

Measuring and Communicating the Uncertainty in Official Economic Statistics

Gian Luigi Mazzi¹, James Mitchell², and Florabela Carausu³

Official economic statistics are uncertain even if not always interpreted or treated as such. From a historical perspective, this article reviews different categorisations of data uncertainty, specifically the traditional typology that distinguishes sampling from nonsampling errors and a newer typology of Manski (2015). Throughout, the importance of measuring and communicating these uncertainties is emphasised, as hard as it can prove to measure some sources of data uncertainty, especially those relevant to administrative and big data sets. Accordingly, this article both seeks to encourage further work into the measurement and communication of data uncertainty in general and to introduce the *Comunikos* (COMMunicating UNCertainty In Key Official Statistics) project at Eurostat. *Comunikos* is designed to evaluate alternative ways of measuring and communicating data uncertainty specifically in contexts relevant to official economic statistics.

Key words: Measurement error; data revisions; official economic statistics; data communication.

1. Introduction

Official economic statistics are inevitably uncertain or, put another way, subject to “errors”, even if not always interpreted or treated as such. Data uncertainty can affect the economic historian’s view of the past and policymaker’s decisions in the present.

Statistical (or measurement) “error” is commonly defined as the difference between the estimate produced by the statistical office and the “true” population value, which is typically unobserved. As Eurostat (2015) states: “Accuracy refers to the closeness of computations or estimates to the exact or true values that the statistics were intended to measure”. But, as Groves and Lyberg (2010) discuss, the notion of a *true* value is itself subject to debate. Accordingly, to define, yet alone measure, “uncertainty” remains elusive. For a general discussion of uncertainty as a concept see Van der Bles et al. (2019). Nevertheless, the idea that the “error” is the difference between the estimated and actual value (even if this itself is unobserved) is at least widely agreed; for example see Dungan et al. (2002) and Fuller (1987).

¹Freelance consultant. Email: glmazzi@pt.lu

²Federal Reserve Bank of Cleveland, 1455 E 6th St, Cleveland, Ohio, 44114, USA. Email: James.Mitchell@clev.frb.org

³GOPA Luxembourg, Rue Luxembourg, Bereldange, 7240, Luxembourg. Email: florabela.carausu@gopa.lu

Acknowledgments: Thanks to the Editor (Francesca Di Iorio), an Associate Editor, two referees, Dario Buono, George Kapetanios, Edwin de Jonge and Massimiliano Marcellino for helpful comments. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

Accordingly, statisticians and statistical offices have sought to categorise and communicate data uncertainties in various ways. This reflects a long history, dating back at least to [Kuznets \(1948\)](#) and [Morgenstern \(1950\)](#), that emphasises the uncertainty of economic statistics. For example, Principle 12 of Eurostat’s European Code of Practice states that “Sampling errors and non-sampling errors are measured and systematically documented according to the European standards” (see [Eurostat 2017](#)); and also “Revisions are regularly analysed in order to improve statistical processes”. Despite this, arguably, statistical errors have received limited attention in the seventy years since Kuznet’s discussion of ‘margins of errors’ and Morgenstern’s comprehensive catalogue of errors that affect economic data; for example see [Bagus \(2011\)](#). Certainly, as [Manski \(2015\)](#), [Manski \(2018\)](#) and [Van der Bles et al. \(2019\)](#) emphasise, headline statistical estimates tend to be presented as point estimates, arguably conveying a misleading degree of reliability in these data.

This lack of communication of economic data uncertainty is common across national statistical offices and, in turn, in the media when they disseminate statistical office data. Over more recent years, following the encouragements of Manski and others, several statistical authorities and organisations have started investing in identifying ways to measure and communicate data uncertainty; these include: the use of fan charts at the Bank of England and the Riksbank to communicate historical GDP data uncertainties; work by CBS Netherlands on ‘Visualising uncertainty’ and on the inventarisation of uncertainty sources; and UK Government Statistical Service guidance on ‘Communicating Uncertainty and Change’ (see [ONS 2018a](#)).

Measuring uncertainty is a complex and challenging task that can involve the use of sophisticated statistical and econometric techniques (classical or Bayesian) and subjective judgement to quantify the data uncertainties. However, as challenging as the quantification of data uncertainties *per se*, is how to communicate them – ideally in a way that is both ‘comprehensive’, in terms of capturing fully the uncertainties, but also ‘understandable’ so that different users and readers of these data correctly infer and interpret the uncertainties communicated to them.

Accordingly, in late 2018 Eurostat launched the *Comunikos* (COMmunicating UNcertainty In Key Official Statistics) project. *Comunikos* explores and analyses tools for measuring and communicating data uncertainties. The aim is to formulate proposals and make recommendations for the most appropriate ways to measure and communicate uncertainties for official statistics. As the risk of misleading or indeed confusing users is arguably high, but providing clear and ‘accurate’ uncertainty measures may enhance the relevance and credibility of official statistics, *Comunikos* investigates the *pros* and *cons* of communicating uncertainties to users of official statistics. In particular, it shows the potential for state-space models to measure uncertainty and produce confidence intervals or densities for official statistics, mainly for intra-annual time-series data ([Kapetanios et al. 2021](#)); analyses methods to measure uncertainty in cross-sectional data focusing on poverty measures ([Laureti and Benedetti 2020](#)) and considers verbal, quantitative and visual tools to communicate uncertainty in official statistics reports ([De Jonge 2020](#)). The aim of *Comunikos* is that by providing additional uncertainty information, users of official statistics will be able to make better decisions, in particular at times of heightened data uncertainty that we might expect to occur precisely (e.g. at business cycle ‘turning points’) when users are most interested in the data.

In this article, to help further the Comunikos project agenda and more generally encourage work measuring and communicating data uncertainty, we provide a methodological review and categorisation of uncertainty measures and their sources for economic statistics. We focus on quantitative economic data. Despite this qualification, the discussion below also has relevance for qualitative, ordinal data, although we refer to [Piccolo and Simone \(2019\)](#) for a more specific discussion of uncertainties for ordinal data. We exploit the fact that statisticians commonly categorise uncertainties to reflect non-sampling and sampling errors. Non-sampling errors apply to administrative records and surveys, including censuses, whereas sampling errors apply only to sample surveys. In principle, therefore, the *total* uncertainty associated with statistical output comprises both sampling error and non-sampling error. Though, in practice, the measurement of the total survey error is difficult, given the complexity of estimating and quantifying both sampling and, in particular, non-sampling errors. As [Boumans \(2012\)](#) discusses, while sympathetic to Morgenstern's call for the use of errors to accompany economic statistics, [Kuznets \(1950\)](#) emphasised the challenges implied by [Morgenstern \(1950\)](#) for measurement. [Kuznets \(1950\)](#) argued that economic statistics are better thought of as the products of evolving institutions, rather than making analogies, as Morgenstern did, with scientific data from controlled experiments. [Groves and Lyberg \(2010\)](#) discuss the conceptual history of total survey error over more than seventy years. The total survey error of an estimate is considered as an indicator of data quality.

As there are various ways in which these *errors* can be classified, a recent distinction and the proposed typology of [Manski \(2015\)](#) is also introduced and discussed. This is helpful in explaining how the different types of uncertainty can be communicated quantitatively – and ideally they then comprise part of the *statistical output*. Other frameworks, and proposals to classify errors have also been proposed including [Morgenstern \(1950\)](#) and [Verma et al. \(2010\)](#). [Morgenstern \(1950\)](#), as drawn on throughout this review, distinguishes ten sources of uncertainty:

- (1) lack of designed experiments;
- (2) hiding of information, lies;
- (3) the training of observers;
- (4) errors from questionnaires;
- (5) mass observations;
- (6) lack of definition and classification;
- (7) errors of instruments;
- (8) the factor of time;
- (9) observations of unique phenomena;
- (10) interdependence and stability of errors.

These are nested in the classification below. It is also common, as discussed in [Van der Bles et al. \(2019\)](#), to distinguish 'aleatory uncertainty', due to the fundamental indeterminacy or randomness in the world, from 'epistemic' uncertainty. Epistemic uncertainty is arguably what matters for statistical data that generally, but not always, seek to measure past or present (via a nowcast) phenomena. That is, our focus is on numbers that we currently do not know but could, at least in theory, know if only the information set were more complete. In contrast, 'aleatory uncertainty' generally relates to future events which we cannot know for certain.

This article is structured as follows. In Subsections 2.1 and 2.2 we review the traditional typology of data uncertainty that distinguishes sampling from nonsampling errors. In each case, we further break down the sources of sampling and nonsampling uncertainties, emphasising the importance of nonsampling uncertainties for administrative and big data sources. Throughout the importance of measurement – of the elements comprising uncertainty – is emphasised. As without measurement, it is hard if not impossible to gauge the consequences and importance of these uncertainties. Section 3 then introduces the new typology of uncertainties of [Manski \(2015\)](#). It is argued this confers some conceptual advantages when it comes to measuring and quantifying the different elements of total uncertainty. Section 4 then discusses the consequences, known and unknown, of data uncertainties. Section 5 concludes. In so doing it discusses the future agenda by repeating the call of [Manski \(2015\)](#) for more empirical research to study the effects of data uncertainty on the public's understanding, interpretation and use of official statistics.

2. The Traditional Typology of Data Uncertainties

2.1. Sampling Errors

Unlike the natural sciences, as [Morgenstern \(1950\)](#) emphasised, official economic statistics are not produced via repeated experiments. Instead, surveys are often run by national statistical institutes to measure the economic variables of interest. In Subsection 2.2.1, we turn to consider the growing use of administrative and big data sources in official statistics.

Sampling error is the most commonly reported measure of statistical uncertainty. This is because, unlike nonsampling errors discussed in Subsection 2.2 below, sampling error can be quantitatively estimated for many – but not all – sample surveys. Sampling error is the uncertainty or variability in an estimate that results from using a sample from a population rather than conducting a census or complete enumeration of the population.

If a sample from the population is chosen randomly, for example, then each random sample will involve sampling some different units and imply that each sample will produce different sample 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 is high. Then there is a larger chance that the survey estimate is far from the true population value. In a census when the entire population is surveyed there is no sampling error, but nonsampling errors still exist.

As summarised by the Office for National Statistics (ONS) in the United Kingdom (see [ONS 2017b](#)), standard errors are typically influenced by a number of factors that include:

- the survey sample size – a larger sample size will reduce standard errors;
- the variability in the population – when measuring a more variable characteristic, standard errors will be larger;
- the survey sample design – for example, any stratification or clustering used; and
- the estimation method used.

