

Data uncertainties: their sources and consequences

Gian Luigi Mazzi, James Mitchell, Florabela Carausu

2019 edition



**Data uncertainties:
their sources and
consequences** | **2019 edition**

Manuscript completed in July 2019.

Printed by the Publications Office in Luxembourg.

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

Luxembourg: Publications Office of the European Union, 2019

© European Union, 2019

Reuse is authorized provided the source is acknowledged.

The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p. 39).

Copyright for the photographs: Cover © Valentin Valkov/Shutterstock

For any use or reproduction of photos or other material that is not under copyright of the European Union, permission must be sought directly from the copyright holders.

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

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.

Abstract

Official economic statistics are uncertain even if not always interpreted or treated as such. From a historical perspective, this paper reviews different categorisations of data uncertainty, specifically the traditional typology that distinguishes sampling from nonsampling errors and a newer typology of Manski (2015, *Journal of Economic Literature*). Throughout the importance of measuring and communicating these uncertainties is emphasised, as hard as it can prove to measure especially some sources of data uncertainty relevant for administrative and big datasets. Accordingly, this paper both seeks to encourage further work into the measurement and communication of data uncertainty in general and introduce the COMUNIKOS project ⁽¹⁾ sponsored by Eurostat. COMUNIKOS is designed to evaluate alternative ways of measuring and communicating data uncertainty specifically in contexts relevant for official economic statistics.

Keywords: Official Statistics, Data uncertainty, Communication, New Metrics

Authors: Florabela Carausu ⁽²⁾, Gian Luigi Mazzi ⁽²⁾, James Mitchell ⁽³⁾,

Acknowledgement: This work has been carried out by under the supervision of the Eurostat project managers Dario Buono and Maria Serfioti. Special thanks goes to George Kapetanios, Edwin de Jonge and Massimiliano Marcellino for helpful comments.

¹ More info about Comunikos project can be found at: https://ec.europa.eu/eurostat/cros/content/communicating-uncertainty-key-official-statistics_en

² GOPA, Luxembourg

³ Warwick Business School, University of Warwick (james.mitchell@wbs.ac.uk); Economic Statistics Centre of Excellence

Table of contents

Data uncertainties: their sources and consequences.....	3
Abstract	4
1. Introduction.....	7
2. The traditional typology of data uncertainties.....	10
2.1 Sampling Errors	10
2.1.1 Measures of sampling error.....	10
2.1.2 Challenges measuring and quantifying sampling errors.....	11
2.2 Nonsampling Errors	12
2.2.1 Specification error.....	13
2.2.2 Coverage (or frame) error.....	13
2.2.3 Nonresponse error.....	13
2.2.4 (Unit-level) measurement error: response error and interviewer error	14
2.2.6 Measures of nonsampling error	15
3. Manski's (2015) typology of data uncertainties	18
3.1 Transitory statistical uncertainty	18
3.2 Permanent statistical uncertainty	18
3.3 Conceptual uncertainty	18
3.3.1 A Bayesian approach	19
3.4 Measures of transitory, permanent and conceptual errors	19
3.4.1 Measurement of transitory uncertainty: real time datasets	21
3.4.2 Measurement of conceptual uncertainty: seasonal adjustment.....	21
4. Consequences of uncertainty	23
4.1 Revisions: real-time data analysis	23
4.2 Case-study on GDP: cross-country comparisons measuring data revisions	24
Figure 3: OECD (Zwijnenburg, 2015) estimates of the bias to GDP revisions	26
4.3 Trade-off between timeliness and accuracy	27
5. Conclusions	28
5.1 Future agenda: and the COMUNIKOS project.....	28
5.2 The need for more empirical research: public misunderstanding and misuse of official statistics	28

Abbreviations

BMA	Bayesian Model Averaging
COMUNIKOS	Communicating Uncertainty in Key Official Statistics
CV	Coefficient of Variation
EU	European Union
GDP	Gross Domestic Product
MSE	Mean Squared Error
OECD	Organisation for Economic Co-operation and Development
RSE	Relative Standard Error

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 defined as the difference between the estimate produced by the statistical office and the ‘true’ population value, which is typically unobserved.⁽⁵⁾ Accordingly, statisticians and statistical offices have sought to categorise and communicate, in various ways, these uncertainties. This reflects a long history, dating back at least to Kuznets (1948) and Morgenstern (1950), that emphasises the uncertainty of economic statistics.⁽⁶⁾ Although as Manski (2015), Manski (2018) and van der Bles et al. (2019) emphasise, headline statistical estimates are often 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, by 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 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’.

Measuring uncertainty is a complex and challenging task, which can involve the use of sophisticated statistical and econometric techniques (classical or Bayesian) and subjective judgement to quantify the data uncertainties. But 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 project (COMmunicating UNcertainty in Key Official Statistics). Following a review of the main sources and types of uncertainty, as summarised in this paper, COMUNIKOS aims to conduct an in-depth methodological and empirical

⁵As Eurostat state (<https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Accuracy>) “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 & 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; e.g. see Dungan et al. (2002). See Fuller (1987) for a textbook discussion, defining the observed estimate as equal to the sum of the latent true value and a measurement error.

