Research article

On the inaccuracies of macroeconomic observations

Peter A.G. van Bergeijk*

International Institute of Social Studies (ISS), Erasmus University Rotterdam, Kortenaerkade 12, 2518, AX, The Hague, the Netherland

* Correspondence: Email: bergeijk@iss.nl.

Abstract: Transparency about measurement errors in macroeconomic statistics is lacking. Inaccuracy is usually acknowledged only during extraordinary situations or under specific policy demands. Despite advances in technology and methodology, the overall accuracy has not improved due to economic changes and new complex activities. Even striking historical efforts did not change the persistence of substantial inaccuracies and biases. While the pioneers of national accounting addressed inaccuracies by using multiple accounting approaches and transparently reported error estimates, current statistical offices and other data producers rarely publish error measures, thereby creating an illusion of precision. Best practices and new measurement methods, such as satellite observation, Big Data, and Artificial Intelligence, promise improvements but face significant challenges without clear standards and an awareness of measurement error and bias. Regular and transparent reporting of measurement errors is and will be essential to improve the data reliability. Empowering data users—funding agencies, academics, and journalists—through error reporting and education can challenge the status quo.

Keywords: measurement error; gross planetary product; national accounting; International Monetary Fund; empowerment; education

JEL Codes: A1, E01, C18

1. Introduction

Why is it so difficult to be transparent about measurement errors in macroeconomic statistics? Measurement error is pervasive in macroeconomics, as in most applied sciences; however, the explicit recognition of macroeconomic inaccuracies is by and large limited to extraordinary times, such as during the lockdown periods of the COVID-19 pandemic or, to isolated cases where specific political and policy demands stood at the basis of evaluations of measurement error. The ultimate impact of these surges of recognition and the ensuing methodological improvements do not have a lasting impact. Indeed, measurement errors and biases remain at similar levels. Of course, the registration of certain items improved due to the enormous increase in computing power, the development of new and advanced methodologies, and the advent of new and exciting modes of observation. However, regarding balance, the accuracy of economic observations does not show a tendency to improve. This sobering observation reflects Simpson's Paradox: even when observations of all parts of an economic system improve, the aggregates can deteriorate due to structural changes. In this case, the shifts are from easily observed and accurately measured activities to invisibles and difficult to measure activities (Griliches, 1994). Moreover, statistics are social constructs, the underlying concepts change due to macro-behavioral and institutional change, and the dynamics of modern economies mean that new activities constantly emerge—new activities are by their very nature more difficult to observe. Therefore, a macro-measurement error was, is and will be a significant, and economically relevant problem. Progress is slow, as illustrated by a perspective that Zoltan Kenessey (Director of the International Statistical Institute) provided on Morgenstern's (1950) *on the accuracy of economic observations*, which is a classic but forgotten and ignored diagnosis of macroeconomic and macrosocial statistical observations: at the minimum, one feels that the increases in accuracy and timeliness of economic statistics are less than one would have expected about fifty years ago when Morgenstern was working on his book. Moreover, this is more surprising in light of the profound electronic revolution which occurred over this period (Kenessey, 1997, p.248).

In the quarter of a century, much has been achieved by the statistical community since this appraisal in terms of timeliness. However, on the eve of the 75th anniversary of Morgenstern's on the accuracy, macro-measurement errors continue to be at a similar level as observed in the 1950s. Two important and widely discussed cases may illustrate this point: first, the consumer price index (CPI) and its derivative inflation which, according to Reiss (2008, 61), provides for the only systematic discussion of measurement issues in economics; and second, the balance of payments registration in the 1980s led to a concerted action for improved measurements under the aegis of the IMF (International Monetary Fund).

The influential Boskin (1996) report on the US CPI appeared in a period when central banks were moving towards inflation-targeting, and policy makers wanted to reevaluate Social Security spending and retirement and compensation programs that were based on an upwardly biased CPI. The interest of academics in consumer price statistics reflects the core role of price in the economic theory; the official inquiry responds to the fact that the CPI is an important input of economic policy and wage policy, and because price indices are necessary for the deflation of nominal data into 'real' data. The Boskin committee's finding of an upward bias of ¾ to 1½ percentage points and its proposals for improvement were influential, thus stimulating the best practices in both the US and abroad (Moulton, 2018). The enthusiastic engagement of statisticians, professional economists, and academic researchers stemmed hopeful. However, according to Feldstein (2017) '(…) despite the attention for

this subject in the professional literature, there remains an insufficient understanding of just how imperfect the existing official estimates actually are'. While helping to address and reduce sources of measurement error, the Boskin report did not eliminate the CPI bias. Indeed, the upward bias of the US CPI persists at comparable levels due to new measurement challenges related to (de)globalization, the changing role of services, and the growing share of internet transactions (Goolsbee and Klenow, 2018; Matolcsy et al., 2020).