2.1.1. Measures of Sampling Error

Measures of sampling error associated with an estimate are typically based on estimates of the standard error and the mean squared error. The latter can be decomposed into the sum

of the square of the sampling bias and the sampling variance. The standard error is often used to compute the coefficient of variation or margin of error, both of which are related measures of the amount of uncertainty in the estimate.

Standard errors to measure sampling error can be computed – given the actual or assumed nature of the survey sample design. When sampling biases are zero or close to zero the standard error can be taken to represent total sampling error. This assumes that “population uncertainty” does not exist, which in practice is a strong often unrealistic assumption. Population uncertainty (see [Plumper and Neumayer 2012](#)) arises from the reality that ‘random samples’ from a given population may not be random, when there is uncertainty about who forms part of the population. As discussed in [Plumper and Neumayer 2012](#)), oversampling and sample selection corrections can be used to tackle population uncertainty.

There exist many approaches (e.g. see [Goedeme 2013](#)) to estimate standard errors. For an “easy-reading” discussion see [Peters \(2001\)](#). *Direct* estimators, which rely on analytic variance formulae, can be distinguished from *indirect* or resampling methods, like the bootstrap. Bootstrap methods involve taking a large number of draws from the original sample to mimic the actual sampling process: the sampling distribution of the target statistic across these bootstrap draws then measures the uncertainty. Whichever approach is used, the sampling process and the estimation procedure should ideally be acknowledged; for example see [Goedeme \(2013\)](#) who considers the complexities involved in estimating standard errors specifically for index numbers from complex surveys, and the discussion in [Seljak \(2006\)](#). For a detailed analysis of how uncertainty can be measured for mid-year population estimates given knowledge of the sampling processes and an understanding of the qualitative importance of different sources of uncertainty, see [ONS \(2017a\)](#).

As another example, [ONS \(2017a\)](#) consider how the bootstrap can be used to quantify the main uncertainties associated with UK migration data. Their method acknowledges the different data sources (including census, survey and administrative data) that are used to measure migration. Also see [Mevik \(2004\)](#) for a detailed study of sampling errors, from the Norwegian Business Tendency Survey, that contrasts “design-based” measures of the standard error with “model-based” ones that again make use of the bootstrap. For a general discussion of design-based versus model-based methods see [Koch and Gillings \(2006\)](#).

2.1.2. Challenges Measuring and Quantifying Sampling Errors

As discussed, standard errors can be computed – given the actual or assumed nature of the survey sample design. For statistical estimates of variables like GDP, the complexities involved in measuring the components, whether on the income or expenditure side, mean that it is not obvious what the survey design is and how this can lead to analytical expressions for the standard error.

To quote [ONS \(2019\)](#): “The estimate of GDP . . . is currently constructed from a wide variety of data sources, some of which are not based on random samples or do not have published sampling and non-sampling errors available. As such, it is very difficult to measure both error aspects and their impact on GDP. While development work continues in this area, like all other G7 national statistical institutes, we don’t publish a measure of the sampling error or non-sampling error associated with GDP”. This quotation is, in fact, remarkably reminiscent of the discussion in [Kuznets \(1948, p.176\)](#): “The treatment of

margins of error is most difficult for the national income and product statistics. The totals are a composite of a great variety of data, which differ in reliability from sector to sector of the economy. The margin of error in the composite totals is thus a complex amalgam of errors in the parts whose magnitude is not easily determined”.

But, as [Kuznets \(1948\)](#) emphasised, this complexity should not imply that attempts to measure these margins of error should not be made. [Kuznets \(1948\)](#), in fact, sought to quantify the uncertainties in GDP via expert judgement – famously concluding that there was a 10% margin of error associated with GDP.

However, it is possible to provide data-based and quantitative indications of “transitory” statistical uncertainties associated with GDP estimates by analysing historical revisions. National statistical offices and central banks accordingly often now publish realtime data vintages and analyse the implied revisions (e.g. see [Croushore and Stark 2001](#)). Other sources of uncertainty, for example due to limitations of the survey methodology, are not represented; and methodological work on measuring non-sampling errors continues (e.g. see [Manski 2016](#)).

There is also a long tradition (dating back at least to [Stone et al. 1942](#)) of exploiting the fact that for some variables there are multiple measures – albeit perhaps ones based on different sampling approaches. In particular, as a leading example, GDP can be estimated by the production, expenditure and income approaches. In principle, all three of these measures should be equal; but they are not in practice, given that they are calculated from different samples. But comparison of these approaches, assuming they all seek to measure the latent variable “true GDP”, can be used to produce so-called balanced or reconciled estimates of “true GDP” that also quantify the “statistical” or “measurement” error, as it is commonly referred to in this literature (e.g. see [Smith et al. 1998](#); [Aruoba et al. 2016](#)). Another cross-country example of how measurement errors can be quantified by comparing alternative estimates is how a specific country’s trade balance statistics can be compared with estimates from their trading partners: one country’s exports are another country’s imports.

2.2. *Nonsampling Errors*

It is more challenging to categorise, and certainly to measure and quantify, nonsampling errors for official statistics. Nonsampling errors stem from the design, data collection and processing methods used. As also seen in the typology of [Morgenstern \(1950\)](#), these errors often stem from lack of knowledge of the “nature of the data” given that the data are typically not measured by designed experiments. Nonsampling errors affect administrative (such as census) data as well as survey-based statistics. In general, sampling errors decrease as the sample size increases, but non-sampling errors increase as the sample size increases.

A common typology of nonsampling errors (e.g. see [Biemer and Lyberg 2003](#); [Eurostat 2019](#); [US Census Bureau 2018](#); [Statistics Canada 2017](#); [Australian Bureau of Statistics 2013](#); [National Science Foundation 2018](#)) is to decompose the nonsampling errors into the five elements listed in the typology below:

- (1) **Specification error:** Survey questions often cannot and/or do not perfectly measure the concept which they are intended to measure. For example, if asked to report whether they have a disability, respondents may have different subjective views of

what constitutes a *disability* and accordingly they provide different answers; as another popular example, the number of patents does not perfectly measure the quantity of invention in a macroeconomy. As emphasised in Manski's typology below, there can also be *classification* errors (perhaps reflecting conceptual uncertainties), for example, reflecting whether to classify some expenditure component of GDP as investment or consumption. Difficulties, and therefore errors, can also arise when, for instance, classifying economic activity to different industries. Economic activity is rarely confined to one specific industry.

- (2) **Coverage (or frame) error:** Coverage error occurs when the sample (frame) is inaccurate or incomplete, as a unit in the sample is erroneously excluded or included (e.g. duplicated), leading to *under* or *over* coverage errors. These errors make the survey less representative of the underlying population. The correction of coverage errors can be expensive, involving survey redesign and undertaking new surveys. A specific source of error that might be interpreted as stemming from a coverage error (albeit one known to the statistical office), arises in a mixed frequency data environment when statistical offices use temporal disaggregation methods to interpolate missing data at the higher frequency using observed data on higher frequency *indicator* variables. In effect, the temporal disaggregation methods fill in the gaps left by the incomplete survey evidence at the higher frequency. For example, monthly estimates of GDP are not commonly published by national statistical offices. So temporal disaggregation methods (from univariate models such as [Chow and Lin \(1971\)](#) to multivariate dynamic extensions, such as [Mitchell et al. \(2005\)](#) and [Frablet et al. 2011](#)) have been used to estimate monthly GDP based on the monthly movements of a range of observed indicator variables believed to relate to (unobserved) monthly GDP. Importantly, these methods impose the constraint that the interpolated monthly estimates for GDP add up to the quarterly totals published by the statistical office.
- (3) **Nonresponse error:** Nonresponse error occurs when not all units of the sample respond to the survey. This leads to a difference between the statistics computed from the collected or observed data and those that would be computed if there were no missing values. Two types of nonresponse can be delineated: (1) unit nonresponse, that is, when no data are collected about a population unit; and (2) item nonresponse, that is, when data on some but not all the survey data items are collected for a given population unit. Nonresponse can cause nonresponse bias (as well as nonresponse variance) when the observed sample differs systematically from those who do not respond (the unobserved sample). For example, complete or partial nonresponse is often more likely among lower-income or less-educated respondents or firms facing serious financial difficulties. The nonresponse rate can usually be accurately measured – as the ratio of the number of completed surveys to the total number of sample units. In turn, response rates therefore indicate the proportion of sample units that respond to the survey. However, these nonresponse rates do not help the user of the statistic directly infer, for example, the standard error of the estimate. They are therefore of limited direct use, as ideally the user would be provided with an estimate of the standard error. Subsection 2.2.1 below considers recent work in econometrics that has sought to quantify nonresponse errors directly.

- (4) **Unit-level measurement error:** response error and interviewer error: Measurement errors stem from what is observed or measured by the survey differing from the actual values for the sample units. Measurement errors, as defined here, relate to the accuracy of measurement at the unit level. In turn, measurement errors can be broken down into response errors and interviewer errors. Response errors arise when respondents knowingly or unknowingly provide inaccurate responses. These errors might arise due to inherent cognitive biases (e.g. a tendency for a respondent to give an answer that they believe is correct or will please the interviewer) and poorly designed survey questionnaires that lead to misunderstandings about what is being asked. Interviewer errors arise when the person undertaking the survey, whether on purpose or not, records incorrect responses or consciously or unconsciously influences the respondent with the effect that they provide inaccurate responses.
- (5) **Processing error:** Processing errors include errors in recording, checking, coding and preparing survey data. They can include interpolation and extrapolation errors for missing or, what are believed to be, *inconsistent* data. In some contexts (some of) these errors can be measured and quantified. [Van der Loo et al. \(2017\)](#) consider the variance caused by data cleaning. They note that survey data sets, $X = \{x_i\}_{i=1}^n$, often suffer from missing values, outliers and incorrect values that preclude the applicability of a simple estimator $f(X)$ such as the sample mean. Although, as touched on below, robust estimators may be more appropriate. So data editing processes are used to transform the raw data set X into a new data set, Y , which is then used for estimation. The population estimator is also given by $f(Y)$. But as Y is a transformation, the variance of the estimator after editing is no longer simply the variance of the new data set but should also reflect the *extra* variance induced by the data editing processes. This extra variance may comprise estimation uncertainties as well as sampling variance. As [Van der Loo et al. \(2017\)](#) explain, in general – in real-life practical examples of interest to official statistics – it is hard to obtain analytical expressions for this composite uncertainty; and they therefore suggest a computational approach to measuring the variance that uses the bootstrap. But in other contexts isolating and removing processing errors (without a warning or help from the statistics office) simply by inspecting a published time-series can be challenging. It amounts to having to define and then isolate *outliers*. This raises identification challenges, since an outlier could be due to variability in the sampling processes rather than a processing (or measurement) error. Interestingly, given that some statistical estimators are more robust to outliers (or more generally to uncertainty whatever the source) than others, the use of robust estimators may offer promise when communicating data in the presence of uncertainties. For example, the median rather than the mean offers a robust measure of central tendency.

