

The national accounting paradox: how statistical norms corrode international economic data

European Journal of
International Relations
2021, Vol. 27(2) 403–427
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1354066120936339
journals.sagepub.com/home/ejt



Daniel Mügge¹  and Lukas Linsi²

¹University of Amsterdam, Amsterdam, Netherlands

²University of Groningen, Groningen, Netherlands

Abstract

The transnationalization and digitization of economic activity has undermined the quality of official economic statistics, which still center on national territories and material production. Why do we not witness more vigorous efforts to bring statistical standards in line with present-day economic realities, or admissions that precision in economic data has become increasingly illusive? The paradoxical answer, we argue, lies in the norms underpinning global statistical practice. Users expect statistics to draw on unambiguous sources, to allow for comparison over time and across countries, and they prize coherence—both internally and with holistic macroeconomic models. Yet as we show, the ambition of the transnational statistical community to meet these norms has in fact undermined the ability of economic data to represent economic life more faithfully. We base our findings on interviews with two dozen leading statisticians at international economic organizations, archival research at the International Monetary Fund and a thorough review of debates among statistical experts.

Keywords

Economic measurement, international organizations, constructivism, balance of payments, norms, statistics

Corresponding author:

Daniel Mügge, Universiteit van Amsterdam, Nieuwe Achtergracht 166, Amsterdam 1018WV, Netherlands.
Email: d.k.muegge@uva.nl

There is a growing appreciation that the statistical compilation tools and accounting frameworks designed and developed over the last 60 years . . . may reflect a world that no longer exists.

Nadim Ahmad, Head of Trade and Competitiveness Statistics Division, OECD
(Ahmad, 2018: 1)

Introduction

Statistics are the bedrock of economic policymaking and debate. They allow computation, comparison, historical analysis, and future forecasting. Without such data, “the economy” would remain an intractable abstraction for policymakers, citizens, and analysts alike.

Yet, the quality of ubiquitous economic data is much worse than their users typically acknowledge (Damgaard and Elkjaer, 2017; International Monetary Fund, 1987, 1992; Linsi and Mügge, 2019; Morgenstern, 1963; UNECE, Eurostat and OECD, 2011). If economic data fail to capture what they purportedly claim to represent, public deliberation, economic policy, and academic analysis drawing on them all suffer.

Statistical quality has deteriorated because of a widening gap between the concepts international economic data claim to capture and the measurements that find their way into official databases—a phenomenon we call the *concept–measurement gap*. Indicators had been devised for economic structures clustered in national territories and focused on material production—the industrial economies in the Global North that we associate with the decades following the Second World War. Today, these structures are transnationally integrated, and intangible production and assets—services, derivatives, knowledge, licenses, and so on—are central. But while the transnationalization and digitization of economic activity has undermined the conceptual validity of key economic indicators (Ahmad, 2018; Lipsey, 2006), our statistical concepts have hardly changed. This is true for many macroeconomic figures, yet it particularly affects Balance of Payment (BOP) statistics, which measure cross-border flows of goods and capital, collected following the *Balance of Payments Manual* (BPM) issued by the International Monetary Fund (IMF).

Statisticians who craft the standards for BOP statistics are keenly aware of the problems an increasingly transnational and intangible economy poses (Ahmad, 2018; Bloch and Fall, 2015; Moulton and van de Ven, 2018; UNECE et al., 2011). Yet their attempts to address the concept–measurement gap have thus far been remarkably ineffective. A priori, we might expect statisticians to respond in two possible ways. They could either overhaul statistical standards to match the new economic structures. Or they could incorporate ambiguity in their published statistics, for example by using uncertainty margins, or by simply admitting that we lack meaningful figures. But we observe neither. The production of data continues largely unchanged, leaving most data users with the erroneous impression of high-quality figures. Why, we ask, is the widening concept–measurement gap neither narrowed by reforming standards nor reflected in the data itself? What explains the skewed statistical representations that surround us and guide economic debates and policy?

We argue that the stickiness of statistical standards stems from the norms that underpin macroeconomic statistics as a field of transnational knowledge production. Our analysis highlights four norms that create a strong conservative bias in international statistical standards. We call them *comparability* (the desire to compare statistics across countries),

continuity (the ambition to build time-series datasets), *certitude* (the predilection for reliably quantifiable data), and *coherence* (the aspiration to integrate separate statistical domains into one overarching representation of “the economy”). Considered in isolation, these norms seem common sense. But adhering to them also keeps measurement standards from accommodating present-day economic realities, which increasingly resist unambiguous quantification and pigeonholing in national accounts. Deference to these statistical norms damages the economic figures that populate our databases, politics, and news.

Our argument builds on a growing scrutiny in International Relations (IR) of the production and use of quantitative information in global politics—for example in the form of rankings or indicators (Broome and Quirk, 2015; Broome et al., 2018; Cooley and Snyder, 2015; Davis et al., 2012; Honig and Weaver, 2019; Kelley and Simmons, 2015, 2019). From this literature we take the question why indicators are produced the way they are, and specifically, in our case, why Balance of Payments statistics continue to suffer from the concept–measurement gap we identify. At the same time, our emphasis on the norms underlying statistics production harks back to sociologically informed analyses of international institutions more generally (Babb, 2007; Barnett and Finnemore, 2004; Chorev and Babb, 2009; Kentikelenis and Seabrooke, 2017; Murdoch et al., 2018).

Our empirical investigation centers on the evolution of the IMF’s authoritative *Balance of Payments Manuals* and the key economic indicators defined therein, in particular for trade, foreign direct investment, and portfolio capital flows. We draw on a range of sources to show how deeply ingrained norms skew the production of macroeconomic statistics: specialized reports from statistical agencies and international organizations that produce statistics or oversee standards reveal how much the transnational and intangible economy has dented statistical quality. Two dozen interviews with leading statisticians in Paris, Frankfurt, The Hague, London, Geneva, New York, and Washington offer insights into the concerns, trade-offs, and norms as experienced by central figures in international economic statistics. Documents from the IMF archives in Washington allow us to trace these norms backward through time, at times to the very beginning of systematic BOP statistics.

In the remainder of this article, we first situate our research within broader social science understandings of economic statistics. We then detail how the rise of the transnational and intangible economy has widened the concept–measurement gap to the point where official statistics grossly misrepresent economic relationships and dynamics. Finally, we show how the norms of comparability, continuity, certitude, and coherence structure the statistical field and explain how, paradoxically, they help to reproduce inadequate statistical standards.

Macroeconomic statistics in an IR perspective

Macroeconomic statistics have been an international success story.¹ The global spread of GDP as the universal metric to gauge economic prowess has been thoroughly documented (Fioramonti, 2013; Lepenies, 2013; Masood, 2016; Philipsen, 2015). But international organizations such as the United Nations, the International Labour Organization, the World Bank, and the International Monetary Fund have promulgated a much wider

set of economic statistics (Ward, 2004), including for example poverty measures (Clegg, 2010), government finance statistics, balance of payments statistics, and internationally harmonized unemployment statistics. More recently, international organizations as well as nongovernmental organizations (NGOs) have proactively crafted new indicators and rankings to nudge, name, and shame governments toward different policies (Broome and Quirk, 2015; Broome et al., 2018; Cooley and Snyder, 2015; Honig and Weaver, 2019; Kelley and Simmons, 2015, 2019).

At the same time, economic concepts such as unemployment (Baxandall, 2004; Salais et al., 1986), growth (Pilling, 2018; Schmelzer, 2016), inflation (Mackie and Schultze, 2002; Stapleford, 2009), or debt (Bloch and Fall, 2015) defy straightforward definition and measurement. The concepts macroeconomic statistics purport to capture are best understood as social facts (Searle, 1995)—constructs that derive their power in society and politics from their institutionalization and widespread acceptance (cf. Chwieroth and Sinclair, 2013). Statistics boost the public and societal role of such macroeconomic concepts, because they translate them into concrete numbers that can be tracked, compared, and used for computations. Quantification makes abstract concepts amendable to routinized application, for example in bureaucracies (Desrosières, 1993; Porter, 1995).

As economic concepts are institutionalized through concrete measurement routines, they rigidify. Indeed, by codifying what counts as growth, inflation, or unemployment, and what does not, these standards delineate “the economy” as a governance object itself (Allan, 2017). GDP is no longer a more or less appropriate way to measure economic growth; instead, in public discourse growth *is* whatever GDP measures—even if agreement emerges that the measure is outdated (Stiglitz et al., 2010). If statistical measurement approaches are inflexible, they can thus take public and policy debates hostage, as such debates inadvertently become tied to obsolete codifications of the concept in question.