The IMF (1983, 1987, 1992) reports on the accounting problems of international financial flows on the balance of payments that occurred at a time when the size of the item 'errors and omissions' made it difficult, if not impossible, to glean if the capital account of many OECD countries was either in surplus or in deficit. In 1992, the IMF observed that 'statistical problems have worsened dramatically and may well continue to worsen in the absence of a major effort to improve the data'. (IMF, 1992, p. 1). This sense of urgency made a concrete impact; unfortunately, the improvement did not last as illustrated in Figure 1. Figure 1 reports the world's current account as reported in the most recent vintage of the IMF's World Economic Outlook database. Since our planet does not trade with Moon or Mars (yet), the world's current account should logically be zero. Therefore, its statistical imbalance in the face of non-existing extraterrestrial trade reflects the underlying measurement errors (Grassman, 1983; van Bergeijk (2013, 2024). The implicit measurement errors reveal two phases: first, up till and including 2003, a negative world current account of on average −0.4 percent with periodic movements towards zero; and second, a phase starting in 2004 of a persistent surplus of similar size, but with no apparent drive towards its logical value of zero[1].
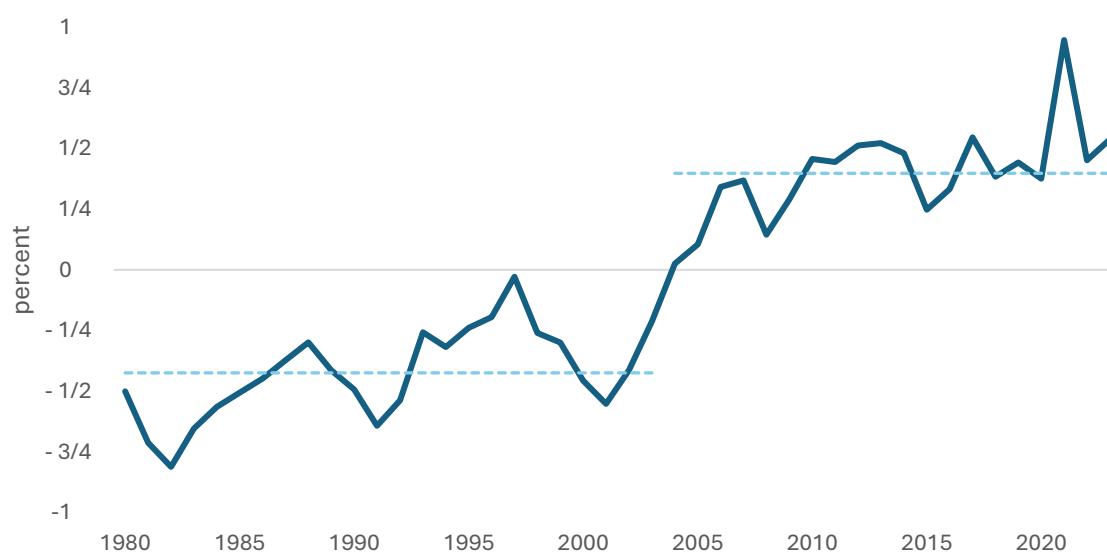

Figure 1. World current account in percent of Gross Planetary Product (1980–2023). Note: period averages 1980–2004 and 2005–2023, respectively. Source: IMF World Economic Outlook database (April 2024 version), available at:

---

[1] This issue can also be analysed from the perspective of the worlds saving-investment balance (the mirror of the current account in an autarkic planetary system, with similar results.

https://www.imf.org/en/Publications/WEO/weo-database/2024/April and accessed June, 1, 2024.

The two cases illustrated the persistent and non-transitory nature of macroeconomic measurement error[2]. These cases are special because of the professional recognition of measurement error. However, broad-based awareness typically appears to be mainly limited to the macroeconomic statistics of the Global South, in particular Africa (Jerven, 2013), and to a critical emerging literature on 'incredible certitude' (Manski, 2011, 2015, 2020), 'pseudo accuracy' (de Jonge, 2020), and 'mock accuracy' (Linsi et al., 2023). A symptom is the remarkable phenomenon of 'digit-inflation' (van Bergeijk, 2024), which is the growing tendency to report statistics with many more digits than can be justified on the basis of the accuracy of the data. On the one hand, these phenomena lack transparency regarding the measurement error; on the other hand, digit-inflation goes hand in hand. This article argues that their self-reinforcing cycle can only be broken if our profession starts investigating and transparently reporting the inaccuracies of observations on a regular basis. This needs to be performed in statistical institutions, in academia, and in policy circles. While the focus of the present article is on National Accounting, the problems in other fields of macroeconomic observations such as (un)employment and inflation are of a similar size and importance.

The remainder of this article is structured as follows. Section 2 starts with a discussion of National Accounting from the perspective that the approach was originally intended to deal with measurement errors and that best practices initially included the explicit and transparent reporting of margins of error of the reported aggregates and/or components. Based on concrete examples, the section argues that it is possible to return to the transparency of the pioneering era of national accounting. Section 3 presents the case of the growth rate of Gross Planetary Product (GPP, aka world production) to illustrate the persistence of non-transient data inaccuracies based on 37 vintages of the IMF's World Economic Outlook database. The section zooms in on the measurements regarding three important global recessions: the Volcker recession at the start of the 1980s, the Great Recession of the Financial Crisis in 2007–2009, and the Great Lockdown of the COVID-19 pandemic in 2020. Section 4 argues for a coherent strategy based on the duty to inform (a professional standard for data producers), a return to transparency on inaccuracies, and empowering data users. Section 5 offers some concluding observations.

2. National accounts and inaccuracy

For the pioneers of national accounting, the inaccuracy, imprecision, and incompleteness of (macro)economic data to the tune of ten to twenty percent were realities, much more than for the current user of the accounts and their most important headline statistic, national income. This awareness is clear from the start of the combination of three lenses or measurements as a method to reduce measurement errors (Stone et al., 1942, see also Comim, 2001) via the use of three different accounting approaches (the income, expenditure and production accounts). Importantly, the teams that developed the first national accounts in the United Kingdom and the United States frankly provided expert opinions on the reliability of (components of) national income (Mitchell et al., 1922; Kuznets, 1941; see also Biddle and Boumans, 2021). These publicly available expert opinions on inaccuracies are best practices that

---

[2] In van Bergeijk (2024) I review the areas studied in Morgenstern (1963) and find that macroeconomic measurement error is still of similar size as in his case studies.

accompany the earliest publications of the national accounts in other countries (see Table 1 on the self-reported error ranges that accompanied the National Accounts in the Netherlands).

Table 1. Self-reported error ranges of National Accounts line items (Netherlands, 1950).

Grade	Examples	Self-reported error range
a	Indirect taxes (receipts), wages	2%–5%
b	Capital goods, profits, current account	5%–10%
c	Household saving, depreciation, inventories	10%–20%
d	Reserves (larger firms)	>20%

Source: Centraal Bureau voor de Statistiek (1953).