2.2.1. Measures of Nonsampling Error

As emphasised, nonsampling errors are typically hard to measure and quantify. But, as [Manski \(2015\)](#) emphasises, this does not justify ignoring them.

Statistically, it can be helpful to consider that nonsampling errors can be classified into two groups: random errors and systematic errors. Random errors are the unpredictable (ideally, independently and identically distributed) errors. They generally cancel out if a

large enough sample is used. They lead to increased variability in the statistic, but no bias. In turn, systematic errors are errors that accumulate. For example, if there is an error in the survey or questionnaire design, this causes errors in respondents' answers, often leading to biases. We first consider, in Subsubsection 2.2.2 how measures of nonsampling error may be computed using partial identification methods. Then, in Subsubsection 2.2.3, we consider the scope to measure nonsampling errors for administrative and big data.

2.2.2. Partial Identification and Measures of Survey Nonresponse

Recent work on “partial identification” (Manski 2016) has shown how, with access to the underlying micro data, more could be said about nonsampling uncertainties (for aggregated data) – and in particular nonsampling errors due to survey nonresponse – than at present is commonplace.

The basic idea is that in the presence of missing data, sample statistics can still be computed. But to measure the nonsampling errors, due to missing data, these statistics can be computed taking into account all the values that the missing data might take. This delivers interval rather than point estimates. The approach of Van der Loo et al. (2017), who as mentioned uses the bootstrap, can also be interpreted within this framework.

In simple terms, Manski (2016) sets out how if one lets $P(y|z = 1)$ denote the distribution of random variable Y for those units who report y (denoted, $z = 1$; $z = 0$ otherwise), then from the law of total probability

$$P(y) = P(y|z = 1)P(z = 1) + P(y|z = 0)P(z = 0) \tag{1}$$

The sample evidence reveals $P(z)$ and the observables $P(y|z = 1)$ when $P(z = 1)$. But the sample evidence is uninformative on $P(y|z = 0)$. Therefore the sample evidence reveals that $P(y)$ lies in the identification region

$$H[P(y)] = [P(y|z = 1)P(z = 1) + \gamma P(y|z = 0)P(z = 0), \gamma \in \Gamma_Y] \tag{2}$$

where Γ_Y denotes the set of all probability distributions on the set Y . As discussed by Manski (2016), the notion of the identification set can then be used for meaningful inference. For example, suppose the statistics office is interested in quantifying the probability that Y falls within some interval or set, B i.e. $P(y \in B)$. Then, again by the law of total probability,

$$P(y \in B) = P(y \in B|z = 1)P(z = 1) + \gamma P(y \in B|z = 0)P(z = 0) \tag{3}$$

and the empirical evidence reveals $P(y \in B|z = 1)$, $P(z = 1)$ and $P(z = 0)$. But it does not reveal $P(y \in B|z = 0)$. However $P(y \in B|z = 0)$ must lie between 0 and 1. This yields the sharp bound on $P(y \in B)$:

$$P(y \in B|z = 1)P(z = 1) \leq P(y \in B) \leq P(y \in B|z = 1)P(z = 1) + P(z = 0) \tag{4}$$

If the statistician is willing and/or able to make assumptions on the nature of the nonresponse, that restrict $P(y|z = 0)$ within some probability space (e.g. to a specific set of density functions), then these bounds can be made tighter.

2.2.3. Nonsampling Error: Administrative and Big Data

The increasing availability and use of administrative and big data, including from new data sources (such as the internet and social media), raises both new challenges and opportunities for the measurement and quantification of uncertainties, especially those coming from nonsampling errors. The hope is to exploit some data source that provides the “true” estimate; comparison of other estimates, say from surveys, with this “true” estimate then provides a clear way to measure the statistical or measurement “error” of the other survey-based estimate.

But many of these administrative and big data sources were designed for purposes other than official data collection. Indeed, many of the data sources are from private companies, raising challenges as to data ownership and privacy. This has prompted experimental research into, for example, the use of apps (see [Gromme et al. 2017](#)) to measure directly data from the population (the “citizens”) rather than rely on third party data. As emphasised by [Hand \(2018\)](#), analysis of administrative data presents new statistical challenges not least that these data are, by definition, typically not random samples but so-called non-probability samples.

As stressed by [Kapteyn and Ypma \(2007\)](#) and [Abowd and Stinson \(2013\)](#), while administrative data and indeed big data sets in general offer the prospect of fewer non-response errors than traditional surveys, they still likely suffer from uncertainties in particular due to “measurement error”. They may not measure exactly the concept a researcher is interested in. And since administrative databases typically link data from different sources there is the possibility of mismatching, due to imperfect linkage information (e.g., errors in social security numbers). [Abowd and Stinson \(2013\)](#) therefore emphasise the errors that are present in all data sources; and, in the tradition of [Stone et al. \(1942\)](#), they specify a so-called prior weight vector used to define the “truth” as a weighted average of both the administrative and the survey data.

Therefore new sources of uncertainties, but also opportunities to reduce these, arise from this practice of matching administrative or big data with existing, more traditional, sources of data collected by the statistics office. Matching involves combining information available in distinct sample surveys about the same target population. For example, work by [Lui et al. \(2011\)](#) sought to match firm-level qualitative survey data from the Confederation of British Industry, a business organisation in the United Kingdom (that provides information on a range of variables not posed in official surveys) with those same firms’ responses to official surveys from the ONS. This sort of matched data set offers the prospect of both better understanding the nature and statistical properties of the non-official data and of cross-checking the accuracy of the new data.

[Kapteyn and Ypma \(2007\)](#) provide a framework to model the errors in administrative data due to mismatching, based on a comparison of the administrative and survey data when estimates from both are available. [Conti et al. \(2012\)](#) also consider the measurement of uncertainty in statistical matching. [Conti et al. \(2012\)](#) set out a model that can be used to estimate the joint distribution of variables observed in separate and independent surveys. Consider two surveys that deliver random variables Y and Z , with observations y and z , respectively. Both surveys also provide information on a set of additional random variables, X , with observations x . Let the two (known) marginal conditional distributions then be denoted, $F(y|x)$ and $F(z|x)$. In the spirit of “partial identification” as set out by [Manski](#)

(2016), Conti et al. (2012) measure uncertainty as the set of probability distributions of the random vector $(Y, Z|X)$ compatible with $F(y|x)$ and $F(z|x)$. Again a bounds-based approach is proposed as a way to quantify the uncertainty. Recent work by Oberski et al. (2017) extends analysis to estimate the extent of measurement errors in administrative data that measures the errors in administrative data, allowing both the administrative data and the survey data to be simultaneously subject to measurement errors.

Coverage errors, as discussed in Subsection 2.2, remain a concern for administrative and big data data. The administrative population is often a proxy for the target population. For example, the employment register in the Netherlands also contains employees that work but do not live in the Netherlands; but it misses Dutch inhabitants who have a job abroad. And with big data, often the observations cannot be identified/linked to a member of the target population; it is then not directly possible to find out how representative the data, often based on non-probability samples, are.

“Hiding of information”, one of the elements in the typology of errors listed by Morgenstern (1950), may afflict at least some administrative data. For example, when replying to the tax authorities, individuals or businesses may in a sense, as Morgenstern puts it, “deliberately lie” or at least obscure the truth. As Morgenstern (1950) writes, there is a long history of apparently venerable institutions falsifying or at least obscuring “facts” for strategic or political purposes. A more recent example of this typology of error is when Greek public finance data were investigated by the European Commission in 2010. In the final report of the European Commission’s investigation (see European Commission 2010) it was recognised that there were political pressures to obstruct accurate data collection. Nevertheless, growing use at statistical offices of individual and firm-level tax return data in the national accounts is delivering improvements both in terms of timeliness and accuracy; that is, the use since 2017 of Value Added Tax data in measuring GDP in the United Kingdom means that turnover data from 630,000 businesses now inform GDP estimates (see ONS 2018b), providing wider coverage.

Some of Morgenstern’s other elements in his typology are also probable sources of uncertainty for big data and administrative data. His “mass observations” may well involve errors that likely accumulate and do not necessarily cancel out. And his “errors of instrument” are also likely to become more important as economic statistics are increasingly collected by machines (e.g. scanners) rather than human beings (via surveys and questionnaires).

Measurement of these uncertainties for administrative and big data is in its infancy. Chambers (2014) considers how “model-based thinking” can help measure nonignorable nonresponse in surveys; and how adaptive surveys can be used to select a sample of nonrespondents to interview or survey at a second wave of the survey so as to minimise the nonresponse bias. ONS (2018a) recommend that for administrative data measures of coverage and completeness, editing rates and imputation rates should be measured and used as quantitative estimates of data uncertainty. And the aforementioned approach of Morgenstern (1950) is attractive in measuring uncertainties without having to assume that either the administrative data or the survey data are accurate. Hand (2018) provides a recent discussion; and calls for research to establish what the ‘generally accepted theory’ might be for the analysis of administrative data. The first of his challenges is, in fact, to consider how to define and communicate uncertainty for administrative data, given that the sources of uncertainty in administrative data are many and diverse, and may not include sampling variation.

3. Manski's (2015) Typology of Data Uncertainties

Manski (2015) re-interprets sampling and nonsampling uncertainty as comprising three elements:

- (1) “transitory” statistical uncertainty;
- (2) “permanent” statistical uncertainty;
- (3) conceptual uncertainty.

Transitory statistical uncertainty stems from publication of early data releases that are revised over time as new information arrives. For example, for many years the Office for National Statistics (ONS) in the United Kingdom published its first – so-called “preliminary” – quarterly GDP estimates around 27 days after the end of the quarter. Because this timeliness was achieved by basing their estimate on 44% of the sample, it is (and should be) no surprise to see the ONS revise these preliminary estimates as more sampling information subsequently became available to them. Interestingly, in the summer of 2018 ONS shifted back its production – so that the new so-called *first* estimate is now available *only* at about 40 days. But this delay *buys* the ONS a higher sampling fraction, and should therefore reduce transitory statistical uncertainty.