Statistical measurement routines are open to challenges from several sides. They may be attacked because they obscure politically salient dimensions of a phenomenon. GDP has been vilified for its omission of unremunerated labour, mostly by women (DeRock, 2019; Hoskyns and Rai, 2007; Waring, 1999) and of environmental destruction (Fioramonti, 2013); unemployment measures can suffer from both racial and gender biases (Alenda-Demoutiez and Mügge, 2019). Challenges also emerge as structural changes in economic life—say, the rise of derivatives in finance or expanding global trade—clash with the conceptual assumptions on which statistical measures had been built. Given that our macroeconomic statistical edifice essentially dates back to the mid-20th century, the question thus is not only which political forces were responsible for the initial statistical choices (Mügge, 2016). It is also why measurement approaches have remained inflexible in spite of such forceful challenges as we describe in this article.

The norms structuring the production of economic statistics

To understand such inertia, we need to analyze the dynamics among the statisticians who set global standards. Over the decades, a tightly knit transnational epistemic community has emerged that dominates global standard setting for macroeconomic statistics (Ward, 2004). Central hubs include the Statistics and Data Directorate at the OECD, the

economic statistics branch at the United Nations Statistics Division (UNSD), Eurostat as a focal point for European statistical expertise, and, for finance statistics in particular, the statistics department of the IMF (Harper, 1998). Overlapping and rotating membership and leadership of standard-setting bodies has generated a small but highly integrated community of statistical experts in charge of setting and reforming international standards.

Much IR scholarship has highlighted how ideas and beliefs can become institutionalized in international organizations and diffused through them (Adler and Haas, 1992). In the economic realm, the focus has been on the IMF and the World Bank in their promulgation of Washington Consensus-inspired policies (Barnett and Finnemore, 2004; Broome, 2010; Chorev and Babb, 2009; Chwioroth, 2009; Kentikelenis and Seabrooke, 2017; Weaver, 2007; Woods, 2006). While we take inspiration from this work, our case is somewhat different: statistical standard setting has much less immediate and obviously distributional effects than, say, IMF's lending decisions. They are therefore less politicized, leaving more room for expert deliberation. Given that statisticians recognize the measurement problems we outline below, we would expect statistical standards to be attuned to contemporary economic circumstances. Yet that is not what we observe. So what keeps statistical standards in place when they face such challenges to their solidity?

Our answer emphasizes four norms underpinning economic statistics: certitude, comparability, continuity, and coherence. These statistical norms capture desirable attributes of statistics—in essence, they define what characterizes “good” data. While we focus on cross-border flows of trade and capital, the relevance of these statistical norms extends beyond the transnational or the economic realm. They apply to public statistics more widely. We therefore first lay out these norms and what motivates them in general terms.

Alongside their apparent simplicity, numbers are attractive in politics and policymaking due to their air of objectivity (Dorling and Simpson, 1999; Porter, 1995; Sætnan et al., 2011). Arguments backed up by numbers carry authority, even if the figures rest on shaky foundations. This emphasis on numbers in policy has only grown as new public management has introduced corporate practices such as auditing and cost–benefit analysis into the public sector (Knafo, 2019; Power, 1997).

Claims to objectivity require reliable techniques to gather and aggregate data: assessments based on individual judgment and experience must give way to indicators that are readily reproducible by building on unambiguously quantifiable information (Daston and Galison, 2007). Thereby statistics introduce a “countability bias” (Mügge, 2019) into public policy, systematically privileging information that can be entered into spreadsheets (student numbers and awarded diplomas, prices of goods, number of people with full-time jobs) at the expense of things that are hard to quantify (student learning, value creation, job security, and satisfaction) (Muller, 2018). Because reliability has a specific meaning in scientific measurement (Krippendorff, 2008), we use the label *certitude* instead to describe the associated norm: statistics should contain as little information as possible that requires subjective interpretation. Although the certitude norm chimes with statisticians' mandate to produce “objective” information, it becomes a problem when the properties we want an indicator to capture grow resistant to straightforward quantification.

To be sure, economic statistics never speak for themselves, even if we ignore how they were put together. They need to be narrated and put in context for us to make sense of them (Beckert, 2016; Leins, 2018; Muniesa, 2014). Rising unemployment may signal the malfunctioning of stifled labor markets (the liberal interpretation) or cyclical gyrations of unchecked capitalism (the critical one). Statistics require policy goals, programs, and interpretative frames to unleash their full force (Abolafia, 2010).

Political and policy narratives that use statistics often have a comparative dimension: they compare units with each other (countries, provinces, schools, and so on) and track the evolution of indicators over time. For statistics to function in this way, measurement standards must be harmonized between units and stay constant over time. If we use different yardsticks in different places or adapt them from one year to the next, observations are no longer directly comparable.

Two norms follow: *comparability* (the interunit comparison) and *continuity* (the constancy of measurement standards over time). Both limit the adaptation of statistical standards to changing economic circumstances. Users who demand continuity—be they policymakers or academics—will object to frequent breaks in time series. Comparability works differently. Once countries have agreed to a shared measurement standard, it will take collective agreement to adapt it. The harder it is to capture the object of measurement in updated standards, the longer such agreement will take. A commitment to harmonized standards thereby retards their adaptation to structural economic changes. It also pushes countries toward relatively unambiguous standards, lest *de jure* harmonization becomes a *de facto* free for all (cf. Aragão and Linsi, 2020).

Certitude, comparability, and continuity are three of the four statistical norms we highlight. They are, indeed, norms. Whether they achieve their goals is a different matter. As has often been noted in measurement theory, concept validity and reliability can be at loggerheads. If the concept validity of a reliable proxy is poor, the ultimate measure may contain little information about the concept it purportedly captures. Certitude as a norm can generate mock-accuracy.

Applying similar standards to diverse countries can generate data that make meaningful comparison impossible, just as sticky measures may fail in the face of societal transformation. Poverty indicators are a good example: many countries define household poverty relative to median household income. As societies grow more affluent, the material deprivation associated with poverty may vanish; at the same time, the condition of occupying the bottom rungs of the societal ladder remains constant. Depending on which dimension of poverty one highlights, keeping the measure constant may frustrate or aid over-time comparison. Either way, it is far from obvious that keeping measures constant aids comparability if the object being measured changes.

The final norm that obstructs the adaptation of statistical standards is *coherence*. In macroeconomic statistics, individual measures do not exist in isolation. Instead, theory or common sense tell us how they relate to each other as variables in a model or representation of “the macroeconomy” (Mankiw, 2017). According to economic theory, GDP, for instance, can be compiled in three ways—based on production, expenditures, or income (Lequiller and Blades, 2006). For the sake of theoretical coherence, the total market value of goods and services (production-side) should be identical to the sum of consumption, investment, government purchases, and net exports (expenditure-side), as

well as the sum of labor and capital income (income-side). Macroeconomic theory specifies how the different quantities relate to each other, and because they interact like cogs in a clockwork, the definition and measurement of one cannot be changed without affecting the others—putting up a formidable obstacle to replacing single parts of what aspires to be a coherent whole.

On their own, these four norms capture desirable characteristics of statistics. Data users—be they policymakers, academics, or private sector professionals—structurally expect figures that conform to them. They are indispensable if data is to be deemed useful in over-time comparisons, policy evaluations, computations, authoritative justification of public policies, and so on. In statistical practice, the question is not whether any individual figure optimally captures a particular phenomenon, but whether it can be processed productively because it conforms to the expectations users have of data.

The mission of statisticians then is not to produce data that is “correct” in any strict sense but data that is useful to the various constituencies that work with it. Note that “usefulness” here does not suggest manipulation. On the contrary, it implies a commitment to abstract statistical norms often meant to *counter* arbitrary manipulation, for example through codification of unambiguous measurement rules or adherence to international best practices—even if those may fail to suit economic realities on the ground (Linsi and Mügge, 2019). Statistical norms thus sit between producers and users of statistics. However, because data producers—statisticians themselves—are tasked with implementing these norms, they are the focus of our empirical investigation.

The four norms act not only as brakes on adapting measures to new circumstances; they also limit the ability of statistical representations to incorporate ambiguity. Certitude clearly privileges specific numbers, just as comparability across units and over time breaks down without them. Coherence, too, demands exact quantification; admitting to ambiguity in one area would infect the whole interlinked system.