This transparency is in sharp contrast with current practices. Notifications such as the following explanation from the UK Office for National Statistics are common: "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 the error aspects and their impact on the GDP. While development work continues in this area, similar to all other G7 national statistical institutes, we do not publish a measure of the sampling error or non-sampling error associated with GDP.[3]

The choice to not publish a measure of inaccuracy had already become common practice in the early 1960s (Morgenstern, 1963, pp. 252–253). This reflects the hesitance of the statistical community to provide error estimates unless these are accurate. However, in this case, 'better is the enemy of good'. Indeed, the illusion of accuracy simply continues if the measurement error is only reported for cases where the inaccuracy could be attributed and estimated with precision. Of course, there is a well-developed standard practice to provide the audience with the headline figures and a discussion of uncertainty, though this suffers from the Caveat Fallacy. Afterall, if the audience sees or hears a number, then the number is the message that it gets whatever the data supplier says and writes on its uncertainty and inaccuracy (Guberek, 2019). Additionally, the inaccuracy of economic statistics is not a regular concern in academia, perhaps because the national accounting procedures have become too complex and difficult to grasp for outsiders (Coyle, 2017). Moreover, it may be that a disinterest in the exact definitions, data collection, measurement issues, methods, and their exact documentation exists (Knibbe, 2019). Anyhow, for most economists, national income is in the words of O'Brien (1998, p. 2) *a fait accompli*, which reflects the general assumption that the definition and measurement of the GDP have been settled long ago on the basis of the 'fundamental notions of economic theory'.

2.1. Transparency is possible

In recent discussions on the reporting of inaccuracies and uncertainties by an official institution, The Bank of England often features as a prominent example (e.g., Manski, 2015; van der Bles et al., 2019). The Bank regularly publishes a band for historical GDP data in its Monetary Policy Report; while this at least helps to communicate that the recent past is statistically speaking more fluent than

---

[3] https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi accessed August 1, 2023.

the users assume, the graphical representation of the GDP inaccuracy (incidentally, already pioneered by Morgenstern, 1963, Figure 10) is that it does not address the important 'Why?'-and 'Does it matter?'-questions. Moreover, there is a fundamental problem with the band because it assumes time stationarity of the data measurement and revision processes (Manski, 2019, p. 7639). This is problematic because the reported historical data are in different stages of revision (it is also unclear which data vintage is actually being reported). As will become clear in Section 3, some years are much more likely to be revised than others and often have no convergence towards a 'final estimate'. Answering 'Why' and 'Does it matter' are important steps towards understanding the macroeconomic data uncertainty. The essential first step to be set is to make the material available that is already collected and evaluated during the process of National Accounting, though this typically remains under the hood. Best practice examples that provide much needed transparency are Akritidis (2002), Harchaoui et al. (2004), and Meader and Tily (2008). Akritidis (2002) reported on the accuracy assessment by the UK Office for National Statistics of the 2000 GDP, which detailed the data sources (surveys, administrative data, models), adjustments (validation, conceptual issues, balancing), quality indicators (response rate, error, coverage), and GDP components that were impacted, and noted a 3.5% adjustment in the income approach amongst others. Harchaoui et al. (2004) discussed the industry data quality in the input-output tables of Statistics Canada and exemplified how and why reliability issues arose from the imputed data, the lack of deflators, or series breaks. They provided comparative analyses of the data series and found that the labor data was mostly reliable, while the capital data was less and that lower aggregation levels were associated with lower reliability. As a part of the plans for the modernisation of the UK National Accounts, Meader and Tily (2008) monitored the quality of quarterly estimates through 2007 and 2008. They provided inter alia information on price coherence adjustments, the size and development of revisions, and real time alignment adjustments and residual errors. They discussed the main surveys and sources that feed into GDP assessments and provided benchmarks against external data. These three examples showed that a lot of information is already available. There is no good reason why this manner of essential 'meta-data' should not become a regular element of standard reporting in the framework of National Accounting.

2.2. The academic approach

Discomfort with timeliness and accuracy, especially during the 'data fog' of the COVID-19 pandemic, has led academics unto a quest for new modes of observations. These new modes include satellite observation of night light intensity (Nordhaus and Chen, 2015; Hu and Yao, 2022), social media (Indaco, 2020), and even seismometers (Tiozzo-Pezzoli and Tosetti, 2022). The increase of measurement rods holds potential for improving the measurements, though it is worrying that conceptual and measurement errors are typically overlooked. At the same time, the academic discourse is strange because it challenges a statistical system that has clear strengths. National accounting has a long history of reliably application, is solidly grounded in theory, and aims to reduce measurement errors by design. This can be compared to the new modes that was comprised of blended opportunity data that was collected without any theory of measurement, quality standards, transparency, and more often than not without a history of application. There still is a lot of simplicity in papers that use administrative data and Big Data. All too often, these contributions focus only on (assumed) promises and opportunities without consideration of the drawbacks. These drawbacks include privacy issues, discontinuities, erratic access, inconsistent methodologies, and bias. Big data users (academics and

policy makers alike) assume that Big Data's coverage is better and much more complete. The practice provides sobering lessons, a prime of which is the so-called Big Data Paradox (Meng, 2018). Paradoxically, increasing the sample size suggests an increased accuracy, as it reduces the confidence intervals; however, at the same time, it magnifies the effect of survey bias: 'the more the data, the surer we fool ourselves'. Additionally, there is a lot of naivete in the hope that artificial intelligence (AI) in combination with the new modes of observation can provide better and more timely estimates of key macroeconomic statistics. The training material for AI models would a priori seem to be compromised because measurement errors are currently not reported; thus, AI models will tend to simply repeat the current bad practices. Related risks are the dart of training materials for AI models and AI-hallucination when data are lacking.

