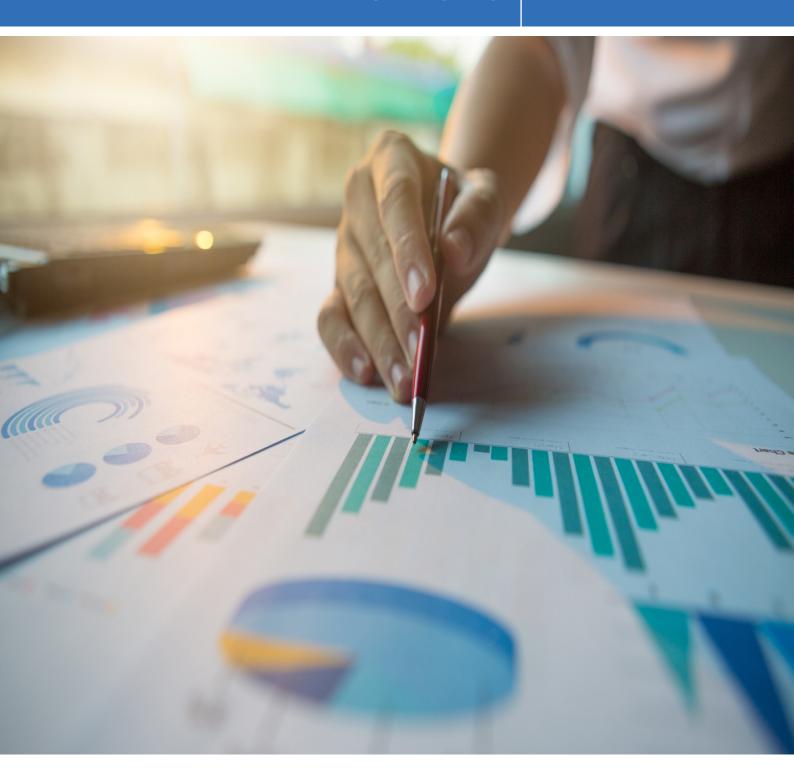
Measuring and communicating uncertainty in official statistics: State of the art and perspectives

GEORGE KAPETANIOS, MASSIMILIANO MARCELLINO, FELIX KEMPF, GIAN LUIGI MAZZI, JANA EKLUND, VINCENT LABHARD

2021 edition





Measuring and communicating uncertainty in official statistics: State of the art and perspectives

GEORGE KAPETANIOS, MASSIMILIANO MARCELLINO, FELIX KEMPF, GIAN LUIGI MAZZI, JANA EKLUND, VINCENT LABHARD 2021 edition

Manuscript completed in December 2020

The Commission is not liable for any consequence stemming from the reuse of this publication.

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021



The reuse policy of European Commission documents is implemented based on Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence

(https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders. The European Union does not own the copyright in relation to the following elements:

For more information, please consult: https://ec.europa.eu/eurostat/about/policies/copyright

Copyright for the photograph: Cover © AnemStyle/Shutterstock

The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Theme: General and regional statistics Collection: Statistical working papers

PDF: ISBN 978-92-76-189116-9 ISSN 2315-0807 doi: 10.2785/452295 KS-TC-20-006-EN-N

Abstract

This paper examines the state-of-the-art methodologies that are available to measure and communicate uncertainties on official statistics in reports, publications and online resources. Our aim is to investigate the available tools for a more accurate representation of inherent uncertainties underlying economic and social statistics. For a complete assessment, we evaluate different categories of uncertainty ranging from sampling and non-sampling ones to conceptual uncertainty. We also briefly address some aspects of uncertainty, not necessarily related to its measurement, which could help in better understanding and interpreting it. Moreover, we discuss the quantification of uncertainty through quantitative methods and we investigate different communication tools applied by prominent European agencies. Overall, our analysis shows that the fan-chart— used by the Bank of England to communicate inherent uncertainty in their forecasts — might serve as a particularly useful tool also for official statistics. Starting from this consideration we present an extensive empirical application aiming at measuring total uncertainty. The methodology is based on recent work undertaken at policy making institutions such as the Bank of England. Our empirical application deals with key macroeconomic indicators of the euro area largest economies. Further, together with some concluding remarks, we formulate a first series of recommendations addressed to statistical authorities.

Keywords: Measuring Uncertainty, Communication Tools, Fan-Chart, Official Statistics.

Authors: George Kapetanios (2), Massimiliano Marcellino (3), Felix Kempf (4), Gian Luigi Mazzi (5), Jana Eklund, Vincent Labhard.

⁽²⁾ george.kapetanios@kcl.ac.uk

⁽³⁾ massimiliano.marcellino@unibocconi.it

⁽⁴⁾ felix.kempf@kcl.ac.uk

⁽⁵⁾ glmazzi@pt.lu

Table of contents

1. Introduction	6
2. Categories and sources of uncertainty	8
2.1 Sampling uncertainty	
2.2 Non-sampling uncertainty	
2.3 Conceptual uncertainty	10
3. Further aspects of uncertainty	12
3.1 Uncertainty and data typology	12
3.2 Uncertainty and data frequency	12
3.3 Uncertainty and the complexity of the production process	13
3.4 Uncertainty, transparency and replicability of the production process	13
4. Quantification of uncertainties	15
4.1 Less-quantitative approaches	15
4.2 Quantitative approaches	
5. Communication	20
5.1 Importance of the communication of uncertainty	20
5.2 Specific communication tools	20
6. Communicating Uncertainty in Practice	23
6.1 Practices in the European Statistical System	
6.2 Bank of England	25
6.3 FED / FOMC	29
6.4 European Central Bank	31
6.5 Swedish Central Bank – Riksbank	32
6.6 Office for National Statistics (ONS) United Kingdom	
6.7 Other disciplines	34
7. Our methodology for measuring uncertainty in macroecono	
7.1 The model	
7.2 Estimation of the state space model	
8. Empirical application	41
8.1 Data	
8.2 Empirical Results	
8.3. Robustness	44
9. Conclusion	45
10. Annex	51

Abbreviations

CBS Centraal Bureau voor de Statistiek

CPIF consumer price index with a fixed interest rate

ECB European Central Bank

ESS European Statistical System

FED Federal Reserve Bank

FOMC Federal Open Market Committee

GDP gross domestic product

GRADE Grading of Recommendations Assessment,

Development and Evaluation

HICP harmonized index of consumer prices

IPCC intergovernmental panel on climate change

IPI index of industrial productionMPC Monetary Policy Committee

MSE mean squared error

ONS Office for National Statistics
RTI index of retail trade volume

SEP Summary of Economic Projections
SIMS Single Integrated Metadata Structure

UR unemployment rate

1

Introduction

Statistical offices and other public agencies producing statistical data usually communicate a variety of official economic and social indicators to the public, in general as single values (normally corresponding to the central point estimate). This is done without explicitly mentioning the associated inherent and unavoidable uncertainty. While the technical documentation associated with such publicly available official data often acknowledges the possible presence of errors, little is done to communicate widely such features, let alone account for the effects that such errors, and, in general the associated uncertainty, have on the decision-making of private and public agents.

While it is difficult to derive a valid scientific or professional explanation for this circumstance, Manski (2019) argues that one possible reason for this status quo lies in the partly political nature of official statistics. He argues that policy makers or other public agencies may be incentivised to express strong certitude in their communication rather than providing further information about the underlying and inherent uncertainty. However, conveying strong certitude about data or economic analysis can be harmful for the development of public policies in multiple ways. For instance, if policy makers incorrectly believe that existing statistical analyses provide an errorless description of the current state of the economy and accurate predictions for future developments, they will not take into proper account the underlying uncertainty when taking their decisions. More generally, policy and decision makers will struggle to properly understand, learn from and deal with inherent uncertainty if the latter is not properly communicated. Last but not least, communicating official statistics with strong certitude leads to further difficulties because of the way that third parties, and in particular the mass-media, disseminate this information to a wider audience, namely by largely taking them at face value, which may lead to further miscommunication. On the other hand official statisticians are often worried about the uncertainty associated to their statistics. More specifically they are worried by the possibility that showing the uncertainty affecting statistics they are producing could lower their credibility. Furthermore, they consider that uncertainty, especially when it is relatively high, could confuse or even mislead users especially policy makers and analysts. This might explain, even if it does not justify completely, the traditional conservative position taken by official statistical agencies. Nevertheless, as we will see later in this document, things are starting to move, even if slowly, and the attention to all aspects related to the uncertainty in official statistics is progressively growing up. The present paper focuses on the review of existing works that categorise sources of errors and uncertainty in official statistics and data releases; provide a quantification of such uncertainty; and promote tools for communicating such uncertainty to private and public agents. While our focus primarily lies on economic data, our review of such work will also include contributions from other scientific disciplines such as meteorology, political and other sciences.

In particular, starting with the first strand of the review, we will distinguish among various sources of error and uncertainty. Such sources include sampling uncertainty, where the use of small or biased samples can lead to errors with respect to an unknown population quantity of interest such as, e.g., national accounts data (consumption, investment) or real activity data (industrial production). Such uncertainty is characterised by a reduction in magnitude as more and more survey and other granular data become available, leading to revisions in official releases of data which are, therefore, referred to as revision errors. A second source of uncertainty is methodological, related for example to the use of more appropriate methodologies to gather and aggregate data over time. Such errors do not diminish

over time for particular data periods and so are separate from revision errors. Finally, we have conceptual uncertainty raising out of the fact that measured data may not fully correspond to the economic concept one is trying to measure and analyse. Such errors cannot easily be analysed, unless one uses an economic model to provide some structure on the data that are available. Since Mazzi et al. (2020) measuring already provided an exhaustive survey of the various sources of error and uncertainty, here we will only provide a short summary, focusing on those aspects that are more relevant for uncertainty quantification and communication.

The second strand of the literature we will explore focuses on quantifying uncertainty. In particular, we will review models and other techniques used by statistical agencies and other policy making institutions to quantify uncertainty and measurement errors. This includes a wide variety of tools and methods, ranging from informal rules of thumb to more fully specified and complex econometric models. Tools from various other disciplines, including hard sciences, will be also reviewed and commented on, focusing on those that could be of use in an official statistics context.

The third strand of the survey will explore the more public facing aspect of dealing with errors and uncertainty, reviewing how statistical agencies and policymakers discuss and present their views and estimates of uncertainty in official documents.

The final contribution of this paper is a detailed empirical application. This considers recent work undertaken at policy making institutions such as the Bank of England, which, also as noted by Manski (2015a), has been leading the way in the past 30 years in analysing and communicating data and forecast uncertainty through the use of tools such as the fan chart for inflation and GDP growth. Here we are extending this approach to deal with official statistics and we apply the model to measuring uncertainty of some key macroeconomic indicators of the euro area and its largest member countries.

The rest of this paper is structured as follows. Section 2 introduces different categories and sources of uncertainty in official statistics. This section provides a high-level overview; then section 3 describes further aspects of the uncertainty which could help in better understanding and interpreting this complex phenomenon; section 4 introduces more quantitative approaches. Section 5 provides an indepth summary of tools used to communicate different uncertainties. Section 6 discusses a number of examples, with a particular focus on the fan-chart used by the Bank of England, as it provides an effective way to communicate uncertainty and it is often mentioned in the media. Finally, sections 7 and 8 present our main methodological contribution and empirical results.

Categories and sources of uncertainty

As already touched on in Mazzi et al. (2020), official economic and social statistics inherently carry uncertainties or errors due to the way they are compiled as well as to the fact that the phenomena they measure are often latent and not directly observed. This inherent uncertainty or error can be defined as the difference between the estimated and the true population value. This section critically discusses and reviews different categories and sources of those uncertainties. While the focus of this report is different from that in Mazzi et al. (2020), parts of this section may display some overlap with previously discussed concepts. To reduce the overlap, we focus on those aspects that are particularly relevant for the quantification of uncertainty and its communication.

Statisticians and statistical offices have sought to categorise and communicate uncertainties in various ways. Morgenstern et al. (1963) presented one of the most comprehensive work on this topic and they emphasised that the examination of accuracy has been more neglected by economists than by natural scientists, while observing the same sources of error or uncertainty in both disciplines. For example, as pointed out by Manski (2015b), Manski (2019) and van der Bles et al. (2019), headline flash estimates, nowcasts and forecasts are often presented as point estimates, arguably conveying a misleading degree of reliability, without explicitly expressing some underlying and inherent uncertainties. Another research worth mentioning is Plümper and Neumayer (2012).

This section explores and critically evaluates this strand of literature in further detail by focusing on the two most commonly established sources of uncertainty: sampling and non-sampling uncertainty, e.g. see Heckman (1979) Mazzi et al. (2020). While sampling errors apply to sample surveys, non-sampling errors apply to administrative records and surveys, including censuses for instance. The total uncertainty associated with a statistic, therefore, in principle, comprises both sampling error and non-sampling error—the total survey error. However, quantifying the total uncertainty or total survey error is complicated, especially with regard to quantifying non-sampling errors. Last but not least, we will also consider conceptual uncertainty which may arise out of the fact that measured data may not fully correspond to the economic concept one is trying to measure and analyse.

It is important to consider these issues as, when it comes to economic statistics, a recurring criticism is that agencies do not always indicate clearly enough, or even explicitly state, the existence of sampling and non-sampling uncertainties.

2.1 Sampling uncertainty

Official economic and social statistics are exposed to sampling uncertainty, which arises whenever an analysis is conducted on a subsample of the population to draw conclusions on the population values. Sample uncertainty inevitably occurs due to a subsample selection and is predominantly driven by the sample size and its bias (e.g. see Heckman (1979)). In economic and social statistics, small or biased samples can lead to errors with respect to an unknown population quantity of interest, such as e.g. national accounts data (consumption, investment), real activity data (industrial production) and many others — for examples see Chang and Li (2018), US Department of Labor, Bureau of Labor Statistics (2019) or UK Office for National Statistics (2019). However, the problem is by no means exclusive to economic and social statistics, but it occurs in many other disciplines, for example in marketing (see,

Assael and Keon (1982)), fluid mechanics (see, Benedict and Gould (1996)) or neuroscience (see, Varoquaux (2018)) to name a few.

In contrast to non-sampling uncertainty, sampling uncertainty can usually be quantified, as pointed out by Manski (2019). Common measures include standard error (SE), relative standard error (RSE) or mean squared error (MSE). However, other measures exist — for example see Goedemé (2013) or Mevik (2004).

Sampling errors inevitably arise because not all units in the population are measured. If a sample from the population is chosen randomly, for example, then each random sample will involve sampling some different units, and this implies that each sample will produce different estimates. When there is great variation among the samples drawn from a given population (i.e. there is greater variability in the population), the sampling error can be high. It follows that there is a larger chance of the survey estimate being further awayfrom the true population value. Sample uncertainty is typically influenced by a number of different factors, including the sample size (larger sample sizes reduces sample uncertainty), variability in the population (greater variability increases sample uncertainty), survey sample design, and the estimation method used.

One key aspect about sampling errors in economic and social statistics is that those statistics can be (and usually are) revised over time. This means that official statistics may get amended retrospectively as more information becomes available. More information evidently means a larger sample size. In other words, the sample size increases over time, resulting in diminishing sampling errors over time, which is why this is referred to as transitory uncertainty. Another factor generating revisions is due to the use of an incomplete set of information, especially when compiling rapid estimates. This means working with a sample covering not the full period of time (e.g. only one or two month of the quarter) or with a sample for which not the involved units have not provided information on time. These generate the so-called trade-off between timeliness and accuracy Mazzi and Ruggeri-Cannata (2017): the public is interested in getting official statistics measuring important economic phenomena as soon as possible, but if the data release happens too early it will be subject to substantial uncertainty and larger revisions. In this respect, a formal quantification of the uncertainty can be helpful both to the public, to understand the reliability of the rapid estimates, and to the official statistical agencies, to understand the proper release time to guarantee a sufficient level of data accuracy and reduce the need of subsequent large revisions.