⁶ For example, Principle 12 of Eurostat’s European Code of Practice <https://ec.europa.eu/eurostat/web/quality/principle12> states that “Sampling errors and non-sampling errors are measured and systematically documented according to the European standards.” 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; see Bagus (2011).

evaluation of a number of alternative approaches to measuring and communicating uncertainty. These will constitute the basis for formulating proposals and 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 on the other hand as providing clear and ‘accurate’ uncertainty measures may enhance the relevance and credibility of official statistics, COMUNIKOS will carry out a detailed investigation of the *pros* and *cons* of communicating uncertainties to users of official statistics. It will consider appropriate tools for measuring and disseminating data uncertainties. A hope 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 paper, 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.⁽⁷⁾ To do so, 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, Simon Kuznets emphasised the challenges implied by Morgenstern (1950) for measurement. Kuznets 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 & 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)⁽⁸⁾ and Verma et al. (2010). These are nested in the classification below. It is also common, as discussed in van der Bles et al. (2019),⁽⁹⁾ 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 paper is structured as follows. In Sections 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

⁷ Despite this qualification, the discussion below also has relevance for qualitative, ordinal data. For a more specific discussion of uncertainties for ordinal data, see Piccolo (2019).

⁸ Morgenstern (1950), as drawn on throughout this review, distinguishes ten sources of uncertainty: i) lack of designed experiments; ii) hiding of information, lies; iii) the training of observers; iv) errors from questionnaires; v) mass observations; vi) lack of definition and classification; vii) errors of instruments; viii) the factor of time; ix) observations of unique phenomena; x) interdependence and stability of errors.

⁹ See also Bolton (2008) for a general “easy reading” discussion.

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.⁽¹⁰⁾

Sampling error is the most commonly reported measure of statistical uncertainty. This is because, unlike nonsampling errors discussed in Section 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.

Sampling error is therefore lower when samples are large. As summarised by the ONS⁽¹¹⁾, 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;
- 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. In turn, the standard error is often used to compute the coefficient of variation (CV) or margin of error, both of which are related measures of the amount of uncertainty in the estimate. For a normal distribution the 95% confidence interval is measured by two standard errors either side of the estimate.

The Standard Error (SE) is a measure of the variation between a sample estimate and the true population value. Since the standard error of an estimated value generally increases with the size of

¹⁰ In section below, we turn to consider the growing use of administrative and big data sources in official statistics.

¹¹ [ONS methodology working paper series no. 9 — Guide to calculating standard errors for ONS Social Surveys](#)

the estimate, a large standard error need not indicate an unreliable estimate. Therefore it is often better to use the Relative Standard Error (RSE) which is the standard error expressed as a proportion of an estimated value. RSEs provide an indication of the relative size of the error likely to have occurred because of sampling. A higher RSE indicates lower confidence that an estimated value is close to the true population value.

Standard errors to measure sampling error can be computed — given the actual or assumed nature of the survey sample design. Assuming a simple random (probabilistic) sample, the standard error of a mean estimate, $\bar{x} = \sum_{i=1}^n x_i$ given a sample $\{x_i\}_{i=1}^n$, is:

$$SE(\bar{x}) = \frac{s^2}{n}, \quad (1)$$

where s^2 is the population variance and n the sample size. When sampling biases are zero or close to zero SE can be taken to represent total sampling error.⁽¹²⁾ This assumes that ‘population uncertainty’ does not exist, which in practice is a strong often unrealistic assumption. Population uncertainty (see Plumper & Neumayer (2012)) arises from the reality that ‘random samples’ from a given population may not be random, as assumed in (1), when there is uncertainty about who forms part of the population. As discussed in Plumper & Neumayer (2012), oversampling and sample selection corrections can then sometimes be used to tackle the ‘population uncertainty’.

But, in principle, there exist many approaches (e.g. see Goedeme (2013)) to estimate standard errors. *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 when SE is computed; e.g. see Goedeme (2013) which considers the complexities involved in estimating SE specifically for index numbers from complex surveys.⁽¹³⁾ For a detailed analysis of how, given knowledge of the sampling processes and importantly also how it is necessary to understand qualitatively the relative importance of different sources of uncertainty, see ONS2017 (2017).

ONS2017 (2017), for example, 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. And see Mevik (2004) for a detailed study of sampling errors, from the Norwegian Business Tendency Survey, that contrasts ‘design-based’ measures of SE with ‘model-based’ ones that again make use of the bootstrap. For a general discussion of design-based versus model-based methods see Koch & Gillings (2006).

Additional intuition on what SE measures is gleaned when the mean SE is decomposed into the sum of two components: the square of the sampling bias and the sampling variance:

$$\text{Mean Squared Error (MSE)} = \text{Variance} + \text{Squared Bias} \quad (2)$$

If and only when the *bias* = 0 does MSE reduce to the variance alone.

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

¹² For an “easy-reading” discussions, explaining how statistics like SE can be calculated and used, see Peters (2001).

¹³ See also the discussion at [Eurostat web link](#).

survey design is and how this can lead to analytical expressions for *SE*.