3. The growth rate of Gross Planetary Product

I am motivated to investigate the growth rate of world production by the impact of the IMF World Economic Outlook (WEO) on academics, policy makers, and the public at large,[4] the fact that these data are produced in a highly professional organization with well-trained and competent staff, and the Fund's detailed knowledge about the quality of underlying national statistics, thanks to its regular Article IV consultations and Dissemination Data Bulletin Board initiative. The use of a planetary variable, rather a national number, reflects the general nature of the topic that I want to address and helps to avoid the imbalance in the debate on measurement error that focusses on problems in the Global South to some extent. Additionally, the choice is guided by an *ex ante* error prognosis that basically recognizes that studying a real annual growth rate of a key concept at a very high level of aggregation with data that are at least three years old and produced by an international institution and an academic consortium can be expected to substantially reduce the measurement error compared to an alternative analysis of a monthly level of a nominal component of GPP that includes the most recent data and is produced by a national government (institute). Basically, this choice avoids six important problems that have been identified in the literature: the choice of the numeraire (that is currency conversion issues (e.g., Obstfeld et al., 2015), seasonal adjustment (e.g. Manski, 2015, Abeln and Jacobs 2022), classification error by country and component (e.g. van Delden et al., 2023, Du Rietz, 2024), depreciation (e.g., Bos, 2009), and strategic over- and/or underreporting by national governments (e.g., Fariss et al., 2022). Thus, we can reasonably expect an investigation of the measurement error in the real annual GPP growth to arrive at a conservative estimate of the measurement error. Figure 2 reports the highest and largest reported annual growth rate for each year in the period from 1991 to 2019 in each of the 37 bi-annual vintages of the WEO database published in April 2006 to April 2024, inclusive. As a practical measure of inaccuracy, I define the Implicit Minimal measurement Error (IMME) for each year as a ratio with the average μ of the observations M_i (the 'best guess') in the denominator and the average distance of the observations M_i from μ in the numerator; therefore, for N vintages, we have the following:

---

[4] World production is mainly used in the publications of the international organizations including World Bank, WTO and UNCTAD (sometimes with slightly different versions). Substantial academic use of GPP occurs in the World System Theory, in macroeconomic modelling where the world's business cycle is important, and particular as one of the controlling variables in empirical studies on world trade and investment.

$$IMME = \frac{\sum |M_i - \mu|/N}{|\sum M_i|/N} \tag{1}$$

Obviously, *IMME* is a conservative estimate of the measurement error compared to other practical measurement error indicators, as it does not assume that one of the sources is correct and because it evaluates the average error at an average observation. For the period between1991–2019, the average median *IMME* was ten percent (median six percent), which against a median growth rate across years and vintages of four percent indicated a meaningful measurement error of 0.2 percentage points. Among these years, 1991 and 2009 stood out as years with particularly large deviations. By way of an external benchmark, the figure also reports two alternative estimates of the GPP growth that can be derived from the Penn World Table (PWT), which is another heavily used global data resource that indicates the Great Recession year 2009 as a year for which even the sign of the GPP growth has been difficult to establish[5].
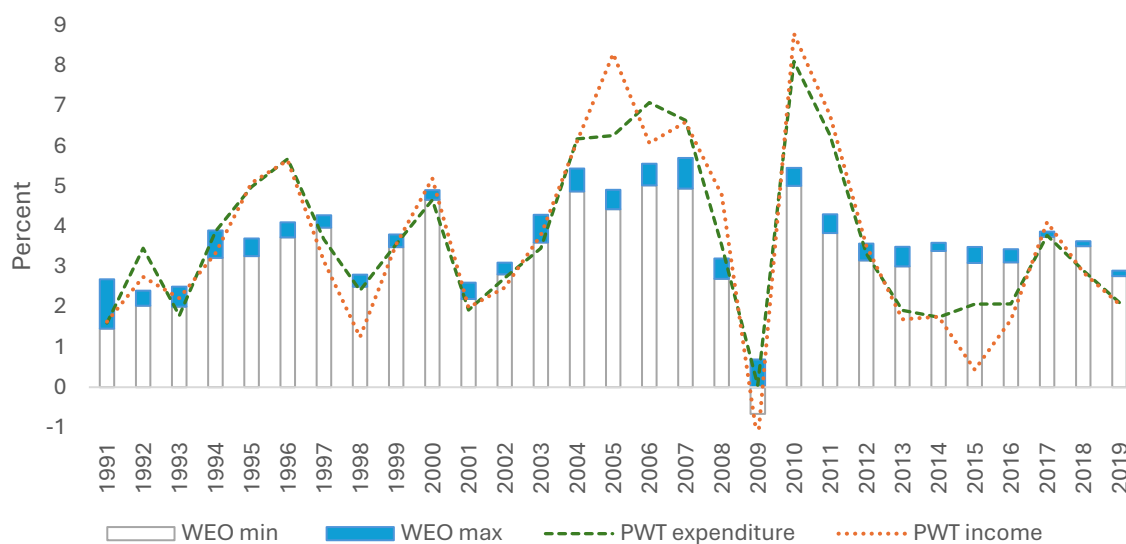

Figure 2. Annual GPP growth rate (largest and smallest estimate) in IMF WEO vintage and benchmark against latest PWT-based estimates (1991–2019). Sources: Calculations based on 37 vintages of the IMF WEO database, available from

---

[5] PWT, a data set that is regularly being updated and most widely available to and used by the economic profession, is currently maintained by a consortium of the universities of California and Groningen. The key variable of interest is GPP at Purchasing Power Parity (PPP). Unlike the IMF's WEO the PWT does not provide an estimate of GPP, but this measure can be derived by aggregating the 182 countries for which data are provided into a world total. Since 2009 PWT distinguishes and reports real GDP measured from the expenditure-side and from the output-side, respectively. For an individual country the difference between the two measures is related to the terms of trade effect, an effect which at the aggregate level of the world does logically not exist of course and thus the two measures for the world should coincide. The analogy is the world current account balance: the world cannot have a current account deficit or surplus and therefore the world's current account imbalances in the IMF WEO represent measurement error (Figure 1). The world likewise logically cannot have a terms of trade effect and therefore the imbalances between expenditure side GPP and output side GPP are useful indications of measurement error for PWT.

https://www.imf.org/en/Publications/SPROLLs/world-economic-outlook-databases. accessed and PWT 10.01 available from https://www.rug.nl/ggdc/productivity/pwt/. Data accessed June 3, 2024.