2.2 Non-sampling uncertainty

A detailed overview of non-sampling uncertainty and its different types has been provided recently in Mazzi et al. (2020). This section, therefore, only briefly summarises previous findings. The preceding section outlined how uncertainty can arise from using samples instead of the entire population. Non-sampling uncertainty, on the other hand, arises from the design, data collection and processing methods used and can therefore be associated with the methodology used, rather than being a function of the sample chosen, Bank of England (2014).

In comparison to sampling uncertainty, non-sampling uncertainty is much harder to quantify, which is arguably one reason why it is hardly ever explicitly stated in the meta-data accompanying official statistics. Yet, the extent of non-sampling errors can be mitigated through good survey design and practices, such as data cleansing, imputation and other techniques. Yet, non-sampling errors are in general independent of the sample size, Morgenstern et al. (1963), actually they can even increase with the sample size. Moreover, they do not diminish over time with the arrival of additional information, and so are conceptually different from revision errors.

A common typology of non-sampling uncertainty is based on a decomposition into five elements, e.g. see Biemer and Lyberg (2003), including specification, coverage, non-response, measurement and processing uncertainty.

Good practices to tackle non-sampling errors include ESA 95, respectively its revision ESA 2010 by Eurostat. Their goal is it to align national accounts in Europe with international standards of the System of National Accounts (SNA) and therefore has a standardising effect.

2.3 Conceptual uncertainty

Last but not least, one should also consider conceptual uncertainty, which arises out of the fact that measured data may not fully correspond to the economic concept one is intending to measure and analyse. Similarly to non-sampling uncertainty, this type of uncertainty is difficult to quantify and analyse. Specific expertise in each respective field is generally needed to minimise this type of uncertainty.

Further aspects of uncertainty

Uncertainty is a complex multidimensional phenomenon which needs not only to be adequately measured and communicated, but also properly understood and interpreted. In particular, looking at the uncertainty phenomenon in official statistics, there is a number of aspects which appear almost orthogonal to the categories discussed above and which do not influence either the measurement or the communication of uncertainty, but the way in which it is perceived and understood. Such as pects are mainly related to data characteristics and specificities (e.g. frequency, flow versus stock data, etc.) and to the characteristics of data production process (i.e. date generating process). These aspects can affect our a priori perception on the uncertainty associated to a given statistical variable and help us in better assessing the realised uncertainty once estimated. For this reason various aspects considered in this section can be used as a set of meta-data associated to the uncertainty measures, providing users with relevant information for better understanding and interpreting the uncertainty associated to given statistics. In this section we shortly discuss some of them. Before going on with our discussion, it is important to clarify one essential point. In most cases, the various aspects we will discuss later in this section are not linked by a clear causal relation with either the presence or the size of uncertainty. Instead they can offer a number of alternative scenarios, depending on a number of factors such as the sampling structure, the applied production methodologies, and so on. For this reason, in our description we only present subjective considerations reflecting our past experience, not pretending to cover all possibilities.

3.1 Uncertainty and data typology

Various socio-economic phenomena that official statistics aim to measure can be characterised by different degrees of variability and volatility. The higher the variability and volatility of a phenomenon, the more complexis its estimation. Consequently, it is possible to assume that statistical measures of a phenomenon characterised by high variability and volatility tend to be more uncertain than those of more regular ones. Looking at the data traditionally produced by statistical agencies, without pretending to make a general assessment, we can say that flow phenomena tend often to be characterised by high variability and volatility, which complicates their estimation, nowcasting and forecasting. This is especially the case for volume or quantity data, such as production, or orders. Examples are represented by the industrial production, the retail trade volume, new orders, etc. On the other hand, prices or stocks, especially population and labour force related stocks, tend to be characterised by smaller variability and volatility, making them relatively easier to be estimated, nowcasted and forecasted. Obviously, this is not an exhaustive analysis of data typologies, but just a short list of examples. We are nevertheless convinced that it is sufficient to show that when assessing any kind of empirical measure of uncertainty, we need to have clear in mind which type of statistical indicators we look at.

3.2 Uncertainty and data frequency

Over the last two-three decades, many statistical agencies have progressively increased the frequency

of their statistical production, mainly from annual to infra-annual (i.e. quarterly and monthly) frequencies. While the production of annual data can be based on detailed surveys, conducted on large samples, this can be realistically more complex to be done at infra-annual frequencies. Furthermore, especially in the past, annual statistics were released with large delays, privileging accuracy rather than timeliness. In this context, statistics became available quite late after the end of the year, being of little utility for policy making, evaluation and monitoring, particularly for policies based on short-term statistics (such as euro area monetary and economic policy, for example). They were built up with the aim of having a quite low degree of uncertainty, even if mainly sampling and methodological problems have prevented, in several cases, the achievement of this goal. To ensure the regular and timely production of quarterly and monthly statistics, official statistical agencies have been confronted to the need of simplifying their surveys, using smaller sample sizes and alternative data sources whenever possible. This has been necessary both to reduce production costs and to lower the burden on individuals, households, firms, etc. entering in the sampling population. Infraannual statistics also require a more intensive utilization of imputation, estimation and filtering techniques than annual statistics. In this context, we can consider that there is a concrete risk of increased sampling and non-sampling uncertainty in infra-annual statistics with respect to annual ones. This can be seen as another side of the so-called trade-off between timeliness and accuracy. Infraannual statistics become available much earlier than annual ones, they are available already during the year providing updated information on the economic evolution largely in advance than annual ones. The price to pay could be represented by a larger amount of uncertainty associated to infra-annual statistics. In principle, we could assume that uncertainty tends to increase with data frequency even if this can be mitigated, cancelled or even reversed by the use of more advanced statistical techniques.

3.3 Uncertainty and the complexity of the production process

Statistics can be built up by means of production processes of very different complexity. Some of them are very simple, involving only a single survey and a few statistical interventions; others can be very complex involving a number of different surveys and several statistical interventions, such as estimation, calibration, imputation, balancing and so on. An interesting question which legitimately users could ask is: 'is there any relation between the complexity of the production process and the uncertainty of statistics?' In other words, could we assume that a complex production process tends to generate more uncertain statistics or vice versa? As in previous cases, our opinion is that there is no uniquely defined relation between the complexity of the production process and the presence and size of uncertainty. In some complex production processes uncertainties associated to various surveys can tend to cumulate each other generating more uncertain outputs. In other cases, they can also tend to, at least partially, cancel out each other with beneficial effect on the output uncertainty. Furthermore, often in complex production processes, statistical procedures explicitly designed to mitigate uncertainty are used in order to enhance the reliability and the accuracy of the final output. The production process of the GDP is a good example of how a complex production process can generate a statistic characterised usually by a moderate degree of uncertainty. Furthermore, looking at production processes of different degrees of complexity, we can say that for simple production processes the uncertainty is much easier to be measured, understood and interpreted than for complex processes. Finally, we can say that, especially in case of complex production processes, their knowledge can be particularly helpful in guiding policy makers, analysts and users in better interpreting and assessing uncertainty associated to their output statistics.

3.4 Uncertainty, transparency and replicability of the production process

Developing as much as possible transparent production processes is a common objective to almost all statistical agencies. By contrast, developing replicable production processes is more a dream than a realistic objective, especially when production processes are very complex. In this respect,

replicability should be seen as the frontier of the transparency. An interesting question is whether or not transparent and even replicable production processes can help in producing less uncertain statistics. Unfortunately, once again, the answer is not easily found, as for example it is possible to make very transparent a low quality production process which produces biased, inconsistent, and too volatile statistics. If we restrict the class of production processes only to good and well-designed ones, based on sound methodologies and robust sampling schemes, the situation can be seen in a different way. As before, in this case the ability of reducing uncertainty is coming from the characteristics of the production process and not from its transparency. Nevertheless, if a good production process is made transparent, this can contribute to show that, with the given data set and tools, it is hard to do better in terms of uncertainty associated to the produced statistics. This fact will be a positive element for statistical agencies and it can increase their credibility and reputation.

Quantification of uncertainties

The previous sections have, on one hand, given an overview of the main issues arising from inherent uncertainties in official statistics. This section, on the other hand, focuses on various different methods — some less quantitative while others more quantitative — which are used to deal with the data uncertainties. When it comes to quantifying uncertainties across various disciplines, different strategies are applied, ranging from rules of thumb to more fully specified and complex econometric models. In the following, we will focus on the former first, before introducing more quantitative models. Moreover, the approaches presented mainly relate to sampling uncertainty, as the other types of uncertainties cannot necessarily be quantified.

Stigler (1986) displays a history of statistical modelling prior to the 20th century. He shows how fundamental tools such as regression, least squares and correlation or association analyses were developed. Those tools also include test statistics, which inherently allow for 'inference under uncertainty', according to the critical values chosen. By referring to other authors such as Groves (2004), Stigler promotes the idea of a *supermodel* which can process and incorporate multiple sources of uncertainty. To uphold this approach, one focus of this section lies on Cunningham et al. (2012b), who introduced a state space approach to extract signals from uncertain data, which is targeted to be applied in economics.

4.1 Less-quantitative approaches

One less-quantitative approach is to multiply predictions with a constant. This is done, for example, in civil engineering where it is well-known that certain loads are inherently statistical / stochastic. For instance, when designing a structure and taking loads resulting from snow into account, the engineer does not know for certain how much snow there will be in the future. However, there are specific standard values depending on the geographic location of the structure which are organised in European norms called Eurocodes. Moreover, because snow is a varying force, the Eurocodes introduce so called partial safety factors which are constant factors by which the standard values are multiplied. The reasoning behind this is that loads have specific unknown distributions implying that extreme values are statistically possible. At the same time, the construction sector only allows for very small probabilities of failure. The standard values can be translated into average values which are then increased by the partial safety factor. This is a very effective mechanism. However, when comparing such an approach to economics it becomes obvious that engineering is only concerned about maximum values, as this equates to the maximum force a structure must with stand. This is inherently different from economics, where we are interested in both directions of predictions.

An approach related to the above-mentioned strategy is used by the Federal Open Market Committee (FOMC) to communicate uncertainty about their predictions in their Minutes of the FOMC, see Money and Banking (2019). In particular, the FOMC uses past prediction errors to quantify future uncertainty about forecasts. To be more specific, suppose that the median projection for the unemployment rate for 2021 is 3.9 percent, with an error range of plus or minus 1.7 percentage points, where this value is based on past forecast errors as explained in more details below. This tells us that, as suming normality and given historical experience, there is a 70 percent chance that in 2021, the unemployment rate will

16

be between 2.2 and 5.7 percent. While at first sight this approach seems very similar to the fan-chart method, which is used by the Bank of England, it is instead distinctively different, as discussed in section 6. Of course, one majordrawback of this approach is that past predictions are by no means an accurate indicator for future accuracy. One could argue that in less-volatile times predictions may be more accurate, indicating less uncertainty. However, if macroeconomic conditions change quickly such apparent past accurate predictions may not be reliable future indicators any more. The confidence interval around the median projected values is based on root mean squared errors of various private and government forecasts made over the previous 20 years. The confidence interval is not strictly consistent with the projections for the federal funds rate, primarily because these projections are not forecasts of the likeliest outcomes for the federal funds rate, but rather projections of participants' individual as sessments of appropriate monetary policy. Still, historical forecast errors provide a broad sense of the uncertainty around the future path of the federal funds rate generated by the uncertainty about the macroeconomic variables as well as additional adjustments to monetary policy that may be appropriate to offset the effects of shocks to the economy. The confidence interval is assumed to be symmetric, except when it is truncated at zero, the bottom of the lowest target range for the federal funds rate that has been adopted in the past by the FOMC. A natural counterpart of this approach in the context of official statistics would be to use past revision errors to provide measures of uncertainty around the various data releases for each variable of interest.(6)

4.2 Quantitative approaches

Clogg and Dajani (1991) argue that it is the main purpose of a statistical model to explicitly capture the sources of uncertainty so that they can be quantified. Taking this philosophy into account, this section provides an overview of more quantitative econometric models for uncertainty quantification.

Knüppel and Schultefrankenfeld (2019) examine the quality of inflation forecast uncertainty data provided by several central banks using the same methods as those employed to analyse survey-based measures of forecast uncertainty. They focus on data of the Bank of England, the Banco Central do Brasil (the Brazilian central bank), the Magyar Nemzeti Bank (the Hungarian central bank) and the Sveriges Riksbank (the Swedish central bank) because they are particularly suited, on account of their nature and scope, for this purpose. Much like the survey-based measures of forecast uncertainty, they find that the central bank data they examine in their paper exhibit certain biases. The statistical evidence for these biases is not particularly strong, however.

The evaluation of density forecasts issued by central banks has mostly focused on the Bank of England with important contributions by Clements (2004), and Mitchell and Hall (2005). The asymmetry incorporated in central banks' density forecasts is studied in Knüppel et al. (2012). However, the dispersion of central banks' forecast densities, i.e. uncertainty forecasts, has not been investigated explicitly yet. But, interestingly, when the Bank of England density forecasts fail to pass tests for correct specification, this is often explained by the excessive width of its fan charts, i.e. its excessively large uncertainty forecasts, as done in Clements (2004), Wallis (2004), and Dowd (2007).

Conditional mean forecasts are typically evaluated with respect to their optimality using measures such as their bias and efficiency. Knüppel and Schultefrankenfeld (2019) do the same for conditional uncertainty forecasts. With respect to bias, they investigate whether the ex-ante uncertainty, i.e. the uncertainty surrounding a central bank's mean forecast, coincides, on average, with the ex-post uncertainty, i.e. the size of the realized forecast errors of this mean forecast. The object to be evaluated is the forecast for the squared error of a corresponding conditional h-step-ahead mean forecast. The expected squared error is given by

$$\sigma_{t+h|t}^2 = E[\hat{e}_{t+h|t}^2 | I_t] = E[(y_{t+h} - \hat{y}_{t+h|t})^2 | I_t], \tag{1}$$

eurostat Measuring and communicating uncertainty in official statistics: State of the art and perspectives

⁽⁶⁾In economics and finance there is a recent surge led to the construction of uncertainty indices like in Baker et al. (2016), Rossi and Sekhposyan (2015) and Jurado et al. (2015), where the latter two actually rely on measures of forecast uncertainty.

where I_t is the information set of the forecaster in period t and y_{t+h} denotes the value of the target variable in period t+h. The variable $\hat{y}_{t+h|t}$ is the corresponding conditional mean forecast made in period t which coincides with $E[y_{t+h}|I_t]$ only in the case of mean forecast optimality. The variable $\hat{e}_{t+h|t}$ is the forecast error of the forecast $\hat{y}_{t+h|t}$. It should be stressed that $\sigma_{t+h|t}^2$ depends on the mean forecast $\hat{y}_{t+h|t}$, implying that all evaluations of forecasts for $\sigma_{t+h|t}^2$ are conditional on $\hat{y}_{t+h|t}$. The forecast for $\sigma_{t+h|t}^2$ is denoted by $\hat{\sigma}_{t+h|t}^2$ In Clements (2014) $\hat{\sigma}_{t+h|t}^2$ is labelled ex-ante uncertainty, while $\sigma_{t+h|t}^2$ is referred to as ex-post forecast uncertainty, because it cannot be assessed before y_{t+h} . If, for instance, the forecast density is normal, the ex-ante uncertainty determines the width of a 95 % interval as $\pm 1.96 \hat{\sigma}_{t+h|t}^2$, and the forecaster expects the forecast error $\hat{e}_{t+h|t}^2$ to lie in this interval with a probability of 0.95. Independently of the distribution of the forecast density, $\hat{\sigma}_{t+h|t}^2$ is the forecaster's prediction for the expected squared forecast error $E[\hat{e}_{t+h|t}^2|I_{_t}]$. The squared forecast error $\hat{e}_{t+h|t}^2$ itself is a noisy measure for the unobservable ex-post forecast uncertainty $\sigma_{t+h|t}^2$, with their relationship being given by $\hat{e}_{t+h|t}^2 = \sigma_{t+h|t}^2 + v_{t+h}$ with $E[v_{t+h}|I_t] = 0$. One approach to forecast uncertainty or future variance is to use a simple Bayesian autoregressive

(AR) model

$$y_{t}=c+\sum_{i=1}^{p}\theta_{i}y_{t-i}+u$$
 (2)

with u_t iid $\mathcal{N}(0,\sigma_u^2)$ and uninformative priors, with a uniform prior on $c,\theta_1,\theta_2,\dots,\theta_p$ and a Jeffrey's prior on σ_n^2 With y_t denoting the last available observation, samples from the joint predictive distribution of $(y_{t+1|t}, y_{t+2|t}, \dots, y_{t+H|t})$ can be generated employing the algorithm described in Karlsson (2013). For example, one can use the variance

$$\sigma_{t+h|t}^2 = var[y_{t+h|t}] \tag{3}$$

as the uncertainty forecast. Each time the model is estimated, the lag length p will be determined by the BIC criterion, with the largest value of p considered being equal to a pre-specified limit. It may be of interest, that the Diebold et al. (1997) test can be used to test for equal accuracy of two competing forecasts.