To quote the ONS in the UK: ‘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’.⁽¹⁴⁾ 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 emphasised, this complexity should not imply that attempts to measure these margins of error should not be made. Kuznets, 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. And national statistical offices and central banks accordingly often now publish real-time data vintages and analyse the implied revisions (e.g. see Croushore & 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 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) and 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 & Lyberg (2003), Eurostat⁽¹⁵⁾, US Census Bureau⁽¹⁶⁾, Statistics Canada⁽¹⁷⁾, the Australian statistical office⁽¹⁸⁾ and the NSF⁽¹⁹⁾) is to

¹⁴ See [ONS weblink](#)

¹⁵ [Eurostat weblink](#)

¹⁶ https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch11.pdf

¹⁷ <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch6/nse-enda/5214806-eng.htm>

¹⁸ <https://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+--+types+of+error>

¹⁹ <https://nsf.gov/statistics/2018/nsb20181/report/sections/appendix-methodology/data-accuracy>

decompose the nonsampling errors into the five elements listed in the typology below:

1. Specification error
2. Coverage (or frame) error
3. Nonresponse error
4. Unit-level measurement error: response error and interviewer error
5. Processing error

Each of these components of nonsampling error in turn is considered in more detail in the ensuing subsections.

2.2.1 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.

2.2.2 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 & Lin (1971) to multivariate dynamic extensions such as Mitchell et al. (2005) and Frale 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.

2.2.3 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 = when no data are collected about a population unit;
2. item nonresponse = when data only 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. But these nonresponse rates do not help the user of the statistic directly infer, for example, the SE of the estimate.

They are therefore of limited direct use, as ideally the user would be provided with an estimate of SE. Section 2.2.6 below considers recent work in econometrics that has sought to quantify nonresponse errors directly.

2.2.4 (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.

2.2.5 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; see also section 2.2.2 above.

In some contexts (some of) these errors can be measured and quantified. Pannekoek 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 dataset, 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 dataset 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 Pannekoek 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⁽²⁰⁾, the use of robust estimators may offer promise

²⁰ For example, the median rather than the mean offers a robust measure of central tendency

when communicating data in the presence of uncertainties.

2.2.6 Measures of nonsampling error

As emphasised, nonsampling errors are typically and in general 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. Systematic errors in turn are errors that accumulate. For example, if there is an error in the survey or questionnaire design, this causes errors in the respondent's answers, often leading to biases.

Partial identification and measures of survey nonresponse Recent work on 'partial identification' (see 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 contemplating all the values that the missing data might take. This delivers interval rather than point estimates. The approach of Pannekoek et al. (2017), that 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, $z = 1$), then from the law of total probability

$$P(y) = P(y|z = 1)P(z = 1) + P(y|z = 0)P(z = 0) \quad (3)$$

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

$$H[P(y)] = [P(y|z = 1)P(z = 1) + \gamma P(y|z = 0)P(z = 0), \gamma \in \Gamma_Y] \quad (4)$$

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,

$$P(y \in B) = P(y \in B|z = 1)P(z = 1) + \gamma P(y \in B|z = 0)P(z = 0) \quad (5)$$

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

$$P(y \in B|z = 1)P(z = 1) \leq P(y \in B) \leq P(y \in B|z = 1)P(z = 1) + P(z = 0) \quad (6)$$

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.

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 & Ypma (2007) and Abowd & 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 & 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 CBI in the UK (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 dataset 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 & Ypma (2007) provide a framework to model the errors in administrative data due to mismatching, based on comparing 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. Let the two (known) marginal 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 which 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 section 2.2.2, 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 that 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), also likely afflicts 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 recent example is the condemnation of Greece by the European Commission in 2010 for falsifying data about its public finances and allowing political pressures to obstruct the collection of accurate statistics; see

[FT weblink.](#)

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 non-ignorable non-response in surveys; and how adaptive surveys can be used to select a sample of non-respondents to interview or survey at a second wave of the survey so as to minimise the non-response bias. ONS2014 (2014) 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.

Each of these three components of uncertainty is considered in turn in the ensuing sub-sections.

3.1 Transitory statistical 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 UK 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 subsequently revised 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.

3.2 Permanent statistical uncertainty

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

3.3 Conceptual uncertainty

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) explains 'the theoretical characteristics of, say, an industry or a 'price', are less well established than those of

a wave length' (p. 35).

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 adopted;⁽²¹⁾
- 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);
- 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 of data and their comparability.

3.3.1 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 model, 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 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 predictive densities $g_{it}(y_t) = \Pr(y_t | S_i, \Omega_t)$, where the weights w_{it} are the model's posterior probabilities, $w_{it} = \Pr(S_i | \Omega_t)$:⁽²²⁾

$$p_t(y_t | \Omega_t) = \sum_{i=1}^N w_{it} g_{it}(y_t); \quad (t = 1, \dots, T) \quad (7)$$

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 forecast.

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.

3.4 Measures of transitory, permanent and

²¹ This decomposition of conceptual uncertainty overlaps with the eighth element in the typology of errors of Morgenstern (1950): "the factor of time". This arises as economic data are often measured on a discrete basis with observations attributed to a specific window of time can lead to errors especially as changes in classification (e.g. changes in the definition of an industry or changes to the characteristics of a specific product).

