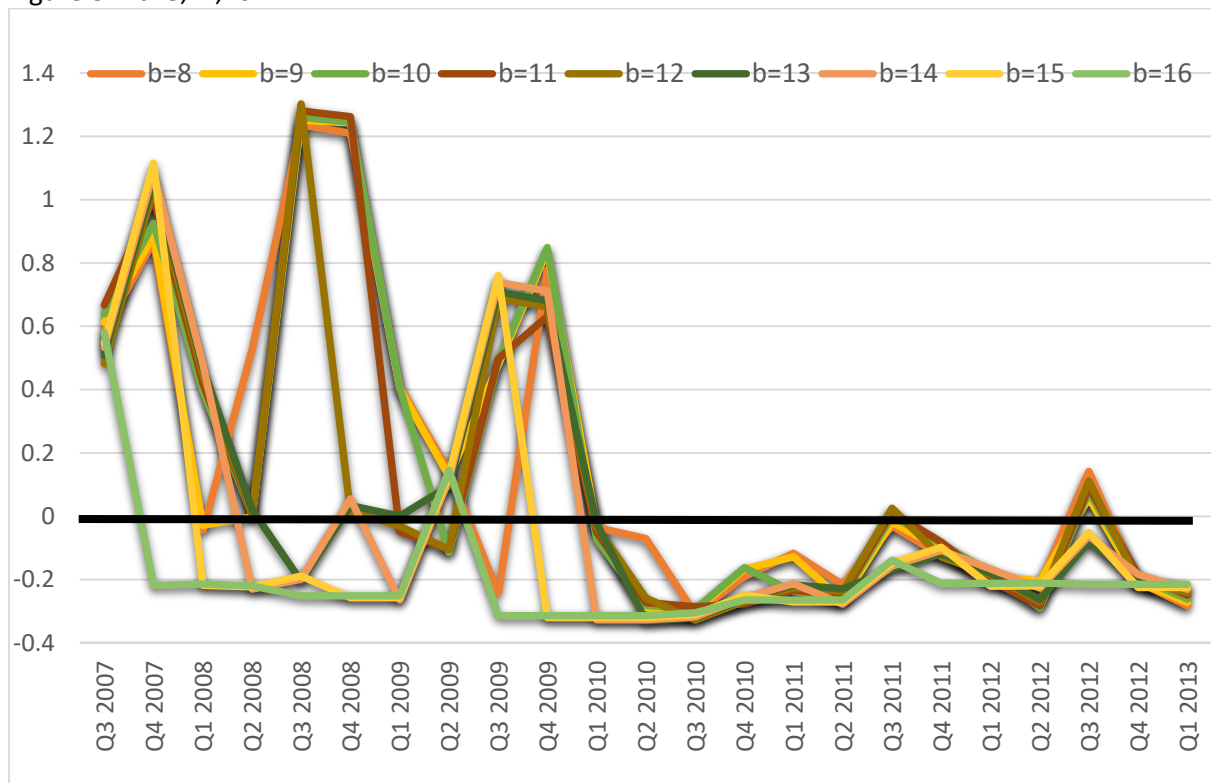
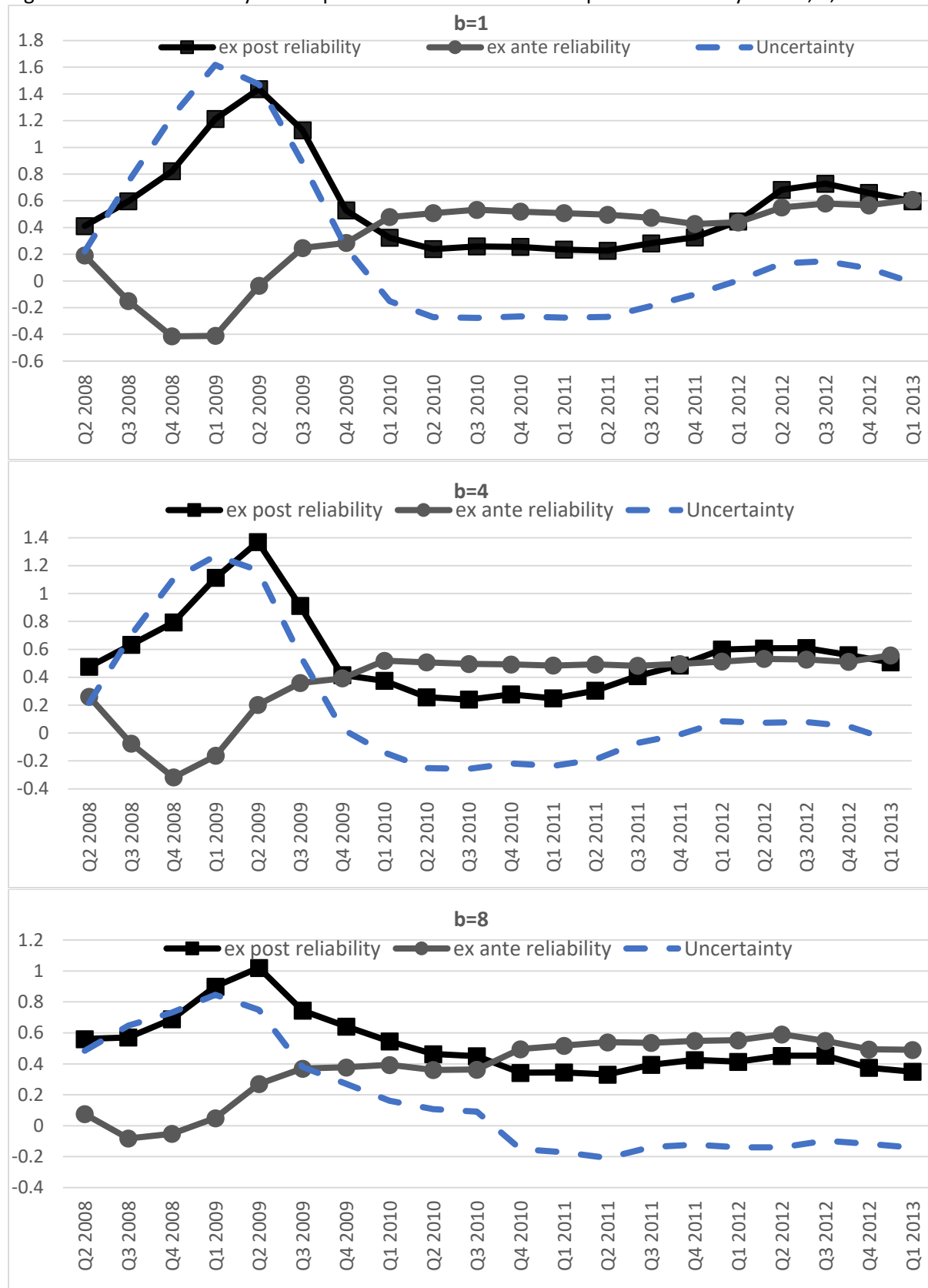