Further below we outline how statisticians’ pursuit of these norms has fueled a growing concept–measurement gap in BOP statistics: measures of international economic transactions are ever less aligned with the theoretical constructs they purport to measure. This trend is widely acknowledged among statistical circles but frequently ignored in academic and public debates. In the next section we present the debate and the key drivers of the growing concept–measurement gap; after that, we turn to the question why so little is done to close it.

Balance of Payment statistics and their discontents

The monitoring of imports and exports has been an obsession already in mercantilist Europe in the 16th and 17th centuries (Lipsey, 2006; McCormick, 2009; Morgenstern, 1963; Studenski, 1958). These efforts only intensified as governments systematized their economic records in subsequent centuries. The first attempt to collect international BOP statistics involved the League of Nations in the 1920s and 1930s. In the aftermath of the Second World War, the responsibility shifted to the IMF (Alves, 1967). As the guardian of international financial and economic stability in the Bretton Woods era, the Fund was responsible for identifying unsustainable imbalances in global financial flows (International Monetary Fund, 1948: 1). In a world of fixed exchange rate regimes, the

original *raison d'être* of BOP accounting was to track changes in countries' official foreign exchange reserves (Cohen, 1969; Machlup, 1950; Meade, 1951). Toward this end, the IMF strove for international conventions on how to collect data on cross-border payments.

The first *Balance of Payments Manual* (BPM1) issued in 1948 (International Monetary Fund, 1948) offered standardized templates for member countries to fill out each year. A slightly expanded version, with more detail about what to include and exclude, followed two years later (International Monetary Fund, 1950). Since then, the IMF's BOP Statistics enterprise has only grown in size and ambition. With the turn to flexible exchange rates and the freeing up of capital mobility in the post-Bretton Woods era, the policy objective of BOP monitoring became increasingly complex (Bryan, 2001; Bryan et al., 2017; Pitchford, 1994).

Users of BOP statistics typically assume the data to be accurate and reliable pieces of information (Linsi and Mügge, 2019: 365). Insiders in the statistical community, however, have been quietly voicing doubts since the 1980s. A 1987 IMF report, for instance, found that

[in] the period after 1979, the available statistics on the world current account began to show a large negative discrepancy. [...] Concern that such discrepancies could lead to inappropriate policy reactions was heightened in 1982, when the excess of reported debits exceeded \$100 billion. [...] Improving the world's data on current account transactions will be a formidable task, especially in an environment where the capacity for statistical measurement is challenged by rapid changes in the technology and forms of international transactions and by budgetary constraints. (International Monetary Fund, 1987: 1)

Five years later, a similar report on capital account discrepancies reached even starker conclusions, finding the "world capital accounts system" to be "in a state of crisis" (International Monetary Fund, 1992: 2). The stakes were clear: "there are strong indications that this body of information on which good economic management depends is undergoing a serious and progressive deterioration" (International Monetary Fund, 1992: 9). That was the IMF's verdict almost 30 years ago.

Our analysis of mirror trade statistics (Linsi and Mügge, 2019) compared the trade or capital flows one country reports sending to another one with the figure this second country reports for incoming flows. In principle, the two should match; in practice, they do not. In 2014, the value of exports of merchandise goods from the Netherlands destined to neighboring Germany was estimated at \$165.6 billion by the Dutch authorities, while official figures from Germany valued imports from the Netherlands at \$96.6 billion; the United States estimated importing goods from China worth \$466.7 billion, while Chinese sources indicated their value to be \$397.1 billion, and so on (own calculations based on IMF DOTS database).

Such discrepancies are nothing unusual. A comprehensive analysis of a global dataset of bilateral merchandise trade flow numbers found that mirror records differ, on average, by no less than a factor of 1.7 (Schultz, 2015: 138). The situation is even worse for capital flows, which are harder to measure than merchandise trade. An IMF analysis of discrepancies in their own bilateral foreign direct investment (FDI) data reported that

for 44 percent of the 1,805 published bilateral economy pairs . . . one economy's number is at least twice as high as the counterpart economy's number, and for almost 10 percent of the pairs, one number is at least 10 times higher than the mirror number. (Damgaard and Elkjaer, 2017: 5–6)

A range of factors can lead countries—even if formally they adhere to the same global statistical standard—to assign different values to the same transaction, such as cross-national differences in data collection practices, differing levels of statistical capacity, the use of different versions of statistical manuals, and so on.

Measurement problems sit deeper than intercountry differences in measurement approaches, however. As a growing number of reports by statisticians acknowledges (Damgaard et al., 2019; International Monetary Fund, 1992; UNECE, Eurostat, & OECD, 2011), official statistics map less and less well onto the economic complexity they purport to capture. They implicitly model the world economy as an interconnected system of semiclosed national economies (Lepenies, 2013; Masood, 2016). Yet this conceptualization is less and less appropriate to capture economic activities in an ever more integrated global economy, in which trade and capital flows crisscross national borders in enormously complex patterns (Oatley, 2019). Massive increases in the volume and complexity of international economic transactions have multiplied the probability that a transaction will escape the nets of statistical measurement, or that it will be misattributed in the national accounts.

Merchandise trade statistics face serious difficulties to distinguish confidently between the places in which cargo is loaded and unloaded, and the locations where it was actually produced or consumed (Ahmad, 2018). And as global production chains deepen, the statistical blend of such conceptually distinct flows increasingly distorts interpretations of the data (interview with Fabienne Fortanier, Head of Trade Statistics at OECD Statistics Directorate, Paris, June 6, 2017).

Statistics on trade in services raise additional questions (Giovannini and Cave, 2005), not least when they struggle to separate actual cross-national transactions from mere MNE-internal accounting procedures. The growing divergence between the geography of corporate activities and the associated accounting practices can lead to situations in which companies' domestic sales are counted as services “trade” merely because they are registered abroad for tax purposes—a phenomenon that Robert Lipsey (2006: 37) refers to as “phantom flows of trade.” Such issues pose a serious challenge for the validity of trade statistics. If left unaddressed, established indicators gradually risk to “lose their meaning” (Lipsey, 2006: 50).

While global companies' use of offshore structures can severely distort trade statistics, the implications for capital flow statistics are even graver. To minimize tax payments, MNEs commonly create special purpose vehicles in low-tax offshore jurisdictions and “book” profits on intellectual property there (Finér and Ylönen, 2017; Shaxson, 2012; Tørsløv et al., 2018). As Maria Borge, Head of Foreign Direct Investment Statistics at the OECD, put it in a paper co-authored with Cecilia Caliandro:

FDI statistics can . . . reflect other factors, such as fiscal optimisation to reduce tax burdens and the increasing sophistication in MNEs' capital structures. This can make it difficult to interpret

FDI statistics, in the sense that they are not “real” and no longer represent “long-term” investments in a country. (Borga and Caliandro, 2018: 1)

To distinguish between long-term productive investments and short-term speculative capital flows (itself a questionable dichotomy, see de Goede, 2005), BPM defines cross-border acquisitions of at least 10% of a company as FDI, with all smaller investment being classified as FPI (Foreign Portfolio Investment). At the same time, the U.S. Bureau of Economic Analysis (Ibarra-Caton and Mataloni, 2014) and Eurostat (2016) have estimated that between one-half and two-thirds of total BOP FDI in- and outflows come from or go to offshore special purpose entities (SPEs) rather than an identified parent or subsidiary company. With their current tools, BOP statisticians cannot determine the purpose or ultimate destination of these flows. Impenetrable ownership structures frustrate distinctions between long-term and speculative investments by for example private equity or hedge funds (Blanchard and Acalin, 2016; interview with U.S. BEA economists, Washington, September 20, 2017), or between genuinely “foreign” investments and corporate inversions. Recent estimates from the IMF indicate that such “phantom investments” amount to \$15 trillion a year, or nearly 40% of all global FDI flows (Damgaard et al., 2019).

FPI statistics face similar problems. Short-term capital flows are channeled through opaque structures of financial intermediaries, which BOP data is unable to track. As a result, official figures are biased toward custodian centers such as Liechtenstein, Luxembourg, or Switzerland (Bertaut et al., 2006; Bryan et al., 2017; Tørsløv et al., 2018), and national statisticians (and tax authorities) struggle to estimate the equity and debt positions of residents who park their assets and liabilities in offshore financial centers (Fichtner, 2017). In a global financial system in which “nationality” is a “tradable attribute of an asset” (Bryan et al., 2017: 52) rather than a physical location, attempts to measure “national” holdings can fundamentally mislead.