Permanent statistical uncertainty arises due to data incompleteness (e.g. nonresponse) or the inadequacy of data collection (e.g. sampling uncertainty due to a finite sample) which does not diminish over time. Therefore, permanent statistical uncertainty comprises elements of both sampling and nonsampling errors, as delineated in the typology above.

Conceptual uncertainty arises from a lack of understanding about what the statistics measure. It arises not from the statistics themselves, as with transitory and permanent statistical uncertainty, but from how the statistics are interpreted. Conceptual uncertainty is of course not a new element in many typologies of the sources of uncertainty, albeit it is one that is often ignored – given the challenges in measuring it. In fact, discussion of conceptual uncertainty again dates back to Morgenstern (1950), who discussed uncertainties arising due to a “lack of definition and classification”. As Morgenstern (1950, 35) explains, “the theoretical characteristics of, say, an industry or a ‘price’, are less well established than those of a wave length”.

It is perhaps helpful to begin to break down conceptual uncertainty by considering the following components or sub-elements:

- conceptual uncertainty due to different definitions and classifications. This overlaps with the eighth element in the typology of errors of Morgenstern (1950): “the factor of time”. Such uncertainties arise as economic data are often measured on a discrete basis with observations attributed to a specific window of time – so there are *errors* especially when there are changes in classification (e.g. changes in the definition of an industry or changes to the characteristics of a specific product);
- conceptual uncertainty due to differences in the compilation process (e.g. direct estimates based on surveys or administrative data versus indirect estimates using temporal disaggregation techniques); and
- conceptual uncertainty due to seasonal adjustment.

Conceptual uncertainty need not produce “errors” in the usual statistical sense – as [Morgenstern \(1950\)](#) explains. But differences of definition, for example, clearly result in uncertainties, revisions to estimates and doubts as to the use and interpretability of data and their comparability.

3.1. A Bayesian Approach

In principle, at a formal level, the approach of [Draper \(1995\)](#) offers a methodological way of understanding some aspects of conceptual uncertainty. If we consider these aspects as part of the “model” used to measure the underlying variable, then [Draper \(1995\)](#) provides an approach to think about both uncertainty about the form of the model (so-called structural uncertainty) and the parameters of the model (so-called parametric uncertainty). This motivates a Bayesian approach.

Bayesian Model Averaging (BMA) offers a conceptually elegant means of dealing with model uncertainty. BMA is an application of Bayes’ theorem, Model uncertainty is incorporated into the theorem by treating the set of models S as an additional parameter and then integrating over S , where $S \equiv \{S_i, i = 1, \dots, N\}$ with N models, and the models S_i are defined as continuous density functions $g_{it}(y_t)$ for the variable of interest y_t .

The posterior density estimate of the variable of interest y_t given ‘data’ Ω_t , $p_t(y_t | \Omega_t)$, is then defined as the weighted average of the densities $g_{it}(y_t) = \Pr(y_t | S_{it}, \Omega_t)$, where the weights w_{it} are the different model’s (or the different estimates’) posterior probabilities, $w_{it} = \Pr(S_{it} | \Omega_t)$:

$$p_t(y_t | \Omega_t) = \sum_{i=1}^N w_{it} g_{it}(y_t); (t = 1, \dots, T), \quad (5)$$

where $w_{it} \geq 0$ and $\sum_{i=1}^N w_{it} = 1$. $p_t(y_t | \Omega_t)$, or, for expositional ease suppressing dependence on the ‘data’ Ω_t when defining the posterior probabilities, equivalently $p_t(y_t)$, is the combined density estimate.

A Bayesian approach, due to its ability to handle multiple sources of uncertainty, also offers promise as a way to provide an integrated measure of total uncertainty – that integrates out uncertainty about sampling and nonsampling errors or transitory and permanent uncertainty. In addition, priors can be used to acknowledge if and when there is additional information that can be used to guide the data in the right direction.

3.2. Measures of Transitory, Permanent and Conceptual Errors

An advantage of the typology of [Manski \(2015\)](#) is that when it comes to actually measuring and quantifying the three elements of total uncertainty, the first element – transitory statistical uncertainty – when relevant, is at least usually measurable. It is measurable by analysing revisions to the statistics as more information becomes available as the statistic is revised. As elaborated on in Subsection 3.2.1, measurement and quantification of transitory statistical uncertainties is now facilitated by the relatively wide availability of real-time (vintage) data sets for macroeconomic variables such as GDP. These real-time data sets let one measure the revisions between successive estimates. Many authors have proposed models of data revisions – using real-time GDP data sets –

to model and forecast this “transitory” GDP data uncertainty (e.g. see [Jacobs and Van Norden 2011](#); [Cunningham et al. 2012](#); [Kishor and Koenig 2012](#); [Galvao 2017](#)).

In turn, measurement of the permanent and conceptual uncertainties is again challenging, as it is for sampling errors (at least for variables like GDP) and nonsampling errors. But attempts can still be made to communicate (at least some of) these uncertainties. A famous example of how uncertainties can be communicated even for a variable like GDP, which as discussed, is usually subject to multiple surveys precluding direct estimates of the data’s standard error, are the fan charts published by the Bank of England; see [Figure 1](#) for an example.

[Figure 1](#) provides an illustrative example of what these fan charts look like, taken from the Bank of England’s *Inflation Report*. Importantly, from [Figure 1](#) we see that the Bank seeks to quantify both future uncertainties but also past or historical data uncertainties. This is emphasised in the words that accompany the fan chart pictures in the Bank’s publications: “(t)o the left of the first vertical dashed line, the centre of the darkest band of the fan chart gives the Committee’s best collective judgement of the most likely path for GDP growth once the revisions’ process is complete” ([Bank of England 2007](#), 39). As the Bank of England explains, these fan charts should be interpreted as “the MPC’s best collective judgement of the most likely path for the mature estimate of GDP growth, and the uncertainty around it, both over the past and into the future.” [Figure 1](#) reveals that the fan becomes progressively narrower as one looks further back in time, as the data revisions’ process is more complete and fewer future revisions are expected to older

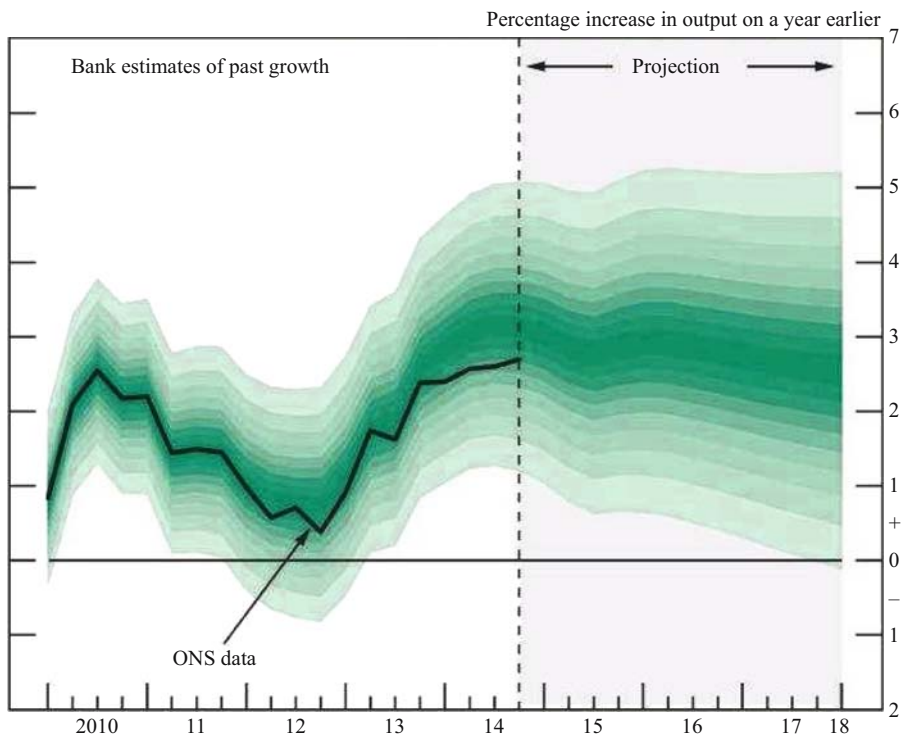


Fig. 1. Illustrative Bank of England fan chart –GDP data uncertainty.

estimates. The ONS's latest estimate of GDP growth is shown in [Figure 1](#) by the solid black line. [Cunningham and Jeffery 2007](#) provide an explanation of the data revisions model, used by Bank staff, that along with judgement from the Bank's Monetary Policy Committee helps shape the form of these backcast fan charts. Their model exploits historical patterns in ONS revisions and information from qualitative business surveys. The Bank assume that data uncertainty is determined by a Gaussian probability density function (see [Bank of England 2007](#)); and the mean of this probability density function does not always equal or have to equal the ONS's latest GDP estimate. This enables the Bank to quantify biases, as well as the variance around the point estimate. For a detailed discussion and an *ex post* calibration analysis of the Bank of England's probabilistic backcasts see [Galvao and Mitchell \(2020\)](#).

3.2.1. Measurement of Transitory Uncertainty: Real-time Data sets

When considering temporary or transitory uncertainty, the availability of a properly defined and updated real-time database constitutes a powerful tool. By a properly defined and updated real-time database we mean one where vintages are recorded by the date of official release. Commonly, in a real-time database, the data for each vintage (the date at which the data are released) are presented in a different column, with the rows referring to the date for which the time series is measured. The final data value shown in each column is therefore the first release of the data point indicated by the row. This means that real-time databases often take the form of data triangles – as more recent data vintages (columns to the right) provide data over longer (more recent) periods of time. Some real-time databases are based on vintages collected at fixed intervals, such as each week or month. In this case, especially when vintages are collected monthly, multiple releases within the month are lost, with only the latest stored. This can compromise the overall quality of the revisions analysis. Returning to the treatment of transitory uncertainty, comparison of the vertical (or column) data vintages can be used to identify characteristics of the revisions process, such as its bias and persistence. In addition, by considering diagonal data vintages (i.e. data down the diagonals of the data triangle) it is possible to study the different degrees of uncertainty associated with various releases as well as empirically check for the existence of a trade-off between timeliness and accuracy. Finally, by comparing horizontal (or row) data vintages it is possible to study uncertainty at specific points in time, such as those related to a period of crisis, to see if their intrinsic uncertainty is different to observations in more *normal* situations.