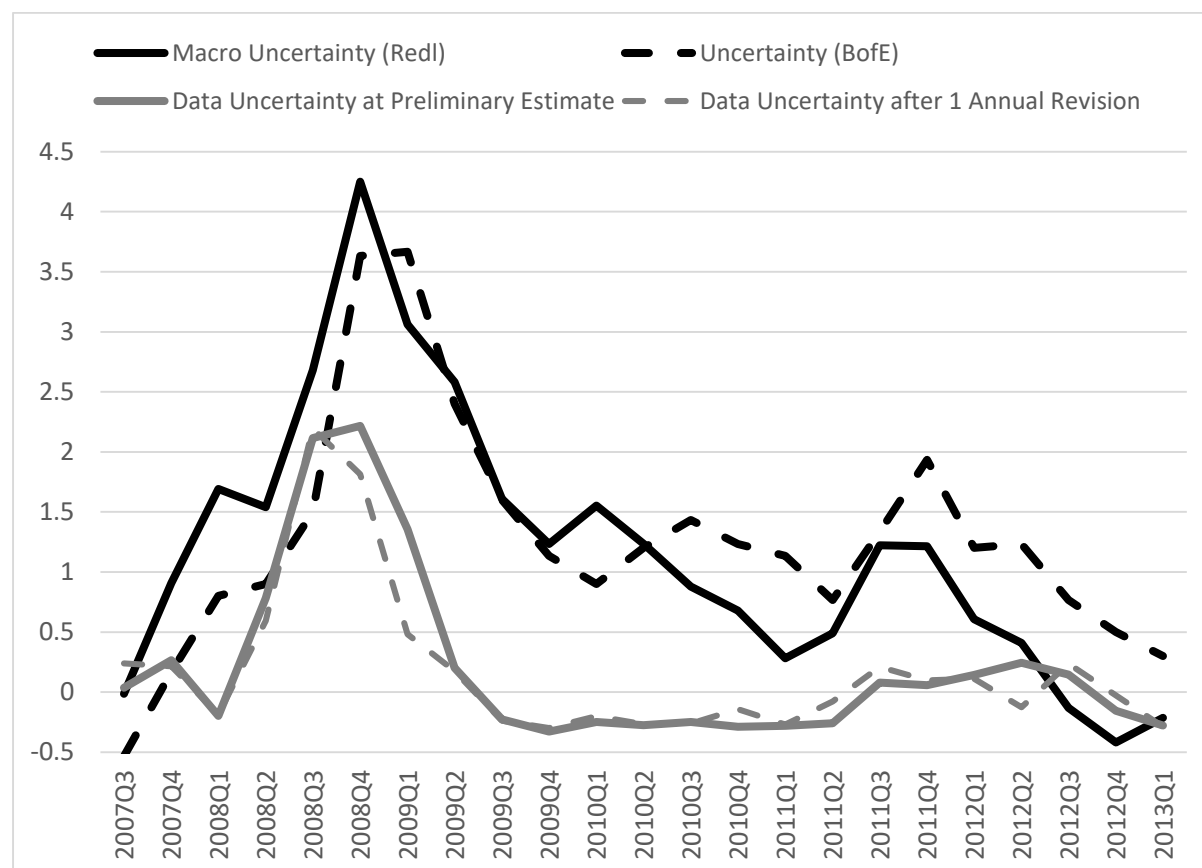