In short, national accounting templates that assume simplistic economic relationships capture our current economic realities less and less well. Denationalized production and opaque corporate and financial structures have undermined the validity and hence usefulness of BOP statistics. Statisticians are well aware of these problems (e.g., Damgaard and Elkjaer, 2017; International Monetary Fund, 1987, 1992; OECD, 2016; UNECE et al., 2011). Indeed, they have been discussing them since the 1950s (International Monetary Fund, 1956; Smith, 1966). One way the international statistical organizations have sought to reduce asymmetries in BOP figures is through facilitating bilateral meetings between national compilers (interview with IMF statisticians, Washington, September 19, 2017). Recent standards seek to capture complex trading activities such as “merchandising” or “goods sent abroad for processing” better. And additional efforts are being pursued to update standards to better reflect present-day realities: the IMF is exploring ways to get a better grip on the measurement of Special Purpose Entities in global financial flows (International Monetary Fund, 2016a), and the OECD has created a Trade in Value Added (TiVA) database that aims to disentangle gross trade flows from actual value creation.

But these efforts have clear limits, statisticians concede:

TiVA offers an interesting complementary perspective. But it is built on data which are not very precise, as compilation involves many imputations and data modelling to fill in the gaps.

Essentially, it is “modelled” data, not real data. (Interview with IMF statisticians, Washington, September 19, 2017; cf. Ahmad, 2018)

Eurostat statisticians feel that

we are only at the very beginning of getting a grip on properly measuring globalisation in a systematic cross-country way in practice. Which parts of the production activities of MNEs are actually “taking place” on the domestic territory of any given country? [H]ow can we distinguish between movements in GDP or its components which are relevant for the domestic economy and those which are driven by the worldwide activities of multinational companies? (Stapel-Weber et al., 2018: 2)

On balance, these initiatives have failed to stem the deterioration of BOP measurement quality—they are “plasters on the holes of a sinking ship,” in the words of one statistician we interviewed (anonymous interview, 25 April 2017).

If the data are as bad as we have outlined and if statisticians are aware of the problems, why do outdated international economic statistics still dominate representations of the global economy? What makes statistical standards so sticky when they increasingly fail to fulfill their goals?

Statistical norms and conservative bias

While we appreciate the practical challenges of producing high-quality statistics, they do not tell the whole story. Clearly, statisticians and the users of statistics could have reformed measurement standards to suit changed circumstances and new perspectives better. Examples from other areas in the statistical field show that this is possible. The World Bank and United Nations have drastically adapted their “development” measures over time (Finnemore, 1996) while many governments have revisited ethnic categories in their censuses (Marquardt and Herrera, 2015; Petersen, 1987). At least in principle, measurement systems can adapt in the face of social and political change.

As an alternative to updating standards, statisticians could have highlighted measurement deficiencies more forcefully. To begin with, they could have refused to report deceptively precise point estimates—single figures—for FDI or FPI. The reporting of data ranges is common in forecasting, for example for different climate change scenarios. While the use of confidence intervals might seem more intuitive for future projections, inaccuracies in the measurement of past economic transactions are frequently substantial enough to warrant similar caution. Were this impossible, statisticians could have abandoned obsolete measures altogether, admitting that we simply do not know the investment relationship between two far-flung countries.

Whereas we do observe bolder experimentations with the adaptation of measurement standards or the creation of new ones in other areas, none of this is happening with more lofty official national accounting figures. To understand the stickiness of these statistical representations, this section evidences how the four norms laid out in general terms above have affected the production of BOP statistics.

Comparability

From its outset, the global statistical enterprise emphasized the need for comparable numbers. Politically, the rise of national accounting systems is strongly tied to the needs and wishes of nation-states. They continue to be the focal points of political authority. That perpetuates pressure to keep producing statistics about “national economies” as the units for which politicians can be held accountable. National economic performance can then be compared to those of other countries—even if both these economic units and the idea that they could be politically controlled top-down are frequently illusory.

This desire to compare stood central since the early days of BOP statistics. The League of Nations tried various strategies to encourage countries to report uniform figures, albeit with limited success (Alves, 1967). The lack of uniformity and cross-country comparability of BOP statistics released by the League were seen as a significant shortcoming. In the eyes of a statistician involved in the elaboration of statistical standards at the IMF in the post-war period, this undermined the whole enterprise: “Because the attempt to achieve uniformity was only partially successful, the usefulness of the figures in the League’s publications is severely limited” (Alves, 1967).

Until today the scope for cross-country comparison is the key attraction of multicountry databases. Indeed, we commonly think of international standards and best practices as improving data quality because they promote interagency learning and facilitate expert debate across borders. Yet it is easy to underestimate the difficulties of achieving the requisite uniformity in numbers, collected as they are by disparate national agencies. The harmonization of accounting-technical standards is challenging, costly, and time-consuming. And because the “sunk costs” of that harmonization demands are so high, the aim of comparability unwittingly retards change. Even when good reform ideas abound, countries struggle to agree on acceptable and implementable standards. To grasp the diversity of cross-border investment flows better, the OECD’s 4th edition of the *Benchmark Definition of FDI* (developed over several years with the IMF and the UN) asked countries to compile separate figures for greenfield, merger and acquisition, and special purpose entity inward FDI flows since 2008 (OECD, 2017). But despite the obvious improvements promised by this distinction, progress has been frustratingly slow (interview with OECD statistician, phone call, 30 May 2017). So long as only a few countries report such data, they cannot be included in cross-national databases.

Comparability as a statistical norm also limits the sensible adaptation of standards to national circumstances, a tension noted by statisticians since the inception of BOP data collection. In an internal letter dated June 22, 1953, A.B. Hersey from the Board of Governors of the Federal Reserve System wrote to the IMF: “Though flexibility is desirable, so is uniformity. The Fund’s problem is how best to reconcile the two objectives” (Hersey, 1953). Rules for FDI statistics offer countries alternative options to value inward FDI stocks, such that countries can pick that which fits their situation best. But national statisticians may privilege convenience over quality of conceptual fit when they choose among the alternatives (Aragão and Linsi, 2020). As high-level Eurostat officials recently argued regarding national accounts, “any new indicator or breakdown, particularly in a European context, should be comparable across countries and not be seen as a GDP or GNI ‘a la carte’ for each country to choose from under specific circumstances” (Stapel-Weber et al., 2018: 1).

Data compilation is done by national authorities with their own organizational structures and legal traditions. Even when “the concepts are exactly the same, [...] the ways in which they are measured can be different” (Interview with Fabienne Fortanier, Head of Trade Statistics at OECD Statistics Directorate, Paris, June 6, 2017). Hence, if the comparability of figures is the goal, room for national discretion must shrink. As the authors of a bilateral asymmetry study of Germany and Portugal have pointed out, to eliminate measurement errors, “harmonization of theoretical concepts is not sufficient. Essential is a common approach to the practical application and interpretation of concepts and definitions” (Deutsche Bundesbank, 1997: 6).

Surveys on data collection practices by national statistical offices (IMF and OECD, 2003; United Nations Statistics Division, 2006) and bilateral reconciliation exercises have highlighted a large number of factors that can undercut the cross-national comparability of figures, such as at-odds currency conversions, the use of dissimilar valuation techniques, or differences in classification decisions for transactions that fall into a gray area.

In response, international organizations have pushed further to narrow national compilers’ room for interpretation and discretion in data gathering and reporting. But the most recent BPM compilation guide concedes that there are real limits:

Articulating balance of payments and IIP [International Investment Position] compilation methodology is difficult because economies have developed procedures independently, and each national methodology may be considered unique. Some patterns emerge, but different national experiences have created different approaches as to the most appropriate methodology. Consequently, it is not possible to present a single methodology suitable in all cases. Instead, the Guide outlines various options that may be available. (International Monetary Fund, 2014: 2)

Adherence to the harmonization norm thus means that statistical standards change slowly. Revised versions of the BPM and the System of National Accounts (SNA) are typically published every decade or two, and even then the changes are rarely radical. (The last major overhaul of the SNA dates from 1993; the currently used 2008 version only updated relatively small elements). Still, given the diversity of economic developments around the world, for example different degrees of digitization, the resulting standards may still be at odds with circumstances in any particular country. Given the potential for misuse, overly flexible standards are no solution either. In short, the ambition to have comparable data retards the evolution of measurement standards, and thereby actually dents measurement quality.

Continuity

While the *comparability* norm requires countries to have a common yardstick, the *continuity* norm seeks to ensure that we can capture developments over time. One of the great attractions of statistics is their claim to track macrosocial or economic developments (Trewin, 2007). If measurement approaches change and we are unable to adjust past measurements retrospectively, diachronic comparability is lost.