Figures 3 and 4 illustrate some of the patterns of the data vintages for specific years. Figure 3 reports the growth rate for the years 1991, 1995, 2000, 2005, 2010, and 2015. For the year 1991 (dark triangles), the pattern is not monotonic: up to April 2010, the reported GPP growth rate for 1991 decreased from 2.4 to 1.5, then jumped to 2.2, and from October 2014 moved in a fluctuating pattern towards a value of 2.7. For the other years, revisions also followed haphazard patterns. It is worth noting that fluctuations continued even in periods for more than five years after the fact.
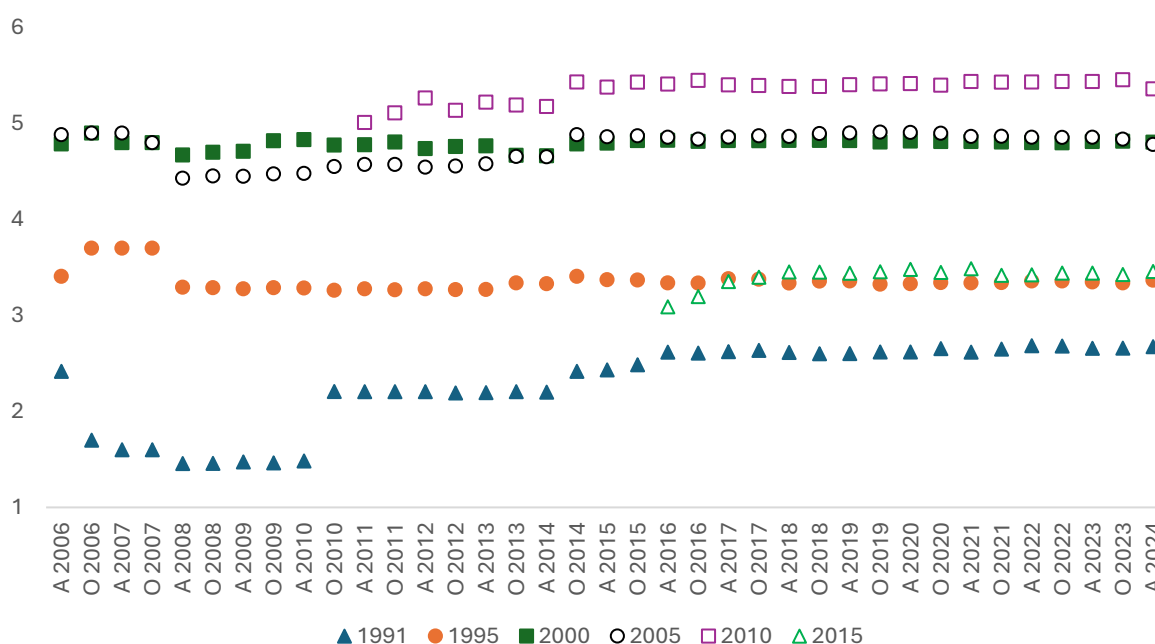

Figure 3. Pattern of revision for six years in 37 IMF WEO vintages. Source: see Figure 2.

Figure 4 zooms in on how statistical reporting on three major post Second World War recessions changed following the first estimate of the recession year. The historical GPP growth rates continued to fluctuate. Note that Figure 4 reports fluctuations during the first seven and a half years (i.e. 15 revisions on the horizontal axis), thus underreporting fluctuations in revision 16 and higher. For example, in the 37 vintages that were published since October 2006, the reported GPP growth rate for the Volcker Recession in 1982 fluctuated between 0.5 and 1.2. Figure 4 suggests that the three major recessions of the world economy were initially overestimated. Upward adjustments of the initial estimate by half a percentage point materialized in the 6th (Volcker recession) and 7th revisions (Great Recession and in the sixth year even to positive territory); however, for the Great Lockdown, an upward adjustment was already evident from the 2nd revision. The revision for the Great Lockdown year 2020 between April 2021 (−3.3) and April 2024 (−2.7) amounted to 0.6 percentage points (an *IMME* of ten percent). Despite these revisions and possible future adjustments, the Great Lockdown growth rate could be expected to remain in negative territory. For the Great Recession, the jury is still out with the most recent vintage reporting a shrinkage of 0.1 percent, which couldn't be considered to be

significantly different from zero against the background of the history of revisions. Nowadays, the Volcker recession appears to be a year with low growth – but not the originally reported recession.
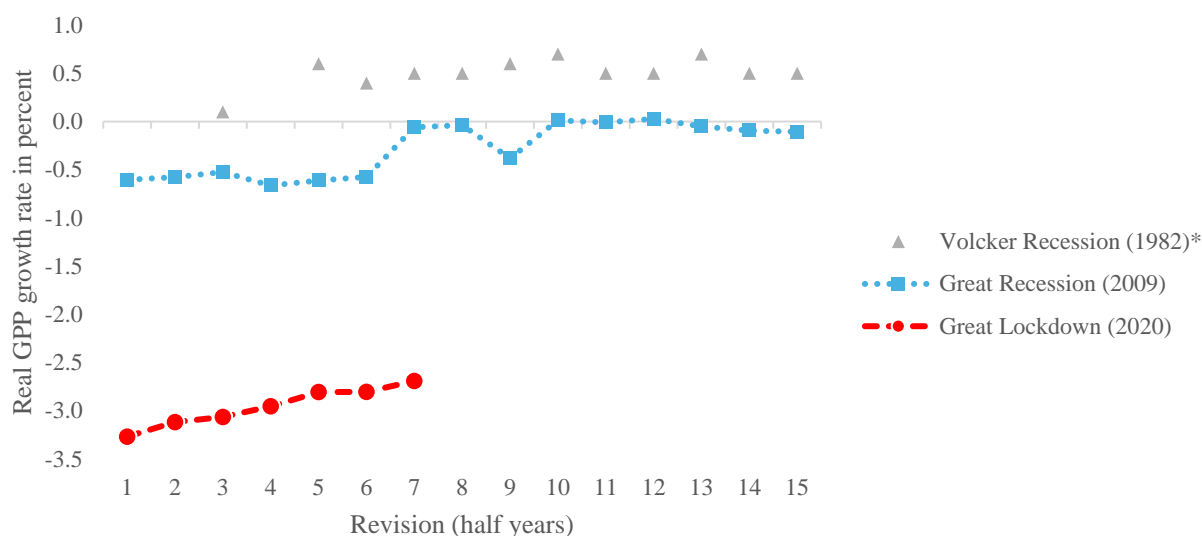

Figure 4. Revisions of GPP growth for three major post Second World War recessions. Sources: For the Volcker recession, data were collected from the 13 spring and autumn issues of the World Economic Outlook, typically from Table A1 in the statistical appendix, but since early editions sometimes do not report a GPP estimate but rather a G7 or advanced economies number in the statistical appendix occasionally data were collected from a table in the main text when the statistical appendix did not provide the world aggregate (for example, Table I-1 in the April 1984 issue). The data for 2008 and 2020 are derived from 23 post-2008 biannual World Economic Outlook Database vintages. Note: * The data for the Volcker recession are not available in a WEO data base and were collected from the print versions of each of the issues that appeared from April 1983 to October 1990, inclusive. No number for world production growth in 1982 is reported in the April 1984 issue. In 1983 no estimate for GPP growth was published.