The increased accessibility of real-time data sets, in particular in the United States as maintained by the Federal Reserve Banks of Philadelphia and St. Louis, has enabled statisticians and economists to analyse and model data revisions; (for example, see [Croushore and Stark 2001, 2003](#); [Croushore 2011](#); [McCracken and Ng 2016](#)). Real-time databases, especially for GDP, are now available for various countries. This includes the Euro area (as maintained by the European Central Bank; see [Giannone et al. 2012](#)), Switzerland ([Bernhard 2016](#)), the United Kingdom (as maintained by the Bank of England), Canada (as maintained by Statistics Canada), New Zealand (as maintained by the Reserve Bank of New Zealand), and the OECD. These databases thus provide a means to measure and quantify transitory statistical uncertainty.

3.2.2. Measurement of Conceptual Uncertainty: Seasonal Adjustment

Conceptual uncertainty involves both subjective and objective components. Measurement of the former is more challenging; and there has been little or no work on it to date. Measurement, perhaps qualitative, would appear to require the design and use of new surveys to gauge, for example, the public's interpretation of GDP data – do the public understand correctly what GDP measures? For example, experimental evidence from 3000 members of the UK public suggests that a large fraction of the public do not correctly understand what GDP measures; see [Galvao et al. \(2019\)](#). But as the sub-classification of conceptual uncertainty suggests, some aspects of conceptual uncertainty can be measured quantitatively (*ex post*, i.e. after the revision) by examining revisions to statistics.

Revisions to seasonally adjusted real-time data can be decomposed into two separate but related sources. The first source is the application of the method used for seasonal adjustment. As seasonal adjustment involves application of a filter to the underlying series, with the passage of time as new data accumulate the weights attached in the filter to specific observations change and there are revisions to the seasonally adjusted estimates; see [Wallis \(1982\)](#) for further details and analysis. [Burrige and Wallis \(1985\)](#) discuss how the variance of the seasonally adjusted series can be calculated when the seasonal adjustment filter is recast as an optimal filtering problem in an unobserved components framework.

The second source of revisions is that for many series, like GDP, the unadjusted data are themselves revised by the statistics office. Attempts to quantify these revisions, paralleling the literature that has examined the revisions properties of output gap estimates (cf. [Orphanides and Van Norden 2002](#)), have involved recursive real-time application of the seasonal adjustment filter to the real-time unadjusted series. For example, [Mehrhoff \(2008\)](#) considers the empirical quantification of both these sources of uncertainty for selected German time series, using the real-time database of the Deutsche Bundesbank.

Additional uncertainties arise when calculating seasonally adjusted estimates for aggregated variables, such as Euro area GDP. This is because Euro area GDP involves the aggregation of GDP data for the member countries. The question then arises of when the series should be seasonally adjusted. One can distinguish two approaches. The “direct” approach consists of seasonally adjusting the raw data of the aggregate itself. The “indirect” approach consists of seasonally adjusting the raw data corresponding to the sub-components (national GDP) and then aggregating. In general, the direct and indirect adjustment for an aggregate series are not identical. Only, under certain conditions, for so-called uniform seasonal adjustment filters, such as X-11, does the order of seasonal adjustment and aggregation not matter. But when the filters differ, as they will when an optimal signal extraction method is used like an unobserved components model, the order is crucial. For further discussion see [Ghysels \(1997\)](#). For completeness we note that one can also distinguish a “multivariate” or simultaneous approach that has certain optimality properties, although since it is computationally demanding and requires hard choices to be made about the appropriate information set this approach is rarely considered; see [Geweke \(1978\)](#).

4. Consequences of Uncertainty

While the impact of sampling errors (in the first typology) or transitory statistical uncertainty (in Manski's typology) can be measured and quantified – at least for some

variables – as emphasised, measuring and quantifying nonsampling errors and permanent statistical uncertainties is much harder. This means it is not generally possible to measure quantitatively, for a specific variable, the relative importance of the different elements or components of total survey error as delineated in the two typologies. However, perhaps in part subjectively formed, attempts can still be made to communicate the total error, as the Bank of England's fan charts illustrate.

It is an open question whether it is better to try and communicate these data uncertainties or not; and if so, how? Is it best to communicate data uncertainties quantitatively, like in the Bank's fan charts, or qualitatively perhaps via textual caveats and qualifications that emphasise that the data are uncertain?

Understanding, and certainly measuring, the consequences of uncertainty requires a cross-disciplinary approach, involving the intersection of psychology, behavioural and decision science and statistics. Consistent with the conclusions of [Manski \(2015\)](#), it requires new empirical research to study the effects of uncertainty – and its communication or lack of – on users' understanding, interpretation and use of statistics. A start is made in [Galvao et al. \(2019\)](#) who conduct randomised controlled experiments and a targeted expert survey in order to assess if and how the UK public and experts interpret and understand GDP data uncertainty. They find that the majority of the UK public do understand that there is uncertainty inherent in GDP point estimates. But communicating uncertainty information to them improves their understanding of why data revisions happen but does not decrease trust in the data. Their results indicate that it is particularly helpful to communicate uncertainty information quantitatively via intervals, density strips and bell curves.

In the absence of (to date) a body of published research both across countries and economic variables on this neglected issue, here three ways of understanding and measuring the consequences of data uncertainty are discussed. First, we review the growing literature, especially in economic statistics, that has sought to analyse and model data revisions. Secondly, we provide a case-study illustrating how for GDP growth the size of data revisions – of transitory statistical uncertainty – varies both across time and countries. Thirdly, we emphasise how the effects of uncertainty, particularly of transitory statistical uncertainty, relate to the trade-off between the timeliness and accuracy of statistics.

4.1. Revisions: Real-Time Data Analysis

As [Croushore \(2011\)](#) reviews, over the last 15 years there has been a growing literature, especially in applied macroeconomics, on if and how data revisions matter. [McKenzie \(2006\)](#) delineates seven reasons for “revisions” including updated sample information, correction of errors, benchmarking, updated base period for constant price estimates and changes in statistical methodology.

Research has examined the properties of data revisions, how structural macroeconomic modelling is affected by data revisions, how data revisions affect forecasting, the impact of data revisions on monetary policy analysis, and the use of real-time data when nowcasting. This research has been supported by the increased, but still imperfect, availability of realtime data sets by central banks and statistical offices. Importantly, as [Croushore \(2011\)](#)

concludes, until these data sets became more widely available most economists thought that data revisions were likely to be small and did not matter. But this view has been shown to be misplaced by real-time research: data revisions are often found to be large and have important implications, including for policymakers like central banks.

The general framework often used to measure and analyse the properties of these data revision uncertainties is twofold. First, studies typically report the mean (or bias) of the revisions and test if these are statistically significant. Secondly, to provide more information on the nature of the uncertainties and the ensuing revisions, studies discriminate between *news* and *noise* revisions following the approach of [Mankiw and Shapiro \(1986\)](#).

Tests for whether revisions are *news* or *noise* are based on so-called forecast efficiency regressions:

$$(y_t^{new} - y_t^{old}) = \beta_0^{news} + \beta_1^{news} y_t^{old} + \varepsilon_t \quad (6)$$

$$(y_t^{new} - y_t^{old}) = \beta_0^{noise} + \beta_1^{noise} y_t^{new} + \varepsilon_t \quad (7)$$

where y_t^{new} denotes the latest or new estimate of variable y at time t , and y_t^{old} denotes the previous or older estimate of variable that is revised.

The null hypothesis that data revisions add information (they contain *news*) implies $\beta_1^{news} = 0$. If data revisions remove the measurement error (*noise*) in the initial release then $\beta_1^{noise} = 0$. If data revisions reduce noise, then data revisions are predictable. For additional details on the application of these tests see [Clements and Galvão \(2010\)](#) and references therein.

4.2. Case-study on GDP: Cross-country Comparisons Measuring Data Revisions

To illustrate the importance, or otherwise, of transitory statistical uncertainties we review recent cross-country comparisons, from existing studies, that have sought to compare GDP data revision *errors* across countries. These papers build on the pioneering work of [Mankiw and Shapiro \(1986\)](#) and [Faust et al. \(2005\)](#). [Faust et al. \(2005\)](#) found that in the G7 economies, revisions to GDP announcements are large – many revisions in quarterly GDP growth are over a full percentage point at an annualised rate. Moreover, they found that while US GDP revisions are largely unpredictable, as predicted by the *news* model, for Italy, Japan and the United Kingdom, about half the variability of subsequent revisions can be accounted for by information available at the time of the preliminary announcement – so there was evidence for *noise*.

While studies that measure and then quantify data uncertainties, due to data revisions, are a helpful method to measure transitory statistical uncertainties, as ever we should recall the last of the ten sources of uncertainty listed in the typology of [Morgenstern \(1950\)](#): “the interdependence and stability of errors”. Measures of data uncertainty based on historical revisions measure just that: “historical” data revisions. They are therefore only a good guide to current data uncertainties to the extent that we expect history to repeat itself. If the statistical office, for example, has improved its measurement processes over time we might well expect current data uncertainties to be less than historical ones.

We should emphasise that different ways of producing and estimating GDP across countries no doubt affect the balance or relative importance of the different sources of GDP data uncertainty.

In more recent work, [OECD \(2015\)](#) compares GDP revisions across OECD countries. [OECD \(2015\)](#) uses the mean revision to measure the importance of data revisions. [Figure 2](#), taken from [OECD \(2015\)](#), provides estimates of the mean revisions for a range of OECD countries, importantly using different measures of the *outturn* – the *final* estimate against which the first GDP estimate is compared. Of course, as data revisions are an ongoing process the true or final estimate is never in reality observed, so an assumption has to be made. In applied macroeconomic studies, it is common to take the $t + 2$ year or $t + 3$ year as the *final* estimate. This is based on the assumption that revisions after this date are more unpredictable often reflecting, for example, benchmarking revisions.

[Figure 2](#) shows that most countries make upward revisions to their initial GDP estimates and this is so across different measures of the outturn. This implies that countries tend to underestimate GDP growth in their early estimates. An obvious exception is the United States that initially overstates GDP. Belgium, France, Germany, Italy, Norway, Korea and the United Kingdom make the lowest mean revisions.

However when these revisions are tested for bias, using statistical significance tests (see [Figure 3](#)), [OECD \(2015\)](#) concludes that short-term revisions (up to five months and after one year) are random and centered around zero for most countries. Focusing again on the Year-on-Year (YoY) growth rates, only Belgium, Australia, Norway, Denmark and the Netherlands experience statistically significant revisions. However, there is more evidence for bias in the longer term.