The IMF already worried in the 1950s: “[the] Fund should . . . insure that continuity of the series is not disturbed” (International Monetary Fund, 1956). Even after the gradual switch to *BPM6* in the 2000s, the IMF continued to receive requests for continuous time-series data covering the past decades up to the present day (Shrestha et al., 2016: 6). Indeed, series continuity is a key argument in statistical disputes. As the global financial system shed its post-war shackles in the late 1960s, IMF statisticians agonized over what they called “empty shell” holding companies. The UK’s central statistical office concluded that

while . . . we would not dissent from the view which the I.M.F. [sic] say they have expressed that, in principle the statistics should ideally relate to the final origin or destination of investment, we do not believe that this goal is obtainable in practice. Any partial move in this direction would involve serious discontinuities, from which we might lose more than we should gain. (Stanton, 1967; emphasis added)

One way to meet the continuity norm while updating statistics is to develop parallel statistics: to begin a new series while continuing with the old one for the time being. Although the figures from the TiVA initiative offer conceptually somewhat better figures than conventional trade statistics, they are not directly integrated into the BOP system to avoid breaking the series. Continuity over time is more important for some series and users than others. Academics using regression analysis, for example, typically rely on temporally extended series to disentangle the effects of multiple variables or to observe delayed effects. Users of BOP data often use models with many variables to accommodate differences between countries—variables that frequently do not shift very much from one year to the next (think of economic growth rates, sectoral profiles, and so on). Long time series then become essential for strong statistical inference. While the importance of a break in the series will vary from case to case, the need to continually adapt indicators to changing economic realities clearly diminishes the comparability of data over time—to the extent that authorities may stick with indicators even when they are becoming obsolete.

Certitude

Statistical systems have in-built preferences for measures that minimize the scope for subjective judgments or gross manipulation, an intuition that dovetails with good statistical practice. Statistics’ claim to objectivity—and their status as neutral arbiters in public affairs—hinges on reliable measures that follow constant routines. To denote the resulting penchant toward “hard” measurement procedures we use the label *certitude*. The norm of *certitude* in turn privileges elements that can be unambiguously quantified, that are directly observable, and require no further interpretation.

The demand for *certitude* is particularly high where official, public statistics are concerned. BOP statistics do not have the immediate distributive implications of for example inflation measures, which can directly feed into uprating for pensions or inflation-adjusted wages. Nevertheless, given the political salience of macroeconomic developments in general, suspicions that statistics would be manipulated or just

guesswork have to be avoided. Only then can they function as seemingly neutral arbiters both in domestic and international political disputes (Mügge, 2019).

Curtailing statisticians' room for subjective judgment and maneuver has costs in terms of validity (cf. Schedler, 2012). Even where informed estimates might generate the best figures, they may be eschewed in favor of measurement procedures that rely on hard, unambiguous data. FDI statistics, for example, hang on the "nationality" of domestic firms' foreign owners. But corporate "nationality" is a complex construct, especially when several owners from various jurisdictions channel investment through multicountry tax structures. In such instances, subjective judgment would be useful, for example by showing the holdings of Amazon Luxembourg to be mostly American investments. In practice, however, statistical standards opt for an unambiguous but ultimately misleading classification: the "legal residence" of an investor, which makes Amazon Europe a Luxembourg company.

Similar problems apply to capital flow statistics that try to distinguish between predominantly "financial" (FPI) and "productive" investments (FDI). Earlier versions of the *Balance of Payments Manual* relied on the qualitative judgment of national accountants (International Monetary Fund, 1961: 120; International Monetary Fund, 1977a: 138). The IMF subsequently abandoned this approach in favor of an unambiguous threshold: all foreign investments that involve at least 10% of a company's voting stock are to be counted as FDI (Linsi, 2018). Although this rule was always arbitrary, it has become more so in recent decades. Activist hedge funds increasingly buy and sell large corporate stakes for quick financial gain—directly contravening the assumption that large investments are automatically also long-term (interview with U.S. BEA economists, Washington, September 20, 2017). What makes this rule attractive despite its shortcomings is that it can be uniformly applied. It is reliable: a different person repeating the same procedure would get similar figures. But consistent rules are less flexible than qualitative judgments and risk ignoring changing circumstances. In a trade-off familiar to social scientists, reliability comes at the cost of validity. In public life, we can therefore find a tension between data that is useful in political practice—because it abides by the certitude norm—and data that does justice to the phenomenon it tries to capture.

These examples are indicative of a broader trend: globalization, digitization, and financialization have reduced the number of points at which we can more or less directly gauge economic quantities of interest. With globalized production, the complexity derives from ever longer and increasingly intermeshed production chains. In other instances, corporations erect complex legal facades to shape outsiders' perceptions, irrespective of how these facades relate to productive activities on the ground or financial connections between the ultimate beneficiaries.

Certitude as a norm not only entails unambiguous measurement procedures; it also means a preference for point estimates—data in the form of single numbers—that obscure the uncertainty underlying statistics. Rather than claiming that, say, the trade deficit of the United States with Mexico in 2017 was \$70.952 billion (United States Census Bureau, 2017), it might be more honest to say that "We, the U.S. Census Bureau, have a sense that last year the deficit was somewhere between \$65 and \$75 billion." But such a presentation of economic statistics is currently not considered acceptable; to retain credibility as social facts, they need to perpetuate the pretense of certitude.

Point estimates are also necessary for interpolation, imputations, and other statistical operations, including regressions. At the same time, as the former deputy director of the Dutch statistical office put it:

Statistical institutions have to guard the authority of their statistics. Therefore they will be reluctant to emphasise the shortcomings or to develop competing (conflicting) information. The authority of a set of statistics grows with the duration of its use. This encourages official statistical institutes to maintain existing statistics, and thus to be conservative in developing substitutes. (van Tuinen, 2007: 267)

This need to safeguard the incontrovertible image of statistics is also appreciated by Eurostat statisticians: “Given the potential impacts on macroeconomic statistics across countries, and the *adverse reaction of users to ‘surprises’ in data*, [globalization] presents a major challenge to official statisticians” (Stapel-Weber et al., 2018: 3; emphasis added). Peter van de Ven, head of national accounts at the OECD, and Brent Moulton, former head of national accounts at the U.S. Bureau of Economic Analysis (2018: 18), equally worry that quirks in macroeconomic data—for example, the on-paper relocation of economic activity that led to a 23% jump in Irish GDP—can be abused to disqualify statistics more generally. They are therefore dismissed as idiosyncratic anomalies, instead of being acknowledged as symptoms of more deeply rooted data problems.

Certitude as a statistical norm thus puts macroeconomic statistics in a double bind. On the one hand, it is felt necessary to sustain the figures’ credibility. On the other, it stands in the way of both a more creative and flexible adaptation of statistical standards to new economic circumstances and more open admission of the increasing uncertainty that underlies macroeconomic figures.

Coherence

The final norm is coherence: individual economic measures should fit into a larger, coherent whole that depicts “the economy” in its entirety. Individual components of BOP statistics are meant to offer an encompassing image of intercountry economic exchanges. FDI and FPI statistics, for example, are direct complements that together are meant to capture cross-border investment.

Statisticians and econometricians played key roles in the elaboration and practical implementation of John Maynard Keynes’ ideas, especially that of the national economy as a system of logically interrelated parts. Jacques Polak, director of the IMF’s research department from 1958 to 1979, developed the “Polak model” relating key domestic macroeconomic variables such as GNP growth and domestic credit of the banking system to cross-border economic variables such as foreign exchange reserves and trade (Polak, 1997; Woods, 2006). Later theoretical refinements formalized the relationship between the balance of payments, changes in the domestic money supply, and developments in the real economy (International Monetary Fund, 1977b). Rather than isolated macroeconomic quantities to be observed individually, the constituent elements of the balance of payments came to be seen as building blocks of a larger integrated whole. This suggested for example that a net surplus or deficit in cross-border flows must imply a depletion or

increase in net foreign reserves, such that the latter could be imputed from knowledge of cross-border financial flows.

Due to these developments, the BOP system does not stand on its own but is a part of the System of National Accounts. The “linkage of the . . . balance of payments accounts to the . . . System of National Accounts (SNA) is strengthened and harmonized to the maximum extent possible” (International Monetary Fund, 2009). The different sectoral accounts are communicating vessels: a change in one account must be accounted for elsewhere—a trade deficit comes with a capital account surplus, while export revenues in the BOP are also someone’s income in the SNA. Revisions of the BOP and the SNA thus progressed in parallel in the 1980s: the “need for compatibility between the two standards is one reason both are being revised” (International Monetary Fund, 1992: 16). Such linkages complicate efforts to update statistical definitions and procedures: changes in one place have knock-on effects elsewhere. They trigger a “train of adjustment” (International Monetary Fund, 1956).