It is worth repeating that the measurement inaccuracy that we illustrated in this section should probably be seen as the minimal kind of measurement error that economists should expect. First, the data producers that we studied were in high regard for their professionalism and experience—other less advanced data producers could be expected to perform not as well. Second, the analysis was based on an ex ante error prognosis that reduced many problems of measurement, including seasonal adjustment, choice of numeraire, delays in information availability, and classification. For other expressions of GPP (e.g., quarterly, nominal, very recent statistics or components,) data discrepancies will be much larger.

Our analysis showed that the IMF's statistics on the historical real growth rate of GPP certainly did not live up to the implied accuracy that was suggested by the use of three decimals in the *WEO* data base. If anything, a maximum should be set at 1 decimal and should work with quarters and halves of a percentage point to be a best practice response to the *IMME*s that were estimated. A band of 0.2 percentage points would be appropriate to take the implied minimal measurement error into

account. Reporting this kind of indeterminacy is perfectly doable at low costs, would do justice to the extent of measurement errors inherent to the aggregation of national account estimates, and would not undermine the economic narrative of world growth over the past decennia. For a limited number of individual years, the story line could be unclear, as illustrated by the example of the GPP rate for the Great Recession year 2009. Additionally, turning points in the cycle may occasionally be associated with different years. These indeterminacies should actually be considered as important benefits of the proposed procedure because this avoids the erroneous representation of these events as established facts in a manner that ignores their fundamental indeterminacy.

Moreover, the findings are sobering from a general statistical viewpoint. The measurement errors do not follow classical assumptions. Estimates of growth rates for a specific year often continue to fluctuate without an apparent convergence to a 'final estimate'. Additionally, nodes in the time series can be identified where both internal and external validations for the data indicate that specific years are measured with numerous problems than other years. A large component of measurement error appears to be non-transient.

However, the key take-away from this section is that the reporting of assessments of the measurement error based on vintages could become a standard procedure of the release of new data on national income and production without much effort.

4.   Empowering data users

None of the above could have come as a big surprise, as the stylized facts of macroeconomic data inaccuracies are well known in the statistical community, experienced policy makers, and well-versed applied researchers in the social sciences. Despite this general awareness, producers of statistics worldwide do not report on inaccuracies and suffering from digit-inflation, which is an unwarranted situation that makes data reliability opaque and hampers the very basis of evidence-based policy making. Therefore, neither the extent of the measurement error nor 'the everyday practice of ignoring it seems sustainable even though this situation has lasted for decades.' (Coyle, 2017, p. S228). The key question is how to organize a change for the better. This task should not be underestimated. It would be unrealistic to ignore the powerful forces that hinder the reporting of the measurement errors as a standard item. The proposed new best practice would be a paradigmatic change as many supposed 'facts' may be either non-existent or turn out to be insignificant. Older generations of economists may oppose because the established procedures and findings will be challenged with the risk of jeopardizing their reputations. This in itself creates a strong countervailing force. Additionally, interest-based statistics have become powerful policy and lobbying tools and may generate a resistance to change. In short, vested interests in theories, analyses, policies, and politics based on questionable 'facts' can be expected to oppose a change for the better. For three quarters of a century, critics of economic data inaccuracies have asked for fundamental reform in an attempt to make governments and international organizations problem-owners of measurement errors. Morgenstern (1963, p. 120) insisted that governments should be forced to stop publishing figures 'with the pretence that they are free from error'. Moreover, others have put the responsibility to report the measurement error on data producers (Manski, 2015). Unfortunately, the first best solution does not work: aggregate inaccuracies continue, digit-inflation increases, and the measurement errors are hardly ever provided. Recognizing that we live in a second-best world opens up the possibility that data users can become a force for the good.

Moreover, from this data user perspective, the task still amounts to a fundamental transformation. To alter the incentives, funding agencies, academics, editors, reviewers, institutions, journalists, and educators have to change how they deal with statistics. Funding agencies should require a transparent reporting of measurement errors in funded research projects. Editors and academics must insist that findings without measurement error indicators are unacceptable. Institutions should lead by example, and journalists should extend their work against fake news to include a hunt for misleading statistics. The economic curriculum should significantly emphasize the inaccuracies of available data from start to finish during the educational journey, and students should receive training on managing imprecision in their professional lives. This new paradigm would provide incentives and opportunities for a new generation of economists who are aware of, report, and, where possible, reduce measurement errors, thus ultimately enhancing the reliability of our economic knowledge and understanding. The power to enforce change might seem unbalanced: data producers hold a monopoly since others rarely have the capability to collect data of a comparable quality and scope. However, the monopoly on inaccuracies (which leads to the non-reporting of measurement errors) can be challenged, as illustrated in the previous section. The key issue is to move beyond the quest for the single most precise observation to combinations of measurements of an economic phenomenon, to provide a simple method to assess inaccuracies that are in reach, and to those that are not skilled in advanced econometrics. This might look as a step backward from the frontiers of current achievements, but it is a step forward because it will provide insight in the extent and behavior of the measurement error (that as we discovered already in the previous section runs counter to mainstream assumptions with the risk of exacerbating bias). The IMME proposed in Equation 1 would provide such a tool and Figure 5 provides a rough and somewhat subjective snapshot of the current state of the (in)accuracy. Judged by this standard, the IMF data on the GPP (Figures 2 and 3) would be judged as 'fair', but for individual global recession years as 'poor'.
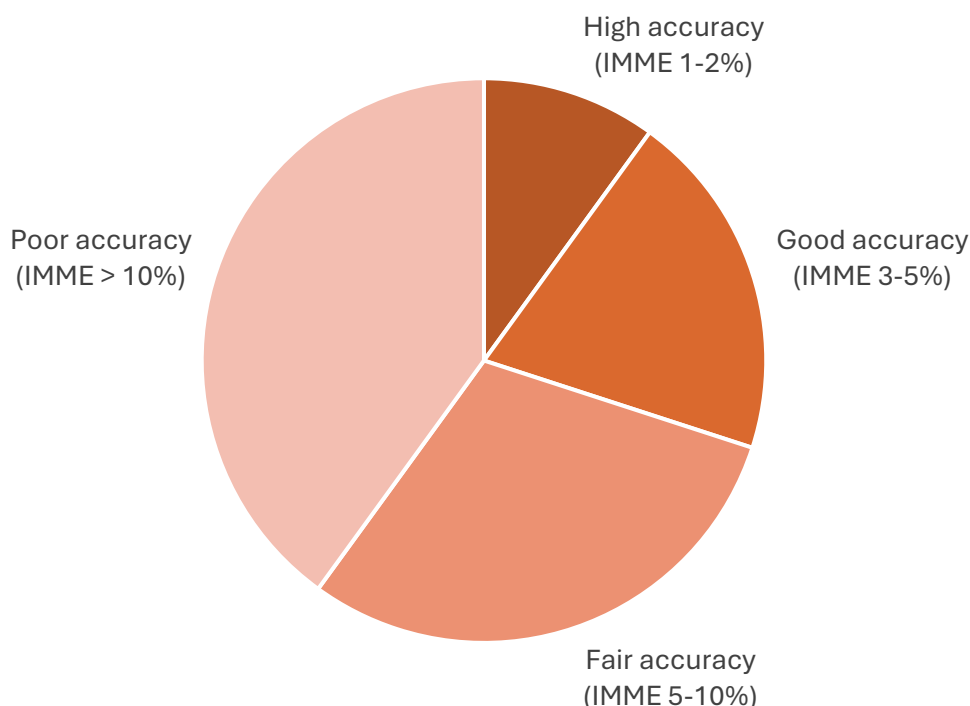