Another approach is to introduce explicit representatives of uncertainties in econometric models. Building on preliminary work by Howrey (1978), Cunningham et al. (2012b) present a model which is built on the fact that statistical agencies revise their initial publication in the light of newly available information or methodological advancements. It follows that, as long as revisions tend to improve data estimates — i.e. moving them closer to the truth — the problem boils downto predicting the cumulative impact of revisions on the latest estimates of current and past activity. The model draws on the revisions history to proxy the uncertainty surrounding the latest published estimates. For example, it establishes the extent to which prior views on economic activity should evolve in light of new data and any other available measures, such as business surveys. The model produces estimates of the 'true' value of the variable of interest, a backcast, that can be used as a cross-check of the latest published official data, or even to substitute for those data in any economic applications. Here we are not going to provide a detailed description of the model, but we only concentrate on how uncertainty is quantified within the model at this point.

Let the m dimensional vector of variables of interest that are subject to data uncertainty at time t be denoted by \mathbf{y}_t , t=1,2,...,T. The vector \mathbf{y}_t contains the unobserved true value of the economic concept of interest. We assume that the model for the true data \mathbf{y}_t is given by

$$\mathbf{y}_{t} = \mathbf{\mu} + \sum_{i=1}^{q} \mathbf{A}_{i} \mathbf{y}_{t-i} + \mathbf{\varepsilon}_{t}, \tag{4}$$

where $\mathbf{A}_1,...,\mathbf{A}_q$ are m x m matrices, $\mathbf{A}(L)=\mathbf{I}-\mathbf{A}_1L-...-\mathbf{A}_qL^q$ is a lag polynomial whose roots are outside the unit circle, $\mathbf{\mu}$ is a vector of constants, $\mathbf{\epsilon}_t=(\epsilon_{1t},...,\epsilon_{mt})$, and $E[\mathbf{\epsilon}_t,\mathbf{\epsilon}_t]=\mathbf{\Sigma}_{\epsilon}$ where the main diagonal of $\mathbf{\Sigma}_{\epsilon}$ is denoted by $\mathbf{\sigma}_{\epsilon}=(\sigma_{\epsilon_1},...,\sigma_{\epsilon_m}^2)$ and $\mathbf{A}_1,...,\mathbf{A}_q$ are diagonal.

Let \mathbf{y}_t^{t+n} denote a noisy estimate of \mathbf{y}_t published by the statistical agency at time t+n, where n=1,2,...,T-t. The model for these published data is

$$\mathbf{y}_{t}^{t+n} = \mathbf{y}_{t} + \mathbf{c}^{n} + \mathbf{v}_{t}^{t+n}, \tag{5}$$

where \mathbf{c}^n is the bias in published data of maturity n and \mathbf{v}_t^{t+n} the measurement error associated with the published estimate of \mathbf{y}_t made at maturity n. One of the main building blocks of the model is the assumption that revisions improve estimates so that official published data become more accurate as they become more mature. Reflecting this assumption, both the bias in the published estimates and the variance of measurement errors are allowed to vary with the maturity of the estimate, as denoted by the n superscript. The constant term \mathbf{c}^n is included to permit consideration of biases in the statistical agency's dataset. Specifically, it is modelled as

$$\mathbf{c}^{n} = \mathbf{c}^{1} (1+\lambda)^{n-1},\tag{6}$$

where c^1 is the bias in published data of maturity n=1 and λ describes the rate at which the bias decays as estimates become more mature $(-1<\lambda<0)$. This representation assumes that the bias tends monotonically to zero as the estimates become more mature. Moreover, the model assumes that the measurement errors \mathbf{v}_t^{t+n} are distributed normally with finite variance. The model also allows for serial correlation in \mathbf{v}^{t+n} . Specifically, the serial correlation in the errors attached to the data in any release published at t+n are modelled as

$$\mathbf{v}_{t}^{t+n} = \sum_{i=1}^{p} \mathbf{B}_{i} \mathbf{v}_{t-i}^{t+n} + \mathbf{\varepsilon}_{t}^{t+n}, \tag{7}$$

where \mathbf{B}_i are $m \times m$ matrices, $\mathbf{B}(L) = \mathbf{I} - \mathbf{B}_1 L - \ldots - \mathbf{B}_p L^p$ is a matrix lag polynomial whose roots are outside the unit circle, and $\mathbf{\epsilon}_t^{t+n} = (\mathbf{\epsilon}_{1t}^{t+n}, \ldots, \mathbf{\epsilon}_{mt}^{t+n})^{'}$ and $E[\mathbf{\epsilon}_t^{t+n}(\mathbf{\epsilon}_t^{t+n})^{'}] = \mathbf{\Sigma}_{\epsilon}^n$ as the models allows for heteroscedasticity in measurement errors with respect to n. This imposes some structure on \mathbf{v}_t^{t+n} because the model assumes a finite AR model whose parameters do not depend on maturity. The representation picks up serial correlation between errors attaching to the various observations within each data release. Furthermore, the model assumes that $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are diagonal. Moreover, it is assumed that $\mathbf{\epsilon}_t^{t+n}$ and therefore \mathbf{v}_t^{t+n} display heteroscedasticity with respect to n. Specifically, the main diagonal of $\mathbf{\Sigma}_{\epsilon}^n$ is modelled as $\mathbf{\sigma}_{\epsilon}^2 \mathbf{e}_1^{-(\sigma_{\epsilon}^n n, \ldots, \sigma_{\epsilon}^n)}$, where $\mathbf{\sigma}_{\epsilon}^2 \mathbf{e}_i^{n+n} = \mathbf{E}[(\mathbf{\epsilon}_{it}^{t+n})^2]$. The model for $\mathbf{\sigma}_{\epsilon}^2 \mathbf{n}$ is given by

$$\sigma_{\mathbf{\epsilon}n}^2 = \sigma_{\mathbf{\epsilon}1}^2 (1+\delta)^{n-1},\tag{8}$$

Where $\sigma_{\epsilon 1}^2$ is the variance of measurement errors at maturity n=1 and δ describes the rate at which

variance decays as estimates become more mature $(-1 < \delta < 0)$. This representation imposes structure on the variance of measurement errors, because it assumes that the variance declines monotonically to zero as the official published estimates become more mature which is supported by Kapetanios and Yates (2004b). Further information and empirical implementation are provided in Task-03.

Moreover, as mentioned in section 2.1, the measurement of error may also change due to an increased sample size, and therefore national statistics often times revise previous releases as more data becomes available. As already mentioned, the main focus of this report is on economic statistics but, whenever possible, we try to provide a wider view, looking also at other branches of statistics, especially social statistics.

5 Communication

5.1 Importance of the communication of uncertainty

As mentioned in the introductory section 1, the communication of uncertainty in economic and social statistics is of fundamental importance. To illustrate the reason whythis is so, one could consider the opposite case, namely a situation in which official reports or similar publications intentionally neglect uncertainties and therefore do not communicate their existence at all. Such a scenario would lead to wrong conclusions about the certitude of the matter. This in turn may lead to subsequent political, financial, social or economic decisions which are not correctly factoring in those uncertainties and consequently may lead to increased risk. Even more so, it would raise questions such as why was a certain statistic revised after some time if it was communicated with certitude in the first instance? and therefore may cause distrust in future publications.

Thus, communicating uncertainty explicitly provides a clearer picture of the matter and is in the interest of all stakeholders. There are not only benefits for the *consumer* of economic and social statistics (more holistic picture) but also for the publishing body itself, as the clear communication of uncertainty enables the publishing body to reduce its exposure to criticism, enables the wider public to engage in a more meaningful interpretation and debate, and promotes transparency in general.

However, the motivation for why the communication is important may depend on the specific publishing authority or agency. For example, central banks may have an interest in controlling market volatilities while national statistical agencies may strive for genuine accuracy and consequently may have little political motives.

Moreover, through the communication it must also be clear that there is a fundamental difference between communicating uncertainty which may be inherent in a phenomenon and communicating uncertainty which relates to the measurement of a phenomenon. Or, in other words, has the matter of interest an intrinsic uncertainty, or is simply the measurement of the matter uncertain? It follows that a clear communication of uncertainty should tackle all the issues mentioned above and also make clear to the *consumer* of economic and social statistics that measuring uncertainty is inherently different from communicating it. In the following we concentrate on the communication aspect, as previous sections introduced concepts of the measurement.

5.2 Specific communication tools

The communication of uncertainty depends on what kind of uncertainty one is concerned about. In general, one can differentiate between the two fundamental levels of uncertainty: direct and indirect, as Bles et al. (2019) argues. Direct uncertainty refers to the uncertainty about a fact, number or scientific hypothesis. It follows, that it can be communicated either in absolute quantitative terms, for instance a probability distribution or confidence interval, or even expressed relative to alternatives, such as likelihood ratios, or given an approximate quantitative form, verbal summary and so on. On the other hand, indirect uncertainty refers to the uncertainty in terms of the quality of the underlying knowledge that forms a basis for any claim about a fact, number or hypothesis. This will generally be

communicated as a list of caveats about the underlying sources of evidence, possibly amalgamated into a qualitative or ordered categorical scale.

One can further categorise the different forms of communication for direct uncertainty, as done by der Bles et al. (2019). Those categories include full explicit distribution, summary of a distribution, rounded number range or an order of magnitude assessment, predefined categorisation of uncertainty, qualifying verbal assessment, list of possibilities and scenarios, informally mentioning the existence of uncertainty, no mentioning of uncertainty, explicit denial of existing uncertainty.

Relative uncertainty about competing hypotheses or values for a measure can also be expressed in different forms. Verbal comparisons include statements of the form 'A is more likely than B', while numerical expressions include likelihood ratios for comparing facts and scientific hypotheses, likelihood functions for relative support for different numbers, and comparative measures of model adequacy, such as the Akaike Information Criterion or Bayesian Information Criterion. P-values are a measure of conflict between data and a hypothesis, and are certainly not direct expressions of a probability of hypotheses. However, in many circumstances they correspond to a specific confidence interval for a numerical parameter.

Methods for communicating the quality of the underlying evidence do not give quantitative information about absolute values or facts, but summarize the subjective confidence we have in any claim. In order to attempt to assess indirect uncertainty, a number of fields have established checklists to try to assess the quality of evidence in as objective a way as possible. These may relate to either an individual claim, such as the CONSORT system, for determining the characteristics of the claims resulting from a randomized controlled trial, and the Maryland Scale of Scientific Methods, for determining the strength of a crime prevention study, or the totality of evidence, attempting to take into account the quality, quantity and consistency of multiple studies to give an overall assessment of the confidence we can have in a particular assertion; see West et al. (2002) or van der Bles et al. (2019) for reviews These tools provide the basis for systems that attempt to communicate overall quality of evidence (although the distinction between methods of assessment and methods of communication of indirect uncertainty is rarely made).

Many methods of communicating indirect uncertainty have been developed in different fields. Limitations in the underlying evidence might be summarized by qualitative verbal caveats, or an ordered set of categories (which may be communicated numerically, graphically or verbally). For example, the GRADE Working Group has established a scale for communicating the quality of the evidence underlying claims about the effects of medical interventions, which ranges from 'Very low quality', graphically represented as a single plus symbol and/or circle, to 'High Quality', graphically represented as 4 plus symbols and/or circles, see Balshem et al. (2011). Other examples are the 'padlock' ratings used by the UK Educational Endowment Foundation, or the US National Intelligence Council's recommendation that intelligence analysts provide a qualitative assessment of analytic confidence on a high/medium/low scale. In effect, such ordered scales provide a form of 'star-rating' for the conclusions.

In van der Bles et al. (2019) it is argued that uncertainty can be expressed in one (or a combination) of three different formats: visual, numerical and/or verbal. The appropriate format in part depends on the medium of communication, which might be written and printed official reports, online websites, smart phone applications, print media, television, or spoken in person or on the radio. They therefore consider these two aspects of format and medium together. However, these different formats have the potential to carry different levels of information and therefore choosing one is not simply a design choice — it can influence the type of expression of uncertainty available and its potential effect on the audience

Whereas numerical (numbers) and verbal (words) communication are relatively constrained in their design, there are a variety of ways to communicate uncertainty visually. Examples of common ways to visualize epistemic uncertainty around a number, expressed as an estimate with a range, include error bars, diamond, violin or fan plots as well as density strips. Error bars are widely used in scientific and other publications to illustrate the bounds of a confidence interval, but provide no indication of the underlying distribution of the number. Other visualizations attempt to give an (approximate) idea of this

underlying distribution: for example, diamonds, which are often used when considering treatment effects in a medical meta-analysis, or violin plots, which are designed to give a more accurate idea of the underlying distribution. Fan plots are designed to show the bounds of several different confidence intervals (often coloured to emphasize the changing probability density going further from the point) and are used, for example, by the Bank of England when communicating past and forecasted future GDP estimates. Finally, density strips are the most accurate representation of the underlying probability distribution around the point estimate.

Communicating Uncertainty in Practice

The previous sections of this report gave an overview over models which can be used to quantify inherent uncertainties across various disciplines, and tools to communicate the resulting uncertainties. In this section, we first provide a synthetic and non exhaustive picture of ongoing activities in a number of statistical offices part of the European Statistical System (ESS). Then we present a number of examples illustrating how uncertainties are communicated to the respective audiences by other main public agencies, such as the Federal Reserve Bank of the USA, the Bank of England, the UK Office for National Statistics (ONS), the Swedish central bank (Riksbank). Finally, we discuss some experiences in other disciplines.

6.1 Practices in the European Statistical System

OVERVIEW

We start with a synthetic and non exhaustive picture of ongoing activities related to communicating uncertainty in a number of statistical offices within the ESS. The discussion is based on short reports provided by these institutions, following direct contact established by the research team. As already mentioned, the initiatives mentioned here do not display a comprehensive picture, but should only be read as examples of current practices.

In general, it turns out that statistical institutes document their statistics through a framework comprising information about the production process, its quality and the quality of the output, including sources of errors and uncertainty. Except for sampling surveys, for which variances and standard errors are generally calculated, the information about the sources of errors and uncertainty is rather descriptive and not shown directly in the tables or figures disseminated. The quality and meta-data reports, and the related meta-data structure (i.e. SIMS) were mentioned by ESS statistical offices as main framework for explaining and communicating statistics uncertainty, at a descriptive level.

Although no specific initiatives or projects on measuring and communicating uncertainty on time series data were reported, statistical offices acknowledged the importance of measuring and communicating uncertainty for time series like seasonal adjustment, forecasting and flash estimates, but also more in general in the current production of official statistics. For instance, in national accounts, but also for other short-term statistics, the statistical production process may be seriously affected by an erroneous approach by not considering model uncertainty. Theoretical references were suggested in this regard, namely: Chatfield (1995), JRSS series A, 158, Part 3, pp.419–466 Model Uncertainty, Data Mining and Statistical Inference, but also a research initiative by Statistics Norway (von Brasch et al, 2017) on Productivity dispersion and measurement errors.