Figure 9: CRPS Uncertainty Decomposition into Realised and Expected Reliability at $b=1, 4, 8$



Notes: The decomposition uses a window of 4 observations, so reported values are averages of the last three quarters and the current quarter. Data uncertainty is defined as $Unc_{t|t+b}^{CRPS}$ as seen in Figure 8. We gratefully acknowledge use of code from Tatevik Sekhposian to decompose the CRPS.

Figure 10: Data Uncertainty and Macroeconomic Uncertainty as measured by Redf (2017) and the Bank of England (*Inflation Report*, Aug 2016)



Notes: Data uncertainty is defined as $Unc_{t|t+b}^{CRPS}$ as seen in Figures 8 and 9; and is reported at both $b=1$ (for the 'preliminary' estimate) and $b=4$ (after one annual revision). The uncertainty measures plotted have their own units of measurement and so focus should be on correlations not scale.

A Online Appendix

A.1 US GDP Data Revisions

In this Appendix we replicate the revisions’ analysis in Section 2 using US GDP data. The US real-time dataset was obtained from the FRB Philadelphia’s Real-Time Dataset for Macroeconomists. Both the UK and the US national statistical offices, over our period of study, provided first releases of quarterly GDP growth around 30 days after the end of the observation quarter. In earlier research Faust, Rogers and Wright (2005) found that while both UK and US first estimates of GDP growth are biased, it is harder to predict US data revisions in real-time; this is also consistent with Aruoba (2008). The apparent (relative to the US) predictability of UK data revisions may help explain why the MPC chose to provide probabilistic backcasts; we do not currently see similar expressions of data uncertainty produced in the US.

Table A1 shows the characteristics of the revision process $(y_t^{t+17} - y_t^{t+1})$ for US data. It matches results in Table 1 in the main text. If we compare the UK statistics in Table 1 with the US values in Table A1, we find that the UK revision size, as measured by the standard deviation of $(y_t^{17} - y_t^{t+1})$, is not very different to US values (only 25% larger) when computed over the longer 30 year period (1983-2013). For the most recent period (2007-2013), however, we find that UK GDP growth revisions are twice as large as in the US. If instead we measure revision size using the mean absolute error, as in Office for National Statistics (2016), as indicated in the third column for each panel in Table 1, we find similar results: UK revisions, measured as $(y_t^{17} - y_t^{t+1})$, are similar in size to US values over the longer periods. This is consistent with the results in Office for National Statistics (2016). But they are twice as large if computed between 2007 and 2013.