²² All probabilities are implicitly conditional on the set of all models S under consideration.

conceptual errors

An advantage of this 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 at the statistic is revised. As elaborated on in section 3.4.1, measurement and quantification of transitory statistical uncertainties is now facilitated by the relatively wide availability of real time (vintage) datasets. These real time datasets let one measure the revisions between successive estimates. Many authors have proposed models of data revisions — using real time datasets — to model and forecast this 'transitory' GDP data uncertainty (e.g. see Jacobs & van Norden (2011), Cunningham et al. (2012), Kishor & Koenig (2012) and 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 SE, are the fan charts produced by the Bank of England; see Figure 1.

Figure 1: Metadata framework – coherence and comparability

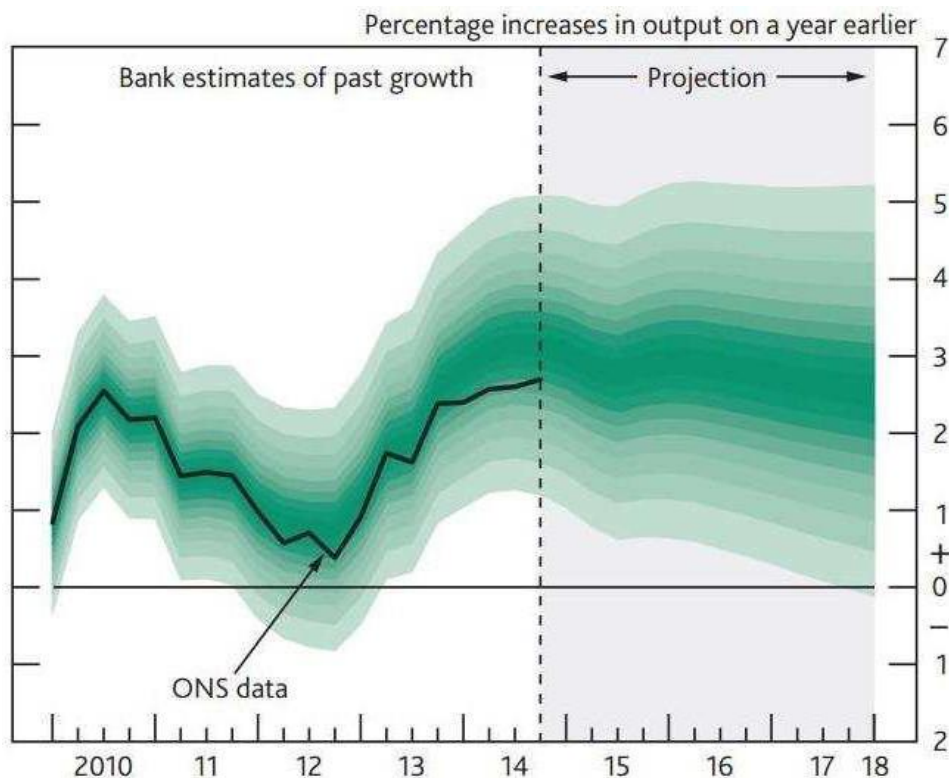


Figure 1 provides an illustrative example of what these fan charts look like taken from the Bank of England's *Inflation Report*. Importantly, in Figure 1 we see that the Bank seeks to quantify both future uncertainties but also past or historical data uncertainties. This is emphasised in the words that 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.' (November 2007;

Inflation Report, p. 39). As the Bank of England explain, 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 estimates. The ONS's latest estimate of GDP growth is shown in Figure 1 by the solid black line. Cunningham & Jeffery (2007) provide an explanation of the data revisions model, used by Bank staff, that along with MPC judgment helps shape the form of these backcast fan charts. Their model exploits historical patterns in ONS revisions and information from qualitative business surveys. 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 & Mitchell (2019).

3.4.1 Measurement of transitory uncertainty: real time datasets

The accessibility of real time datasets, in particular in the US as maintained by the Federal Reserve Banks of Philadelphia and St. Louis with more recently real-time databases for the Euro Area compiled by the ECB (see Giannone et al. (2012)), has enabled statisticians and economists to analyse and model data revisions; see Croushore & Stark (2001), Croushore & Stark (2003), Croushore (2011) and McCracken & Ng (2016).⁽²³⁾ This provides a means to measure and quantify transitory statistical uncertainty.

3.4.2 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? But as the sub-classification of conceptual uncertainty in section 3.3 suggests, 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. Burridge & 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 (see Orphanides & van Norden (2002)), has involved recursive real-time application of the seasonal adjustment filter to the

²³ Real-time data bases, especially for GDP, are now available for various countries. This includes Switzerland (Bernhard (2016)), the UK (as maintained by the Bank of England), Canada (as maintained by Statistics Canada at <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610049101>), New Zealand (as maintained by RBNZ at <https://www.rbnz.govt.nz/research-and-publications/research-programme/additional-research/a-real-time-database-for-gdp>), the OECD (maintained at https://stats.oecd.org/Index.aspx?DataSetCode=MEI_ARCHIVE).

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.⁽²⁴⁾ 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 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).

²⁴ 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 therefore requires a cross-disciplinary approach, involving the intersection of psychology, behavioural and decision science and statistics. Consistent with the conclusions of Manski (2015), it also requires new empirical research to study the effects of uncertainty — and its communication or lack of — on users' understanding, interpretation and use of statistics.