THE CASE OF CENTRAAL BUREAU VOOR DE STATISTIEK (CBS, NETHERLANDS)

It is worth mentioning the recent initiative of CBS Netherlands who developed a Primary Observation Quality Report Manual including measures targeting the mitigation of sources of uncertainty. Four main

dimensions are distinguished in the quality report: concepts, representation, measurement and output (Groves et al, 2009). For each dimension, two or three indicators were identified, which can be complemented by questions that provide users with additional information on which they can assess other quality aspects. The proposed questions and indicators are not separate from each other, but the indicators are the primarily signals of a risk of inferior quality, with the focus of the quality manual being on systematic errors.

Furthermore, work carried out in CBS proposes a method to incorporate uncertainty measures for population forecasts. It is argued, that it has long been realized that deterministic population forecasts are of limited use without an indication of their uncertainty. A deterministic point forecast only gives the most probable value of, for instance, the population size in 50 years time, but no information on how close to this value the actual population size is likely to be. Stochastic techniques for estimating uncertainty intervals for demographic forecasts are still not widely used by statistical institutes. One reason for this may be the complexity of the Monte Carlo approach, which requires the calculation of a thousand or more variants of the deterministic forecast. CBS work is based on a technique to estimate forecast intervals for demographic indicators from only 6 variants. Good agreement is found between the forecast intervals obtained with this technique and from a stochastic forecast. This method does not provide a full probability distribution for the forecasts, however. Also, it cannot be used to compute intervals for demographic flows, although intervals for time-cumulated flows can be obtained.

Despite this long history, stochastic techniques are still not widely used in official population forecasts. Most statistical agencies produce only deterministic forecasts. An exception is Statistics Netherlands, which has been publishing stochastic forecasts since 1998. More recently, the United Nations has started publishing probabilistic forecasts. Their 2012 en 2015 forecasts take into account uncertainty in fertility and mortality, but not in international migration.

One reason the official population forecasts are still largely deterministic may be the perceived complexity of the stochastic approach. CBS has been investigating if it is possible to reproduce the main results of stochastic forecasts, forecast intervals for stock variables such as number of residents or grey pressure, using the familiar variant approach. This is done by adding two elements. First a method to construct time paths for the input assumptions of high and low fertility-, mortality- or migration variants in such a way that their output range for stock variables is similar to the confidence interval obtained from a stochastic forecast (in which only the uncertainty in the particular component is taken into account). Second: a method to combine the results from the 6 fertility, mortality and migration variants into a single, quasi stochastic, uncertainty interval.

Both methods seem useful to determine uncertainties in population forecasts. In particular, extensive Monte Carlo simulations can be used to determine 67 % confidence intervals.

Unlike a real stochastic forecast, the quasi stochastic approach does not provide a full probability distribution for the forecast results. Also, it cannot be used to compute intervals for demographic flows, but intervals for time-cumulated flows can be obtained. The agreement with the stochastic intervals is better for larger age groups and for shorter forecast horizons. Also, it is better for population counts than for ratios of counts. The method works equally well for 67 % and 95 % forecast intervals. An area where this approach can be useful is in more complex forecasts with many degrees of freedom or time consuming matching mechanisms, like subnational forecasts which employ housing market models. For this category of models, a stochastic forecast can be unfeasible, because it takes too long to run the simulations thousands of times. The quasi stochastic method can then be used to estimate approximate forecast intervals from a limited number of variants. The method was used in this way for the 2016-based regional population and household forecast of Statistics Netherlands and the Netherlands environmental assessment agency. So to summarise, CBS also uses visualisations derived from simulations to communicate uncertainty.

With respect to statistical institutes, it seems that the discussion about formal communication of uncertainty has been more relevant in central banks. Actually, as pointed out by Knüppel and Schultefrankenfeld (2019), central banks already reported measures of forecast uncertainty long before the Great Recession, but their motivation for doing so is regarded as different from the recent

motivation in academia. While academics mainly appear to be interested in the effects caused by changes in uncertainty, like in Bloom (2009), several central banks have mainly intended to communicate that future realizations should not be expected to coincide exactly with the central banks' point forecasts. Central banks like the European Central Bank or the Deutsche Bundesbank resort to reporting unconditional measures of forecast uncertainty for this purpose. Others, however, have tried to provide additional information by publishing conditional measures, in line with the idea of Jurado et al. (2015) who suggests relating macroeconomic uncertainty to conditional forecast uncertainty, i.e., to the conditional volatility of the unforecastable component of macroeconomic time series.

From previous sections, it becomes apparent that central banks or statistical agencies should have an interest in unambiguously communicating the fact that rapid estimates and forecasts inherently carry uncertainty with them. However, forecasts of central banks are often only considered as point forecasts, capturing expected future values of the target variables. Especially the central banks' point forecasts for inflation frequently turn out to outperform other competing forecasts in terms of mean-squared forecast errors, as documented, for instance, in Groen et al. (2009) and Faust and Wright (2009). Yet, if the central bank does not have a quadratic loss function, these results must be interpreted with caution, because then, in general, point forecasts do not represent the mean of the forecast density. In fact, empirical studies often find that several characteristics of central banks' forecasts are suggestive of more complicated loss functions, featuring asymmetry, state dependence, or time-variation, like in Capistrán (2008), Wang and Lee (2014), or Patton and Timmermann (2007).

Given the existence of many potential loss functions, it might actually be preferable to issue (and to evaluate) density forecasts. Each user can then infer the point forecast corresponding to her loss function from this density forecast, and the evaluation can rely either on the entire density or only on elements of interest. The evaluation of density forecasts is sued by central banks has mostly focused on the Bank of England, with important contributions by Clements (2004), Wallis (2004), and Mitchell and Hall (2005). The asymmetry incorporated in central banks' density forecasts is studied in Knüppel et al. (2012). However, the dispersion of central banks' forecast densities, i.e., uncertainty forecasts, has not been investigated holistically yet. Similar considerations apply to all kind of rapid or preliminary estimates of statistical indicators having time-series representations (economic statistics with particular focus on macro-economic ones).

We now present a number of examples illustrating how uncertainties are communicated to the respective audiences by a few public agencies, starting with the Bank of England because of their mentioned rather advanced approach.

6.2 Bank of England

A well-known and well-established tool to communicate uncertainty through visualization in the context of economic forecasting is the *fan-chart*, which is used in the Inflation Report by the Bank of England. Since February 1996 the Bank has applied this tool with regard to forecasting inflation, and since November 1997 it has also been applied to visualize GDP growth backcasts, nowcasts (flash estimates), and forecasts.

Since February 1996, the Bank's inflation forecast has been published explicitly in the form of a probability distribution — presented in what is now known as the *fan-chart*. As pointed out by Britton et al. (1998), the aim of the fan chart has been to convey to the reader a more accurate representation of the Bank's subjective assessment of medium-term inflationary pressures, without suggesting a degree of precision that would be spurious. An exemplary visualization is displayed in figure 1.

26

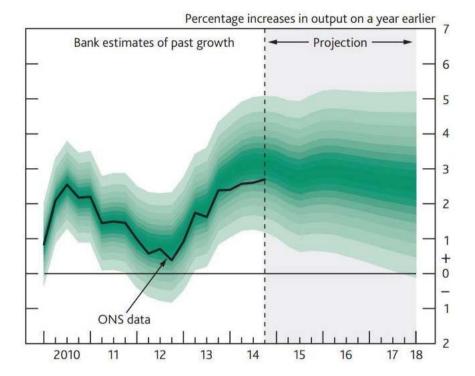


Figure 1: Exemplary Bank of England fan-chart, e.g. see Bank of England (2019)

By doing so, the fan-chart portrays a probability distribution — opposed to a point estimate — which approximates the Bank's subjective assessment of economic developments. The motivation to introduce the fan-chart naturally arose from the fact that the Bank of England introduced an inflation-targeting regime for UK monetary policy in 1992. This, therefore, placed much larger importance on forward-looking inflationary pressure, which inevitably is best described by producing and presenting an explicit economic forecast. Moreover, it was hoped that the fan-chart would promote further discussions of the risks to the economic outlook, and thus that the chart would contribute to a wider debate about economic policy. The fan-chart helps to make it more clear that monetary policy is about making decisions under uncertainty, and that the Monetary Policy Committee (MPC) does not pretend to know with certainty the exact economic figures in two years' time. The process used to produce the fan chart has also had a major impact on the Bank's approach to forecasting. The process forces the MPC to consider not just a single possible outcome for the economy, but a range of possibilities in areas where the central view is most likely to be wrong. In turn, this should promote better economic analysis of the underlying issues, and a necessary focus on the shocks hitting the economy.

So how is the fan-chart derived? There is a series of meetings between the MPC and the Bank staff. At the first meeting, roughly a month before Report publication, the key assumptions, the main issues and the starting-point for the risk assessment are discussed. At this stage, no forecast is presented. Following this meeting, the forecast team maps the decisions of the MPC onto a central projection and risk distribution. As econd meeting with the MPC considers this draft forecast. The quantification of the mapping from each assumption and risk assessment is reviewed, new data are incorporated and changes are requested. Athird meeting gives the MPC an opportunity to fine-tune the revised forecast distribution and bring it up to date. The final forecast, published in the Report, includes adjustment in response to the advent of market-related data in the period up to the relevant monthly MPC meeting, and reflects any change in interest rates made by the Committee in that meeting.

The fan-chart portrays a probability distribution that approximates the MPC's subjective assessment of inflationary pressures evolving through time, based on a central view and the risks surrounding it Whatever the mix of judgment and statistics used in this assessment, the process needs to be as rigorous as possible: the MPC needs to be able to explain exactly why the chart looks as it does, and why it changes between Reports. This is vitally important both for the consistency of policy-making

and for the presentation of the analysis. For any particular forecast, one can think of the projection as being based on a model that maps choices about economic assumptions onto an inflation forecast. To generate the probability distribution, one would ideally like to evaluate all the possible alternative assumptions numerically using this model. In practice, this process is approximated by assuming a known functional form for the distribution and evaluating a limited number of alternative assumptions. These alternatives are sufficient to calibrate the key parameters of the distribution.

One might perceive the possible outcomes for inflation or GDP as being roughly symmetrically dispersed around a central, most probable value, with the values closer to the centre being more likely than those further away. That would suggest that the forecast distribution should be based on the normal 'bell-shaped' distribution widely used in statistical analysis. However, the assessment of likely alternative outcomes sometimes suggests that forecast error is more likely to be in one direction than the other. This has led to the choice of a particular form of statistical distribution (a 'two-piece' normal), which has a degree of asymmetry in the form of a variable skew. Mathematically, the distribution can therefore be summarised as

$$pdf = \frac{2}{((1/\sqrt{1-\gamma}) + (1/\sqrt{1+\gamma}))} \frac{1}{2\pi\sigma^2} e^{\left[\frac{-1}{2\sigma^2}((x-\mu)^2 + \gamma(\frac{x-\mu}{|x-\mu|})(x-\mu)^2)\right]},$$
 (9)

with $_{\mu}$ being the mean of the distribution, σ^2 its variance, while γ is a parameter accounting for the skewness which lies between -1 and +1. Finally, x is a normally distributed random variable.

Therefore, it follows that to derive the MPC's forecast distribution, three parameters need to be evaluated. First, a measure of the central tendency for the economic variable of interest (e.g. inflation or GDP growth) — usually expressed as a particular projected path. Second, a view on the degree of uncertainty— the variance. Third, a view on the balance of the risks, to get a measure of the skew.

No single projection of inflation at a future date has much chance of matching the subsequent outcome. Policy discussions need to take account of the full range of possibilities. The Bank's preference has been to start with a set of key assumptions consistent with the most likely view of developments in the economy. The central projection of inflation is then interpreted as being the 'mode' of the statistical distribution-it is the single most likely outcome based on current knowledge and judgment, even if the actual chance of it matching the eventual outcome is small. This central projection is based on a consistent set of assumptions about economic behaviour that provide the foundations for subsequent assessment of how the economy is evolving relative to the forecast. If the risks are symmetrically distributed around the central view, this will also provide a view of the average outcome (or mean forecast). But when the risks are unbalanced, the average of all the alternatives is unlikely to be the same as the single most likely case, and the mean forecast will differ from the mode. We return to this point with an example when discussing the balance of risks (the skew of the distribution).

Neither the most likely nor the average view will necessarily split the forecast distribution in half. The point in the distribution that has 50 % probability on either side is known as the median. The relationship between the mode, median and mean is important in interpreting the fan chart. The central projection represents a mapping of the central assumptions onto an inflation projection, using an economic model. In order to understand the issues of particular relevance in any one forecast round, the MPC considers several different ways of looking at the economy before selecting the set of relationships — or model — that represent its view for that forecast round. Hence, the econometric model used to ensure consistency of variables in the forecast is not set in stone, but changes from one forecast to the next.

The uncertainty in the subjective assessment of inflation relates to how likely it is that the future events will differ from the central view. It is therefore a forward-looking view of the risks to the forecast, not a mechanical extrapolation of past uncertainty. Nevertheless, the initial calibration of uncertainty is based on the experience of forecast errors from the previous ten years. So the fan chart approach then requires the MPC to form a view as to whether or not uncertainty looking forward is greater or less than in the past. The degree of uncertainty (the degree of dispersion in the distribution) can be measured by a variety of statistics such as variance, mean absolute error or inter-quartile range. The

Bank uses a variance measure. The variance of inflation can be derived from the underlying variances of the basic shocks, using the mapping provided by the economic model. To make this tractable, simulations are used to identify the contribution of the relevant basic variances to the variance of the inflation forecast. For independent shocks, the inflation forecast variance could be obtained by a weighted sum of the individual variances. But rather than add up all the variances, the past inflation forecast error variance is taken as a starting point and then adjusted upwards or downwards, based on changes to a limited number of variance assumptions. By adjusting the basic variances, the forecast variance of inflation is thus changed to match the degree of uncertainty as viewed by the MPC.

Only one number is needed to summarise the degree of skewness (the balance of risks). Just as with the central view and the degree of uncertainty, there is more than one possible choice of parameter. The Bank's analysis focuses on the difference between the mean and the mode of the forecast distribution to be presented in the Report. This difference is of interest as a summary statistic of the balance of risks, and it provides a practical way of calibrating the distribution.

Once all three parameters are gathered, the fan-chart can be constructed: The skew and variances are evaluated for one year ahead and two years ahead projections and then interpolated for the quarterly fan chart. This gives the distribution to plot. So what does all of this mean in practice?

The distribution to be plotted is generated by an iterative procedure, given the central projection as a mode, the variance and balance of risks. The chosen distribution is adjusted until the required variance and the required difference between mean and mode are obtained, and appropriately adjusted to ensure that the probability area is equal to one, as required by a probability distribution. The skew and variances are evaluated for one-year-ahead and two-year-ahead projections, and then interpolated for the quarterly fan chart. This gives the distribution to plot. Having obtained the forecast distribution for inflation at each point up to nine-quarter-ahead, its graphical representation remains a matter of choice. The fan chart was chosen to meet the criterion that it should give information on the whole of the forecast distribution, without claiming a spuriously high degree of precision.

The fan chart itself is best understood by interpreting it as corresponding bird's-eye view of the underlying probability density function (pdf) for a period. The height of the pdf is proportional to the probability of inflation being a particular value in that time period. Hence, the central projection corresponds to the peak of the distribution, as it is associated with the mode. The style of the chart is to make it reflect the relative probability of inflation lying in a particular band.

To draw the bands, the following rule is used. Two points of equal probability density are shown, one on either side of the mode. The two points are then moved away from the centre simultaneously, keeping the values of the probability density the same, until there is 10 % of the distribution in a single central band, with these two points marking the outside edges. That band is coloured the deepest shade of red. The two points are moved outwards again on either side of the first band (still keeping equal probability density) until another 10 % of the distribution has been added, this time marking a pair of bands, one on either side of the centre. These two bands are shaded the same colour as each other, but are lighter than the central band. Pairs of bands continue to be added until 90 % of the distribution is covered. The fan chart always has the following features. There is an equal number of red bands on either side of the central band (eight). Each pair of bands covers 10 % of the distribution but, if the risks are unbalanced, the same colour bands are not of equal width (representing unequal probability intervals). The distribution is truncated, so that there is an implicit ninth and final pair of bands, occupying the white space outside the 90 % covered. The central projection is, by construction, always in the deepest red band since it is associated with the mode. For heavily unbalanced risks, the mean and median may not be in the deepest red band, however.