We also estimate the UC-SV model in (3) using US data, to see if these temporal changes to the data revision process are specific to UK data. Figure A1 plots the posterior mean estimates and 90% credible intervals for the time-varying revision mean and both the measurement error and the revision mean volatilities for $l = 17$ and $b = 1$. The plots also include shaded NBER recession phases; so we can see how the estimated data revision components relate to business cycle phases. The time-variation is smaller in the US data: the revision mean fluctuates between 1.5% and -2% and its sign exhibits more variability, so it may be harder to predict than in the UK. An important similarity with the UK is that the data revision mean is also large and negative during the 2008/2009 recession. This is also associated with an increase in revision

mean volatility from 0.2 to 0.4. There is no evidence of negative revision mean or increasing revision mean volatility in the 1990/91 and the 2001 recessions. There is very little evidence of time-variation in the measurement error volatility.

These results indicate that data revisions were also larger in the US during the Great Recession. Because there is no evidence that data revisions are more sizeable in other recessionary periods in the sample, we speculate that large data revisions might be more strongly linked with recessions originating in the financial sector. An alternative possible explanation is that data revisions are related to the depth of the recession.

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

Here, we start by presenting additional details about estimation of the UC-SV model in Section 2.

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 A2 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 17th 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 A2 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 (3) as:

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

After estimating this modified specification for $l = 17$ and $b = 1$, we found that ρ is estimated to be about 0.05 so, qualitatively, the results in Figure 4 are unchanged.

We then allowed the innovations to have a t-distribution, so that the one-step-ahead predictive 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 4 using the normal distribution.

Online Appendix: Tables and Figures for Galvao and Mitchell “Measuring Data Uncertainty: An Application using the Bank of England's Fan Charts for Historical GDP Growth”

Table A1: Summary Statistics for GDP Growth Data Revisions with US Data (BEA, Philadelphia Fed real-time database)

Obs Period:	$y^{17} - y^1$			$y^{17} - y^{13}$			$y^{latest} - y^{17}$		
	mean	std	MAE	mean	Std	MAE	mean	std	MAE
1983Q2-2013Q1	-0.141	0.740	0.597	0.012	0.188	0.097	0.349	0.402	0.391
1993Q2-2013Q1	-0.335	0.734	0.644	0.011	0.132	0.076	0.186	0.265	0.230
2007Q3-2013Q1	-0.504	0.781	0.607	0.019	0.175	0.107	0.020	0.249	0.126

Notes: Values in bold indicate that the data revision mean is statistically different from zero using robust (Newey-West) standard errors. MAE is the mean absolute error, or mean absolute revision. Because we have to wait 4 years to observe our measure of “mature” data, the last observation for which we observe a 17th release is 2013Q1. We use the July 2017 data vintage to measure y^{latest} .

Table A2: Bayes’ factor validating stochastic volatility components

Model for:	$y^{17} - y^1$	$y^{17} - y^4$	$y^{17} - y^8$	$y^{17} - y^{12}$
Measurement	1.55	2.60	0.8	1.60
Error	(0.06)	(0.13)	(0.09)	(0.14)
Bias	3.28	1.70	0.1	1.2
	(0.13)	(0.07)	(0.07)	(0.12)
Both	24.53	48.6	43.5	70.00
	(1.20)	(3.85)	(3.03)	(3.77)

Notes: Entries are averages over 5 chains; standard deviations are in parentheses.

Table A3: Performance of the UC-SV model backcasts: point and density forecast evaluation statistics at b=1, 4, 8, 12, 16. Evaluation statistics reported using ONS data after 4 years ($l = 17$) as the outturn (prediction horizon $= l - b$)

	Bias	RMSE	LogS	CRPS	Int: 90%	Int: 75%	Int: 50%
b=1	0.179	1.542	2.103	0.872	74%	52%	30%
b=4	0.153	1.444	2.567	0.803	77%	65%	31%
b=8	-0.027	1.098	3.063	0.706	53%	53%	27%
b=12	-0.012	0.858	2.917	0.428	59%	47%	32%
b=16	0.005	0.283	3.138	0.129	84%	84%	82%

Notes: out-of-sample evaluation samples: b=1 backcasts: 2007Q3-2013Q1; b=4 backcasts 2006Q4-2013Q1; b=8 backcasts 2005Q4-2013Q1; b=12 backcasts 2004Q4-2013Q1; b=16 backcasts 2003Q4-2013Q1. Model estimated using data from 1983Q2 up to the date available in real-time. We use the posterior mean of μ_T to compute h-step-ahead forecasts for revised values assuming a Gaussian density with predicted variance of $var(\varepsilon_T) * h * \exp(w_h^2) + var(\eta_T) * h * \exp(w_g^2)$, where h is the forecast horizon ($l - b = h$).

Figure A1: US Estimated Data Revision Components $rev^{16} = (y^{17} - y^1)$