Thinking of national economic and BOP statistics as an integrated whole also means that statistical concepts are often deductively defined. The conceptual coherence of accounts tempts us to impute values for concepts that are not directly observable. Once we accept axiomatically that $a + b = X$ and we have values for a and X , we can impute b and report it as a known quantity. But in the process, all kinds of measurement problems with a and X disappear. The quality of b as a data point is going to be no better than those of a and X .

A false sense of symmetry can also be fomented when the theoretical interlinkages of the system are used to force statistics into balance. The residual “errors and omissions” column in many statistical tables suggests that the values reported for other variables—imports, exports, different kinds of capital flows, and so on—are reliable and that measurement problems have somehow been distilled out of them.

The theoretical elegance of the models underlying the collection of national accounts data stands central in today’s BOP statistics. While the IMF’s original efforts to collect macroeconomic data merely sought to assemble national statistics from various sources, the project has evolved into an intellectual enterprise to integrate the figures into a theoretically coherent whole. The style and substance of the IMF’s *Balance of Payments Manuals* mirror this development: the pioneering *BPM1* (International Monetary Fund, 1948), less than 50 pages long, simply provided a set of tables to be filled out by national statisticians. In stark contrast, the most recent version, *BPM6* (International Monetary Fund, 2009), is a highly didactic document of almost 400 pages, accompanied by a separate 600-page *Compilation Guide* (International Monetary Fund, 2014). As the authors of the preceding *BPM5* highlighted, the manual

not only defines and describes the content of the categories employed but also attempts to explain their rationale. [...] With these amendments, the Manual has become as much an introduction to the principles of balance of payments accounting as a guide to reporting. (International Monetary Fund, 1995: 2)

Statisticians rightly take pride in the sophistication of the models they have developed over the years. Most national accountants are trained economists, and in view of the

mathematical fetish that dominates the discipline (Fourcade, 2010), the theoretical models underlying contemporary national accounts play an important role in granting legitimacy to macroeconomic statistics. As such, statisticians' modeling of the world economy as a logically coherent, internally balancing system buttresses the authority of economic expertise that builds on it. But the same ambition simultaneously represents a monumental obstacle for attempts to reform statistical standards since "solving asymmetries in one item may create new asymmetries in another one" (International Monetary Fund, 2016b: 21). In this way, the coherence norm further reinforces statistics' stickiness to theoretically elegant but otherwise outdated statistical standards.

Conclusion

The statisticians we interviewed were painfully aware of the problems discussed in this article. Yet data users normally take BOP figures for granted and spend little time dissecting them. Macroeconomic statistics cement properties of national economies as social facts—so much so that gaps between the concept and the measurement routines fade from view.

Such concept–measurement gaps in BOP statistics have swelled over the past decades. But economic headline figures take little heed of growing problems, as they might have done through overhauled definitions, substitution of outdated concepts, or simple admissions that uncertainty and ambiguity have risen substantially. Unless one pores over footnotes in statistical yearbooks, the published numbers continue to project a level of accuracy that is at odds with the ambiguities in the data.

The mounting concept–measurement gap has severe ramifications for the quality of economic policy, public debate, and academic analysis. Policies that try to stem capital outflows or to combat tax evasion or financial instability lose effectiveness when the figures on which they build are increasingly distorted. Public debates, for example about trade imbalances, become increasingly vacuous when they are fed with data that hide rather than reveal global interdependencies and complex flows of value added. And as we have shown through a replication exercise of quantitative IPE research elsewhere (Linsi and Mügge, 2019), heeding BOP data defects in regression analyses can seriously affect our inferences.

So why do outdated international economic statistics continue to dominate representations of the global economy? Our analysis has focused on four norms underpinning national accounting: comparability, continuity, certitude, and coherence. To be sure, on their own these norms are intuitive and plausible enough. Dennis Trewin, former head of the Australian Bureau of Statistics, argued that to "be useful, international statistics must be relevant, of good quality and consistent across countries and across time" (Trewin, 2007: 308). Here we find, in a nutshell, the four norms we have discussed: what we have called certitude is seen as a way to quality, while consistency across countries and time is what we have called comparability and continuity. From a policy perspective, relevance is derived from statistics' commensurability with macroeconomic concepts and models as used by policymakers.

But fashioning statistical standards after these norms paradoxically damages measurement quality: abiding by them stands in the way of adapting statistical measures to

amorphous and quickly changing economic dynamics—in particular the digitization of economic activity, ever more complex value and wealth chains (Seabrooke and Wigan, 2017), and the erosion of national borders as economic boundaries. Unless we are willing to compromise on norms such as international comparability or the predilection for long time series, we are destined to live with statistical representations that are increasingly poor guides to global economic dynamics.


Acknowledgements

Earlier versions of this article were presented at the 2018 EWIS Workshops in Groningen, the 2018 SASE Annual Convention in Kyoto, and the 2018 SPERI-PETGOV Workshop in Amsterdam. We are grateful for the comments and suggestions we received there, particularly from Jasper Blom, Greg Fuller, Andrew Hindmoor, Saliha Metinsoy, and Liam Stanley. We thank Takeo David Hymans for editing this manuscript. Hanna Dose provided excellent research assistance. We are deeply indebted to the many statisticians who generously shared their insider perspectives in research interviews. All errors remain our own.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Our work has been supported by the ERC Starting Grant FICKLEFORMS (grant # 637883), the NWO Vidi project 016.145.395, and an Early Postdoc Mobility Grant from the Swiss National Science Foundation (grant P2SKP1_168289).

ORCID iD

Daniel Mügge  <https://orcid.org/0000-0001-9408-7597>

Note

1. We use the label “statistics” to mean systematic, quantitative data—not the branch of mathematics concerned with the analysis of such data.

References

- Abolafia MY (2010) Narrative construction as sensemaking: how a central bank thinks. *Organization Studies* 31(1): 349–367.
- Adler E and Haas P (1992) Conclusion: epistemic communities, world order, and the creation of a reflective research programme. *International Organization* 46(1): 367–390.
- Ahmad N (2018) Accounting frameworks for global value chains: extended supply-use tables. In: *CRIW conference: The challenges of globalization in the measurement of national accounts*. Bethesda, MD, 2018, pp.1–31.
- Alenda-Demoutiez J and Mügge D (2020) The lure of ill-fitting unemployment statistics: how South Africa’s discouraged work seekers disappeared from the unemployment rate. *New Political Economy* 25(4): 590–606. Available at: <https://www.tandfonline.com/doi/full/10.1080/13563467.2019.1613355>
- Allan B (2017) From subjects to objects: knowledge in International Relations theory. *European Journal of International Relations* 24(4): 841–864.
- Alves J (1967) Progress towards uniformity in balance of payments presentation. In: *Research and Statistics Department, Assistant Director Arie Bouter Files, Box #1 File #8*. Washington, DC: IMF Archives.