There are a few common misconceptions about the fan-chart which are worth mentioning. First, as explained above, the 'fan' does not cover 100 % of the probability. Second, the central projection (the mode) is not necessarily the centre of the deepest red ban — although it is always within it and is usually close to the centre. Third, though the fan chart could be used to represent a forecast distribution generated by purely statistical methods, such as stochastic simulation of a model, the Bank's approach is to represent a subjective distribution for its inflation projections based on economic analysis and the judgment of the MPC.

6.3 FED / FOMC

In the United States, the Federal Open Market Committee (FOMC) of the Federal Reserve Bank (FED) uses a variety of means to communicate uncertainty. A key aspect of FOMC communication is the explicit publication of the policymakers' range of uncertainty about the future path for the policy rate. It can be found near the end of the FOMC minutes, published three weeks after the Summary of Economic Projections (SEP) release.

In the past few years, the two key policy questions for the FOMC have been: What should be the normal level of interest rates? And, how quickly should we get there? The SEP usefully reflects the variety and evolution of FOMC participants' responses to these two questions. Following President Evans, the contrast is between Odyssean and Delphic communication — where Delphic communications as those associated with a well-functioning, well-understood monetary policy framework while Odyssean communications arise when unexpected events expose weaknesses and shortcomings in a Delphic framework, e.g. see Campbell (2013). The more the policy rate path depends on the evolution of the economy, the greater is the unavoidable uncertainty. Today, not only is that uncertainty substantial, but it is just as important to communicate it as the interest rate commitment was half a dozen years ago. This brings us to an important, but largely overlooked element of the FOMC's communication framework; historical projection error ranges that are now included in the FOMC minutes three weeks after the initial SEP release. Based on the work of Federal Reserve Board economists David Reifschneider and Peter Tulip, Table 2 in the minutes reports estimates of error ranges (measured as the root-mean-squared prediction error) for projections of real GDP growth, the unemployment rate, inflation, and the short-term interest rate. An example from the minutes from March 2019 is shown below.

To see how informative these error ranges are about the uncertainty associated with the FOMC projections, consider the version from the March 19–20, 2019 meeting minutes. It shows that, for the unemployment rate, the median projection for 2021 is 3.9 percent, with an error range of plus or minus 1.7 percentage points. This tells us that, given historical experience and assuming normality and a 70 percent confidence interval, there is a 70 percent chance that in 2021 the unemployment rate will be between 2.2 and 5.7 percent. For inflation, the median is 2.0 percent, with an error range of plus or minus 1.1 percentage points, so the confidence interval goes from 0.9 to 3.1 percent. For GDP growth, the median projection is 1.8 percent, with a standard error of 1.9 percent — that is, the 70 percent confidence interval is from -0.1 to +3.7 percent. Uncertainty regarding the future level of unemployment and inflation (and real growth) translates directly into uncertainty about the path of future policy. Here, again, the FOMC is remarkably clear about the degree of imprecision. In March 2019, the error range for the 2021 projection of the short-term interest rate is plus or minus 2.5 percentage points. Given the median projection of 2.6 percent, this means that the Committee believes there is a 70 percent chance that, in 2 years, the target interest rate will be between 0.1 and 5.1 percent.

The figures above highlight a potential problem when communicating not only the point forecast or flash estimate but also a measure of uncertainty around it: the resulting interval forecast can be so wide as to make it virtually uninformative. However, while the statistician should try to make the forecast interval as short as possible by enhancing the forecast efficiency, not communicating the uncertainty around the point forecast would simply misguide the decision makers about the reliability of the forecasts. If there is substantial uncertainty, it is better to say so and let the decision makers decide whether or not to base their decisions on this forecast.

Figure 2: Exemplary FED projections, e.g. see Federal Open Market Committee (2019)

Table 2. Average historical projection error ranges

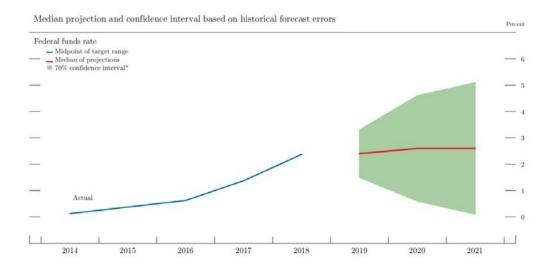
refeelinge points					
Variable	2019	2020	2021		
Change in real GDP ¹	±1.4	±1.9	±1.9		
Unemployment rate ¹	±0.5	±1.3	± 1.7		
Total consumer prices ²	±0.9	± 1.0	±1.1		
Short-term interest rates ³	±0.9	± 2.0	±2.5		

NOTE: Error ranges shown are measured as plus or minus the root mean squared error of projections for 1999 through 2018 that were released in the spring by various private and government forecasters. As described in the box "Forecast Uncertainty," under certain assumptions, there is about a 70 percent probability that actual outcomes for real GDP, unemployment, consumer prices, and the federal funds rate will be in ranges implied by the average size of projection errors made in the past. For more information, see David Reifschneider and Peter Tulip (2017), "Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve's Approach," Finance and Economics Discussion Series 2017-020 (Washington: Board of Governors of the Federal Reserve System, February), https://dx.doi.org/10.17016/FEDS.2017.020.

- 1. Definitions of variables are in the general note to table 1.
- 2. Measure is the overall consumer price index, the price measure that has been most widely used in government and private economic forecasts. Projections are percent changes on a fourth quarter to fourth quarter basis.
- 3. For Federal Reserve staff forecasts, measure is the federal funds rate. For other forecasts, measure is the rate on 3-month Treasury bills. Projection errors are calculated using average levels, in percent, in the fourth quarter.

Since 2017, the FOMC has also published a figure following each SEP that helps visualize the uncertainty about the interest rate path, see Federal Open Market Committee (2019). It makes clear that, while the median suggests little change in the target path over the next 2-plus years (in red), there is considerable uncertainty that increases with the forecast horizon. An example is displayed in figure 3.

Figure 3: Exemplary FED Federal fund rates projections, e.g. see Federal Open Market Committee (2019)



Central banks that produce inflation reports typically use them both to communicate and to guide their deliberations. On the former, the reports have both a backward- and forward-looking function. Retrospectively, they provide an evaluation of how policymakers have performed, allowing parliamentarians, financial market participants and the public at large, to hold independent central bankers accountable for their actions. Prospectively, the reports provide projections to help people understand how policymakers will likely respond to changing conditions. Publication of the information in the reports also allows participants on a monetary policy committee to focus public attention on the changing issues that foster uncertainty. By creating accountability and transparency, inflation reports have a powerful influence on internal committee dynamics. The obligation to publish both an expected value and a range for projections of the state of the economy and policy focuses internal discussions. In particular, the need to reach a consensus significantly influences the nature of committee discussions.

While the SEP uncertainty table and fan charts strongly resemble what other central banks publish, the FOMC does not engage in the consensus building associated with the production of a Bank of England-like inflation report. The uncertainty measures that are described above are based on historical forecast errors, rather than an agreement among the FOMC members. Furthermore, unlike the Bank of England's nine-member Monetary Policy Committee, who all sit in the same building, it seems unrealistic to think that the 19-strong FOMC, spread among 13 locations from Boston to Atlanta to San Francisco, are prepared to forge a consensus four times per year to publish a Bank of England-style report.

6.4 European Central Bank

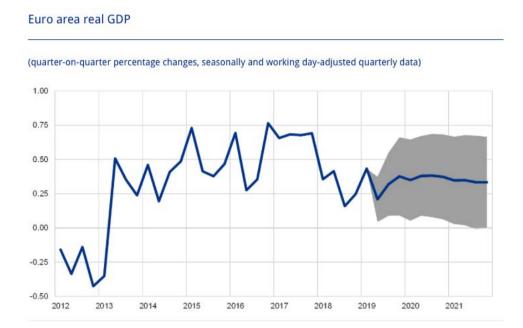
ECB/Eurosystem staff projections are presented in the form of ranges. The use of ranges acknowledges the inevitable uncertainty surrounding macroeconomic projections, see European Central Bank (2009). Various methods have been employed by ECB staff over time to compute these ranges. From the first publication of ECB/Eurosystem staff projections in December 2000 until the publication of Eurosystem staff projections in June 2008, the published projection ranges for each variable and horizon represented twice the mean absolute projection error constructed on the basis of an analysis of historical projection errors. These published ranges were derived using a short sample of projection errors, which had not been updated over time. Another, model-based method was used for the ECB/Eurosystem staff projections published from September 2008 to September 2009. In the meantime, the sample of projection errors has been updated and further analyses have been undertaken, on the basis of which a revised projection error-based approach to constructing ranges is now implemented, embodying the track record of both Eurosystem and ECB staff projections over a longer sample of years and a correction for outliers.

The width of the new Eurosystem and ECB staff projection ranges is derived from errors made in the projections conducted by Eurosystem and ECB staff in the past. More specifically, the mean absolute errors have been computed as the mean of the absolute differences between the projection for a given year and the realisation as available at the time of the June exercise of the following year. In computing the mean absolute errors, large errors are excluded from the sample of projection errors considered on the basis of a statistical outlier correction procedure. The published ranges have a width equal to twice the outlier-corrected mean absolute error.

The ECB/Eurosystem staff macroeconomic projections are conducted each quarter and annual growth rates are reported. As a result, the length of the projection horizon depends on the timing of the projection. For example, projecting the current year in March and June of the same year involves differing projection horizons and corresponding errors. Therefore, the width of the published projection range varies according to the timing of the projection exercise within a year. For all the published variables, four different sets of ranges have to be used over the course of the year.

Ranges are also specific to the variable considered. Some variables are less predictable and more volatile than others, thereby resulting in larger projection errors, and thereby ranges. This is the case particularly for investment and trade variables. The width of the ranges is consistent with a 57.5 % confidence interval for all variables. It is planned that the whole set of ranges will be updated every year at the time of the June projection exercise. An exemplary visualisation from June 2019 can be found in figure 4.

Figure 4: Exemplary ECB projection chart, e.g. see European Central Bank (2019)



However, there are also more indirect measures of uncertainty, as demonstrated by Sauter et al. (2013) who analyse the sentiment derived from official press releases and compare them to a benchmark. Their paper aims for an assessment of central bank communication by semantically analysing ECB press conference statements. The analysis differs from previously adopted approaches in that their indicator expresses the willingness of the central bank to communicate uncertainty to the public by explicitly making use of uncertainty-related words. The central bank is free to choose when and to what extent it wants to communicate uncertainty to the market. While this might be negligible during times of financial stability, it can be a decisive feature of central bank policy in times of economic turmoil such as the recent financial crisis, since communication can be a powerful tool to reduce market uncertainty.

In general, the level of uncertainty communicated by the ECB is closely linked to the market uncertainty measure VSTOXX, Sauter et al. (2013) find. However, the ECB's communicated uncertainty reacted in a timelier manner than the VSTOXX did before the crisis, then became substantially lower following the Lehman bankruptcy in late 2008. Sauter et al. (2013) argue that this is due to an overall appeasement approach by the ECB to reduce overall uncertainty in financial markets. Accorresponding pattern can be observed during the European sovereign debt crisis, in which communicated uncertainty was reduced around severe events that had the potential to destabilise financial markets. The findings thus point to a switch in the ECB uncertainty communication strategy after 2008. However, this switch from countercyclical alerts to pro-cyclical appeasement risks a loss in reputation and trust in the central bank. To support their reasoning, Sauter et al. (2013) employ a Markov-switching analysis that identifies two different communication regimes. These regimes can be described as 'stable' and 'crisis' regimes. The findings indicate that during turbulent times, the 'crisis' regime — with a higher variance in the use of uncertainty-related terms — prevails.

6.5 Swedish Central Bank – Riksbank

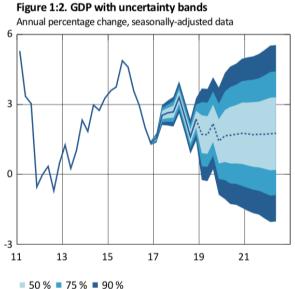
The Riksbank communicates its view of the economic outlook to the public in its Monetary Policy Report. The Riksbank's Monetary Policy Report is published sixtimes a year. The report describes the deliberations made by the Riksbank's executive board when deciding what would be an appropriate monetary policy to conduct. The report also contains a description of the future prospects for inflation and economic activity based on the interest rate path that the Riksbank's executive board currently

considers will provide a well-balanced monetary policy. Each report also contains a description of the new information received since the previous report and an assessment of how the Riksbank views the current economic situation. The purpose of a monetary policy report is to produce background material for monetary policy decisions, and to spread knowledge about the Riksbank's assessments.

The report contains forecasts with uncertainty bands around the forecast. For policy purposes, having uncertainty bands around forecasts is useful for several reasons. First and foremost, they serve to illustrate that the inflation forecast is inherently uncertain. The uncertainty is both about the shocks that will affect the economy as well as uncertainty about both the qualitative and quantitative nature of the transmission mechanism. Second, the bands serve to present the Riksbank's view of the balance of risks to the public and to market participants. In particular, it allows the Riksbank to communicate with a minimum of ambiguity whether the risk is believed to be higher that inflation will be below the forecast than that it will be above, as was the case in Inflation Report 1998:2. Third, the construction of the bands helps to focus internal discussion in the Riksbank about the sources of inflation uncertainty and about their quantitative importance.

In particular, they provide uncertainty bands for the repo rate, GDP and Consumer Price Index with a fixed interest rate (CPIF), e.g. see Sveriges Riksbank (2019a). An example of GDP projections is shown in figure 5. The uncertainty bands for the repo rate are based on the Riksbank's historical forecasting errors and the ability of risk-premium adjusted forward rates to forecast the future repo rate for the period 1999 up to the point when the Riksbank started to publish forecasts for the repo rate during 2007. The uncertainty bands do not take into account the fact that there may be a lower bound for the repo rate. Outcomes are daily rates and forecasts refer to quarterly averages. The uncertainty bands for GDP are based on the Riksbank's historical forecasting errors. The reported outcomes for GDP are also uncertain, as the National Accounts figures are revised several years after

Figure 5: Exemplary Riksbank projection chart, e.g. see Sveriges Riksbank (2019b)



Note. The uncertainty bands are based on the Riksbank's historical forecasting errors. The reported outcomes for GDP are also uncertain, as the National Accounts figures are revised several years after the first publication.

Sources: Statistics Sweden and the Riksbank

the first publication. In the same fashion, the uncertainty bands for CPIF are also based on the Riksbank's historical forecasting errors.

In comparison to the FED / FOMC, the Riksbank's Monetary Policy Report splits the uncertainty bands into 50 %, 75 % and 90 % confidence bands providing a more detailed picture than that of the FOMC and closer resembles the Bank of England's fan-chart.

6.6 Office for National Statistics (ONS) United Kingdom

The Office for National Statistics (ONS) acknowledges the complications of inherent uncertainty as described in previous sections in this report. For that reason it has published a guide on how to deal with uncertainty. The guidance provides a common approach to aid the clear communication of uncertainty and change. It can be applied to all sources of information, including surveys, censuses, administrative data and other sources, as well as estimates derived from a combination of these. It includes examples of good practice, as well as standard wording to be used when appropriate. The guidance does not cover how to measure uncertainty and change, as this is very much dependent on the specifics of particular data collection scenarios. Links to information which maybe helpful on this topic are provided at the end of this guidance, see Government Statistical Service (2014). It lists the Bank of England fan-chart as a prime example of good communication.

6.7 Other disciplines

Communicating uncertainty is not an issue exclusive to economics but it appears across various different disciplines. Examples can be found in history, public health, biology, climate change, politics, intelligence, legal epidemiology, law, physics and many more. Across all these disciplines, one can find different types of communication ranging across all categories introduced by van der Bles et al. (2019). Examples include history (e.g. see King et al. (2014)), public health (e.g. Harris et al. (2011)), biology (e.g. Ronquist et al. (2012)), law Stuart Sumner (2017) or physics Hooft (2016).