These results are supported by the more recent cross-country results reported in [Walton \(2016\)](#); see [Figure 4](#). [Figure 4](#) again shows that most countries (again with the notable exception of the United States) have made upwards revisions to their early GDP estimates, and this result holds across alternative measures of the outturn. This reinforces the finding that data uncertainties matter – and that these data uncertainties can involve mean (bias) terms as well as variance components. An important question for future research is why some countries have high(er) revisions and why some have low(er) and whether, for example, institutional factors and different statistical systems help explain the observed cross-country variations. Most OECD countries release their first GDP estimates around 30 to 45 days after the end of the quarter; so (large) cross-country differences in timeliness

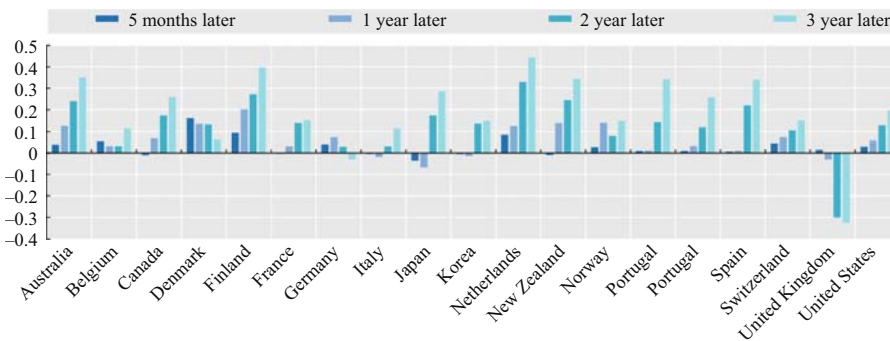


Fig. 2. OECD (2015) cross-country comparison of the importance of GDP revisions: Mean revision to first published YoY quarterly GDP growth rates (in percentage points) from 1994Q4. Source: OECD (2015). Reproduced with permission of the OECD.

Country	5 months later		1 year later		2 year later		3 year later	
	QoQ	YoY	QoQ	YoY	QoQ	YoY	QoQ	YoY
Australia	0.04	0.04	0.02	0.12	0.07	0.24	0.11	0.35
Belgium	0.03	0.06	0.04	0.03	0.03	0.03	0.07	0.11
Canada	0.01	-0.01	-0.01	0.07	0.03	0.17	0.05	0.26
Denmark	0.16	0.16	0.08	0.13	0.10	0.13	0.09	0.06
Finland	0.05	0.10	0.07	0.20	0.08	0.27	0.13	0.40
France	-0.02	0.00	-0.02	0.03	0.00	0.14	0.02	0.15
Germany	0.04	0.04	0.03	0.07	0.04	0.03	0.02	-0.03
Italy	0.02	-0.01	-0.01	-0.02	0.01	0.03	0.04	0.11
Japan	-0.02	-0.04	-0.01	-0.07	0.00	0.17	0.10	0.29
Korea	0.00	-0.01	0.00	-0.01	0.04	0.14	0.07	0.15
Netherlands	-0.01	0.09	0.06	0.12	0.07	0.33	0.10	0.44
New Zealand	0.01	-0.01	0.00	0.14	0.06	0.24	0.11	0.34
Norway	-0.03	0.03	-0.01	0.14	0.00	0.08	-0.01	0.15
Portugal	-0.01	0.01	-0.13	0.01	-0.04	0.14	0.01	0.34
Spain	0.00	0.01	0.01	0.03	0.01	0.12	0.07	0.26
Switzerland	0.00	0.01	-0.02	0.01	0.03	0.22	0.08	0.34
United Kingdom	0.02	0.04	0.01	0.07	0.02	0.11	0.05	0.15
United States	0.01	0.02	0.01	-0.03	-0.06	-0.30	-0.07	-0.33
Average	0.02	0.03	10.01	0.06	0.03	0.13	0.06	0.20

Statistical significance levels 1% 5% 10%

Fig. 3. OECD (2015) estimates of the bias to GDP revisions: Mean revision and statistical significance at different revision intervals for QoQ and YoY quarterly GDP growth rates (in percentage points) from 1994Q4–2013Q4.

Source: OECD (2015). Reproduced with permission of the OECD.

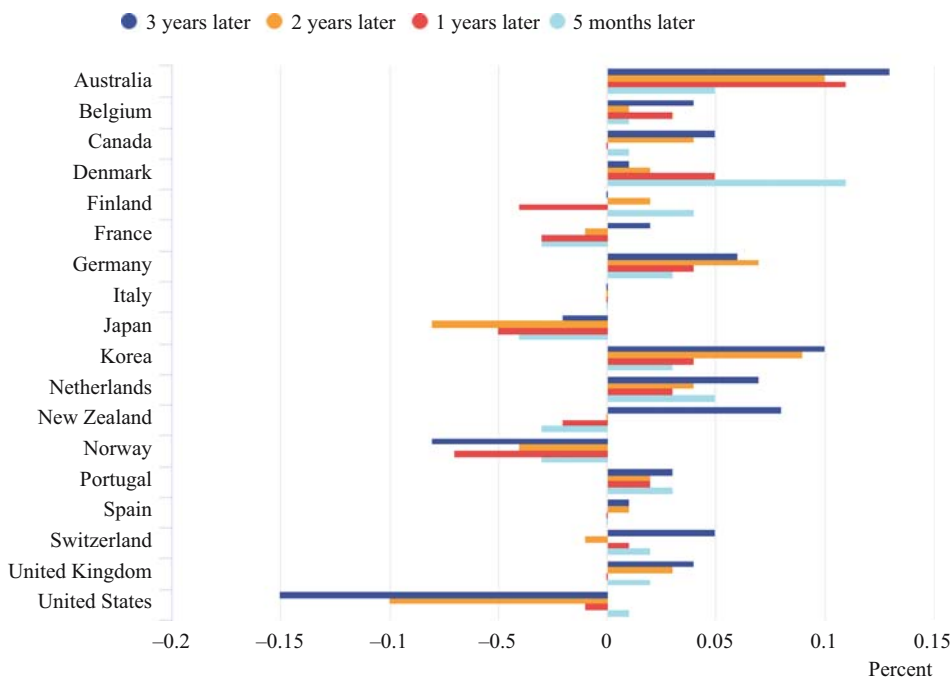


Fig. 4. ONS (taken from Walton 2016) estimates of the bias to GDP revisions for selected OECD and G20 countries: Mean revisions to GDP growth (chain volume measure, quarter on quarter) from 1998Q4–2015Q2. Source: Walton (2016). Reproduced with permission of the ONS.

are likely not a key factor in explaining the observed cross-country variations in data revisions bias.

4.3. Trade-Off Between Timeliness and Accuracy

As discussed, in particular when discussing transitory statistical uncertainty, official measures of variables like GDP, from national statistical offices, are revised as new information is received and methodological improvements are made. So the aforementioned move by the ONS in the United Kingdom to wait 13 days longer before publishing its first GDP estimate means their estimates will now be based on a higher data content than previously. This is expected to deliver more reliable GDP estimates, subject to fewer revisions – emphasising the important trade-off between the timeliness and accuracy of many statistical estimates. That is, statistics with fewer uncertainties can often be produced by delaying publication until more sampling and nonsampling information becomes available. But this delay may impede policy decisions.

As emphasised by [Ruggeri Cannata and Mazzi \(2017\)](#), the choice of the optimal point on the timeliness-accuracy trade-off *curve* has always been open to debate. Decisions made have depended on the characteristics of the statistical system in a given country and users' and policymakers' requests. For example, in the United States, macroeconomic variables tend to be very timely thanks to an efficient and centralised statistical system. By contrast, as [Ruggeri Cannata and Mazzi \(2017\)](#) go on to discuss, in Europe focus was historically more on accuracy although, with the advent of monetary union, pressure (from users and policymakers) for more timely and higher frequency data increased. However, greater decentralisation made it more difficult for Europe to speed up extant production systems. So new statistical systems, such as the construction of flash estimates using econometric methods, were adopted to meet the need for more timely statistics. As national statistical offices now transition to an era of big data, it is increasingly evident that the historical trade-offs between timeliness and accuracy will change, as timely big data are used to complement traditional sources of information. Accuracy may well increase with timeliness.

5. Conclusion

This article emphasises the data uncertainties present in official economic statistics. It accordingly reviews different categorisations of uncertainty, specifically the traditional typology that distinguishes sampling from nonsampling errors and the newer, but complementary, typology of [Manski \(2015\)](#) that distinguishes transitory, permanent and conceptual uncertainties. The importance of nonsampling or permanent uncertainties for administrative and big data sources is explained. Throughout, the article aims to emphasise the importance of measuring and then communicating these uncertainties, as hard as this can prove in some instances. Thereby, the article seeks to introduce and motivate the *Comunikos* project at Eurostat and emphasise the need for more empirical research to minimise public misunderstanding and misuse of official statistics.

To stress once again the relevance of measuring and communicating uncertainty in official statistics, we emphasise the point made by [Manski \(2015, 2019\)](#) and [Van der Bles et al. \(2019\)](#): that reporting official statistics as point estimates projects *incredible certitude*. This may lead to sub-optimal decision making. In other words, this practice may

encourage users to treat statistics as known with certainty. Or they may then make their own, perhaps misleading, (private/subjective) estimates of the degree of uncertainty associated with the point estimates presented to them. In short, in the absence of a body of empirical research seeking to study the impact of data uncertainties on the public's and experts' use and interpretation of official statistics, it is impossible to say, with any confidence, if and how known and unknown data uncertainties do have an impact.

Moreover, as emphasised by [Morgenstern \(1950\)](#) and consistent with a more recent literature in econometrics (cf. [Granger and Pesaran 2000](#)), what surely matters when assessing the importance of uncertainty is how this uncertainty affects decisions. This calls for a joint analysis of how uncertainty matters for decisions made in specific contexts; that is, uncertainty cannot be really understood free from the context in which the uncertain data are used. It calls for empirical and experimental studies following [Kloprogge et al. \(2007\)](#) and [Van der Bles et al. \(2019\)](#) – like [Van der Bles et al. \(2018, 2020\)](#), [Manclossi and Ayodele \(2016\)](#), and [Galvao et al. \(2019\)](#) – that consider, for a given measure of uncertainty, how best this data uncertainty should be communicated.