Figure 5. General overview of macroeconomic data by accuracy. Source: van Bergeijk (2024).

Empowering data users has not been explored as a strategy in the literature concerning the (non)communication of uncertainty and measurement errors. For instance, Van der Bles et al. (2019) only acknowledged a role for experts and official entities that 'own' the uncertainty, along with professional communicators (experts, communication departments, press).

This strategy can be significantly strengthened by incorporating three key elements. First, it is crucial to address the frequent lack of essential information in papers, even those published in leading academic journals. Details on basic data characteristics such as the number of observations, the standard deviations of variables, and specifics about data series can be included. To remedy this, economics should adopt a standardized data reporting format that must always be adhered to, thus ensuring that data sections in papers are explicit, detailed, and comprehensive. Second, change logs and accuracy assessments of statistics should be made fully available and easily accessible. This is a critical responsibility for statistical offices, as well as for academic data collectors. Third, integrating *ex ante* error prognoses into the research strategy is essential.

For data producers, the incentive is the fulfilment of their duty to inform, not just as a moral obligation, but as a professional standard. Transparency about uncertainties and inaccuracies is a fundamental aspect of knowledge. For policymakers, understanding the extent of error enhances the robustness and reliability of the evidence base, thus leading to better decision-making. Academic data users will be motivated by the opportunity to conduct superior research and contribute to the advancement of knowledge.

5. Concluding observation

It is worth thinking about how behavioral and institutional influences have created a low transparency equilibrium. Macroeconomic statistics are influenced by academic research and a strive for best practices at the statistical institutions. Past efforts, such as the Boskin report on CPI and the IMF's work on international financial flows, have influenced current practices, but did not successfully reduce macroeconomic inaccuracies in the end. This manner of hysteresis relates to the constant emergence of new economic activities in a dynamic economy that are difficult to accurately measure, and this adds to the challenge of maintaining accurate statistics; however, this is also caused in a very fundamental way the consequence of the non-reporting of measurement errors. It is time to change this bad practice.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

All author declares no conflicts of interest in this paper.

References

Abeln B, Jacobs JPAM (2022) COVID-19 and seasonal adjustment. *J Bus Cycle Res* 18: 159–169. https://doi.org/10.1007/s41549-022-00071-z

Akritidis L (2002) Accuracy assessment of National Accounts statistics. *Economic Trends*, 38–53. Available from: https://escoe-website.s3.amazonaws.com/wp-content/uploads/2020/01/01233815/ET-589-Accuracy-assessment-of-national-accounts-statistics-L-Akritidis-Dec-2002.pdf.

Biddle J, Boumans M (2021) Exploring the History of Statistical Inference in Economics: Introduction. *Hist Polit Econ* 53: 1–24. https://doi.org/10.1215/00182702-9414747

Bos F (2009) *The National Accounts as a Tool for Analysis and Policy: In View of History, Economic Theory and Data Compilation Issue*.

Boskin MJ, Dulberger ER, Gordon RJ, et al. (1996) *Toward a more accurate measure of the cost of living: final report to the Senate Finance Committee*, Advisory Commission to Study the Consumer Price Index, United States Congress, Washington D.C. Available from: https://www.ssa.gov/history/reports/boskinrpt.html.

Centraal Bureau voor de Statistiek, Statistics Netherlands (1953) Nationale jaarrekeningen 1948–1950, CBS: Voorburg.

Comim F (2001) Richard Stone and measurement criteria for national accounts. *Hist Polit Econ* 33: 213–234. https://doi.org/10.1215/00182702-33-Suppl_1-213

Coyle D (2017) The future of the national accounts: Statistics and the democratic conversation. *Rev Income Wealth* 63: S223–237. https://doi.org/10.1111/roiw.12333

de Jonge E (2020) Communicating uncertainties in official statistics—A review of communication methods. Available from: https://ec.europa.eu/eurostat/documents/3888793/12135507/KS-TC-20-011-EN-N.pdf/b2ee60fe-ed5b-ad3e-05bd-56f6b87dfec0?t=1609752126728.