To introduce a more specific example, let's take a look at climate change. The intergovernmental panel on climate change (IPCC) has a relativelylong history of exploring how to effectively express different forms of uncertainty in their reports but has only recently started to begin incorporating insights from behavioural science. At present, uncertainty in the IPCC assessments is communicated using two metrics. Firstly, quantified measures of direct (absolute) uncertainty are expressed in verbal and probabilistic terms based on statistical analyses of observations, models or expert judgement, corresponding to a pre-defined categorization. Secondly, indirect (underlying) uncertainties are expressed through a qualitative expression of confidence in the validity of a finding based on the type, amount, quality and consistency of evidence (which can include theory, models and expert judgement), see Stocker (2014). Likelihood provides calibrated language for describing quantified uncertainty for a single event, a climate parameter, an observed trend or projected future change. Importantly, the likelihood table is not preferred when a full probability distribution is available instead.

Confidence level is based on the scientific evidence (robust, medium, limited) and working group agreement (high, medium, low). Robustness of evidence is measured by the degree of consistent and independent (multiple) lines of high quality inquiry.

To illustrate the above-mentioned concepts with a written example from the fifth IPCC report: It is certain that global mean surface temperature has increased since the late 19th century. [...] For average annual Northern Hemisphere temperatures, the period 1983–2012 was very likely the warmest 30-year period of the last 800 years (high confidence) and likely the warmest 30-year period of the last 1400 years (medium confidence).

7

Our methodology for measuring uncertainty in macroeconomic indicators

As the large majority of official statistics, also most macroeconomic indicators and data are uncertain—they are obtained as estimates rather than perfect measures of the target phenomenon. Measurement errors may arise because data are based on incomplete samples or because many variables—for example, in-house software investment—are not easily observable. This necessitates the use of proxies, imputation techniques, etc. Without objective measures of data quality, it is difficult to gauge the potential for measurement errors. One symptom of data uncertainties is the propensity of statistical agencies to revise their estimates in the light of new available information or methodological advances. In this section we will explore new methodologies for analysing and measuring data uncertainty. We start from recent work undertaken at policy making institutions such as the Bank of England, which, also as noted by Manski (2015a), has been leading the way in the past 30 years in analysing and communicating data and forecast uncertainty through the use of tools such as the fan chart for inflation and GDP growth.

The methodology will be focused on state space modelling. Such models provide a natural avenue since they are based on the presence of unobserved variables that can proxy for the true process that statistical agencies and other policy making bodies are trying to measure. Such models have an added benefit of allowing for the consideration of particular economic structures that can inform the quantification of conceptual uncertainties as discussed in 2. Below we provide a description of the model class we are proposing to expand and analyse as well as the adopted estimation strategy.

7.1 The model

In this section, we present in detail a state space representation of the signal extraction problem following the work of Cunningham et al. (2012a). Recognising that the analysis of the latest official data may be complemented by the use of business surveys and other indirect measures, the model allows for an array of measures of each macroeconomic variable of interest. Then, for each variable of interest, the model comprises alternative indicators, a transition law and separate measurement equations describing the latest official estimates. The measurement equation is designed to be sufficiently general to capture the patterns in revisions observed historically.

The model is presented in a vector notation, assuming m variables of interest. However, we simplify estimation by assuming block-diagonal structure throughout the model so that the model can be estimated on a variable-by-variable basis for each of the m elements in turn.

Let the m dimensional vector of variables of interest that are subject to data uncertainty at time t be denoted by \mathbf{y}_t , t=1,...,T. The vector \mathbf{y}_t contains the unobserved true value of the economic concept of interest.

The model for the true data y_t is given by

$$\mathbf{y}_{t} = \mathbf{\mu} + \sum_{i=1}^{q} \mathbf{A}_{i} \mathbf{y}_{t-i} + \mathbf{\varepsilon}_{t}, \tag{10}$$

where $\mathbf{A}_1,...,\mathbf{A}_q$ are $m \times m$ matrices, $\mathbf{A}(L) = \mathbf{I}_m - \mathbf{A}_1 L - ... - \mathbf{A}_q L^q$ is a lag polynomial whose roots are outside the unit circle, $\mathbf{\mu}$ is a vector of constants, $\mathbf{\epsilon}_t = (\mathbf{\epsilon}_{1t},...,\mathbf{\epsilon}_{mt})$ and $\mathbf{E}(\mathbf{\epsilon}_t \mathbf{\epsilon}_t) = \mathbf{\Sigma}_{\mathbf{\epsilon}}$, where we denote the main diagonal of $\mathbf{\Sigma}_{\mathbf{\epsilon}}$ by $\mathbf{\sigma}_{\mathbf{\epsilon}}^2 = (\mathbf{\sigma}_{\mathbf{\epsilon}_1},...,\mathbf{\sigma}_{\mathbf{\epsilon}_m}^2)$. We further assume that $\mathbf{A}_1,...,\mathbf{A}_q$ are diagonal.

Let \mathbf{y}_t^{t+n} denote a noisy estimate of \mathbf{y}_t published by the statistical agency at time $_{t+n}$, where $_{n=1,\ldots T-t}$. The model for these published data is

$$\mathbf{y}_{t}^{t+n} = \mathbf{y}_{t} + \mathbf{c}^{n} + \mathbf{v}_{t}^{t+n} \tag{11}$$

where \mathbf{c}^n is the bias in published data of maturity n and \mathbf{v}_t^{t+n} the measurement error associated with the published estimate of \mathbf{y}_t made at maturity n.

One of the main building blocks of the model is the assumption that revisions improve estimates so that official published data become more accurate as they become more mature. Reflecting this assumption, both the bias in the published estimates and the variance of measurement errors are allowed to vary with the maturity of the estimate — as denoted by the n superscript. The constant term

 \mathbf{c}^n is included in equation (11) to permit consideration of biases in the statistical agency's data set. Specifically, \mathbf{c}^n can be modelled as

$$\mathbf{c}^{n} = \mathbf{c}^{1} (1+\lambda)^{n-1}$$
 (12)

where \mathbf{c}^1 is the bias in published data of maturity n=1 and λ describes the rate at which the bias decays as estimates become more mature $(-1<\lambda<0)$. This representation assumes that the bias tends monotonically to zero as the estimates become more mature.

The measurement errors, \mathbf{v}_t^{t+n} , are assumed to be distributed normally with finite variance. Serial correlation in \mathbf{v}_t^{t+n} is allowed. Specifically, serial correlation in the errors attaching to the data in any data release published at t+n, can be modelled as

$$\mathbf{v}_{t}^{t+n} = \sum_{i=1}^{p} \mathbf{B}_{i} \mathbf{v}_{t-i}^{t+n} + \mathbf{\varepsilon}_{t}^{t+n}, \tag{13}$$

where $\mathbf{B}_1,...,\mathbf{B}_p$ are mxm matrices, $\mathbf{B}_{(L)}=\mathbf{I}-\mathbf{B}_1L-...-\mathbf{B}_pL^p$ is a matrix lag polynomial whose roots are outside the unit circle and $\mathbf{\epsilon}_t^{t+n}=(\mathbf{\epsilon}_{1t}^{t+n},...,\mathbf{\epsilon}_{mt}^{t+n})'$ and $\mathbf{E}(\mathbf{\epsilon}_t^{t+n}(\mathbf{\epsilon}_t^{t+n})')=\mathbf{\Sigma}_{\mathbf{\epsilon}}^n$ as we are allowing for heteroscedasticity in measurement errors with respect to n. Equation (13) imposes some structure on \mathbf{v}_t^{t+n} because we assume a finite AR model whose parameters do not depend on maturity. The representation picks up serial correlation between errors attaching to the various observations within each data release. We further assume that $\mathbf{B}_1,...,\mathbf{B}_n$ are diagonal.

Further, $\mathbf{\varepsilon}_t^{t+n}$ and therefore \mathbf{v}_t^{t+n} can have heteroscedasticity with respect to n. Specifically, the main diagonal of $\mathbf{\Sigma}_{\mathbf{\varepsilon}}^n$ are modelled as $\mathbf{\sigma}_{\mathbf{\varepsilon}}^2 = (\mathbf{\sigma}_{\mathbf{\varepsilon}_1}^n, \dots, \mathbf{\sigma}_{\mathbf{\varepsilon}_m}^n)$, where $\mathbf{\sigma}_{\mathbf{\varepsilon}_i}^n = \mathbf{E}(\mathbf{\varepsilon}_{it}^{t+n})^2$. For future reference, define $\mathbf{\sigma}_{\mathbf{\varepsilon}_i}^n = \mathbf{E}(\mathbf{v}_{it}^{t+n})^2$. The model for $\mathbf{\sigma}_{\mathbf{\varepsilon}_i}^n$ is given by

$$\sigma_{\varepsilon^n}^2 = \sigma_{\varepsilon^1}^2 (1+\delta)^{n-1},\tag{14}$$

where $\sigma_{\epsilon^1}^2$ is the variance of measurement errors at maturity n=1 and δ describes the rate at which

variance decays as estimates become more mature $(-1<\delta<0)$. This representation imposes structure on the variance of measurement errors, because we assume that the variance declines monotonically to zero as the official published estimates become more mature. Amonotonic decline in measurement error variances is consistent with models of the accretion of information by the statistical agency, such as that developed in Kapetanios and Yates (2004a). We put forward three reasons for using this specification. Firstly, this model is parsimonious since it involves only two parameters. Secondly, δ has an appealing interpretation as a rate at which revision error variances decline over time. Thirdly, and perhaps most importantly Kapetanios and Yates (2009) provide empirical evidence in favour of this specification.

Over and above any serial correlation in revisions, measurement errors can be correlated with the underlying true state of the economy, \mathbf{y}_t We specify that $\mathbf{\epsilon}_t^{t+n}$ be correlated with the shock $\mathbf{\epsilon}_t$ to the transition law in equation (10), so that, for any variable of interest

$$cov(\varepsilon_{it}, \varepsilon_{it}^{t+n}) = \rho_{\varepsilon\varepsilon} \sigma_{\varepsilon_i} \sigma_{\varepsilon_i}^n. \tag{15}$$

In principle, the model in equation (11 could be applied to previous releases as well as the latest estimates. One natural question is whether data-users should consider these previous releases as

 $\text{competing measures of the truth} \ __ \ \text{that is, using } \mathbf{y}_t^{t+n-j} \ \text{ alongside } \mathbf{y}_t^{t+n} \ \text{ as measures of } \mathbf{y}_t. \ \text{In contrast}$

with the treatment in much of the antecedent literature, we decide to exclude earlier releases from the set of measures used to estimate 'true' activity, see, for example, Garratt and Vahey (2006). The reason for using only the latest release is pragmatic. In principle, given that empirical work across a variety of data sets has found that revisions appear to be forecastable, using earlier releases should be useful. In practice, however, such a model would be complex. That complexity may be costly in various ways — the model would be more difficult to understand, more cumbersome to produce and potentiallyless robust when repeatedly reestimated. Further, by focusing on the latest release we are able to specify a model that is quite rich in its specification of other aspects of interest, such as heteroscedasticity, serial correlation and correlation with economic activity.

We note that there are circumstances where using onlythe latest release is theoretically optimal. An example of a set of such circumstances is provided in the Appendix of Cunningham et al. (2007). The model developed in that appendix makes a number of assumptions that imply a form of rational behaviour on the part of the statistical agency, which may well not hold in practice. Therefore, we must stress that such a model is restrictive. Further, our modelling approach is obviously parametric and therefore has claims to efficiency only if, on top of rationality on the part of the statistical agency, the specification of the model for the unobserved true variable is correct. On the other hand, note that the use of such a parametric model for the unobserved variable can provide benefits as well. Even if the statistical agency is operating optimally in data collection, our state space model can provide further benefits by positing a model for \mathbf{y}_{r} , since that is not a part of the statistical agency's specification.

In addition to the statistical agency's published estimate, the data-user can observe a range of alternative indicators of the variable of interest. We denote the set of these indicators by

 \mathbf{y}_t^{s} , $t=1,\ldots,T$. Unlike official published estimates, the alternative indicators need not be direct measures of the underlying variables. For example, private sector business surveys typically report the proportion of respondents answering in a particular category rather than providing a direct measure of growth. The alternative indicators could be assumed to be linearly related to the true data

$$\mathbf{y}_{t}^{s} = \mathbf{c}^{s} + \mathbf{Z}^{s} \mathbf{y}_{t} + \mathbf{v}_{t}^{s}. \tag{16}$$

The error term \mathbf{v}_t^s is assumed to be i.i.d. with variance $\mathbf{\Sigma}_{\mathbf{v}}^s$. This, of course, is more restrictive than the model for the official data. Simple measurement equations of this form may not be appropriate for all the alternative indicators used in routine conjunctural assessment of economic activity. More general formulations of equation (16) are possible but they would further complicate estimation of the model, in particular in the presence of a limited number of data points as in our empirical application.

Having completed the presentation of the model, it is worth linking our work to the literature that deals with the presence of measurement error in regression models. A useful summary of the literature can be found in Cameron and Trivedi (2005). This body of work is of interest as it can provide solutions to a number of problems caused by the presence of data revisions. In the context of the following simple regression model

$$z_t = \beta y_t + u_t \tag{17}$$

use of y_t^{t+1} as a proxy for y_t can lead to a bias in the OLS estimator of β . Then, the use of later vintages, y_t^{t+n} , n=2,...,T-t, as instruments in (17) can be of use for removing the bias in the estimation of β . One issue of relevance in this case is whether to use all available vintages as instruments. The rapidly expanding literature on optimal selection of instruments, see, for example, Donald and Newey (2001), suggests useful tools for this purpose. Our analysis provides an alternative method of addressing this problem. In our modelling framework, equation (17) becomes a further measurement equation of the state space model and the overall estimation of the resulting model can provide unbiased estimates of β . However, our current state space formulation is of further interest since on top of giving estimates for relevant parameters it also gives an alternative and possibly superior proxy for the unobserved true series, in the form of an estimate for the state variable. This, can then be used for a variety of purposes including forecasting.

7.2 Estimation of the state space model

In this section, we discuss the strategy adopted in estimating the model. The estimation is performed in two steps: first using the revisions history to estimate equations (11) through (15); and then, as a second step, estimating the remaining parameters via maximum likelihood using the Kalman filter. Approaching estimation in two steps simplifies greatly the estimation of the model and has the additional benefit of ensuring that the model is identified. Were all parameters to be estimated in one step, the state space model would not always satisfy the identification conditions described in Harvey (1989).

Define the revisions to published estimates of an individual variable between maturities n and n+j as

$$w_t^{n,j} = y_t^{t+n+j} - y_t^{t+n}$$
 (18)

For estimation purposes, revisions can be taken over the J quarters subsequent to each observation to be representative of the uncertainty surrounding that measure of activity. If the real-time data set contains W releases of data, and we are interested in the properties of N maturities, we can construct an $N \times (W - J)$ matrix of revisions $\mathbf{Error!}$ over which to estimate the parameters of equations (11) through (15). Each column of the matrix \mathbf{W}^J contains observations of revisions to data within a single data release. Each row describes revisions to data of a specific maturity N. N and N are both choice variables and should be selected to maximise the efficiency of estimation of the parameters driving equations (11) to (15). There is a trade-off between setting N sufficiently large to pick up all measurement uncertainties and retaining sufficient observations for the estimated mean, variance, and serial correlation of revisions and their correlation with mature data to be representative. In the remainder of the paper we set N = J = 20.

We use the sample of historical revisions in matrix \mathbf{W}^J to estimate c^1 and λ trivially. Recall that we assume $\mathbf{B}_1, ..., \mathbf{B}_p$ to be diagonal. As a result, the functions can be estimated for individual variables rather than for the system of all variables of interest. In the remainder of this paper we therefore consider estimation for a single variable and discard vector notation. The sample means of revisions of each maturity n=1 to N are simply the average of observations in each row of \mathbf{W}^J . Denoting the average revision to data of maturity n by $\mathrm{mean}(w^{n,J})$, the parameters c^1 and λ are then estimated from the moment conditions $\mathrm{mean}(w^{n,J})=c^1(1+\lambda)^{n-1}$ via GMM, where $-1<\lambda<0$.