- Aragão R and Linsi L (2020) Many shades of wrong: what governments do when they manipulate statistics. *Review of International Political Economy Early View*: DOI: 10.1080/09692290.2020.1769704
- Babb S (2007) Embeddedness, inflation, and international regimes: the IMF and the early postwar period. *American Journal of Sociology* 113(1): 128–164.
- Barnett M and Finnemore M (2004) *Rules for the World: International Organizations in Global Politics*. Ithaca, NY: Cornell University Press.
- Baxandall P (2004) *Constructing Unemployment: The Politics of Joblessness in East and West*. Hoboken, NJ: John Wiley & Sons.
- Beckert J (2016) *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Cambridge, MA: Harvard University Press.
- Bertaut CC, Grier WL and Tryon RW (2006) Understanding U.S. cross-border securities data. *Federal Reserve Bulletin* 97(May 2006): A59–A75.
- Blanchard O and Acalin J (2016) *What Does Measured FDI Actually Measure?* Washington, DC: Peterson Institute for International Economics.
- Bloch D and Fall F (2015) *Government Debt Indicators: Understanding the Data*. OECD Economics Department Working Papers 1228. Paris.
- Borga M and Caliendo C (2018) Eliminating the pass-through: towards FDI statistics that better capture the financial and economic linkages between countries. In: *CRIW conference: the challenges of globalization in the measurement of national accounts*, Bethesda, MD, 2018.
- Broome A (2010) *The Currency of Power: The IMF and Monetary Reform in Central Asia*. London: Palgrave Macmillan.
- Broome A and Quirk J (2015) The politics of numbers: the normative agendas of global benchmarking. *Review of International Studies* 41(05): 813–818.
- Broome A, Homolar A and Kranke M (2018) Bad science: International organizations and the indirect power of global benchmarking. *European Journal of International Relations* 24(3): 514–539.
- Bryan D (2001) Global accumulation and accounting for national economic identity. *Review of Radical Political Economics* 33(1): 57–77.
- Bryan D, Rafferty M and Wigan D (2017) From time–space compression to spatial spreads: situating national liquidity in global financial liquidity. In: Christophers B, Leyshon A and Mann G (eds) *Money and Finance After the Crisis: Critical Thinking for Uncertain Times*. Hoboken, NJ: John Wiley & Sons, 43–68.
- Chorev N and Babb S (2009) The crisis of neoliberalism and the future of international institutions: a comparison of the IMF and the WTO. *Theory and Society* 38(5): 459–484.
- Chwieroth J (2009) *Capital Ideas: The IMF and the Rise of Financial Liberalization*. Princeton, NJ: Princeton University Press.
- Chwieroth J and Sinclair T (2013) How you stand depends on how we see: International capital mobility as social fact. *Review of International Political Economy* 20(3): 457–485.
- Clegg L (2010) Our dream is a world full of poverty indicators. The US, the World Bank, and the power of numbers. *New Political Economy* 15(4): 473–492.
- Cohen BJ (1969) *Balance-of-Payments Policy*. Baltimore, MD: Penguin Books.
- Cooley A and Snyder J (2015) *Ranking the World. Grading States as a Tool of Global Governance*. New York: Cambridge University Press.
- Damgaard J and Elkjaer T (2017) *The Global FDI Network: Searching for Ultimate Investors*. IMF Working Paper WP/17/258. Washington, DC.
- Damgaard J, Elkjaer T and Johannesen N (2019) The rise of phantom investments. *Finance & Development* 56(3): 11–13.
- Daston L and Galison P (2007) *Objectivity*. New York: Zone Books.

- Davis K, Kingsbury B and Merry SE (2012) Indicators as a technology of global governance. *Law & Society Review* 46(1): 71–104.
- De Goede M (2005) *Virtue, Fortune and Faith: A Genealogy of Finance*. Minneapolis, MN: University of Minnesota Press.
- DeRock D (2019) Hidden in plain sight: unpaid household services and the politics of GDP measurement hidden in plain sight. *New Political Economy*. DOI: 10.1080/13563467.2019.1680964.
- Desrosières A (1993) *La Politique Des Grands Nombres: Histoire de La Raison Statistique*. Paris: Éditions La Découverte.
- Deutsche Bundesbank (1997) *Comparison of Bilateral Balance of Payments Between Portugal and Germany*. IMF Committee on Balance of Payments Statistics Working Papers CBOPWB/97/2.
- Dorling D and Simpson S (1999) *Statistics in Society: The Arithmetic of Politics*. London: Arnold.
- Espeland W and Stevens M (2008) A sociology of quantification. *European Journal of Sociology* 49(3): 401–436.
- Espeland WN and Sauder M (2007) Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology* 113(1): 1–40.
- Eurostat (2016) Special purpose entities within EU direct investment statistics. In: *Twenty-Ninth meeting of the IMF committee on balance of payments statistics BOPCOM-16/05*, Washington, DC, 24–26 October 2016.
- Fichtner J (2017) Perpetual decline or persistent dominance? Uncovering Anglo-America's true structural power in global finance. *Review of International Studies* 43(1): 3–28.
- Finér L and Ylönen M (2017) Tax-driven wealth chains: a multiple case study of tax avoidance in the Finnish mining sector. *Critical Perspectives on Accounting* 48(October 2017): 53–81.
- Finnemore M (1996) *National Interests in International Society*. Ithaca, NY: Cornell University Press.
- Fioramonti L (2013) *Gross Domestic Problem: The Politics behind the World's Most Powerful Number*. London: Zed Books.
- Fourcade M (2006) The construction of a global profession: the transnationalization of economics. *American Journal of Sociology* 112(1): 145–194.
- Fourcade M (2010) *Economists and Societies: Discipline and Profession in the United States, Britain, and France, 1890s to 1990s*. Princeton, NJ: Princeton University Press.
- Giovannini E and Cave W (2005) *The Statistical Measurement of Services*. OECD Statistics Working Papers 2005/2. Paris. DOI: 10.1787/748203537383.
- Griever WL, Lee GA and Warncock FE (2001) The U.S. system for measuring cross-border investment in securities: a primer with a discussion of recent developments. *Federal Reserve Bulletin* 92(October 2001): 633–650.
- Halliday TC and Carruthers BG (2009) *Bankrupt: Global Lawmaking and Systemic Financial Crisis*. Stanford, CA: Stanford University Press.
- Herrera Y (2010) *Mirrors of the Economy: National Accounts and International Norms in Russia and Beyond*. Ithaca, NY: Cornell University Press.
- Hersey AB (1953) Comments on IMF Research Department Paper of May 21 on balance of payments presentation. In: *Research and Statistics Department, Assistant Director Arie Bouter Files, Box #1 File #5*, Washington, DC: IMF Archives.
- Honig D and Weaver C (2019) A race to the top? The aid transparency index and the social power of global performance indicators. *International Organization* 73(3): 579–610.
- Hoskyns C and Rai S (2007) Recasting the global political economy: counting women's unpaid work. *New Political Economy* 12(3): 297–317.
- Ibarra-Caton M and Mataloni RJJ (2014) *Direct Investment Position for 2013*. Survey of Current Business 94(7): 1–17.

- International Monetary Fund and OECD (2003) *Foreign Direct Investment Statistics: How Countries Measure FDI*, Washington DC: IMF
- International Monetary Fund (1948) *Balance of Payments Manual*. 1st edn. Washington, DC.
- International Monetary Fund (1950) *Balance of Payments Manual*. 2nd edn. Washington, DC.
- International Monetary Fund (1956) Verbatim Report of the International Monetary Fund Meeting of Fund Statistical Correspondents on Balance of Payments Discussions held at the Burgundry Room—Sheraton-Park Hotel, Washington D.C. on Thursday, September 27, 1956 at 2:30 pm. In: *Research and Statistics Department, Assistant Director Arie Bouter Files, Box #1 File #1*, Washington, DC, 1956. IMF Archives.
- International Monetary Fund (1961) *Balance of Payments Manual*. 3rd edn. Washington, DC.
- International Monetary Fund (1977a) *Balance of Payments Manual*. 4th edn. Washington, DC.
- International Monetary Fund (1977b) *The Monetary Approach to the Balance of Payments*. Washington, DC.
- International Monetary Fund (1987) *Report on the World Current Account Discrepancy*. Washington, DC.
- International Monetary Fund (1989) Study on the measurement of international capital flows. In: *Bureau of Statistics, Balance of Payments Division, Subject Files, Box #12*, Washington, DC: IMF Archives.
- International Monetary Fund (1992) *Report on the Measurement of International Capital Flows*. Washington, DC.
- International Monetary Fund (1995) *Balance of Payments Compilation Guide*. Washington, DC: International Monetary Fund.
- International Monetary Fund (2009) *Balance of Payments and International Investment Position Manual*. Washington, DC.
- International Monetary Fund (2014) *BPM6 Compilation Guide: Companion Document to the Sixth Edition of the Balance of Payments and International Investment Position Manual*.
- International Monetary Fund (2016a) Compiling data on special purpose entities. In: *Twenty-Ninth meeting of the IMF committee on balance of payments statistics BOPCOM 16-05*, Washington, DC, 24–26 October 2016.
- International Monetary Fund (2016b) Summary of discussions. In: *Twenty-Ninth meeting of the IMF committee on balance of payments statistics*. Washington, DC, 2016, International Monetary Fund.
- Jerven M (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Ithaca: Cornell University Press.
- Kelley JG and Simmons BA (2015) Politics by number: indicators as social pressure in International Relations. *American Journal of Political Science* 59(1): 55–70.
- Kelley JG and Simmons BA (2019) Introduction: the power of global performance indicators. *International Organization* 73(3): 491–510.
- Kentikelenis AE and Seabrooke L (2017) The politics of world polity: script-writing in international organizations. *American Sociological Review* 82(5): 1065–1092.
- Kerner A, Jerven M and Beatty A (2017) Does it pay to be poor? Testing for systematically under-reported GNI estimates. *Review of International Organizations* 12(1): 1–38.
- Knafo S (2019) Neoliberalism and the origins of public management. *Review of International Political Economy* DOI: 10.1080/09692290.2019.1625425
- Krippendorff K (2008) Reliability. In: Donsbach W (ed.) *The International Encyclopedia of Communication*. DOI: <https://doi.org/10.1002/9781405186407.wbiecr029>.
- Leins S (2018) *Stories of Capitalism: Inside the Role of Financial Analysts*. Chicago: University of Chicago Press.