In the absence of (to-date) published research 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.⁽²⁵⁾ 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 real-time datasets by central banks and statistical offices. Importantly, as Croushore (2011) concludes, until these datasets 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.

²⁵ 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.

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 & Shapiro (1986).

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

$$(y_t^{new} - y_t^{old}) = \beta_0^{news} + \beta_1^{news} y_t^{old} + \varepsilon_t \quad (8)$$

$$(y_t^{new} - y_t^{old}) = \beta_0^{noise} + \beta_1^{noise} y_t^{new} + \varepsilon_t \quad (9)$$

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 $\beta_1^{noise} = 0$. For additional details on the application of these tests see Clements & 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 & 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 news model, for Italy, Japan and the UK, 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 Zwijnenburg (2015) compares GDP revisions across OECD countries. Zwijnenburg (2015) uses the mean revision to measure the importance of data revisions. Figure 2 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, 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 reflecting, for example, benchmarking revisions.

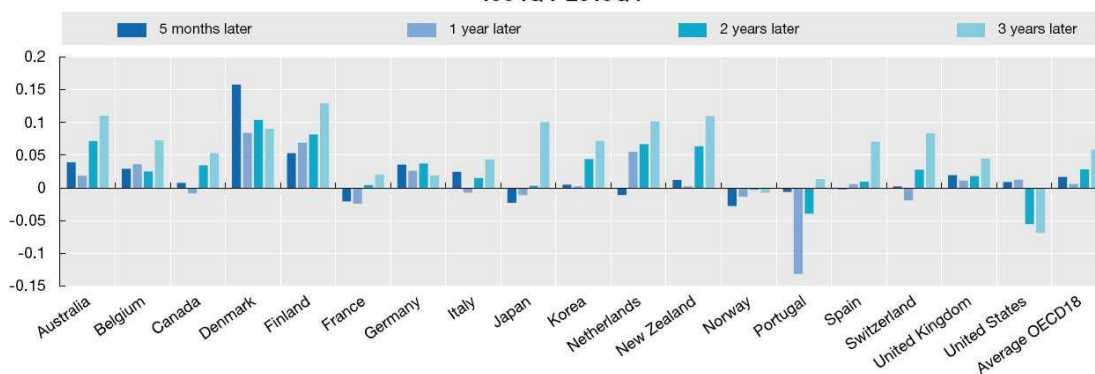
Figure 2 shows that most countries make upward revisions to their initial GDP estimates and this is so across different measures of the outturn. This implies that countries tend to underestimate quarterly GDP growth in their early estimates. An exception is the US. France, Italy, Norway and the UK make the lowest mean revisions for both QoQ and YoY growth rates.

But when these revisions are tested for bias, using statistical significance tests (see Figure 3), Zwijsenburg (2015) concludes that short-term revisions (up to five months and after one year) are random and centered around zero for most countries. For the QoQ growth rates only Australia, Denmark and Germany experience statistically significant revisions; for the YoY growth rates this is the case only for Belgium, Australia, Norway, Denmark and the Netherlands. 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 US) 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; cf. (2).

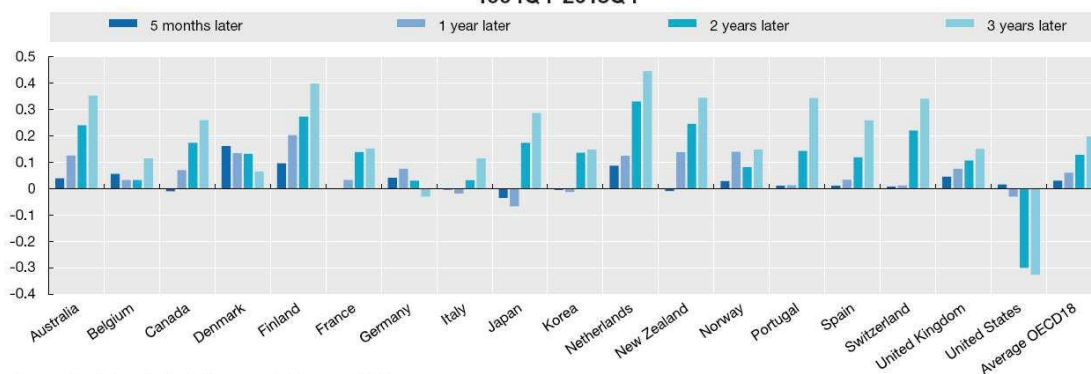
Figure 2: OECD (Zwijsenburg, 2015) cross-country comparison of the importance of GDP Revisions

Figure 1. Mean revision to first published QoQ quarterly GDP growth rates percentage points, 1994Q4-2013Q4



Source: Revisions Analysis Dataset – Infra-annual Indicators.

Figure 2. Mean revision to first published YoY quarterly GDP growth rates percentage points, 1994Q4-2013Q4



Source: Revisions Analysis Dataset – Infra-annual Indicators.