We cannot use historical revisions to estimate $\rho_{\varepsilon\varepsilon}$ directly, because neither ε_t nor ε_t^{t+n} are observable. But we can use the historical revisions to form an approximation of ρ_{yv} — denoted ρ_{yv}^* . The manipulation in obtaining $\rho_{\varepsilon\varepsilon}$ from ρ_{yv}^* , is summarised in the Appendix of Cunningham et al. (2009). We start by estimating ρ_{yv}^* . We can readily calculate the correlation between revisions to data of maturity n and published estimates of maturity J_{t+n} , denoted by $\rho_{yv}^n = \operatorname{corr}(y_t^{t+J+n}, w_t^{n}, J)$. Averaging across the N maturities in \mathbf{W}^J gives an average maturity-invariant estimate of ρ_{yv}^* . When the variance of measurement errors decays sufficiently rapidly, we do not introduce much approximation error by taking this correlation with mature published data as a proxy for the correlation with the true outcome, y_t We do not apply any correction for this approximation because derivation of any correction would require untested assumptions about the relationship between measurement errors across successive releases which we do not wish to impose on the model.

The variance-covariance matrix of historical revisions may be used to jointly estimate both the heteroscedasticity in measurement errors and their serial correlation. This requires us to first express the variance-covariance matrix of measurement errors as a function of the parameters in equations (13) and (14) and then to estimate the parameters consistent with the observed variance-covariance matrix of revisions.

Assuming for simplicity first-order serial correlation in the measurement errors, we can easily build-up a full variance-covariance matrix at any point in time. The variance-covariance matrix of the measurement errors in the most recent N maturities, will be invariant with respect to t and is given by

$$\mathbf{V} = \frac{\sigma_{\epsilon}^{2}}{1 - (1 + \delta)\beta_{1}^{2}} \times$$

$$\begin{pmatrix} 1 & (1 + \delta)\beta_{1} & \cdots & (1 + \delta)^{N-1}\beta_{1}^{N-1} \\ (1 + \delta)\beta_{1} & (1 + \delta) & \cdots & (1 + \delta)^{N-1}\beta_{1}^{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ (1 + \delta)^{N-1}\beta_{1}^{N-1} & (1 + \delta)^{N-1}\beta_{1}^{N-2} & \cdots & (1 + \delta)^{N-1} \end{pmatrix}$$
Error! Bookmark not

A sample estimate of the variance-covariance matrix $\hat{\mathbf{V}}$ can be calculated trivially from the matrix of historical revisions \mathbf{W}^J . Taking the variance-covariance matrix to the data, we can estimate $\beta_1, \sigma_{\epsilon 1}$ and δ_1 via GMM by minimizing $(\text{vec}(\mathbf{V})-\text{vec}(\hat{\mathbf{V}}))^{'}(\text{vec}(\mathbf{V})-\text{vec}(\hat{\mathbf{V}}))$. The derivation of the variance-covariance matrix for higher lag-orders requires some further manipulation, as outlined in the Appendix

of Cunningham et al. (2009). It is worth noting here that there exists an interesting special case where the first step estimation does not affect the second step ML estimation via the Kalman filter. This is the case where the number of available vintages, N, tends to infinity. In this case, the GMM estimation outlined above, results in parameter estimates that are \sqrt{NT} consistent whereas the second step ML estimation is only \sqrt{T} consistent implying that the parameters that are estimated in the first step can be treated as known for the second step and the resulting approximation error associated with the first step estimation is asymptotically negligible.

More generally, the fact that more data are used in the first step implies that the variability of the first step estimates is likely to be lower than that of the second step estimates. However, the use of a two-step estimation procedure implies that, in practice, the variability of the first step estimates is not taken into account when the likelihood based second step variance estimates are obtained. Of course, if the variances of the parameter estimates are of particular interest, a parametric bootstrap can provide a standard avenue for obtaining variance estimates that implicitly take into account the variability arising out of both estimation steps. The parametric bootstrap would have to replicate both steps of the two-step estimation procedure to capture appropriately the parameter uncertainty associated with the first step estimation. However, note that the validity of the bootstrap in this two-step estimation context is not obvious. Further, use of the bootstrap requires the specification of a model for all vintages used in the first step GMM estimation, which may be problematic in practice. For these reasons, we provide standard errors for the estimated parameters obtained from the second estimation step, using standard likelihood based inference.

8 Empirical application

After having described in some details our reference model and the adopted estimation strategy, here we are designing its application to real statistical data produced by Eurostat and some other European statistical agencies. The proposed application aims at deriving reliable, easily understandable and ready for communication uncertainty measure associated to a number of main macroeconomic indicators. The proposed application also allow for cross-countries and cross-variables comparison.

8.1 Data

We consider publicly available data from the OECD vintages database for the four largest economies of the euro area: Germany, France, Italy and Spain and for the euro area. Data is available from January 2000 until December 2018; with monthly vintages starting in March 2018 until March 2019, i.e. 13 releases. We have selected five key macroeconomic indicators: harmonized index of consumer prices (HICP), real gross domestic product (GDP), index of industrial production (IPI), index of retail trade volume (RTI) and harmonised unemployment rate (UR). Where data is available on a monthly basis, we have created a quarterly data set by averaging over the months of a quarter.

Data was transformed into quarterly growth rates, except for unemployment rate, which was transformed to a difference. Among the series, HICP is not being revised or there is 1 revision — a level shift. Retail trade is the series that is revised the most. The other variables, industrial production, unemployment rate and GDP are revised to some extent.

Over the 2000–18 sample, the data for GDP show the mean growth rate to have been highest for Spain and lowest for Italy, with the growth rate for the euro area, Germany and France being in between, and quite similar to each other. The data for France have the lowest volatility over the period, with the data for the other countries up to twice as volatile. The data for all countries show a skew to the downside, reflecting the great recession, that downside skew is greatest for Germany, and smallest for Spain.

Out of the monthly series, industrial production and retail trade are shown to be somewhat more volatile than GDP, even when considered at the quarterly frequency. In broad terms, the patterns across the countries are similar to what is observed for GDP, with the exception of the volatility of retail trade growth which turns out to be higher in the case of Spain than for the other countries.

In contrast to GDP growth, which has been positive on average over the 2000–18 sample for all countries, the growth rates of industrial production and retail trade have not been positive on average for all the countries. Industrial production growth has been negative on average over the sample for Italy, Spain and France, and retail trade growth for Italy and Spain, what might be an indication of a shift in activity from industry and retail trade to other parts of the service sector.

The unemployment rate, finally, has declined on average over the sample in the cases of Germany and the euro area, increasing for the cases of Italy and Spain, and unchanged for France. As GDP growth, the changes in the unemployment rate are shown to have the lowest volatility over the sample in the case of France. The highest volatility is recorded in the case of Spain, even though the volatility

of real GDP growth for that country has been the lowest across the countries.

In terms of revisions, the data suggest that the average revision is of the same order of magnitude for the (quarterly) growth rate of real GDP and retail trade, while revisions to the growth rate of industrial production have tended to be somewhat larger on average, at least for the euro area, Germany and France. The volatility of the revisions is quite similar across the countries, with exceptions to the upside in the case of industrial production for Italy and Spain, and retail trade in the case of Germany.

8.2 Empirical Results

Due to the limited number of releases we set the initial error variance to 1, the parameter beta in serial correlation to -0.2 implying an AR(1) process for v_{t+1} . Initial bias in the statistical agency's data set

and bias decay were both set to zero, as based on an earlier studyby Cunningham et al. (2012a) for UK data. We have experimented with the decay parameter delta, setting it to -0.01 and -0.05 and with the correlation of the measurement errors with the underlying state of the economy. The correlation was set to -0.5; 0 and 0.5 in turn.

The results show that delta = -0.05 and rho = -0.5 perform best in describing the revisions in inflation, GDP for all countries and retail trade for all countries but Spain as measured by Mean Squared Error of the final release and the filtered estimate. For industrial production the same combination of variables performs best for Germany and Italy, while for the remaining three countries moving to a slower decay of delta = -0.01 provides better results. In the case of unemployment the no correlation case for delta = -0.05 works best for all countries but Spain for which delta = -0.01.

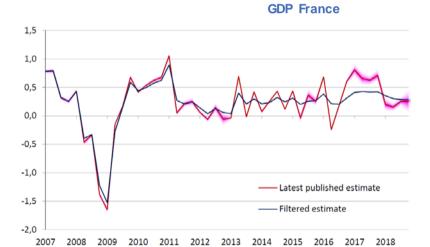
What is interesting to note about those results is that with respect to the correlation of the measurement error, the best-performing specification is the same (rho = -0.5) for the variables expressed as growth rates — GDP, industrial production and retail trade — and another (rho = 0.0) for the unemployment rate, expressed as a simple difference, and irrespective of the decay parameter (whether delta = -0.05 or -0.01). This supports the notion that variables tend to display different patterns of data uncertainty, and therefore require different treatment for data uncertainty, depending on whether they are stationary or trending.

Figures 8.2–8.2 report our results for GDP, for the various countries, together with the associated uncertainty measures. Finally, the Annex presents figures with the results for the other economic indicators for the various countries.

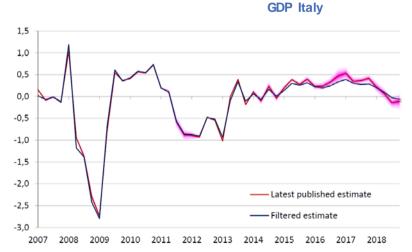




Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.



Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.



8.3. Robustness

The results presented above suggest that it is important in principle to use the appropriate treatment when trying to account for the uncertainty surrounding past data releases. In fact, this is true for both the decay parameter δ and the parameter ρ capturing the correlation of the measurement error also for the performance of the model in terms of the MSE.

Regarding the decay parameter $_{\delta}$, the sensitivity is considerable — setting the parameter not to the optimal value of +0.5 but to 0.0 results in a multiple increase of the MSE for most of the variables (GDP, retail trade ad the unemployment rate) with only one exception (industrial production — the variable for which it is not clear in the first place whether +0.5 or 0.0 is the optimal setting for that parameter.

As for the parameter capturing the correlation of the measurement error, ρ , moving from -0.5 to 0.0 also results in noticeable increases in the MSE. This is true especially for GDP and retail trade, somewhat less for industrial production and the unemployment rate, where the major deterioration in performance occurs as the parameter moves further away from the optimal setting (from 0.0 to +0.5).

Conclusion

In the first part of this paper, after having shortly introduced the major sources and categories of uncertainty in official statistics, we have critically reviewed, analysed and evaluated existing methods for measuring and communicating uncertainties inherent to official statistics. It shows that, while there already exist various different approaches of quantifying these uncertainties, one major issue remains: the identification of a clear and unambiguous method for communicating uncertainty measures, especially numerical ones. Some communication methods are available and in use at various public institutions. Among them, the fan-charts, used by the Bank of England, provide a clear and effective tool for numerical uncertainty communication around backcasts, nowcasts (flash estimates) and forecasts. However, there still remain the risk of misunderstanding. In order to minimise such risk, clear verbal communication strategies should be developed to complement and integrate quantitative communication tools. In addition, we have presented a series of considerations related to the relationships between the characteristics and specificities of statistical indicators' data generating process and the presence of uncertainty, which could lead the user to a better understanding and assessment of this phenomenon. The detailed critical review carried out in this first part of the paper contributed to highlighting a number of key issues and relevant aspects related to the measurement and communication of uncertainties in official statistics. They have constituted the ideal starting point to developing our strategy for measuring and communicating uncertainty in key macroeconomic infraannual statistics. In the second part of this paper we considered a state of the art framework for modelling real time data and quantifying the uncertainty surrounding those data. We applied this framework to a post-crisis sample (2000–18) for the euro area and its four largest countries (Germany, France, Italy, Spain) in order to provide an extensive set of empirical results.

The results indicate that the model-based estimate of the true process in terms of mean squared error (MSE) outperforms the most recent published estimate for all the variables considered (the growth of GDP, industrial production, retail trade, and the change in the unemployment rate). The results further illustrate the sensitivity of the performance of the model in terms of the mean squared error (MSE) with respect to the two model parameters, the decay parameter δ , and the parameter capturing the correlation in measurement error, ρ .

Overall, the results suggest that the treatment for data uncertainty is both feasible and important, and that the appropriate choice of parameters is quite important in order to achieve the best empirical performance. One aspect that must be given some consideration in applying the methodology is the nature of the variables — trending or stationary—that are being estimated. The outcome of this paper shows the possibility of measuring uncertainty around official statistics in a reliable, easily understandable and ready to be communicated way, by using sound and well established methodology. Finally, based on the overall work carried out for this paper, we would like to highlight some points and considerations which, in a further stage, could also constitute the basis for a more formal set of reccomendations addressed to statistical authorities. The major points we would like to highlight are:

- the importance of measuring and communicating uncertainty as sociated to official statistics;
- the need of identifying the most appropriate tools for measuring and communicating

- uncertainty taking into account the characteristics and specificities of the official statistics' generating process;
- the importance of using methodologically sound, well established and reliable tools for measuring and communicating uncertainty;
- the necessity of communicating uncertainty in a clear, unambiguous and meaningful manner providing a positive message to users;
- the importance of identifying clear verbal communication standards to be always used as complements of either numerical or graphical communication tools;
- the importance of clarity and transparency around all activities of measurement and communication of uncertainty by providing a complete set of standard meta-information and documentation.

References

Assael H., Keon J. (1982), 'Nonsampling vs. sampling errors in survey research', *Journal of Marketing*, 46(2), pp. 114–123.

Baker S. R., Bloom N., Davis S. J. (2016), 'Measuring economic policy uncertainty', *The quarterly journal of economics*, 131(4), pp. 1593–1636.

Balshem H., Helfand M., Schünemann H. J., Oxman A. D., Kunz R., Brozek J., Vist G. E., Falck-Ytter Y., Meerpohl J., Norris S. et al. (2011), 'Grade guidelines: 3. rating the quality of evidence', *Journal of clinical epidemiology*, 64(4), pp. 401–406.

Bank of England, Statistics and regulatory data division (2014), 'Data quality framework', https://www.bankofengland.co.uk/-/media/boe/files/statistics/data-quality-framework, Accessed: 2019-06-21.

Bank of England (2019), Inflation reports, https://www.bankofengland.co.uk/inflation-report/inflation-reports, Accessed: 2019-06-21.

Benedict L. H. and Gould R.D. (1996), 'Towards better uncertainty estimates for turbulence statistics', *Experiments in fluids*, 22(2), pp. 129–136.

Biemer P. P., Lyberg L. E. (2003), Introduction to survey quality, volume 335, John Wiley & Sons.

Bloom N. (2009), 'The impact of uncertainty shocks', Econometrica, 77(3), pp. 623–685.

Britton E., Fisher P., Whitley J. (1998), 'The inflation report projections: understanding the fan chart', Bank of England, *Quarterly Bulletin*, 38(1), pp.30.

Cameron A. C., Trivedi P. K (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Campbell J. R. (2013), 'Odyssean forward guidance in monetary policy': A primer, *Economic Perspectives*, 37(4).

Capistrán C. (2008), 'Bias in federal reserve inflation forecasts: Is the federal reserve irrational or just cautious?', *Journal of Monetary Economics*, 55(8), pp. 1415–1427,

Chang A. C., Li P. (2018), 'Measurement error in macroeconomic data and economics research: Data revisions, gross domestic product, and gross domestic income', *Economic Inquiry*, 56(3), pp. 1846–1869.

Clements M.P. (2004), 'Evaluating the bank of England density forecasts of inflation', *The Economic Journal*, 114(498), pp. 844–866.

Clements M. P.(2014), 'Forecast uncertainty ex ante and ex post: Us inflation and output growth', *Journal of Business & Economic Statistics*, 32(2), pp. 206–216.