6. References

- Abowd, J.M., and M.H. Stinson. 2013. “Estimating measurement error in annual job earnings: A comparison of survey and administrative data.” *The Review of Economics and Statistics* 95(5): 1451–1467. DOI: <http://dx.doi.org/10.2139/ssrn.1894690>.
- Aruoba, S.B., F.X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song. 2016. “Improving GDP measurement: A measurement-error perspective.” *Journal of Econometrics* 191(2): 384 – 397. DOI: <http://dx.doi.org/10.1016/j.jeconom.2015.12.009>.
- Australian Bureau of Statistics. 2013. “Types of error.” Available at: [https://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical + language + - + types + of + error](https://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+types+of+error) (accessed December 2019).
- Bagus, P. 2011. “Morgenstern’s forgotten contribution: A stab to the heart of modern economics.” *American Journal of Economics and Sociology* 70(2): 540–562. DOI: <https://doi.org/10.1111/j.1536-7150.2011.00783.x>.
- Bank of England. 2007. *Inflation Report, November 2007*. Available at: <https://www.bankofengland.co.uk/-/media/boe/files/inflation-report/2007/november-2007.pdf?la=en&hash=03AA4FB7EE5CA613952B392EFB00411A7D10195A> (accessed December 2019).
- Bernhard, S. 2016. “A real-time GDP data set for Switzerland.” *SNB Economic Studies* 9/2016. Available at: https://www.snb.ch/n/mmr/reference/economic_studies_2016_09/source/economic_studies_2016_09.n.pdf (accessed December 2019).
- Biemer, P.P., and L.E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: John Wiley Sons. DOI: <https://doi.org/10.1002/0471458740>.
- Boumans, M. 2012: “Observations in A Hostile Environment: Morgenstern on the Accuracy of Economic Observations.” *History of Political Economy* 44: 110–131. DOI: <https://doi.org/10.1215/00182702-1631806>.
- Burridge, P., and K.F. Wallis. 1985. “Calculating the variance of seasonally adjusted series.” *Journal of the American Statistical Association* 80(391): 541–552. DOI: <https://doi.org/10.1080/01621459.1985.10478151>.

- Chambers, R. 2014. "Survey sampling in official statistics - some thoughts on directions." *Proceedings of Statistics Canada Symposium 2014: Beyond traditional survey taking: adapting to a changing world*. Available at: <https://www.statcan.gc.ca/eng/conferences/symposium2014/program/14251-eng.pdf> (accessed December 2019).
- Chow, G., and A.-I. Lin. 1971. "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series." *Review of Economics and Statistics* 53(4): 372–75. DOI: <https://doi.org/10.2307/1928739>.
- Clements, M.P., and A.B. Galvão. 2010. "First announcements and real economic activity." *European Economic Review* 54: 803–817. DOI: <https://doi.org/10.1016/j.eurocorev.2009.12.010>.
- Conti, P.L., D. Marella, and M. Scanu. 2012. "Uncertainty analysis in statistical matching." *Journal of Official Statistics* 28: 69–88. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbe5bf7be7fb3/uncertainty-analysis-in-statistical-matching.pdf> (accessed December 2019).
- Croushore, D. 2011. "Frontiers of real-time data analysis." *Journal of Economic Literature* 49: 72–100. DOI: <https://doi.org/10.1257/jel.49.1.72>.
- Croushore, D., and T. Stark. 2001. "A real-time data set for macroeconomists." *Journal of Econometrics* 105(1): 111–130. DOI: [https://doi.org/10.1016/s0304-4076\(01\)00072-0](https://doi.org/10.1016/s0304-4076(01)00072-0).
- Croushore, D., and T. Stark. 2003. "A real-time data set for macroeconomists: Does the data vintage matter?" *The Review of Economics and Statistics* 85: 605–617. DOI: <https://doi.org/10.1162/003465303322369759>.
- Cunningham, A., J. Eklund, C. Jeffery, G. Kapetanios, and V. Labhard. 2012. "A state space approach to extracting the signal from uncertain data." *Journal of Business and Economic Statistics* 30: 173–80. DOI: <https://doi.org/10.2139/ssrn.1080189>.
- Cunningham, A. and C. Jeffery. 2007. "Extracting a better signal from uncertain data." *Bank of England Quarterly Bulletin 2007Q3*. Available at: <https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2007/extracting-a-better-signal-from-uncertain-data.pdf?la=en&hash=58933100C07CD43ED04DC670F9D6721154AD4F87> (accessed December 2019).
- De Jonge, E. (2020). "Communicating uncertainties in official statistics – A review of communication methods." Available at: <https://ec.europa.eu/eurostat/documents/3888793/12135507/KS-TC-20-011-EN-N.pdf/b2ee60fe-ed5b-ad3e-05bd-56f6b87dfec0?t=1608645519272> (accessed May 2021).
- Draper, D. 1995. "Assessment and propagation of model uncertainty." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 45–97. DOI: <https://doi.org/10.1111/j.2517-6161.1995.tb02015.x>.
- Dungan, J., D. Gao, and A. Pang. 2002. "Definitions of uncertainty." Available at: <ftp://ftp.cse.ucsc.edu/pub/reinas/papers/white.pdf> (accessed December 2019).
- European Commission. 2010. "Report on Greek government deficit and debt statistics." Available at: https://ec.europa.eu/eurostat/documents/4187653/6404656/COM_2010_report_greek/c8523cfa-d3c1-4954-8ea1-64bb11e59b3a (accessed October 2020).
- Eurostat. 2015. "Eurostat statistics explained." Available at: <https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Accuracy> (accessed December 2019).
- Eurostat. 2017. "European statistics code of practice: revised edition 2017." Available at: <https://ec.europa.eu/eurostat/web/quality/principle12> (accessed December 2019).

- Eurostat. 2019. "ESS Standard for Quality Reports." Available at: https://unstats.un.org/unsd/dnss/docs-nqaf/eurostat-esqr_final.pdf (accessed December 2019).
- Faust, J., J.H. Rogers, and J. H. Wright. 2005. "News and noise in G-7 GDP announcements." *Journal of Money, Credit and Banking* 37(3): 403–420. DOI: <https://doi.org/10.1353/mcb.2005.0029>.
- Frare, C., M. Marcellino, G.L. Mazzi, and T. Proietti. 2011. "EUROMIND: a monthly indicator of the Euro Area economic conditions." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174(2): 439–470. DOI: <https://doi.org/10.1111/j.1467-985x.2010.00675.x>.
- Fuller, W. 1987. *Measurement Error Models*. John Wiley and Sons: New York. DOI: <https://doi.org/10.1002/jae.3950030407>.
- Galvao, A.B. 2017. "DSGE models and data revisions." *Journal of Econometrics* 196: 215–232. DOI: <https://doi.org/10.1016/j.jeconom.2016.09.006>.
- Galvao, A. and J. Mitchell. 2020. *Real-Time Perceptions of Historical GDP Data Uncertainty, EMF Research Papers 35, Economic Modelling and Forecasting Group, Warwick Business School*. Available at: <https://ideas.repec.org/p/wrk/wrkemf/35.html> (accessed September 2020).
- Galvao, A. J. Mitchell, and J. Runge. 2019. "Communicating data uncertainty: experimental evidence for UK GDP." Economic Statistics Centre of Excellence Discussion Paper 2019-20. Available at: <https://EconPapers.repec.org/RePEc:nsr:escoe:escoe-dp-2019-20> (accessed December 2019).
- Geweke, J. 1978. "The temporal and sectoral aggregation of seasonally adjusted time series." In *Seasonal Analysis of Economic Time Series* edited by A. Zellner. Washington, DC: US Department of Commerce, Census Bureau. Available at: <https://www.census.gov/ts/papers/Conference1978/Geweke1978.pdf> (accessed December 2019).
- Ghysels, E. 1997. "Seasonal adjustment and other data transformations." *Journal of Business and Economic Statistics* 15: 410–418. DOI: <https://doi.org/10.1080/07350015.1997.10524719>.
- Giannone, D., J. Henry, M. Lalik, and M. Modugno. 2012. "An Area-Wide Real-Time Database for the Euro Area." *Review of Economics and Statistics* 94: 1000–1013. DOI: https://doi.org/10.1162/rest_a_00317.
- Goedeme, 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): 89–110. DOI: <https://doi.org/10.1007/s11205-011-9918-2>.
- Granger, C.W.J., and M.H. Pesaran. 2000. "Economic and statistical measures of forecast accuracy." *Journal of Forecasting* 19: 537–560. DOI: [https://doi.org/10.1002/1099-131x\(200012\)19:7%3C537:aid-for769%3E3.3.co;2-7](https://doi.org/10.1002/1099-131x(200012)19:7%3C537:aid-for769%3E3.3.co;2-7).
- Gromme, F. Ustek-Spilda, E. Ruppert, and B. Cakici. 2017. "Citizen data and official statistics: Background document to a collaborative workshop." Available at: <https://pdfs.semanticscholar.org/4bc0/6a6504c5086a2e81a299abd95ef42617b77b.pdf> (accessed December 2019).
- Groves, R.M., and L. Lyberg. 2010. "Total survey error: past, present and future." *Public Opinion Quarterly* 74: 849–879. DOI: <https://doi.org/10.1093/poq/nfq065>.