Figure 3: OECD (Zwijnenburg, 2015) estimates of the bias to GDP revisions

Table 1. Mean revision and assessment of statistical significance at different revision intervals for QoQ and YoY quarterly GDP growth rates, percentage points, 1994Q4-2013Q4

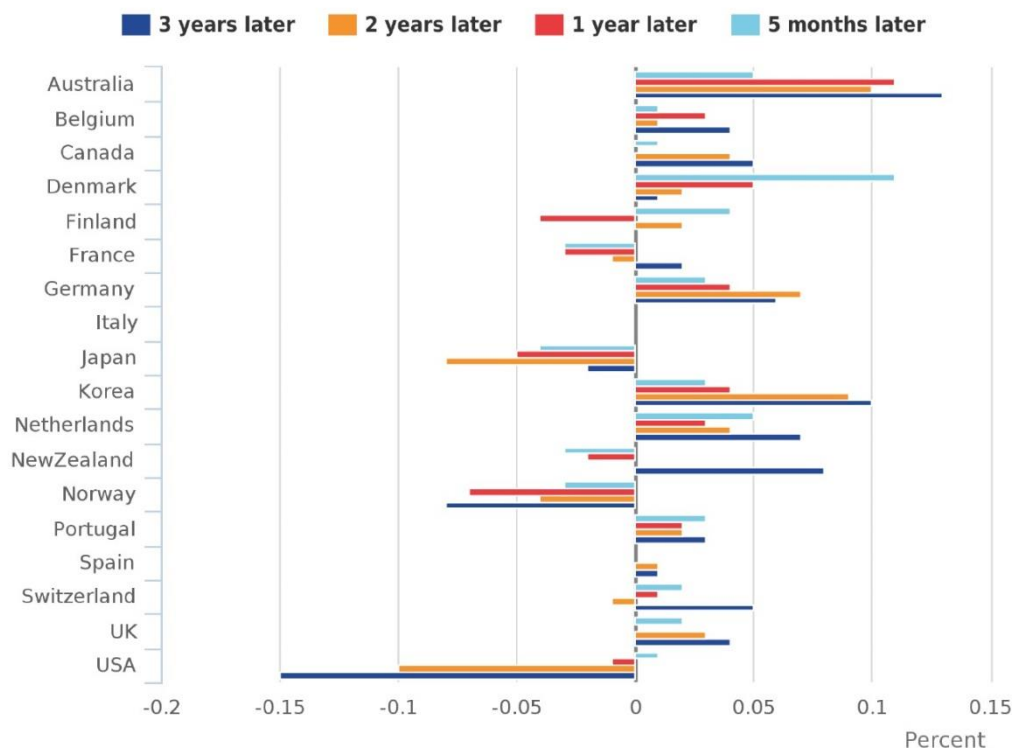
Country	5 months later		1 year later		2 years later		3 years later	
	QoQ	YoY	QoQ	YoY	QoQ	YoY	QoQ	YoY
Australia	0.04	0.04	0.02	0.12	0.07	0.24	0.11	0.35
Belgium	0.03	0.06	0.04	0.03	0.03	0.03	0.07	0.11
Canada	0.01	-0.01	-0.01	0.07	0.03	0.17	0.05	0.26
Denmark	0.16	0.16	0.08	0.13	0.10	0.13	0.09	0.06
Finland	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	0.01	0.06	0.03	0.13	0.06	0.20

Statistical significance levels: 1% (red), 5% (yellow), 10% (green)

Source: Revisions Analysis Dataset – Infra-annual Indicators.

Figure 4: ONS (Walton, 2016) estimates of the bias to GDP revisions

Figure 1: Chained Volume Measure, GDP growth, quarter on quarter, quarter 4 (October to December) 1998 to quarter 2 (April to June) 2015, selected OECD and G20 countries



4.3 Trade-off between timeliness and accuracy

As discussed in section 3.1, 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 Office for National Statistics in the UK 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; so ultimately it is a matter of choice for statistical offices where they view the optimal point on the timeliness-accuracy trade-off *curve* to be.

5

Conclusions

This paper 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 typology of Manski (2015). The importance of nonsampling uncertainties for administrative and big data sources is explained. Throughout the paper aims to emphasise the importance of measuring and then communicating these uncertainties, as hard as this can prove. Thereby the paper seeks to introduce and motivate the COMUNIKOS project.

5.1 Future agenda: and the COMUNIKOS project

Building on this paper, and its call for more research into both the measurement and communication of data uncertainty, the COMUNIKOS project plans to investigate and evaluate alternative ways of measuring and communicating data uncertainty in contexts of interest to statistical authorities producing official economic statistics. This will involve the development and consideration of statistical and econometric methods to estimate and present data uncertainty, and draw on work measuring and communicating uncertainty in other scientific disciplines. Theoretical, empirical and computational aspects of alternative methods will be described, as well as their advantages and drawbacks. A 'proof of concept' will then be developed, in order to demonstrate how alternative methods could be incorporated into the statistical production process to estimate and present the uncertainty associated with key statistical indicators. After this phase, with particular reference to the estimation of consumer price indices using both traditional sources and scanner price data, a case study will be used to show how the data uncertainties can be estimated and presented.