Clogg C. C., Dajani A. N. (1991), 'Sources of uncertainty in modeling social statistics: An inventory', *Journal of Official Statistics*, 7(1), pp.7–24.

Cunningham A., Eklund J., Jeffery C., Kapetanios G., Labhard V. (2007), *A state space approach to extracting the signal from uncertain data*, Bank of England Working Paper, (336).

Cunningham A., Eklund J., Jeffery C., Kapetanios G., and Labhard V. (2009), *A state space approach to extracting the signal from uncertain data*, Queen Mary University of London Working Paper, (637).

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

Cunningham A., Eklund J., Jeffery C., Kapetanios G., and Labhard V. (2012b), 'A state space approach to extracting the signal from uncertain data', *Journal of Business & Economic Statistics*, 30(2), pp. 173–180.

Diebold F. X., Tay A. S., Wallis K. F. (1997), *Evaluating density forecasts of inflation: the survey of professional forecasters*, Technical report, National bureau of economic research.

Donald S. G., Newey W. K. (2001), 'Choosing the number of instruments', *Econometrica*, 69, pp. 1161–1191.

Dowd K. (2007), 'Too good to be true? the (in) credibility of the UK inflation fan charts', *Journal of Macroeconomics*, 29(1), pp. 91–102.

European Central Bank (2009), *New procedure for constructing eurosystem and ECB staff projection ranges*, https://www.ecb.europa.eu/pub/pdf/other/newprocedureforprojections200912en.pdf, Accessed: 2019-06-21.

European Central Bank (2019), Eurosystem staff macroeconomic projections for the euro area, June 2019

https://www.ecb.europa.eu/pub/projections/html/ecb.projections201906eurosystemstaff~8e352fd82a.en.html#toc2, Accessed: 2019-06-21.

Faust J., Wright J. H. (2009), 'Comparing Green Book and reduced form forecasts using a large real-time dataset', *Journal of Business & Economic Statistics*, 27(4), pp. 468–479.

Federal Open Market Committee (2019), *Minutes of the federal open market committee*, https://www.federalreserve.gov/monetarypolicy/fomcminutes20190320ep.htm, Accessed: 2019-06-21.

Garratt A., Vahey S. P. (2006), 'UK real-time macro data characteristics', *The Economic Journal*, 116(509), pp. F119–F135.

Goedemé T. (2013), 'How much confidence can we have in EU-SILC? complex sample designs and the standard error of the Europe 2020 poverty indicators', *Social Indicators Research*, 110(1), pp. 89–110.

Government Statistical Service (2014), Communicating uncertainty and change—guidance for official statistics producers,

https://gss.civilservice.gov.uk/wp-content/uploads/2018/03/Communicating-uncertainty-and-change-Guidance-for-Official-Statistics-4.pdf, Accessed: 2019-06-21.

Groen J. J., Kapetanios G., Price S. (2009), 'A real time evaluation of Bank of England forecasts of inflation and growth', *International Journal of Forecasting*, 25(1), pp. 74–80.

Groves R. M. (2004), Survey errors and survey costs, volume 536, John Wiley & Sons.

Harris R. J., Ramsay M., Hope V. D., Brant L., Hickman M., Foster G. R., De Angelis D. (2011), 'Hepatitis c prevalence in England remains low and varies by ethnicity: an updated evidence synthesis', *The European Journal of Public Health*, 22(2), pp. 187–192.

Harvey A. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Heckman J. J. (1979), 'Sample selection bias as a specification error', *Econometrica: Journal of the econometric society*, pp. 153–161.

Hooft G. (2016), 'The evolution of quantum field theory: From ged to grand unification', *The Standard Theory of Particle Physics: Essays to Celebrate CERN's 60th Anniversary*, pp. 1–27, World Scientific.

Howrey E. P. (1978), 'The use of preliminary data in econometric forecasting', *The Review of Economics and Statistics*, pp. 193–200.

Jurado K., Ludvigson S. C. (2015), Ng S., 'Measuring uncertainty', *American Economic Review*, 105(3), pp. 1177–1216.

Kapetanios G., Yates T. (2004a), 'Estimating time-variation in measurement error from data revisions; an application to forecasting in dynamic models', *Queen Mary University Working Papers*, (520),

Kapetanios G., Yates T. (2009), 'Estimating time-variation in measurement error from data revisions;

an application to forecasting in dynamic models', Journal of Applied Econometrics, forthcoming.

Kapetanios G., Yates T. (2004b), Estimating time-variation in measurement error from data revisions; an application to forecasting in dynamic models.

Karlsson S. (2013), 'Forecasting with bayesian vector autoregression', *Handbook of economic forecasting*, volume 2, pp. 791–897, Elsevier.

King T. E., Gonzalez Fortes G., Balaresque P., Thomas M. G., Balding D., Maisano Delser P., Neumann R., Parson W., Knapp M., Walsh S. et al. (2014), 'Identification of the remains of king Richard III', *Nature communications*, pp. 5–5631.

Knüppel M., Schultefrankenfeld G. (2019), 'Assessing the uncertainty in central banks inflation outlooks', *International Journal of Forecasting*.

Knüppel M., Schultefrankenfeld G. et al. (2012), 'How informative are central bank assessments of macroeconomic risks', *International Journal of Central Banking*, 8(3), pp. 87–139.

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

Manski C. F. (2015b), 'Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern', *Journal of Economic Literature*, 53(3), pp. 631–653.

Manski C. F. (2019), 'Communicating uncertainty in policy analysis', *Proceedings of the National Academy of Sciences*, 116(16), pp. 7634–7641.

Mazzi G.L., Ruggeri-Cannata R. (2017): 'The trade-off between timeliness and accuracy'

Mazzi G.L., Mitchell J., Carausu F. (2020): 'Measuring and communicating the uncertainty in official economic statistics', *Journal of Official Statistics*.

Mevik A.K. (2004): 'Uncertainty in the Norwegian business tendency survey', *Statistics Norway, Statistical Methods and Standards.*

Mitchell J., Hall S.G. (2005): 'Evaluating, comparing and combining density forecasts using the klic with an application to the Bank of England and Niesr fancharts of inflation', *Oxford bulletin of economics and statistics*, 67:995-1033.

Money and Banking. 'Communicating monetary policy uncertainty'. https://www.moneyandbanking.com/commentary/2019/4/21/communicating-monetary-policy-uncertainty, Accessed: 2019-06-21

Morgenstern O. et al. (1963): 'On the accuracy of economic observations', Princeton University Press.

Patton A.J., Timmermann A. (2007): 'Testing forecast optimality under unknown loss'. *Journal of the American Statistical Association*, 102(480):1172-1184.

Plűmper T., Neumayer E. (2012): 'Model uncertainty and robustness tests: Towards a new logic of statistical inference'. *Available at SSRN 2129113.*

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Hőhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. (2012): 'Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space', *Systematic biology*, 61(3):539-542.

Rossi B., Sekhposyan T. (2015): 'Macroeconomic uncertainty indices based on nowcast and forecast error distributions', *American Economic Review*, 105(5):650-55.

Sauter O., Schroff S., Spankowski U., Burghof H. (2013): 'Talking by numbers: communicated uncertainty of the ECB', *Intereconomics*, 48(5):287-292.

Stigler S.M. (1986): 'The history of statistics: The measurement of uncertainty before 1900', *Harvard University Press*.

Stocker T. (2014): 'Climate change 2013: the physical science basis', Working Group I contribution to -the Fifth assessment report of the Intergovernmental Panel on Climate Change, Cambridge

Universtity Press.

Sumner S. (2017): 'GDPR: Confusion reigns as experts disagree as to whether it's already in force'. https://www.computing.co.uk/ctg/news/3006534/gdpr-confusion-reigns-as-experts-disagree-as-to-whether-its-already-in-force, Accessed: 2019-06-21

Sveriges Risksbank (2019a): 'How is inflation measured?'. https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target/how-is-inflation-measured/, Accessed: 2019-06-21

Sveriges Riskbank (2019b): 'Monetary policy report'.

https://www.riksbank.se/globalassets/media/rapporter/ppr/engelska/2019/190425/monetary-policy-report-april-2019.pdf, Accessed: 2019-06-21

UK Office for National Statistics (2019): 'A guide to labour market statistics'.

https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/aguidetolabourmarketstatistics, Accessed: 2019-06-21

US Department of Labor, Bureau of Labor Statistics (2019): 'Reliability of estimates'. https://www.bls.gov/iif/osh_rse.htm, Accessed: 2019-06-21

Van der Bles A.M., van der Linden S., Freeman A.L.J., Mitchell J., Galvao A. B., Zaval L., Spiegelhalter D.J. (2019): 'Communicating uncertainty about facts, numbers and science', *Royal Society Open Science*, 6(5):181870.

Varoquaux G. (2018): 'Cross-validation failure: small sample sizes lead to large error bars', *Neuroimage*, 180:68-77.

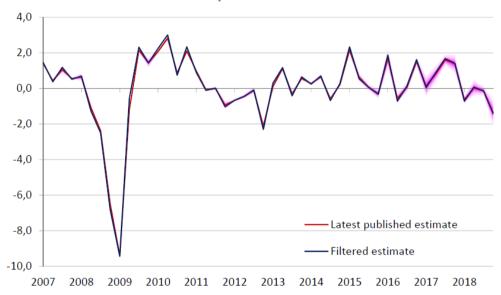
Wallis K.F. (2004): 'An assessment of Bank of England and national institute inflation forecast uncertainties', *National Institute Economic Reviews*, 189(1):64-71.

Wang Y., Lee T. (2014): 'Asymmetric loss in the greenbook and the survey of professional forecasters', *International Journal of Forecasting*, 30(2):235-245.

West S., King V., Carey T.S., Lohr K. N., McKoy N., Sutton S.F., Lux L. (2002): 'Systems to rate the strength of scientific evidence'. *Evidence report/technology assessment*, 47:1-11.

10 Annex

Industrial production euro area



Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.

Industrial production Germany

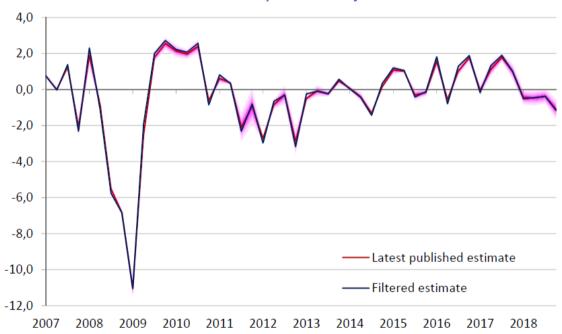


Industrial production France

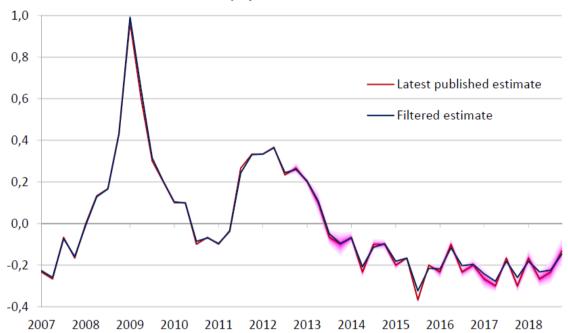


Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.

Industrial production Italy

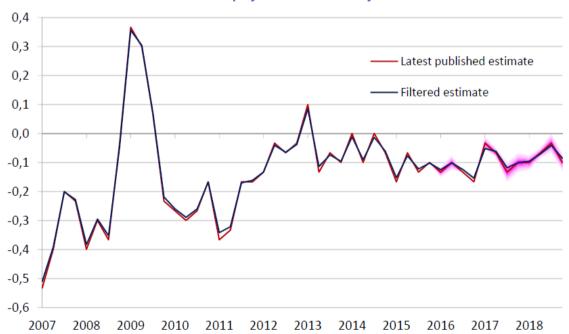


Unemployment rate euro area

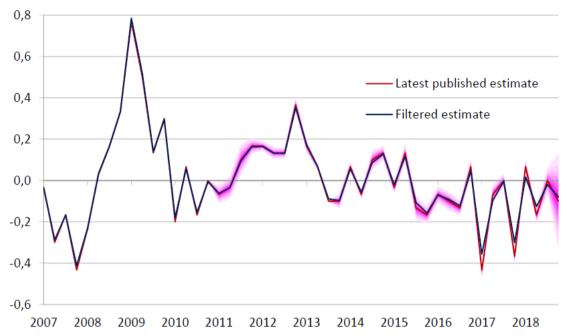


Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.

Unemployment rate Germany

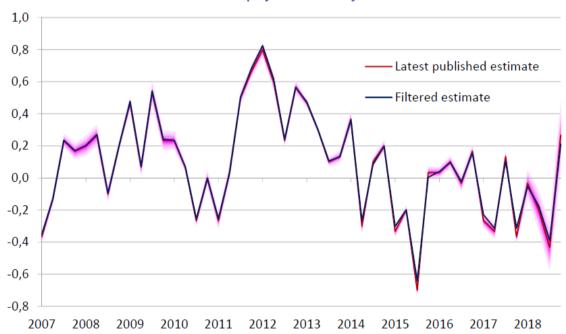


Unemployment rate France

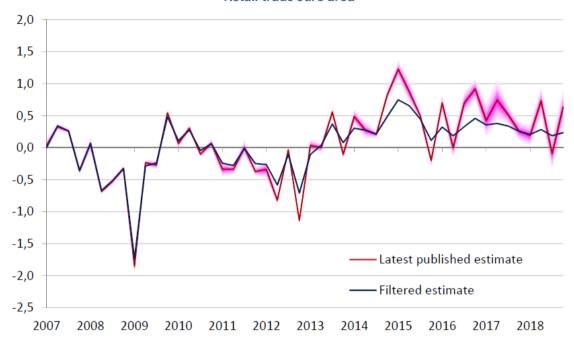


Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.

Unemployment rate Italy



Retail trade euro area

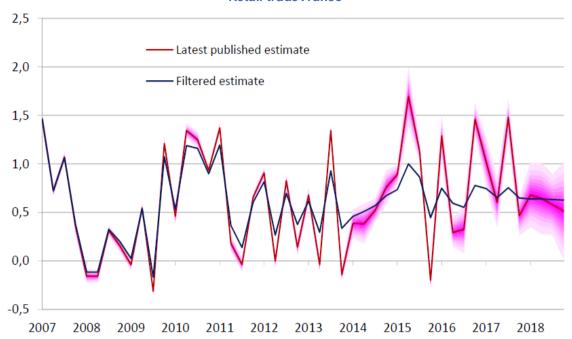


Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.

Retail trade Germany

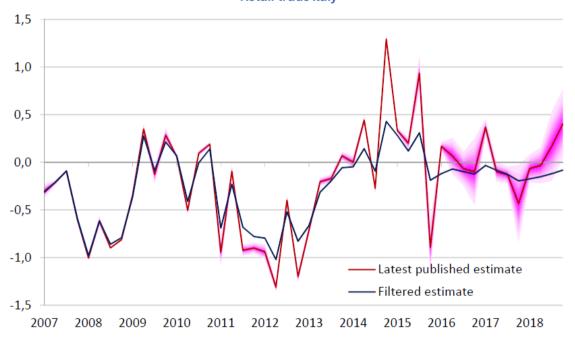


Retail trade France



Note: The pink area shows the confidence interval around the latest published estimates, in steps of 5 percentage points.

Retail trade Italy



GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://op.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

Measuring and communicating uncertainty in official statistics: State of the art and perspectives

This paper examines the state-of-the-art methodologies that are available to measure and communicate uncertainties on official statistics in reports, publications and online resources. Our aim is to investigate the available tools for a more accurate representation of inherent uncertainties underlying economic and social statistics. For a complete assessment, we evaluate different categories of uncertainty ranging from sampling and non-sampling ones to conceptual uncertainty. We also briefly address some aspects of uncertainty, not necessarily related to its measurement, which could help in better understanding and interpreting it. Moreover, we discuss the quantification of uncertainty through quantitative methods and we investigate different communication tools applied by prominent European agencies. Further, together with some concluding remarks, we formulate a first series of recommendations addressed to statistical authorities.

For more information

https://ec.europa.eu/eurostat/