- Hand, D.J. 2018. “Statistical challenges of administrative and transaction data.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(3): 555–605. DOI: <https://doi.org/10.1111/rssa.12315>.
- Jacobs, J.P.A.M. and S. van Norden. 2011. “Modeling data revisions: Measurement error and dynamics of “true” values.” *Journal of Econometrics* 161: 101–109. DOI: <https://doi.org/10.1016/j.jeconom.2010.04.010>.
- Kapetanios, G., M. Marcellino, F. Kempf, G.-L. Mazzi, J. Eklund, and V. Labhard. 2021. “Measuring and communicating uncertainty in official statistics: State of the art and perspectives.” Statistical working papers collection, Eurostat. Available at: <https://ec.europa.eu/eurostat/cros/system/files/ks-tc-20-006-en-n.pdf> (accessed May 2021).
- Kapteyn, A., and J.Y. Ypma. 2007. “Measurement error and misclassification: A comparison of survey and administrative data.” *Journal of Labor Economics* 25(3): 513–551. DOI: <https://doi.org/10.1086/513298>.
- Kishor, N.K., and E.F. Koenig. 2012. “VAR estimation and forecasting when data are subject to revision.” *Journal of Business and Economic Statistics* 30: 181–190. DOI: <https://doi.org/10.1198/jbes.2010.08169>.
- Klopogge, P., J. van der Sluijs, and A. Wardekker. 2007. “*Uncertainty Communication*, Utrecht: Copernicus Institute.” Available at: http://www.nusap.net/downloads/reports/uncertainty_communication.pdf (accessed December 2019).
- Koch, G.G., and D.B. Gillings. 2006. *Inference, Design-Based vs. Model-Based*. American Cancer Society. DOI: <https://doi.org/10.1002/0471667196.ess1235.pub2>.
- Kuznets, S. 1948. “National income: a new version.” *Review of Economics and Statistics* 30(3): 151–179. DOI: <https://doi.org/10.2307/1926746>.
- Kuznets, S. 1950. “Review of *On the accuracy of economic observations*, by Oscar Morgenstern.” *Journal of the American Statistical Association* 45: 576–79. DOI: <https://doi.org/10.2307/2280732>.
- Laureti, T., and I. Benedetti. 2020. “Measuring and communicating uncertainty of poverty indicators at regional level.” Statistical working papers collection, Eurostat. Available at: <https://ec.europa.eu/eurostat/documents/3888793/12137895/KS-TC-20-010-EN-N.pdf/6745684c-c989-b3e5-33ae-7bd7dd89bf92?t=1608634421113> (accessed May 2021).
- Lui, S., J. Mitchell, and M. Weale. 2011. “Qualitative business surveys: signal or noise?” *Journal of the Royal Statistical Society Series A* 174(2): 327–348. DOI: <https://doi.org/10.1111/j.1467-985x.2010.00667.x>.
- Manclossi, S., and V. Ayodele. 2016. “Users understanding of uncertainty measures to describe data quality.” Presented at The 21st Government Statistical Service (GSS) Methodology Symposium Methodology: the key to the door of innovation. Available at: <https://gss.civilservice.gov.uk/wp-content/uploads/2016/07/4.4.3-Silvia-Manclossi-Users-understanding-and-use-of-uncertainty-measures-to-descripdf> (accessed December 2019).
- Mankiw, N.G., and M.D. Shapiro. 1986. “News or noise: An analysis of GNP revisions.” *Survey of Current Business (May 1986)*, US Department of Commerce, Bureau of Economic Analysis: 20–25. DOI: <https://doi.org/10.3386/w1939>.

- Manski, C.F. 2015. "Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern." *Journal of Economic Literature* 53(3): 631–653. DOI: <https://doi.org/10.1257/jel.53.3.631>.
- Manski, C.F. 2016. "Credible interval estimates for official statistics with survey nonresponse." *Journal of Econometrics* 191(2): 293–301. DOI: <https://doi.org/10.1016/j.jeconom.2015.12.002>.
- Manski, C.F. 2019. "Communicating uncertainty in policy analysis." *Proceedings of the National Academy of Sciences* 116(16): 7634–7641. DOI: <https://doi.org/10.1073/pnas.1722389115>
- McCracken, M.W., and S. Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business and Economic Statistics* 34(4): 574–589. DOI: <https://doi.org/10.1080/07350015.2015.1086655>.
- McKenzie, R. (2006), "Undertaking revisions and real-time data analysis using the OECD main economic indicators original release data and revisions database." OECD Statistics Working Papers, 2006/02, OECD Publishing, Paris.
- Mehrhoff, J. 2008. "Sources of revisions of seasonally adjusted real time data." Paper prepared for the Meeting of the OECD Short-term Economic Statistics Working Party (STESWP), 23–24 June 2008, Paris. France. Available at: <http://www.oecd.org/sdd/fin-stats/40671433.pdf> (accessed December 2019).
- Mevik, A.-K. 2004. "Uncertainty in the Norwegian Business Tendency Survey." Available at: https://www.ssb.no/a/english/publikasjoner/pdf/doc_200410_en/doc_200410_en.pdf (accessed December 2019).
- Mitchell, J., R. Smith, M. Weale, S. Wright, and E.L. Salazar. 2005. "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth." *Economic Journal* 115(501): 108–129. DOI: <https://doi.org/10.1111/j.0013-0133.2005.00974.x>.
- Morgenstern, O. 1950. *Introduction to survey quality*, Princeton University Press.
- National Science Foundation 2018. "Data accuracy." Available at: <https://nsf.gov/statistics/2018/nsb20181/report/sections/appendix-methodology/data-accuracy> (accessed December 2019).
- Oberski, D.L., A. Kirchner, S. Eckman, and F. Kreuter. 2017. "Evaluating the quality of survey and administrative data with generalized multitrait-multimethod models." *Journal of the American Statistical Association* 112(520): 1477–1489. DOI: <https://doi.org/10.1080/01621459.2017.1302338>.
- OECD 2015. "Revisions of quarterly GDP in selected OECD countries." *Statistics Brief—July 2015 – No 22 by Zwijnenburg*. Available at: <https://www.oecd.org/sdd/na/Revisions-quarterly-GDP-selected-OECD-countries-OECD22.pdf> (accessed December 2019).
- ONS 2017a. "Methodology for measuring uncertainty in ONS local authority midyear population estimates: 2012 to 2016." Available at: <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/methodologyformeasuringuncertaintyinonslocalauthoritymidyearpopulationestimates20> (accessed December 2019).
- ONS 2017b. "ONS methodology working paper series no. 9 – Guide to calculating standard errors for ONS Social Surveys." Available at: <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/ons->

- methodologyworkingpaperseriesno9guidetocalculatingstandarderrorsforonssocialsu (accessed December 2019).
- ONS 2018a. “Communicating quality, uncertainty and change.” Available at: <https://gss.civilservice.gov.uk/policy-store/communicating-quality-uncertainty-and-change/> (accessed December 2019).
- ONS 2018b. “VAT turnover data in National Accounts: background and methodology.” Available at: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/vatturnoverdatainnationalaccountsbackgroundandmethodology> (accessed December 2019).
- ONS 2019. “Gross domestic product (GDP) QMI.” Available at: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi> (accessed December 2019).
- Orphanides, A., and S. van Norden. 2002. “The unreliability of output-gap estimates in real time.” *The Review of Economics and Statistics* 84: 569–583. DOI: <https://doi.org/10.1162/003465302760556422>.
- Peters, C.A. 2001. “Statistics for Analysis of Experimental Data.” In *Environmental Engineering Processes Laboratory Manual*, edited by S.E. Powers. Champaign, IL: AEESP. Available at: https://www.princeton.edu/~cap/AEESP_Statchap_Peters.pdf (accessed December 2019).
- Piccolo, D., and R. Simone, 2019. “The class of cub models: statistical foundations, inferential issues and empirical evidence.” *Statistical Methods and Applications* 28: 389–435. DOI: <https://doi.org/10.1007/s10260-019-00461-1>.
- Plumper, T., and E. Neumayer. 2012. “Population and sample uncertainty.” *EPSA 2013 Annual General Conference Paper 166*. Available <https://ssrn.com/abstract=2224688> (accessed December 2019).
- Ruggeri Cannata, R., and G.-L. Mazzi. 2017. “The trade-off between timeliness and accuracy: the perspective of a statistical agency.” In *Handbook on Rapid Estimates*, edited by G.L. Mazzi. Eurostat and United Nations, : 123–131. DOI: <https://doi.org/10.2785/317290>.
- Seljak, R. 2006. “Estimation of standard error of indices in the sampling business surveys.” *Proceedings of Q2006 European Conference on Quality in Survey Statistics*. Available at: <https://ec.europa.eu/eurostat/documents/64157/4374310/16-ESTIMATION-OF-STANDARD-ERROR-OF-INDICES-IN-THE-SAMPLING.pdf/5721f63b-9ac8-4767-b557-dd30b0378336> (accessed December 2019).
- Smith, R.J., M.R. Weale, and S.E. Satchell. 1998. “Measurement Error with Accounting Constraints: Point and Interval Estimation for Latent Data with an Application to U.K. Gross Domestic Product.” *Review of Economic Studies* 65(1): 109–134. DOI: <https://doi.org/10.1111/1467-937x.00037>.
- Statistics Canada 2017. “Non-sampling error.” Available at: <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch6/nse-enda/5214806-eng.htm> (accessed December 2019).
- Stone, R., D.G. Champernowne, and J. Meade. 1942. “The precision of national income estimates.” *Review of Economic Studies* 9(2): 111–125. DOI: <https://doi.org/10.2307/2967664>.

- US Census Bureau 2018. “Measures of nonsampling error (chapter 11).” Available at: <https://www.census.gov/ts/papers/Conference1978/Geweke1978.pdf> (accessed December 2019).
- Van der Bles, A.M., D. Spiegelhalter, S. Dryhurst, A. Freeman, M. Pearson, and J. Park. 2018. “Determining and facilitating the clearest ways to visualize uncertainty around estimates, time series and curves.” Report to M2D. Available at: <https://wintoncentre.maths.cam.ac.uk/projects/communicating-uncertainty/> (accessed June 2020).
- Van der Bles, A.M., S. van der Linden, A.L.J. Freeman, and D.J. & Spiegelhalter. 2020. “The effects of communicating uncertainty on public trust in facts and numbers.” *Proceedings of the National Academy of Sciences* 117(14): 7672–7683. DOI: <https://doi.org/10.1073/pnas.1913678117>.
- Van der Bles, A.M., S. van der Linden, A. Freeman, J. Mitchell, A.B. Galvao, L. Zaval, and D. Spiegelhalter. 2019. “Communicating uncertainty about facts, numbers and science.” *Royal Society Open Science* 6: 1–42. DOI: <https://doi.org/10.1098/rsos.181870>.
- Van der Loo, J. Pannekoek, and L. Rijnveld. 2017. “Computational estimates of data-editing related variance.” United Nations Economic Commission for Europe, Conference of European Statisticians, Work Session on Statistical Data Editing (The Hague, Netherlands, 24/26 April 2017). Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2017/mtg2/Paper_7_DataEditingBootstrap.pdf (accessed December 2019).
- Verma, V., G. Betti, and F. Gagliardi. 2010. “An assessment of survey errors in EU-SILC.” *Eurostat: Methodologies and working papers*. Available at: <https://ec.europa.eu/eurostat/documents/3888793/5848229/KS-RA-10-021-EN.PDF/2b5b31ad-3973-48ad-9a87-8645c1d2c16c> (accessed December 2019).
- Wallis, K.F. 1982. “Seasonal adjustment and revision of current data: linear filters for the X-11 method.” *Journal of the Royal Statistical Society A* 145: 74–85. DOI: <https://doi.org/10.2307/2981422>.
- Walton, A. 2016. “International comparison of GDP revisions.” *Office for National Statistics* : 1–34. Available at: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/internationalcomparisonofgdprevisions/2016-04-28> (accessed December 2019).

Received June 2019

Revised December 2019

Accepted November 2020