5.2 The need for more empirical research: 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), Manski (2018) 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 in 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 expert's 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 & 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; i.e. uncertainty cannot be really understood free from the context in which the uncertain data are used. It calls for studies following Klopogge et al. (2007) and van der Bles et al. (2019) like van der Bles et al. (2018), Manclossi & Ayodele (2016) and other recent studies including van der Bles et al. (2018) and Galvao et al. (2019), many of them associated with the ONS in the UK (see ONS2014 (2014)), that consider, for a given measure of uncertainty, how best this uncertainty should be communicated.

6

Bibliography

References

- Abowd, J. M. & Stinson, M. H. (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.
- Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F. & Song, D. (2016), 'Improving gdp measurement: A measurement-error perspective', *Journal of Econometrics* 191(2), 384–397.
- 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.
- Bank of England (2007), *Inflation Report, November 2007*.
- Bernhard, S. (2016), 'A real-time gdp data set for switzerland'. *SNB Economic Studies* 9/ 2016.
- Biemer, P. P. & Lyberg, L. E. (2003), *Introduction to survey quality*, Hoboken, NJ: John Wiley Sons.
- Bolton, P. (2008), *Statistical literacy guide -uncertainty and risk*. House of Commons Library, SN/SG/4836.
- Boumans, M. (2012), 'Observations in a hostile environment: Morgenstern on the accuracy of economic observations', *History of Political Economy* 44, 110–131.
- Burridge, P. & Wallis, K. F. (1985), 'Calculating the variance of seasonally adjusted series', *Journal of the American Statistical Association* 80(391), 541–552.
- 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*.
- Chow, G. & Lin, A.-I. (1971), 'Best linear unbiased interpolation, distribution, and extrapolation of time series by related series', *Review of Economics and Statistics* 53(4), 372–75.
- Clements, M. P. & Galvão, A. B. (2010), 'First announcements and real economic activity', *European Economic Review* 54, 803–817.
- Conti, P. L., Marella, D. & Scanu, M. (2012), 'Uncertainty analysis in statistical matching', *Journal of Official Statistics* 28, 69-88.
- Croushore, D. (2011), 'Frontiers of real-time data analysis', *Journal of Economic Literature* 49, 72–100.
- Croushore, D. & Stark, T. (2001), 'A real-time data set for macroeconomists', *Journal of Econometrics* 105(1), 111–130.

- Croushore, D. & Stark, T. (2003), 'A real-time data set for macroeconomists: Does the data vintage matter?', *The Review of Economics and Statistics* 85, 605–617.
- Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G. & Labhard, V. (2012), 'A state space approach to extracting the signal from uncertain data', *Journal of Business & Economic Statistics* 30, 173–80.
- Cunningham, A. & Jeffery, C. (2007), 'Extracting a better signal from uncertain data', *Bank of England Quarterly Bulletin* 2007Q3.
- Draper, D. (1995), 'Assessment and propagation of model uncertainty', *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1), 45–97.
- Dungan, J., Gao, D. & Pang, A. (2002), 'Definitions of uncertainty'. working paper, University of California Santa Cruz.
- Faust, J., Rogers, J. H. & Wright, J. H. (2005), 'News and noise in g7 gdp announcements', *Journal of Money, Credit and Banking* pp. 403–420.
- Frail, C., Marcellino, M., Mazzi, G. L. & Proietti, T. (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.
- Fuller, W.A., a. . N. (1987), *Measurement Error Models*, John Wiley and Sons.
- Galvao, A. B. (2017), 'DSGE models and data revisions', *Journal of Econometrics* 196, 215–232.
- Galvao, A. & Mitchell, J. (2019), 'Measuring Data Uncertainty: An Application using the Bank of England's 'Fan Charts' for Historical GDP Growth', *ESCoE Discussion Paper* 2019-08 .
- Galvao, A., Mitchell, J. & Runge, J. (2019), 'Communicating and understanding the effects of data uncertainty: experimental evidence for UK GDP'. Working paper, Warwick University.
- Geweke, J. (1978), The temporal and sectoral aggregation of seasonally adjusted time series, in A. Zellner, ed., 'Seasonal Analysis of Economic Time Series', Washington,DC: US. Department of Commerce, Census Bureau.
- Ghysels, E. (1997), 'Seasonal adjustment and other data transformations', *Journal of Business and Economic Statistics* 15, 410–418.
- Giannone, D., Henry, J., Lalik, M. & Modugno, M. (2012), 'An Area-Wide Real-Time Database for the Euro Area', *The Review of Economics and Statistics* 94, 1000–1013.
- 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.
- Granger, C. W. J. & Pesaran, M. H. (2000), 'Economic and statistical measures of forecast accuracy', *Journal of Forecasting* 19, 537–560.
- Gromme, F., Ustek-Spilda, F., Ruppert, E. & Cakici, B. (2017), *Citizen data and official statistics: Background document to a collaborative workshop*. Goldsmiths, University of London.
- Groves, R. M. & Lyberg, L. (2010), 'Total survey error: past, present and future', *Public Opinion Quarterly* 74, 849–879.
- 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.
- Jacobs, J. P. A. M. & van Norden, S. (2011), 'Modeling data revisions: Measurement error and dynamics of 'true' values', *Journal of Econometrics* 161, 101–109.
- Kapteyn, A. & Ypma, J. (2007), 'Measurement error and misclassification: A comparison of survey and administrative data', *Journal of Labor Economics* 25(3), 513–551.