- Lepenies P (2013) *Die Macht Der Einen Zahl: Eine Politische Geschichte Des Bruttoinlandsprodukts*. Frankfurt am Main: Suhrkamp Verlag.
- Lequiller F and Blades D (2014) *Understanding National Accounts*. 2nd edn. Paris: OECD.
- Linsi L (2018) Foreign direct investment. In: Dingwerth K and Weinhardt C (eds) *The Language of World Trade Politics: Unpacking the Terms of Trade*. London: Routledge, 50–63.
- Linsi L and Mügge D (2019) Globalization and the growing defects of international economic statistics. *Review of International Political Economy* 26(3): 361–383.
- Lipsey RE (2006) Measuring international trade in services. *National Bureau of Economic Research Working Paper Series* No. 12271(May).
- Machlup F (1950) Three concepts of the balance of payments and the so-called dollar shortage. *The Economic Journal* 60(237): 46–68.
- Mackie C and Schultze C (2002) *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Washington, DC: National Academies Press.
- Mankiw GN (2017) *Principles of Macroeconomics*. 8th edn. New York: Cengage Learning.
- Marquardt KL and Herrera YM (2015) Ethnicity as a variable: an assessment of measures and data sets of ethnicity and related identities. *Social Science Quarterly* 96(3): 689–716.
- Masood E (2016) *The Great Invention: The Story of GDP and the Making and Unmaking of the Modern World*. New York: Pegasus Books.
- McCormick T (2009) *William Petty and the Ambitions of Political Arithmetic. William Petty: And the Ambitions of Political Arithmetic*. Oxford: Oxford University Press.
- Meade JE (1951) *The Theory of International Economic Policy: The Balance of Payments*. Oxford: Oxford University Press.
- Morgenstern O (1963) *On the Accuracy of Economic Observations*. 2nd edn. Princeton, NJ: Princeton University Press.
- Moulton B and van de Ven P (2018) Addressing the challenges of globalization in national accounts. Washington, DC.
- Mügge D (2016) Studying macroeconomic indicators as powerful ideas. *Journal of European Public Policy* 23(3): 410–427.
- Mügge D (2019) International economic statistics: biased arbiters in global affairs? *Fudan Journal of Humanities and Social Sciences* 13(1): 93–112.
- Muller J (2018) *The Tyranny of Metrics*. Princeton, NJ: Princeton University Press.
- Muniesa F (2014) *The Provoked Economy. Economic Reality and the Performative Turn*. London: Routledge.
- Murdoch Z, Kassim H, Connolly S, et al. (2018) Do international institutions matter? Socialization and international bureaucrats. *European Journal of International Relations* 25(3): 852–877.
- Oatley T (2019) Toward a political economy of complex interdependence. *European Journal of International Relations* 25(4): 957–978.
- OECD (2016) Reducing discrepancies and asymmetries: what can direct investment and trade in services statisticians learn from each other? In: *Twenty-Ninth meeting of the IMF committee on balance of payments statistics BOPCOM-16/11*.
- OECD (2017) *OECD Benchmark Definition of Foreign Direct Investment*. 4th edn. Paris: OECD.
- Petersen W (1987) Politics and the measurement of ethnicity. In: Alonso W and Starr P (eds) *The Politics of Numbers*. New York: Russell Sage Foundation, 187–233.
- Philipsen D (2015) *The Little Big Number: How GDP Came to Rule the World and What to Do about It*. Princeton, NJ: Princeton University Press.
- Pilling D (2018) *The Growth Delusion: The Wealth and Well-Being of Nations*. London: Bloomsbury.
- Pitchford J (1994) The exchange rate, monetary policy and the economy. *Economic Papers: A Journal of Applied Economics and Policy* 13(1): 68–75.

- Polak JJ (1997) *The IMF Monetary Model at Forty*. IMF Working Paper WP/97/49. Washington, DC.
- Porter T (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Power M (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Sætnan AR, Lomell HM and Hammer S (2011) *The Mutual Construction of Statistics and Society*. London: Routledge.
- Salais R, Baverez N and Reynaud B (1986) *L'invention Du Chômage*. Paris: Presses Universitaires de France.
- Schedler A (2012) Judgment and measurement in political science. *Perspectives on Politics* 10(1): 21–36.
- Schmelzer M (2016) *The Hegemony of Growth: the OECD and the Making of the Economic Growth Paradigm*. Cambridge, UK: Cambridge University Press.
- Schultz KA (2015) Borders, conflict, and trade. *Annual Review of Political Science* 18(1): 125–145.
- Seabrooke L and Wigan D (2017) The governance of global wealth chains. *Review of International Political Economy* 24(1): 1–29.
- Searle JR (1995) *The Construction of Social Reality*. New York: Free Press.
- Shaxson N (2012) *Treasure Islands: Uncovering the Damage of Offshore Banking and Tax Havens*. New York: Palgrave Macmillan.
- Shrestha M, Hierro A and Razin T (2016) *IMF Committee on Balance of Payments Statistics 2015 Annual Report*. Washington, DC.
- Smith JS (1966) Asymmetries and errors in reported balance of payments statistics. In: *Research and Statistics Department, Assistant Director Arie Bouter Files, Box #1 File #8*, Washington, DC, 1966. IMF Archives.
- Stanton HC (1967) Copy of a letter dated 5th October 1967 from H.C. Stanton, Board of Trade, to J.W.S. Walton: Overseas Investment - Classification to Countries. In: *Research Department, Balance of payments Division Subject Files, Box#2 File #2*, Washington, DC, 1967. IMF Archives.
- Stapel-Weber S, Konijn P, Verrinder J, et al. (2018) Meaningful information for domestic economies in the light of globalization—will additional macroeconomic indicators and different presentations shed light? In: *Paper presented at the CRIW conference 2018*, Washington, DC.
- Stapleford T (2009) *The Cost of Living in America: A Political History of Economic Statistics, 1880–2000*. Cambridge: Cambridge University Press.
- Stiglitz JE, Sen A and Fitoussi J-P (2010) *Mismeasuring Our Lives: Why GDP Doesn't Add Up*. New York: New Press.
- Studenski P (1958) *The Income of Nations: Theory, Measurement and Analysis Past and Present*. New York: New York University Press.
- Tørsløv T, Wier L and Zucman G (2018) *The Missing Profits of Nations*. NBER working papers 24701. Cambridge, MA: NBER. Available at: <http://www.nber.org/papers/w24701>.
- Trewin D (2007) An effective global statistical system: a prerequisite for consistency in global measurements? In: OECD (ed.) *Statistics, Knowledge and Policy 2007: Measuring and Fostering the Progress of Societies*. Paris, 307–316.
- UNECE, Eurostat and OECD (2011) *The Impact of Globalization on National Accounts*. New York and Geneva: United Nations.
- United Nations Statistics Division (2006) An overview of national compilation and dissemination practices. In: *International Merchandise Trade Statistics: Supplement to the Compilers Manual*. New York: UNSD.
- United States Census Bureau (2017) *Trade in Goods with Mexico*. Washington, DC: United States Census Bureau.

- van Tuinen HK (2007) Innovative statistics to improve our notion of reality. In: OECD (ed.) *Statistics, Knowledge and Policy 2007: Measuring and Fostering the Progress of Societies*. Paris: OECD, 257–296.
- Ward M (2004) *Quantifying the World: UN Ideas and Statistics*. New York: Indiana University Press.
- Waring MJ (1999) *Counting for Nothing: What Men Value and What Women Are Worth*. Toronto: University of Toronto Press.
- Weaver C (2007) The world's bank and the bank's world. *Global Governance* 13(4): 493–512.
- Woods N (2006) *The Globalizers. The IMF, the World Bank, and Their Borrowers*. Ithaca, NY: Cornell University Press.

Author biographies

Daniel Mügge is Professor of Political Arithmetic at the Amsterdam Institute for Social Science Research of the University of Amsterdam, Netherlands. There he directs the FickleFormulas project, which is funded by the European Research Council and the Netherlands Organization for Scientific Research.

Lukas Linsi is Assistant Professor of International Political Economy at the University of Groningen, Netherlands. His expertise covers international trade and foreign direct investment as well as macroeconomic statistics and international political economy more broadly.