Du Rietz S (2024) Accounting for GDP—A study of epistemic strategies when calculating the quarterly economy. *Account Organ Soc* 112: 101522. https://doi.org/10.1016/j.aos.2023.101522

Fariss CJ, Anders T, Markowitz JN, et al. (2022) New estimates of over 500 years of historic GDP and population data. *J Confl Resolut* 66: 553–591. https://doi.org/10.1177/00220027211054432

Feldstein M (2017) Underestimating the real growth of GDP, personal income, and productivity. *J Econ Perspect* 31: 145–164. https://doi.org/10.1257/jep.31.2.145

Goolsbee AD, Klenow PJ (2018) Internet rising, prices falling: Measuring inflation in a world of e-commerce. *AEA papers and proceedings* 108: 488–492. https://doi.org/10.1257/pandp.2018103

Grassman S (1983) Comments, In: von Furstenberg, G.M. (ed.), *International money and credit: the policy rules*, IMF, Washington DC, 248–252. https://doi.org/10.5089/9780939934270.071

Griliches Z (1994) Productivity, R&D, and the Data Constraint. *Am Econ Rev* 84: 1–23.

Guberek T (2019) Data Dilemmas: The Science and Politics of Communicating Uncertainty in Human Rights Information, PhD thesis University of Michigan: Ann Arbor.

Harchaoui TK, Kaci M, Maynard JP, et al. (2004) The Statistics Canada Productivity Program: Methodology 2000, Statistics Canada, Ontario.

Hu Y, Yao J (2022) Illuminating economic growth. *J Econom* 228: 359–378. https://doi.org/10.1016/j.jeconom.2021.05.007

Indaco A (2020) From twitter to GDP: Estimating economic activity from social media. *Reg Sci Urban Econ* 85. https://doi.org/10.1016/j.regsciurbeco.2020.103591

International Monetary Fund (1983) Statistical asymmetry in global current account balances, In: *World Economic Outlook*, IMF, Washington DC, 16–17.

International Monetary Fund (1987) Report on the World Current Account Discrepancy. IMF, Washington, DC.

International Monetary Fund (1992) Report on the Measurement of International Capital Flows. IMF, Washington, DC.

Jerven M (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*, Cornell University Press: Ithaca.

Kenessey Z (1997) A perspective on the accuracy of economic observations. *Int Stat Rev* 65: 247–259.

Knibbe M (2019) *Macroeconomic measurement versus macroeconomic theory*, London: Routledge. https://doi.org/10.4324/9781351136709

Kuznets S (1941) *National Income and its composition, 1918–1938*, NBER: New York.

Kuznets S (1950) Review of On the Accuracy of Economic Observations by Oskar Morgenstern. *J Am Stat Assoc* 45: 576–579. https://doi.org/10.1080/01621459.1950.10501149

Linsi L, Tanaka S, Giumelli F, et al. (2023) Measuring the unmeasurable: The politics of guestimates, 2022 International Political Economy Society Conference, Pittsburg.

Manski CF (2011) Policy analysis with incredible certitude. *Econ J* 121: F261–F289. https://doi.org/10.1111/j.1468-0297.2011.02457.x

Manski CF (2015) Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *J Econ Lit* 53: 631–653. http://dx.doi.org/10.1257/jel.53.3.631

Manski CF (2020) The lure of incredible certitude. *Econ Philos* 36: 216–245. https://doi.org/10.1017/S0266267119000105

Matolcsy G, Nagy M, Palotai D, et al. (2020) Inflation in the digital age: Inflation measurement and bias in the 21st Century. *Financ Econ Rev* 19: 5–36. https://doi.org/10.33893/FER.19.1.536

Meader R, Tily G (2008) Monitoring the quality of the National Accounts. *Econ Lab Market Rev* 2: 24–33. https://doi.org/10.1057/elmr.2008.39

Meng XL (2018) Statistical paradises and paradoxes in big data: law of large populations, big data paradox, and the 2016 us presidential election. *Ann Appl Stat* 12: 685–726. https://doi.org/10.1214/18-AOAS1161SF

Mitchell WC, King WI, Macaulay FR (1922) *Income in the United States, Its Amount and Distribution, 1909–1919: Detailed report*, Brace: Harcourt.

Morgenstern O (1950) *On the Accuracy of Economic Observations*, Princeton University Press: Princeton.

Morgenstern O (1963) *On the Accuracy of Economic Observations*, (2nd completely revised edition), Princeton University Press: Princeton.

Moulton BR (2018) The measurement of output, prices, and productivity: What's Changed Since the Boskin Commission?, The Hutchins Center on Fiscal and Monetary Policy, Brookings Institution: Washington, DC.

Nordhaus W, Chen XI (2015) A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. *J Econ Geogr* 15: 217–246. https://doi.org/10.1093/jeg/lbu010

O'Brien ES (1998) Contested accounts: the evolution of the meaning and measurement of national income, PhD thesis University of Notre Dame: Notre Dame, Indiana.

Obstfeld MO, Hemmati M, Milesi-Ferretti GM, et al. (2015) The choice of numeraire matters when calculating world GDP growth.

Reiss J (2008) *Error in Economics: Towards a More Evidence-Based Methodology*, Routledge.

Stone R, Champernowne DG, Meade JE (1942) The precision of national income estimates. *Rev Econ Stud* 9: 111–125. https://doi.org/10.2307/2967664

Tiozzo-Pezzoli L, Tosetti E (2022) Seismonomics: Listening to the heartbeat of the economy. *J R Stat Soc Ser A-Stat Soc* 185: S288–S309. https://doi.org/10.1111/rssa.12912

van Bergeijk PAG (2013) *Earth Economics: An Introduction to Demand Management, Long-Run Growth and Global Economic Governance*, Edward Elgar: Cheltenham.

van Bergeijk PAG (2024) *On the inaccuracies of economic observations: Why and How We Could Do Better* Edward Elgar: Cheltenham.

van Delden A, Scholtus S, Burger J, et al. (2023) Accuracy of Estimated Ratios as Affected by Dynamic Classification Errors. *J Surv Stat Methodol* 11: 942–66. https://doi.org/10.1093/jssam/smac015

van der Bles AM, van der Linden S, Freeman ALJ et al. (2019) Communicating uncertainty about facts, numbers and science. *R Soc Open Sci* 6: 181870. https://doi.org/10.1098/rsos.181870