- Kishor, N. K. & Koenig, E. F. (2012), 'VAR estimation and forecasting when data are subject to revision', *Journal of Business and Economic Statistics* 30, 181–190.
- Klopprogge, P., van der Sluijs, J. & Wardekker, A. (2007), 'Uncertainty communication, note=Copernicus Institute, Utrecht'.
- Koch, G. G. & Gillings, D. B. (2006), *Inference, Design-Based vs. Model-Based*, American Cancer Society.
- Kuznets, S. (1948), 'National income: a new version', *Review of Economics and Statistics* 30(3), 151–179.
- Lui, S., Mitchell, J. & Weale, M. (2011), 'Qualitative business surveys: signal or noise?', *Journal of the Royal Statistical Society Series A* 174(2), 327–348.
- Mannclossi, S. & Ayodele, V. (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.
- Mankiw, N. G. & Shapiro, M. D. (1986), 'News or noise: An analysis of GNP revisions', *Survey of Current Business (May 1986)*, US Department of Commerce, Bureau of Economic Analysis pp. 20–25.
- Manski, C. F. (2015), 'Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern', *Journal of Economic Literature* 53(3), 631–653.
- Manski, C. F. (2016), 'Credible interval estimates for official statistics with survey non-response', *Journal of Econometrics* 191(2), 293–301.
- Manski, C. F. (2018), 'Communicating uncertainty in policy analysis', *Proceedings of the National Academy of Sciences*.
- McCracken, M. W. & Ng, S. (2016), 'FRED-MD: A Monthly Database for Macroeconomic Research', *Journal of Business & Economic Statistics* 34(4), 574–589.
- 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) in Paris, 23-24 June 2008.
- Mevik, A.-K. (2004), 'Uncertainty in the norwegian business tendency survey'. Statistics Norway Statistical Methods and Standards.
- Mitchell, J., Smith, R., Weale, M., Wright, S. & Salazar, E. L. (2005), 'An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth', *Economic Journal* 115(501), 108–129.
- Morgenstern, O. (1950), *Introduction to survey quality*, Princeton University Press.
- Oberski, D. L., Kirchner, A., Eckman, S. & Kreuter, F. (2017), 'Evaluating the quality of survey and administrative data with generalized multitrait-multimethod models', *Journal of the American Statistical Association* 112(520), 1477–1489.
- ONS2014 (2014), 'Communicating uncertainty and change. guidance for official statistics producers'. Government Statistical Service.
- ONS2017 (2017), 'Methodology for measuring uncertainty in ONS local authority midyear population estimates: 2012 to 2016'. Office for National Statistics, UK.
- Orphanides, A. & van Norden, S. (2002), 'The unreliability of output-gap estimates in real time', *The Review of Economics and Statistics* 84, 569–583.
- Pannekoek, J., van der Loo, M. & Rijnveld, L. (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).
- Peters, C. A. (2001), *Statistics for Analysis of Experimental Data*, S. E. Powers, Ed. AEESP, Champaign, IL.
- Piccolo, D. (2019), 'A paradigm for rating data models'. URL: <https://ec.europa.eu/eurostat/cros/system/files/piccolontts2019.pdf>
- Plumper, T. & Neumayer, E. (2012), 'Population and sample uncertainty'. EPSA 2013 Annual General Conference Paper 166. Available at SSRN: <https://ssrn.com/abstract=2224688>.
- Smith, R. J., Weale, M. R. & Satchell, S. E. (1998), 'Measurement Error with Accounting Constraints: Point and Interval Estimation for Latent Data with an Application to U.K. Gross Domestic Product', *The Review of Economic Studies* 65(1), 109–134.
- Stone, R., Champernowne, D. G. & Meade, J. (1942), 'The precision of national income estimates', *Review of Economic Studies* 9(2), 111–125.
- van der Bles, A. M., Spiegelhalter, D., Dryhurst, S., Freeman, A., Pearson, M. & Park, J. (2018), 'Determining and facilitating the clearest ways to visualize uncertainty around estimates, time series and curves'. Working Paper, Cambridge University.
- van der Bles, A. M., van der Linden, S., Freeman, A., Mitchell, J., Galvao, A. B., Zaval, L. & Spiegelhalter, D. (2019), 'Communicating uncertainty about facts, numbers and science', *Royal Society Open Source* 6, 1–42.
- Verma, V., Betti, G. & Gagliardi, F. (2010), 'An assessment of survey errors in eu-silc', Eurostat: *Methodologies and working papers*.
- 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.
- Walton, A. (2016), 'International comparison of gdp revisions', *Office for National Statistics* pp. 1–34.
- Zwijnenburg, J. (2015), 'Revisions of quarterly gdp in selected oecd countries', *OECD Statistics Brief* 22.

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/contact>

On the phone or by e-mail

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 electronic mail via: <https://europa.eu/contact>

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>

EU Publications

You can download or order free and priced EU publications from EU Bookshop at: <http://bookshop.europa.eu>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see <https://europa.eu/contact>)

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: <https://eur-lex.europa.eu>

Open data from the EU

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

Data uncertainties: their sources and consequences

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

For more information

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